

HIV sequence. Because this average number of copies is small, the actual number of copies in a given reaction can vary considerably. Since no HIV DNA target was added to reactions 1 and 5, these reactions served as negative controls.

All eight reactions were subjected to thermocycling using a Perkin Elmer Cetus DNA thermocycler as follows: denature at 96°C, hold for 1 minute, anneal at 64°C, hold for 1 minute. This profile was repeated for 29 cycles, during which the flanking primer pair, RH171 (SEQ ID NO: 1) and RH176 (SEQ ID NO: 2), efficiently annealed and were used in amplification, while the nested primer RH182 (SEQ ID NO: 3), which does not efficiently anneal at 64°C, was not used in efficient amplification. This was followed by denaturation at 96°C, hold for 1 minute, annealing at 52°C, hold for 1 minute. This profile was repeated for 2 cycles, during which all three primers efficiently annealed and were extended in amplification such that products were made using either RH171 (SEQ ID NO: 1) and RH176 (SEQ ID NO: 2), or RH171 (SEQ ID NO: 1) and RH182 (SEQ ID NO: 3). Because the use of a third, nested primer, increases product specificity, products made using RH171 (SEQ ID NO: 1) and RH182 (SEQ ID NO: 3) were more likely to be HIV specific. These cycles were followed by denaturation at 86°C, hold for 1 minute, anneal at 52°C, hold for 1 minute. This profile was repeated 18 times, during which, products that included the GC-rich primer RH176 (SEQ ID NO: 2), both HIV specific and non-specific, did not efficiently denature at 86°C and, therefore, did not amplify efficiently, while the amplified HIV sequences made using the nested primer RH182 (SEQ ID NO: 3) and RH171 (SEQ ID NO: 1) did efficiently denature and amplify.

All eight reactions were analyzed, when completed, by gel electrophoresis and photographed (see Figure 3). Reactions 2-4 and 5-8 were shown to contain a product of the expected size (approximately 200 bp) as the predominant band on the gel. Reactions 1 and 5, the negative controls, contained no such product. However, reactions 2-4, which were not given a "hot-start," could be seen to contain DNA fragments of other than the expected size. These other DNA fragments were also visible in reaction 1, indicating that they are not derived from HIV sequences. These other DNA fragments were not visible in reactions 5-8, indicating that use of the "hot-start" had enhanced the specificity of these reactions.

Eight additional reactions were performed as described above, except that all were given a "hot-start" as described above. The positive control DNA was diluted in these eight reactions, numbered 1 through 8, such that on average, each contains half an HIV target molecule. Since a molecule cannot be divided, this means that some reactions should contain a target molecule and some should not. If this experiment were repeated many times, the fraction of reactions that do contain a target will vary considerably, but should be on average about half. Those that do contain a target molecule are most likely to contain a single target molecule. Upon completion of the reactions, all eight were analyzed by gel electrophoresis and photographed (see Figure 4). The result was that two of the eight reactions, numbers 1 and 8, displayed a DNA fragment of the expected size (approximately 200 bp) as the predominant band on the gel, with no other bands that migrated into the gel visible except a band corresponding to the primers. Reactions 2-6 displayed no such bands nor any other DNA fragment bands due presumably to the absence of the specific HIV template sequence.

Example II

Demonstration of Increased Specificity Via "Drop-in/Drop-out" Protocol

The experiment described in Example I was repeated in such a way as to demonstrate that the thermocycling parameters had the expected effect on amplification specificity.

Nine reactions were set up as described in Example I, except that, on average, ~20 copies of HIV plasmid were present in each reaction mix and each reaction did not contain the nested primer RH182 (SEQ ID NO: 3). In Figure 5, this primer is indicated as primer B, and the flanking primers RH176 (SEQ ID NO: 2) and RH171 (SEQ ID NO: 1), which were included in the reaction mixtures, are indicated as primer A and primer C, respectively. These nine reactions will be referred to here as "A and C" reactions. Nine other reactions were set up in the same way except containing primers RH182 (SEQ ID NO: 3) and RH171 (SEQ ID NO: 1) only. These reactions will be referred to, as "B and C" reactions.

Nine additional reactions were set up in the same way except using all three primers, RH176 (SEQ ID NO: 2), RH182 (SEQ ID NO: 3), and RH171 (SEQ ID NO: 1). These reactions will be referred to as "A, B, and C" reactions.

All reactions were begun using a "hot-start" as described in Example I. Three reactions each of the "A and C," "B and C," and "A, B, and C" groups were amplified using conditions that prevented the "drop-in" primer "B" from priming DNA synthesis. These were amplified in a thermocycler using 49 cycles at a denaturation temperature of 96°C and an annealing temperature of 64°C. This annealing temperature was

above the calculated T_m of primer B (RH182 (SEQ ID NO: 3)). As expected, in the "B and C" group, in which primer B was the only "left" primer present, there was no detectable PCR product when aliquots of the three amplifications were run out on agarose gel electrophoresis as shown in Figure 5. When A and C or A, B, and C were present, product was observed that was of the size expected for HIV sequences bound by primers A and C, as well as other, non-specific product.

Three reactions from each of the three groups were also amplified using conditions that prevented DNA synthesized from primer "A" (RH176 (SEQ ID NO:2)), which contains the GC rich, non-complementary, 5' "tail," from denaturing and amplifying. These conditions were a denaturation temperature of 86 °C and an annealing temperature of 52 °C for 49 cycles. This annealing temperature was low enough to allow annealing and extension of primer B (RH182 (SEQ ID NO: 3)). As expected, no amplification product was seen upon gel electrophoresis of the "A and C" reactions amplified under these conditions. However, amplification products of a size corresponding to HIV sequences bound by primers B and C, as well as many extra, non-specific DNA products, were seen in amplifications "B and C" and "A, B, and C" (see Figure 5).

Finally, three reactions of each of the three groups were amplified under conditions that allowed "drop-in/drop-out" amplifications as diagrammed in Figure 2. These conditions were 29 cycles using a 96 °C denaturation and a 64 °C annealing, two cycles using a 96 °C denaturation and a 52 °C annealing, and 18 cycles using 86 °C denaturation and 52 °C annealing. Only the first two thermocycling phrases allowed utilization of primer A as an efficient amplification primer; these 31 cycles were not enough to detectably amplify the few HIV target molecules present in the sample by gel electrophoresis and ethidium bromide staining. Consequently, as expected, no PCR was product formed with primers A and C only. Also as predicted, the complete "drop-in/drop-out" protocol using all three primers gave the most specific amplification with the greatest yield of product (Figure 5). The only other major band seen upon gel electrophoresis was attributable to the respectively long primer A (RH176 (SEQ ID NO: 2)).

Surprisingly, an increase in specificity and yield was also seen under these amplification conditions using only primers B and C. As discussed above, this was due to inefficient, yet highly specific, amplification during the 29 cycles at an annealing temperature greater than the T_m of primer B. The additional 20 cycles at the lower annealing temperature were sufficient to amplify this inefficiently, yet specifically, produced product.

Sequence Listing

5 INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 bases

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other Nucleic Acid

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCAGGCCAGA TGAGAGAACC AAGGGG 26

INFORMATION FOR SEQ ID NO: 2:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 bases

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

30 GCGGGCAGGG CGGCGGGGGC GGGGCCGAAC CGGTCTACAT AGTCTCTAAA GG 52

INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 17 bases

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: Other Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGTCCCTGTC TTATGTC 17

45

Although the foregoing invention has been described in some detail for the purpose of illustration, it will be obvious that changes and modifications are practical within the scope of the appended claims by those of ordinary skill in the art.

Claims

1. A method for nested amplification of a sequence within a target nucleic acid in a sample, the method comprises:

- 55 (a) mixing said sample in an amplification reaction mixture containing an outer primer pair and an inner primer pair, wherein said outer primer pair is capable of amplifying a segment of said target nucleic acid to provide an amplified target sequence, and said inner primer pair is capable of amplifying a subsequence within said target sequence;

- (b) treating the amplification reaction mixture of step (a) in amplification reaction at a temperature for annealing and extending said outer primer pair on said target nucleic acid and at a temperature for denaturing the extension products of said outer primer pair to provide an amplified target sequence, wherein said temperature for annealing and extending said outer primer pair is higher than the temperature for annealing and extending said inner primer pair to said target nucleic acid; and
- (c) treating the mixture of step (b) in an amplification reaction at said temperature for annealing and extending said inner primer pair on said amplified target sequence, and at a temperature for denaturing the extension products of said inner primer pair to provide an amplified subsequence.
2. The method of claim 1 wherein said amplification reaction mixture contains three primers, where in the first primer is a member of said outer primer pair, the second primer is a member of said outer primer pair and said inner primer pair and the third primer is a member of said inner primer pair.
 3. The method of claim 1 or 2 wherein after step (b) said temperature for annealing and extending said outer primer pair is lowered to provide said temperature for annealing and extending said inner primer pair at step (c).
 4. The method of any one of claims 1-3 wherein said temperature for annealing and extending said inner primer pair is between 1° and 30°C lower than said appropriate temperature for annealing and extending said outer primer pair.
 5. The method of any one of claims 2-4 wherein at step (a) said first primer is present in said amplification reaction mixture at a limiting concentration.
 6. The method of claim 5 wherein said limiting concentration is between 1% and 50% of the concentration of said second and third primers in said amplification reaction mixture.
 7. The method of any one of claims 1-6 wherein at step (c) said temperature for denaturing the extension products of said inner primer pair is suitable for denaturing the extension products of said outer primer pair and the method further comprises:
 - (d) amplifying said subsequence, at an appropriate temperature for annealing and extending said inner primer pair on said amplified target sequence, and at a temperature for denaturing only the extension products of said inner primer pair to provide an amplified subsequence, wherein said temperature for denaturing only the extension products of said inner primer pair is lower than said temperature for denaturing the extension products of said outer primer pair.
 8. The method of claim 7 wherein at step (d) the temperature for denaturing only the extension products of said inner primer pair is from 1° to 20°C lower than the temperature for denaturing the extension products of said outer primer pair.
 9. The method of claim 7 or 8 wherein said first primer comprises a 5' G-C tail.
 10. The method of any one of claims 1-9 that further comprises determining if amplification has occurred.
 11. The method of claim 10 wherein said amplification reaction mixture further comprises a detectable DNA binding agent.
 12. The method of claim 11 wherein said binding agent is ethidium bromide.
 13. The method of claim 11 that comprises, after step (b) and prior to step (c), determining if amplification has occurred.
 14. The method of any one of claims 1-13 wherein said target nucleic acid is present at one copy per 70,000 cells.
 15. The method of any one of claims 1-14 wherein said target nucleic acid is an infectious agent.
 16. The method of any one of claims 1-13 wherein said target nucleic acid is a genetic marker and said

outer primer pair hybridizes to a sequence known to be present.

17. The method of claim 16 wherein said inner primer pair hybridizes to an allele-specific sequence.
- 5 18. The method of claim 17 wherein said allele-specific sequence is indicative of a particular disease state or useful for tissue or blood typing.
19. A method of improved specificity in a PCR that comprises:
- 10 (a) mixing a sample containing a target nucleic acid sequence in an amplification reaction mixture containing a primer pair for specifically amplifying said target;
- (b) amplifying said target sequence at a temperature for annealing and extending said primer pair that is between 1° and 10° C higher than the T_m for said primer pair; and
- (c) amplifying the amplification reaction mixture of step (b) at a temperature for efficiently annealing and extending said primer pair to provide an amplified target sequence.
- 15 20. The method of claim 19 that further comprises determining if amplification has occurred.

20

25

30

35

40

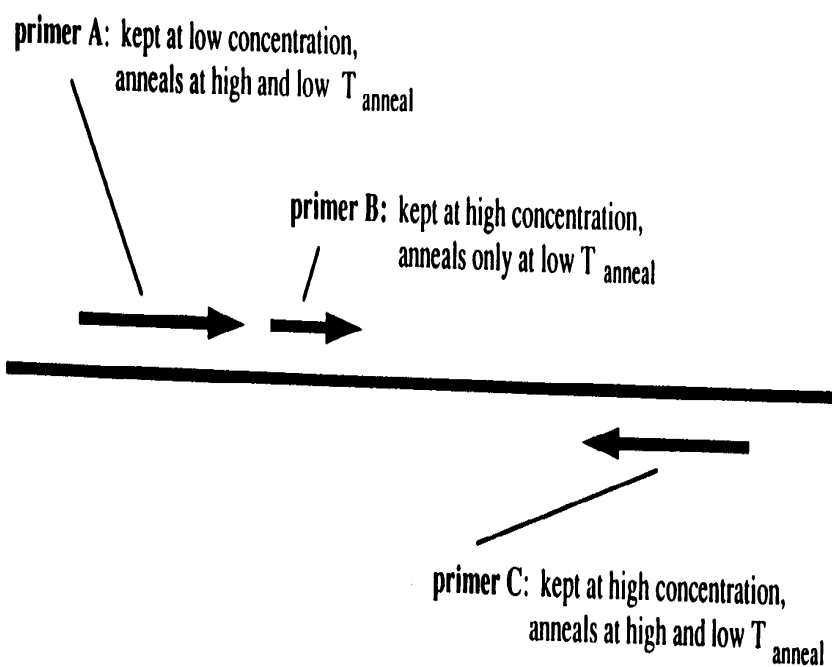
45

50

55

Figure 1 – Primer 'Drop-in' PCR

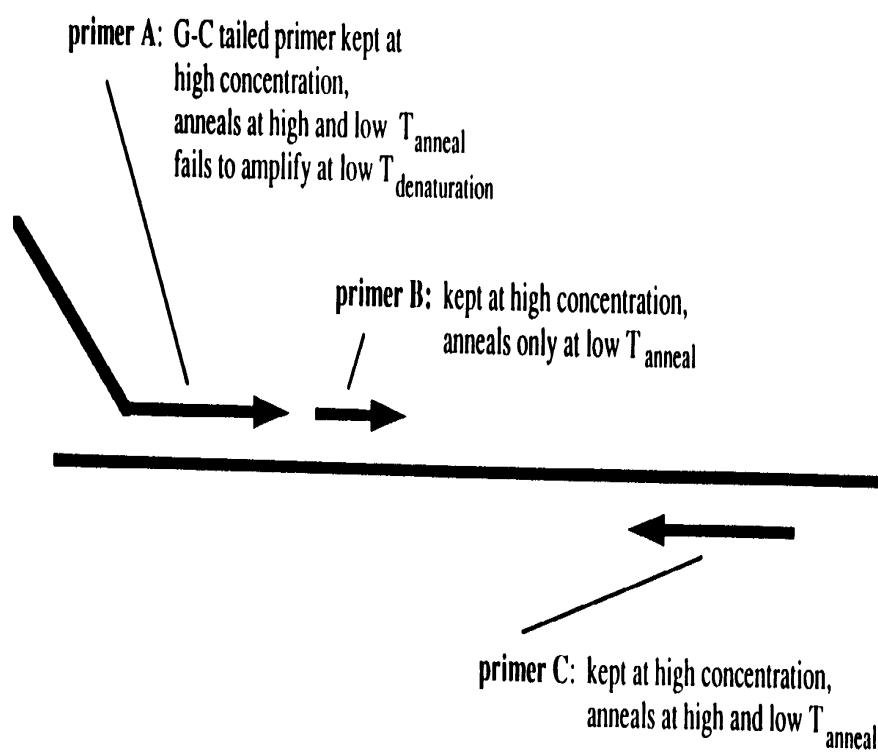
High T_{anneal} PCR $\rightarrow$ Low T_{anneal} PCR



EP 0 519 338 A1

Figure 2- Primer 'Drop-in/Drop-out' PCR

High $T_{\text{denaturation}}$ High T_{anneal} PCR $\longrightarrow$ High $T_{\text{denaturation}}$ Low T_{anneal} PCR $\longrightarrow$
 Low $T_{\text{denaturation}}$ Low T_{anneal} PCR



EP 0 519 338 A1

'Drop-in/Drop-out' nested primer
detection of ~4 HIV copies

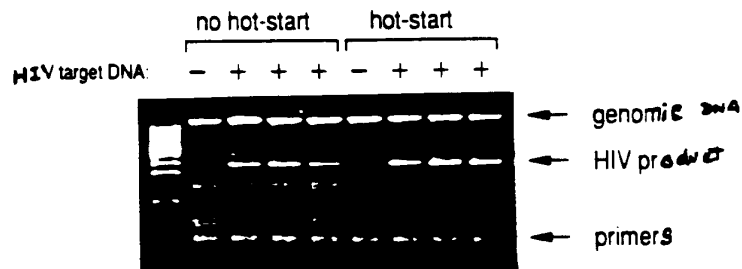


Figure 3

Detection of single HIV copies
using 'drop-out/drop-in'
nested primers and 'hot start'

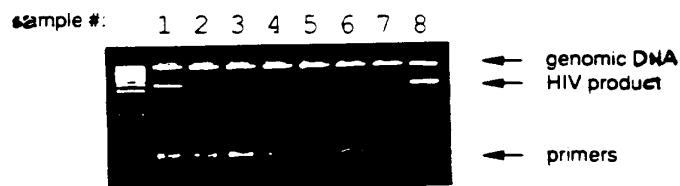


Figure 4

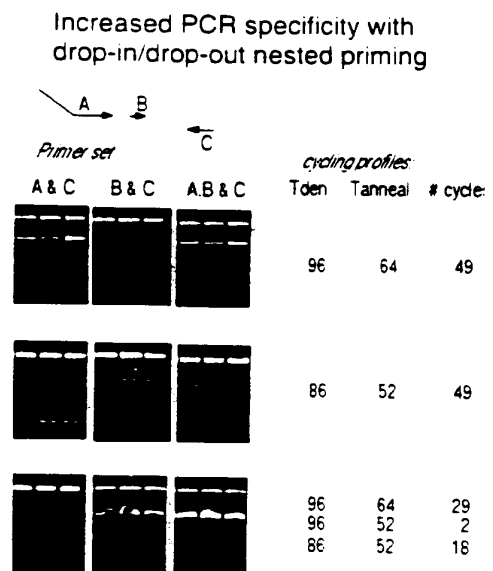


Figure 5



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number

EP 92109886.9

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
D, A	US - A - 4 683 195 (MULLIS et al.) * Abstract; claims 1-16 *	1	C 12 Q 1/68 /C 07 H 21/00
A	CHEMICAL ABSTRACTS, vol. 114, no. 11, March 18, 1991 Columbus, Ohio, USA N. GONZALES-CADAVID et al. "Automated direct sequencing of polymerase chain reaction- amplified fragments of the human Ha-ras gene.", page 190, abstract-no. 96 224x & Anal. Biochem. 1990, 191(2), 359-64	1	
D, A	US - A - 4 965 188 (MULLIS et al.) * Abstract; claims 1-35 *	1	
D, A	US - A - 4 683 202 (MULLIS et al.) * Abstract; claims *	1	TECHNICAL FIELDS SEARCHED (Int. Cl.5)
D, A	US - A - 4 358 535 (FALKOW et al.) * Abstract *	1	C 12 Q C 07 H
The present search report has been drawn up for all claims			
Place of search VIENNA		Date of completion of the search 28-09-1992	Examiner SCHNASS
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03.82 (P0401)



(11) **EP 1 448 581 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
12.11.2008 Bulletin 2008/46

(51) Int Cl.:
C07H 21/02 (2006.01) C12Q 1/68 (2006.01)

(21) Application number: **02763230.6**

(86) International application number:
PCT/US2002/021278

(22) Date of filing: **03.07.2002**

(87) International publication number:
WO 2003/003810 (16.01.2003 Gazette 2003/03)

(54) **METHODS FOR DETECTION AND QUANTIFICATION OF ANALYTES IN COMPLEX MIXTURES**

VERFAHREN ZUM NACHWEIS UND ZUR QUANTIFIZIERUNG VON ANALYTEN IN KOMPLEXEN
MISCHUNGEN

PROCEDES DE DETECTION ET DE QUANTIFICATION D'ANALYTES DANS DES MELANGES
COMPLEXES

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
IE IT LI LU MC NL PT SE SK TR**

(30) Priority: **03.07.2001 US 898743**

(43) Date of publication of application:
25.08.2004 Bulletin 2004/35

(73) Proprietor: **The Institute for Systems Biology
Seattle, WA 98103-8904 (US)**

(72) Inventors:
• **DIMITROV, Krassen
Seattle, WA 98133 (US)**
• **DUNAWAY, Dwayne
Seattle, WA 98115 (US)**

(74) Representative: **UEXKÜLL & STOLBERG
Patentanwälte
Beselerstrasse 4
22607 Hamburg (DE)**

(56) References cited:
**WO-A-99/47643 WO-A-99/62926
WO-A1-99/19515 WO-A2-02/056014
US-A- 5 401 847 US-A- 5 487 973**

- **HIRSCHHORN JOEL N ET AL: "SBE-TAGS: An array-based method for efficient single-nucleotide polymorphism genotyping" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, DC, US, vol. 97, no. 22, 24 October 2000 (2000-10-24), pages 12164-12169, XP002158215 ISSN: 0027-8424**
- **BRENNER SYDNEY ET AL: "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays" NATURE BIOTECHNOLOGY, NATURE PUB. CO, NEW YORK, NY, US, vol. 18, no. 6, June 2000 (2000-06), pages 630-634, XP002199492 ISSN: 1087-0156**
- **COLLINS M L ET AL: "A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 25, no. 15, 1997, pages 2979-2984, XP002208495 ISSN: 0305-1048**
- **CHEN J ET AL: "A MICROSPHERE-BASED ASSAY FOR MULTIPLEXED SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS USING SINGLE BASE CHAIN EXTENSION" GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, WOODBURY, NY, US, vol. 10, no. 4, April 2000 (2000-04), pages 549-557, XP000927257 ISSN: 1088-9051**
- **TONG ANTHONY K ET AL: "Combinatorial fluorescence energy transfer tags for multiplex biological assays" NATURE BIOTECHNOLOGY, NATURE PUB. CO, NEW YORK, NY, US, vol. 19, no. 8, August 2001 (2001-08), pages 756-759, XP002193783 ISSN: 1087-0156**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

- BONETTA L: "Gene expression: One size does not fit all" NATURE METHODS 2006 UNITED KINGDOM, vol. 3, no. 5, 2006, pages 401-409, XP008070014 ISSN: 1548-7091 1549-1676

Description**BACKGROUND OF THE INVENTION**

- 5 **[0001]** This invention relates generally to the field of genomics and, more specifically to detection, identification, and quantification of target analytes in mixtures.
- 10 **[0002]** Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression are at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in normal cells compared to diseased cells can aid the discovery of new drug and diagnostic targets.
- 15 **[0003]** Nucleic acids can be detected and quantified based on their specific polynucleotide sequences. The basic principle underlying existing methods of detection and quantification is the hybridization of a labeled complementary probe sequence to a target sequence of interest in a sample. The formation of a duplex indicates the presence of the target sequence in the sample and the degree of duplex formation, as measured by the amount of label incorporated in it, is proportional to the amount of the target sequence.
- 20 **[0004]** This technique, called molecular hybridization, has been a useful tool for identifying and analyzing specific nucleic acid sequences in complex mixtures. This technique has been used in diagnostics, for example, to detect nucleic acid sequences of various microbes in biological samples. In addition, hybridization techniques have been used to map genetic differences or polymorphisms between individuals. Furthermore, these techniques have been used to monitor changes in gene expression in different populations of cells or in cells treated with different agents.
- 25 **[0005]** In the past, only a few genes could be detected in a complex sample at one time. However, DNA microarrays, devices that consist of thousands of immobilized DNA sequences present on a miniaturized surface, have made this process more efficient. Using a microarray, it is possible in a single experiment to detect the presence or absence of thousands of genes in a biological sample. This allows researchers to simultaneously perform several diagnostic tests on one sample, or to observe expression level changes in thousands of genes in one experiment. Generally, microarrays are prepared by binding DNA sequences to a surface such as a nylon membrane or glass slide at precisely defined locations on a grid. Then nucleic acids in a biological sample are labeled and hybridized to the array. The labeled sample DNA marks the exact position on the array where hybridization occurs, allowing automatic detection.
- 30 **[0006]** Unfortunately, despite the miniaturization of array formats, this method still requires significant amounts of the biological sample. However, in several cases, such as biopsies of diseased tissues or samples of a discrete cell type, the biological sample is in limited supply. In addition, the kinetics of hybridization on the surface of a microarray is less efficient than hybridization in small amounts of aqueous solution. Furthermore, microarrays require a large dynamic range of detection to account for large difference in abundance of the different molecular species. This results in decreased sensitivity since there is a trade-off between sensitivity and dynamic range. A further problem with microarray methods is that the output is quantitative analog data that has undergone several intermediary transformations. In microarrays, the amount of nucleic acid hybridized to each spot is determined by measuring its label and so any nonlinear correlation between the amount of DNA hybridized and the amount of the label detected will skew the data output. Such non-linearity has been widely documented.
- 35 **[0007]** Collins et al. (Nucleic Acids Research 25(15): 2979- 2984 (1997)) disclose target-specific oligonucleotides called label extenders which include a non-target extension which bind to a preamplifier which in turn binds to many amplifiers so as to generate signal amplification; the same label is bound to all label extenders, wherein detection is based on immobilization (localization by immobilization). The method disclosed by Collins et al. does not provide for simultaneous (multiplexing) signal amplification for direct detection of nucleic acids, in particular when present in complex mixtures.
- 40 **[0008]** Thus, there exists a need for accurate and sensitive detection, identification and quantification of analytes in complex mixtures. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

- 50 **[0009]** The invention provides a diverse population of uniquely labeled probes, containing about thirty or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid. Also provided is a method of detecting a nucleic acid analyte. The method consists of (a) contacting a mixture of nucleic acid analytes with the diverse population of uniquely labelled probes of the invention under conditions sufficient for hybridization of the probes to target nucleic acid analytes; (b) measuring the resulting signal from one or more of the target specific probes hybridized to an analyte, where the signal uniquely identifies the analyte.
- 55

BRIEF DESCRIPTION OF THE DRAWINGS**[0010]**

5 Figure 1 shows components of a specifier. Figure 1A shows an association between a target and a labeled specifier. Figure 1B shows a structure of a genedigit. Figure 1C shows an association between a genedigit and labeled anti-genedigit.

10 Figure 2 shows a DNA probe molecule (specifier) that contains several different types of spatially separated labels.

Figure 3 shows DNA probe molecules containing a label on one end that have been stretched on a coverslip using a flow-stretch technique.

15 Figure 4 shows labeled DNA molecules that have been aligned using an electro-stretch technique. Figure 4A shows labeled DNA probe molecules in a flow cell with the voltage off. Figure 4B shows the labeled DNA probe molecule in the flow cell with the voltage on.

Figure 5 shows a schematic of probes being aligned by constriction in the flow of a liquid containing the probes in the presence of an oscillating electric field.

20

DETAILED DESCRIPTION OF THE INVENTION

25 **[0011]** This invention is directed to a diverse population of uniquely labeled probes, comprising thirty or more target-specific nucleic acid probes, each attached to a unique label bound to a nucleic acid, each nucleic acid probe being a nucleic acid molecule comprising a target-specific region and a region comprising genedigits, said genedigits being nucleic acid regions of predetermined nucleotide sequences that are specifically bound by complementary anti-genedigit nucleotide sequences, each said anti-genedigit nucleotide sequence being attached to a label monomer or a combination of label monomers, wherein the population is in solution, and wherein the label has a detectable signal that is fluorescence, luminescence, colorimetric, voltage, current, or a magnetic field can be used for the detection, identification, and direct quantification of a wide variety of target analytes. The diverse population is advantageous in that it provides a large number of unique labels of about the same unit signal starting from just a small number of different labels. Enough labels are provided so that each analyte in a complex mixture can be uniquely bound by a label and thus identified. The labels are designed so that they can be used in a small volume of solution which increases the efficiency of the binding reaction and is useful when only small amounts of sample are available for analysis. After the individual molecules in a sample are identified, they can be directly counted resulting in a digital read-out of each molecular species in a mixture. Since the diverse labels are based on a small number of starting labels, this allows the detection method to operate in a narrow dynamic range resulting in improved sensitivity of the system since the trade-off between sensitivity and dynamic range is avoided. The methods of the invention therefore provide an accurate and sensitive system for the detection, identification and quantitation of analytes in a mixture.

30 **[0012]** In one embodiment, the invention is directed to detecting nucleic acid analytes in a complex mixture by first contacting the mixture with the diverse population of probes of the invention under conditions sufficient for hybridization of the probes to target nucleic acid analytes. These target specific nucleic acid probes, called specifiers, contain a target specific region and a region containing one or more unique "genedigit" sequences. The genedigits are nucleic acid regions of predetermined nucleotide sequences that are specifically bound by complementary anti-genedigit nucleotide sequences, each said anti-genedigit nucleotide sequence being attached to a label monomer or a combination of label monomers, wherein the label has a detectable signal that is fluorescence, luminescence, colorimetric, voltage, current, or a magnetic field.

35 **[0013]** As used herein, the term "bound" when referring to a unique label or nucleic acid is intended to mean that a label monomer is attached to a nucleotide in a 1:1 correspondence. A label monomer as used herein is intended to mean an individual measurable moiety, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other moiety known in the art that is measurable by analytical methods. A label monomer can be directly attached to a nucleotide using methods well known in the art. Nucleotides can also be chemically modified or derivitized in order to attach a label monomer. For example, a fluorescent monomer such as a fluorescein molecule can be attached to dUTP (deoxyuridine-triphosphate) using a four-atom aminoalkynyl group. Each label monomer is attached to a nucleotide making a label monomer:nucleotide complex. This label monomer:nucleotide can be incorporated into nucleic acids in a variety of ways. For example, a label monomer:nucleotide can be incorporated at only one location within a nucleic acid or at two or more locations within a nucleic acid. A nucleotide can be attached to a label monomer first and then be incorporated into a nucleic acid, or an existing nucleic acid can be labeled by attaching a label monomer

55

to a nucleotide within the nucleic acid. In addition, for example, a label monomer:nucleotide can be incorporated into a nucleic acid and a different type of label monomer:nucleotide can be incorporated into the same nucleic acid.

[0014] As used herein, "analyte" or target when referring to an analyte is intended to mean any nucleic acid molecule whose presence is measured. A nucleic acid analyte molecule can be essentially any nucleic acid molecule for which a detectable probe or assay exists or can be produced by one skilled in the art. Measurement can be quantitative or qualitative. An analyte can be part of a sample that contains other components or can be the sole or major component of the sample. Therefore, an analyte can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The analyte can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also the analyte can have either a known or unknown structure or sequence.

[0015] As used herein, the term "target specific" is intended to mean an agent that binds to a target analyte selectively. This agent will bind with preferential affinity toward the target while showing little to no detectable cross-reactivity toward other molecules. For example, when the target is a nucleic acid, a target specific sequence is one that is complementary to the sequence of the target and able to hybridize to the target sequence with little to no detectable cross-reactivity with other nucleic acid molecules. A nucleic acid target could also be bound in a target specific manner by a protein, for example by the DNA binding domain of a transcription factor.

[0016] The term "complementary" refers to two nucleotides that can form multiple thermodynamically favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if the nucleotides of the first sequence are complementary to the nucleotides of the second sequence. The percent of complementarity (i.e. how many nucleotides from one strand form multiple thermodynamically favorable interactions with the other strand compared with the total number of nucleotides present in the sequence) indicates the extent of complementarity of two sequences.

[0017] As used herein, the term "repeat sequence" is intended to mean two or more copies of a core element. A repeat sequence can have direct repetition of the core sequence without any intervening sequence, or the repeat sequence can have non-consecutive repetition of the core element with intervening sequences. A core element can be made of nucleic acids such as an oligonucleotide or an aptamer. If, for example, the core element is a nucleic acid sequence of 8 base pairs then three direct repeats of this sequence would be a 24 base sequence. A "complimentary repeat sequence" is a sequence that binds specifically to the repeat sequence. For the example above where the repeat sequence is repetition of a nucleic acid core element, the complimentary repeat sequence can contain one or more copies of the complementary strand of the core element that will hybridize specifically to the repeat sequence.

[0018] As used herein, the term "genedigit" is intended to mean a region of pre-determined nucleotide sequence that serves as an attachment point for a label. The genedigit can have any structure including, for example, a single unique sequence or a sequence containing repeated core elements. Each genedigit has a unique sequence which differentiates it from other genedigits. An "anti-genedigit" is a nucleotide sequence or structure that binds specifically to the gene digit. For example, if the genedigit is a nucleic acid, the anti-genedigit can be a nucleic acid sequence that is complementary to the genedigit sequence. If the genedigit is a nucleic acid that contains repeated core elements then the anti-genedigit can be a series of repeat sequences that are complementary to the repeat sequences in the genedigit. An anti-genedigit can contain the same number, or a lesser number, of repeat sequences compared to the genedigit as long as the anti-genedigit is able to specifically bind to the genedigit.

[0019] As used herein, the term "specifier" is intended to mean the linkage of one or more genedigits to a target specific sequence. The genedigits can be directly linked or can be attached using an intervening or adapting sequence. A specifier can contain a target specific sequence which will allow it to bind to a target analyte. An "anti-specifier" has a complementary sequence to all or part of the specifier such that it specifically binds to the specifier.

[0020] As used herein, the term "mixture" is intended to mean a composition that contains more than one molecule. A mixture can be homogenous, containing a single species, or heterogeneous, containing different species. Examples of homogeneous samples include, for example, isolated populations of nucleic acids. Heterogeneous mixtures include extracts from tissues, cells, lysates and fractionated portions thereof.

[0021] As used herein, the term "label" is intended to mean a molecule or molecules that render an analyte detectable by an analytical method. An appropriate label depends on the particular assay format and are well known by those skilled in the art. For example, a label specific for a nucleic acid molecule can be a complementary nucleic acid molecule attached to a label monomer or measurable moiety, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other moiety known in the art that is measurable by analytical methods. In addition, a label can include any combination of label monomers.

[0022] As used herein, "unique" when used in reference to label is intended to mean a label that has a detectable signal that distinguishes it from other labels in the same mixture. Therefore, a unique label is a relative term since it is dependent upon the other labels that are present in the mixture and the sensitivity of the detection equipment that is used. In the case of a fluorescent label, a unique label is a label that has spectral properties that significantly differentiate

it from other fluorescent labels in the same mixture. For example, a fluorescein label can be a unique label if it is included in a mixture that contains a rhodamine label since these fluorescent labels emit light at distinct, essentially non-overlapping wavelengths. However, if another fluorescent label was added to the mixture that emitted light at the same or very similar wavelength to fluorescein, for example the Oregon Green fluorophore, then the fluorescein would no longer be a unique label since Oregon Green and fluorescein could not be distinguished from each other. A unique label is also relative to the sensitivity of the detection equipment used. For example, a FACS machine can be used to detect the emission peaks from different fluorophore-containing labels. If a particular set of labels have emission peaks that are separated by, for example, 2 nm these labels would not be unique if detected on a FACS machine that can distinguish peaks that are separated by 10nm or greater, but these labels would be unique if detected on a FACS machine that can distinguish peaks separated by 1 nm or greater.

[0023] As used herein, the term "signal" is intended to mean a detectable, physical quantity or impulse by which information on the presence of an analyte can be determined. Therefore, a signal is the read-out or measurable component of detection. A signal includes, for example, fluorescence, luminescence, colorimetric, density, image, sound, voltage, current, magnetic field and mass. Therefore, the term "unit signal," as used herein is intended to mean a specified quantity of a signal in terms of which the magnitudes of other quantities of signals of the same kind can be stated. Detection equipment can count signals of the same type and display the amount of signal in terms of a common unit. For example, a nucleic acid can be radioactively labeled at one nucleotide position and another nucleic acid can be radioactively labeled at three nucleotide positions. The radioactive particles emitted by each nucleic acid can be detected and quantified, for example in a scintillation counter, and displayed as the number of counts per minute (cpm). The nucleic acid labeled at three positions will emit about three times the number of radioactive particles as the nucleic acid labeled at one position and hence about three times the number of cps will be recorded.

[0024] As used herein, the term "dendrimer" is intended to mean a branched nucleic acid. These structures are composed of layers of nucleic acid, each layer being composed of partially single-stranded heteroduplexes which are called dendrimer monomers. The outermost layer of a given dendrimer can have multiple single-stranded arms capable of hybridization with a complementary nucleic acid sequence. Dendrimer monomers have the property that sequential addition of monomers yields a three-dimensional structure composed of nucleic acid. Various configurations of nucleic acid molecules can give rise to a large number of differently shaped dendritic structures. For example, using commercially available dendrimeric synthons, a dendrimer with 1 stem and 81 branches can be synthesized. Also fork-like, comb-like and bubbled structures are possible.

[0025] As used herein, the term "nucleic acid" is intended to mean a single- or double-stranded DNA or RNA molecule including, for example, genomic DNA, cDNA and mRNA. The term is intended to include nucleic acid molecules of both synthetic and natural origin. A nucleic acid molecule of the invention can be of linear, circular or branched configuration, and can represent either the sense or antisense strand, or both, of a native nucleic acid molecule. A nucleic acid molecule of the invention can further incorporate a detectable moiety such as a radiolabel, a fluorochrome, a ferromagnetic substance, a luminescent tag or a detectable moiety such as biotin.

[0026] As used herein, the term "hybridizing" is intended to mean joining different components together. Any number of components can be joined together, for example two components can be joined together to make a duplex, three components can be joined together to make a triplex, and so on. Nucleic acids can form a hybrid or duplex, for example, by hydrogen bonding between complementary nucleotides. The formation of nucleic acid hybrids is dependent on several conditions known in the art, including temperature, salt concentration, and pH.

[0027] As used herein, "complexity" refers to the degree of repeated elements between two nucleic acids that are being hybridized together in a solution (see Anderson, M.L.M., Nucleic Acid Hybridization, Springer-Verlag, New York (1999)). When the nucleic acid molecules that are to be hybridized contain repeated core elements or homopolymeric regions, there are many possible pairing opportunities and so the hybridization proceeds quickly. When the nucleic acid molecules that are to be hybridized do not contain any repeated core elements then there is only one way that the two sequences can be hybridized and so the hybridization proceeds more slowly. Sequences that hybridize quickly are said to have a low complexity, while sequences that take longer to hybridize have a higher complexity. For example, a 40 base pair genedigit sequence made up of five direct repeats of an 8 base pair core element, can be hybridized to a 24 base pair anti-genedigit containing three repeats of the 8 base pair core repeat in three different registers (see, for example, Figure 1C). Thus the anti-genedigit can hybridize to the 40 base pair genedigit through a 24 base pair sequence that only has a complexity of an 8 base pair sequence.

[0028] The invention provides a diverse population of uniquely labeled probes providing a large number of unique labels of about the same unit signal starting from just a small number of different labels. Enough labels are provided so that each analyte in a complex mixture can be uniquely bound by a label and thus identified. These labels can be used in very small volumes which improves kinetics in the binding reaction. In addition, the design of these labels allows for improved accuracy in detection.

[0029] A large diversity of unique labels can be desirable in order to provide a unique label to each species in a complex mixture. The invention provides methods for combining different labels in pre-determined ratios to generate a large

diversity of unique labels. The labels are designed in a modular fashion which allows for flexibility in the number of unique labels that are generated. For example, if a large number of modules are used, a large number of ratios of the different labels is possible which leads to a large number of unique labels. The number of labels generated can be adjusted to cover mixtures with different numbers of species.

[0030] The invention provides a diverse population of labels that contains thirty or more unique labels where each unique label is bound to a nucleic acid. A diverse population of labels is a mixture of distinct label species. This population can have as few as about thirty distinct label species or as high as 10^{17} distinct label species. The actual number of molecules of each label species can vary as long as at least one molecule of the label species is present. In addition, the invention provides a diverse population of labels that contains 40, 60, 80, 100, 120, 140, or about 150 unique labels.

A portion of this population can be made up of different individual label monomers. The invention also provides unique labels made from combinations of different labels which can increase the number of unique labels substantially.

[0031] The labels of the invention are bound to nucleic acids. In particular, the labels are bound to nucleic acids through the attachment of a label monomer to a nucleotide within a nucleic acid in a 1:1 correspondence. A nucleic acid can contain several label monomers, however each label monomer is attached directly to a nucleotide.

[0032] A label monomer can be attached to any nucleotide including both natural and non-natural nucleotides. A nucleotide contains three parts, a phosphate group, a pentose five-carbon sugar molecule, and an organic base. In RNA, the pentose is ribose and in DNA it is deoxyribose and so nucleotides for incorporation into RNA are called ribonucleotides and nucleotides for incorporation into DNA are called deoxyribonucleotides. Three bases adenine, guanine, and cytosine are found in both DNA and RNA while thymine is normally found only in DNA and uracil is normally found only in RNA. Nucleotides can have one, two or three attached phosphate groups and are sometimes referred to as nucleoside phosphates. Nucleotides can contain modified nucleosides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl ribosyl, 2'-O-methoxyethyl ribosyl, 2'fluoro ribosyl, 2'amino ribosyl, and the like). An example of non-natural bases that are used in the art are isocytidine and isoguanine.

[0033] A label monomer as used herein is intended to mean an individual measurable moiety, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other moiety known in the art that is measurable by analytical methods. A label monomer can be attached to a nucleotide using methods well known in the art and exemplified herein.

[0034] Radioisotopes are an example of label monomers that can be utilized by the invention. Several radioisotopes can be used as label monomers for labeling nucleotides including, for example, ^{32}P , ^{33}P , ^{35}S , ^3H , and ^{125}I . These radioisotopes have different half-lives, types of decay, and levels of energy which can be tailored to match the needs of a particular experiment. For example, ^3H is a low energy emitter which results in low background levels, however this low energy also results in long time periods for autoradiography. Radioactively labeled ribonucleotides and deoxyribonucleotides are commercially available. Nucleotides are available that are radioactively labeled at the first, or α , phosphate group, or the third, or γ , phosphate group. For example, both $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ and $[\gamma\text{-}^{32}\text{P}]\text{dATP}$ are commercially available. In addition, different specific activities for radioactively labeled nucleotides are also available commercially and can be tailored for different experiments.

[0035] Another example of label monomers that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, and Texas Red. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum. Also different formulations of the same fluorophore have been produced for different applications. For example, fluorescein, can be used in its isothiocyanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or as isomeric dichlorotriazine forms of fluorescein (DTAF). These labels are chemically distinct, but all emit light with a peak between 515-520 nm. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein. In addition, different formulations or related dyes have been developed around other fluorophores that emit light in other parts of the spectrum.

[0036] Amine-reactive and thiol-reactive fluorophores are available and used for labeling nucleotides and biomolecules. Generally, nucleotides are fluorescently labeled during chemical synthesis, for example, incorporation of amines or thiols during nucleotide synthesis permit addition of fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum.

[0037] Fluorescent dyes that can be bound directly to nucleotides can also be utilized as label monomers. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These fluorescently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-ddGTP and ROX-ddUTP, are commercially available.

[0038] Non-radioactive and non-fluorescent label monomers are also available. For example, biotin can be attached

directly to nucleotides and detected by specific and high affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin labeled nucleotides can also similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

[0039] Very small particles, termed nanoparticles, also can be used as label monomers to label nucleic acids. These particles range from 1 - 1000nm in size and include diverse chemical structures such as gold and silver particles and quantum dots.

[0040] When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle.

Four to five different particles in close proximity will each scatter monochromatic light which when superimposed will give a specific, unique color. The particles are being manufactured by companies such as Genicon Sciences. Derivatized silver or gold particles can be attached to a broad array of molecular probe molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivitized to allow attachment to a nucleotide.

[0041] Another type of nanoparticle that can be used as a label monomer are quantum dots. Quantum dots are fluorescing crystals 1-5nm in diameter that are excitable by a large range of wavelengths of light. These crystals emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties. These particles have been used in the semiconductor industry for several years, but are just now being applied to molecular biology applications.

[0042] Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to desired size and/or emission wavelengths. Use of quantum dots for labeling particles, in the context of the present invention, is new, but is old in the art of semiconductors. Two examples of earlier references in which quantum dots are embedded within intrinsic silicon epitaxial layers of semiconductor light emitting/ detecting devices are United States Patent Nos. 5,293,050 and 5,354,707 to Chapple-Sokol, et al.

[0043] Due to their very small size the quantum dots can be coupled into oligonucleotides directly without affecting the solubility or use of the oligonucleotide. The invention requires that only one oligonucleotide molecule is coupled to each nanoparticle. To synthesize an oligonucleotide-nanoparticle complex in a 1:1 ratio by conventional batch chemistry, both the oligonucleotide and the nanoparticle require a single reactive group of different kinds that can be reacted with each other. For example, if an oligonucleotide has an amino group and a nanoparticle has an aldehyde group, these groups can react to form a Schiff base. An oligonucleotide can be derivitized to attach a single amino or other functional group using chemistry well known in the art. However, when a nanoparticle is derivitized, it is covered with a chemical reagent which results in coating the entire surface of the nanoparticle with several functional groups.

[0044] In this context, a method of coupling one oligonucleotide to one nanoparticle by chemically coupling the oligonucleotide on a solid surface such as the glass support used for the oligonucleotide synthesis is provided. For example, commercially available resins for oligonucleotide synthesis such as long chain alkylamino controlled pore glass (lcaa CPG) can be used. Alternatively, a flat surface such as a derivitized microscope slide can be used. The surface density of the nascent oligonucleotide chains should be lower than the diameter of the nanoparticle. This can be achieved by either choosing a glass support with low surface density of the reactive groups, or by using diluted reagent for the first step of the oligonucleotide synthesis so that the surface is not saturated. Another point of consideration when using the standard glass matrices for oligonucleotide synthesis is to use a pore diameter higher than the nanoparticle diameter to ensure the flow of the reagents. For example, an oligonucleotide can be synthesized on a diluted basis relative to the solid support, for example one tenth of a normal synthesis, to ensure good spacing of the oligonucleotides on the glass support. After the oligonucleotide is synthesized with a reactive functional group, for example, an amino group, derivitized nanoparticles are passed over the glass support to react with the oligonucleotides. A sufficiently large pore size of the glass support can be chosen to prevent clogging with nanoparticles. For example, a pore size of about 200 nm can be used. After the reaction is complete, un-reacted groups on the nanoparticle can be blocked and the complexes can be uncoupled from the glass support.

[0045] The labels of the invention are bound to nucleic acids through nucleotides in the nucleic acid. A nucleotide can be attached to a label monomer first and then the label monomer:nucleotide can be incorporated into a nucleic acid, or an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid.

[0046] A label monomer can be attached to a nucleotide using a variety of methods well known in the art and described herein. For example, the label monomer can be directly attached to the nucleotide in a 1:1 correspondence by incorporation of a radioactive phosphate into the phosphate backbone of the nucleotide. Also, for example, a general method for labeling phosphates with a fluorescent label that employs an imidazole derivative prepared from a BODIPY FL hydrazide has been reported (Wang and Giese, Anal. Chem. 65: 3518 (1993)).

[0047] Depending on the labeling moiety used, it can be desirable to derivitize or chemically modify a nucleotide in

order to bind the label monomer. These methods and chemistries are known in the art. In addition, a linker can be used to attach a label monomer to a nucleotide in a 1:1 correspondence. For example, a fluorescently labeled nucleotide such as fluorescein-12-dUTP can have a fluorophore monomer attached via a four-atom aminoalkynyl group to the dUTP molecule.

5 [0048] These nucleotides attached to label monomers can be incorporated into a nucleic acid using several methods for labeling nucleic acids well known in the art. For example, enzymes such as DNA or RNA polymerases, Taq polymerases, terminal deoxynucleotidyl transferases, or reverse transcriptases can be used to incorporate labeled nucleotides into nucleic acids.

10 [0049] Labeled nucleotides can be incorporated into nucleic acids, for example, by nick translation. In this procedure DNase I is used to create single-strand nicks in double stranded DNA and then the 5' to 3' exonuclease and 5' to 3' polymerase actions of E. coli DNA polymerase I are used to remove stretches of single stranded DNA starting at the nicks and replace them with new strands made by incorporation of labeled nucleotides. Nick translation can utilize any labeled nucleotide including radioactively labeled nucleotides and biotinylated or digoxigenin labeled nucleotides. In a similar way T4 DNA polymerase can be used to incorporate labeled nucleotides. In addition, labeled nucleotides can be incorporated into nucleic acids using the polymerase chain reaction (PCR) and Taq polymerases. The degree of labeling can be controlled by including one, or up to all four labeled nucleotides. In addition, the degree of labeling can be controlled by increasing or decreasing the concentration of the labeled nucleotide(s).

15 [0050] Other methods for labeling nucleic acids include generating single-stranded cDNA from RNA by using a reverse transcriptase in the presence of labeled nucleotides. In addition, DNA can be cloned into a vector with SP6 or T7 polymerase sites. Transcription in the presence of SP6 or T7 RNA polymerase and labeled nucleotides results in a labeled RNA transcript. The transcript can be labeled to different degrees by including one or more labeled nucleotides. In addition, several nucleotides within a nucleic acid can be labeled, for example, by cloning DNA into a bacteriophage M13 based vector. Then the Klenow fragment of DNA polymerase I and the M13 universal probe primer can be used to synthesize the complementary strand with incorporation of labeled nucleotides.

20 [0051] Several methods are described above for incorporation of labeled nucleotides into newly synthesized nucleic acids. Existing nucleic acids can also be labeled using several methods known in the art. For example, RNA or DNA can be end-labeled with [γ -³²P]ATP and T4 polynucleotide kinase. This kinase can be used to transfer the radioactive phosphate of ATP to a free 5' OH group in either DNA or RNA. The enzyme also has a phosphatase activity and so two reactions are possible. In the forward reaction, the enzyme catalyzes phosphorylation following removal of 5' terminal phosphates with alkaline phosphatase (or other phosphatase). In the exchange reaction, the kinase catalyzes the exchange of an existing 5' phosphate with the third or γ phosphate of ATP. The latter reaction is carried out in the presence of excess ATP and ADP for efficient phosphorylation. Using this method the radioactive phosphate of ATP is transferred to the end of the nucleic acid molecule.

25 [0052] Nucleic acids can also be labeled with terminal deoxynucleotidyl transferase which adds labeled nucleotides onto the 3' end of DNA fragments. Both single and double-stranded DNAs are substrates for this enzyme. The large (Klenow) fragment of E. coli DNA polymerase I can also be used to label the ends of nucleic acids. Since this enzyme has a 5' to 3' polymerase activity it can be used to "fill in" the 3' ends of DNA fragments opposite of 5' extensions or overhangs with labeled nucleotides. End-labeling of nucleic acids using polynucleotide kinase or terminal deoxynucleotidyl transferase results in the incorporation of one label per nucleic acid. The "fill in" reaction can be used to label the nucleic acid at one nucleotide per nucleic acid or at more than one nucleotide per nucleic acid.

30 [0053] In addition, nucleic acids can be labeled by modification of nucleotides within the nucleic acid. For example, cytidine residues in DNA and RNA can be modified by reaction with sodium bisulfite to form sulfonate intermediates that can then be directly coupled to hydrazides or aliphatic amines. Virtually any of the fluorescent, biotin or other hydrazides or aliphatic amines can be used in this reaction. The bisulfite-activated cytidylic acid can also be coupled to aliphatic diamines such as ethylenediamine. The amine-modified DNA or RNA can then be modified with any of the amine-reactive dyes. In addition, phosphate groups can be targeted in nucleic acids for labeling. Although phosphate groups of nucleotides are not very reactive in aqueous solution, their terminal phosphate groups can react with carbodiimides and similar reagents in combination with nucleophiles to yield labeled phosphodiester, phosphoramidate and phosphorothioate. For example, DNA can be reacted quantitatively with carbonyl diimidazole and a diamine such as ethylenediamine to yield a phosphoramidate that has a free primary amine and that this amine can then be modified with amino-reactive reagents. Fluorescent or biotinylated amines have been coupled to the 5' phosphate of tRNA using dithiodipyridine and triphenylphosphine.

45 [0054] The bond between labels and nucleic acids can be covalent bonds or non-covalent bonds that are stable to hybridization and washing conditions. The labels can be bound to a nucleic acid in a sequence specific manner, for example by the incorporation of a labeled nucleotide into DNA that has been digested by a restriction enzyme. Alternatively the labels can be bound to a nucleic acid in a non-sequence specific manner, for example by the incorporation of a label onto the terminal phosphate of a nucleic acid using [γ -³²P]ATP and T4 polynucleotide kinase.

50 [0055] Several types of nucleic acids can be utilized with this invention including a single- or double-stranded DNA or

RNA molecule which can include, for example, genomic DNA, cDNA and mRNA. Nucleic acid molecules of both synthetic and natural origin can be used. In connection with the present invention, a nucleic acid molecule can be of linear, circular or branched configuration, and can represent either the sense or antisense strand, or both, of a native nucleic acid molecule. Nucleotide bases not found in nature such as isocytidine and isoguanine can be incorporated into the nucleic acid.

[0056] Labels can bind to nucleic acids in a number of diverse ways. For example, a particular label monomer can be bound to a nucleic acid at only one position in the nucleic acid or at many positions in the nucleic acid. In addition, a particular label monomer can be bound to a nucleic acid and one or more other label monomers can also bound to the same nucleic acid. In this case the label can contain a mixture of two or more different labels. Furthermore, nucleic acids labeled with any or all of these combinations can be bound to another nucleic acid through hybridization.

[0057] Additional diversity is introduced when the nucleic acid is branched. One example of a branched nucleic acid is a dendrimer. Dendrimers are composed of layers of nucleic acid, each layer being composed of partially single-stranded heteroduplexes yielding a three-dimensional structure composed of nucleic acid. Various configurations of nucleic acid molecules can give rise to a large number of differently shaped dendritic structures including, for example, a dendrimer with 1 stem and 81 branches, or dendrimers with fork-like, comb-like or bubbled structures. The outermost layer of a given dendrimer can have multiple single-stranded arms capable of hybridization with a complementary nucleic acid sequence. Due to the relatively large size of nucleic acid molecules, nucleic acid dendrimers can contain numerous labels with limited steric hindrance. Use of dendrimers can multiply the signal generated by a labeled nucleic acid by a pre-determined factor equal to the number of branches.

[0058] Several unique combinations of labels can be formed using branched nucleic acids. For example, by using different chemical protective groups, one label monomer can bind to one branch while one or more other label monomers bind to other branches. Furthermore, labeled nucleic acids can be attached in various combinations to the branches of a dendrimer through hybridization.

[0059] The invention provides a diverse population of 30, or about 40, 60, 80, 100, 120, 140, or 150 unique labels bound to a nucleic acid. Part of this population can be made up of different individual label monomers. The invention also provides unique labels made from a combination of two or more different labels. This can increase the number of unique labels substantially.

[0060] A unique label is a label that generates a signal that is distinguishable from other labels in the same mixture. Therefore, designation as a unique label is dependent upon the sensitivity of the detection equipment that is used. For example, where fluorescent or nanoparticle labels are used a CCD camera can be used to detect the labels. The sensitivity of this equipment depends on the manufacturer, model, and design of the equipment. In addition, several parameters can be set by the user in order to achieve maximum sensitivity. For example, the use of different filter sets can increase the sensitivity of detection for certain experiments.

[0061] The ability to distinguish different labels also depends on the particular properties of the label. For example, some fluorophores emit light within a broad peak or range of wavelengths while other fluorophores emit light within a narrow peak. Fluorophores that emit light within a broad peak can obscure neighboring peaks. In addition, the shapes of the emission peaks from different fluorophores varies. For example, some fluorophores have a peak that has a sharp ascent, but a broad trailing end that can obscure neighboring peaks. If two peaks cannot be distinguished then the two labels associated with those peaks cannot be considered unique. Where fluorescent labels are used, the labels would emit fluorescent light at distinct, essentially non-overlapping wavelengths distant from each other by at least 1nm, 5nm, 10nm, 15nm, 20nm, 25nm, and preferably 30nm, 35nm, 40nm, 45nm, and more preferably by at least 50nm. For example, the emission peak of dye #1 could be 585nm, and the peak emission of dye #2 could be 630nm.

[0062] A unique label is a label that generates a signal that is distinguishable from other labels in the same mixture. A unique label, therefore, is dependent on the other labels that are included in the mixture. For example, fluorescein which has an emission spectra peak at 518nm and rhodamine red with an emission spectra peak of 590nm have clearly distinct emission peaks and hence both are considered to be unique labels when included together in a mixture. However, if another fluorophore is added that has an emission peak that overlaps with fluorescein or rhodamine red, these labels would not be unique. For example, if Oregon green (emission spectra peak 522nm) is added to the above mixture the emission peak of fluorescein and Oregon Green can overlap so much that, depending on the detection equipment used, they cannot be clearly distinguished from each other and hence neither fluorescein nor Oregon Green can act as unique labels in that particular mixture.

[0063] Often several formulations of the same label or related labels are synthesized for use in different applications, and while these labels have different chemical properties, they are not distinct in terms of detection. For example, fluorescein, can be used in its isothiocyanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or isomeric dichlorotriazine forms of fluorescein (DTAF). These labels are chemically distinct, but all emit light with a peak between 515-520 nm and hence would overlap so as to appear identical on most currently available detection equipment. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon

Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein.

[0064] A limited number of different label monomers are known that can be used together in a mixture and still provide unique signals. For example, five analytes can be distinctly labeled using the BODIPY fluorophore set from Molecular Probes (Eugene, OR). These fluorophores have the following distinct emission peaks: BODIPY FL(513), BODIPY R6G (550), BODIPY TMR (574), BODIPY 581/591 (592) and BODIPY TR (617). In order to obtain clear results using currently available detection equipment, the number of different fluorophores that can be used is less than thirty.

[0065] Unique labels also can be generated by combining two or more different label monomers to make a new label. The signal from the resulting label must be distinguishable from the signals of other labels used in the same experiment in order to be a unique label. For example, a nucleic acid labeled with both fluorescein and rhodamine will emit light at a different wavelength than a nucleic acid labeled with either fluorescein or rhodamine alone.

[0066] In this invention, various ratios of different label monomers bound to nucleic acids can be combined to generate a diverse population of unique labels that can include up to 10^{17} or more unique labels. For example, a nucleic acid labeled with two fluorescein labeled nucleotides and three rhodamine labeled nucleotides will emit light at a different wavelength compared to a nucleic acid labeled with three fluorescein nucleotides and two rhodamine nucleotides. In another example, a nucleic acid could be labeled with different ratios of three or more label monomer:nucleotides which greatly increases the variety of unique labels that can be generated.

[0067] The signal generated by each of the label monomers bound to the nucleotides can be normalized to have about the same unit signal. For example, if fluorescent monomer A is known to emit a different quanta of light as fluorescent monomer B, the signal from unique labels that contain a pre-determined number of these monomers can be normalized based on the known signal properties of the label monomers and the number of each monomer present in the unique label. Different numbers of labels can be used in the invention and so different multiples of the same unit signal can be utilized by the invention. For example, a nucleic acid can be labeled with two fluorophores and another nucleic acid can be labeled with six fluorophores. The second nucleic acid will have three times the signal of the first nucleic acid. Since the number of label monomers attached to each nucleic acid is known, the signals from the labeled nucleic acids can be normalized based on the number of label monomers present. For example, the signal from the nucleic acid with six fluorophores can be divided by three which normalized the signal relative to the signal from the nucleic acid with two fluorophores.

[0068] This method of producing labels results in important advantages over existing techniques such as microarray formats. Since each analyte is uniquely identified by a label with about the same unit signal, it allows the labels to be directly counted resulting in a digital read-out of each molecular species in a mixture. In contrast, microarray data must undergo several intermediate transformations to quantitate the number of molecules which results in a less precise analog output. In addition, using a diversity of labels with the same unit signal allows the detection method to operate in a narrow dynamic range resulting in improved sensitivity of the system since the trade-off between sensitivity and dynamic range is avoided.

[0069] The invention provides a diverse population of uniquely labeled probes. This population of probes contains about 30 or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid. In addition, the invention provides a diverse population of uniquely labeled probes containing a diversity of 50, 100, 200, 500, 1,000, 2,000, 5,000, 1×10^4 , 3×10^4 and about 1×10^5 or more different labels. As described above, these unique labels can contain a mixture of two or more different labels and comprise about the same unit signal or multiple thereof.

[0070] The invention provides a diverse population of uniquely labeled probes in which a target specific nucleic acid contains a nucleic acid bound to a unique label. In addition, the invention provides a diverse population of uniquely labeled probes containing two attached populations of nucleic acids, one population of nucleic acids containing thirty or more target specific nucleic acid probes, and a second population of nucleic acids containing a nucleic acid bound by a unique label.

[0071] A target specific probe is intended to mean an agent that binds to the target analyte selectively. This agent will bind with preferential affinity toward the target while showing little to no detectable cross-reactivity toward other molecules.

[0072] The target analyte is a nucleic acid. For example, a target can be a nucleic acid that is recognized and bound specifically by a complementary nucleic acid including for example, an oligonucleotide or a PCR product, or a non-natural nucleic acid such as a locked nucleic acid (LNA) or a peptide nucleic acid (PNA).

[0073] Aptamers are nucleic acid sequences that have three dimensional structures capable of binding small molecular targets including metal ions, organic dyes, drugs, amino acids, co-factors, aminoglycosides, antibiotics, nucleotide base analogs, nucleotides and peptides (Jayasena, S.D., Clinical Chemistry 45:9, 1628-1650, (1999)) incorporated herein by reference. Nucleic acids can be directly attached to the aptamer or attached through the use of hybridization.

[0074] The population of labels of the invention is produced by a method consisting of synthesizing a population of nucleic acids bound to a predetermined ratio of at least two different labels. The method involves incorporating labeled nucleotides into a repeated nucleic acid structure using a DNA polymerase. The repeated nucleic acid structures can

be designed to allow incorporation of a pre-determined ratio of labels. Using this method several unique labels can be generated from a small number of starting labels.

[0075] A specific example of this method where ten unique labels are made from two different labels is provided in Example 1. Briefly, ten unique templates of a 220 base pair single-stranded DNA are synthesized. The templates consist of a pre-determined ratio of the following 20 base pair repeats where $n+m=11$: 5' (ACTCTCTCTCTCTCTCTC)_n (GCTCTCTCTCTCTCTCTC)_m 3' (SEQ ID NO:1). The second strand is synthesized using the primer GAGAGAGAGA (SEQ ID NO:2), Klenow polymerase, DNA ligase, dGTP, dATP, and dCTP and dUTP each labeled with a different fluorophore. The labeled nucleotides will be incorporated into the DNA in a unique ratio determined by the ratio of the two repeats. In this example, the end result is ten uniquely labeled nucleic acids where the set ratio of the two fluorophores is 1:10, 2:9, 3:8, 4:7, 5:6, 6:5, 7:4, 8:3, 9:2, and 10:1.

[0076] In Example 1, two different labels resulted in ten unique labels. Using the same protocol, three different labels would result in 30 unique labels, four different labels would result in 60 unique labels, five different labels would result in 100 unique labels, and so on. Several variations of the method will be apparent to one skilled in the art. For example, the number of repeats can be changed to be less than or greater than ten. An increased number of repeats will increase the number of unique ratios possible. This will result in an increase in the number of unique labels that can be generated from the same number of starting different labels. Also in Example 1, the ratio between the two fluorophores can include 0:11 and 11:0 which results in two additional labels that contain one fluorophore or the other.

[0077] One skilled in the art will recognize that the sequence of the templates can differ from that shown above. For example, the repeat sequence in the template can be (GA)_n instead of (CT)_n. In addition, the repeat sequence could be a single nucleotide homopolymer such as (A)_n. With a homopolymer template, three labeled nucleotides can be incorporated in different ratios thus increasing the number of unique labels that can be generated. It is possible using the claimed method to generate a large number of unique labels including 40, 60, 80, 100, 120, 140, 150, 200, 500, 2,000, 5,000, 1×10^4 , 3×10^4 , 1×10^5 or more labels. Again as described above, these unique labels can comprise about the same unit signal or multiple thereof.

[0078] Another possible modifications of the method is to change the length of the repeat, for example, to less than or greater than 20 base pairs. The repeat serves to separate the labeled nucleotides and therefore to decrease possible interference between the labels. In the case of a fluorescent label, this can decrease quenching between fluorophores. In addition, the protocol for incorporating the labeled nucleotides into the DNA can be modified as would be clear to one skilled in the art and as described herein.

[0079] In one embodiment, the labeled DNA described above can be attached to a dendrimer. Oligonucleotide tags can be synthesized at the branches of the dendrimer to allow binding of the labeled DNA. For example, a linker can be attached to the labeled DNA described above that corresponds to an oligonucleotide tag on the branches of the dendrimer. Several types of linkers are known to one skilled in the art. For example, a restriction enzyme linker can be attached to the labeled DNA. These linkers are double-stranded oligonucleotides that contain the recognition sequence of a particular restriction enzyme. These linkers can be ligated onto double-stranded DNA using a DNA ligase and digested using the appropriate restriction enzyme. The result is an overhanging single stranded sequence that is available to hybridize to another nucleic acid.

[0080] The labeled DNA described above can be directly attached to a target specific probe. In addition, the labeled DNA can be indirectly attached to a target specific probe, for example, through use of a bridging nucleic acid. One or more of these labels can be attached to each target specific probe. Binding of a uniquely labeled target specific probe to a target analyte results in the unique tagging of that analyte. This tagging allows identification of the target analyte from a mixture of analytes.

[0081] The number of unique labels can be further increased by combining the unique labels described above in different combinations. In connection with the invention, a label to a nucleic acid probe may be attached by a method, the method comprising hybridizing a nucleic acid probe having a genedigit to an anti-genedigit having a label. The anti-genedigits that contain the labels described above are hybridized to genedigits. Genedigits can be linked together in unique combinations creating an even larger number of unique labels. The modular design of the genedigits allows for flexibility in the number of unique labels that are generated. For example, if a large number of genedigit modules are used, a large number of unique templates will be available for attachment of unique labels. The number of templates generated can be adjusted to cover the number of species in the mixture.

[0082] A genedigit can be a region of pre-determined nucleotide or amino acid sequence that serves as an attachment point for a label. The genedigit can have any sequence including, for example, a single unique sequence or a sequence containing repeated core elements. However, each genedigit has a unique sequence which differentiates it from other genedigits. When added to a complex mixture of targets, a nucleic acid genedigit can contain non-natural bases such as isocytidine and isoguanine which can reduce hybridization to naturally occurring target sequences. The sequence, length of a core element, and number of repeated core elements can be varied according to the particular requirements of an experiment and will be clear to one skilled in the art. For example, a nucleic acid core element can be between about 5 and 12 base pairs in length and the core element can be repeated once to about ten times.

[0083] Different genedigits can be synthesized that have a core unit that differs from the core units of other genedigits, for example, by at least two bases. By combining these unique genedigits in various combinations, a highly diverse number of structures can be synthesized. For example, 50 genedigits with unique sequences can be synthesized and split into five groups containing ten genedigits in each group. The genedigits of each group can be synthesized to have a short tag on each end. One genedigit from each group is then linked together using an adapter oligonucleotide that is complimentary to the tags that are common for each group. In this example, 1×10^5 ($10 \times 10 \times 10 \times 10 \times 10$) unique combinations are possible.

[0084] Genedigits serve as attachment points for the unique labels described above. Since genedigits can be linked together in unique combinations this greatly increases the number of unique labels. By linking together genedigits, a large number of unique labels can be generated including 200, 500, 2,000, 5,000, 1×10^4 , 1×10^4 , 1×10^5 or more unique labels.

[0085] In connection with the invention, a label may be attached to a nucleic acid probe by a method, the method comprising hybridizing a nucleic acid probe having a genedigit to an anti-genedigit having a label where the genedigit comprises a set of three or more repeat sequences, and the anti-genedigit comprises a cognate set of at least two complimentary repeat sequences where the anti-genedigit specifically hybridizes to the genedigit through a sequence having a complexity less than the number of hybridized base pairs.

[0086] An example of a nucleic acid genedigit is given in Example 2. Briefly, in this example an eight base pair core element containing the non-natural bases isocytidine and isoguanine is repeated five times. This results in a 40 base pair genedigit. In this example the anti-genedigit consists of a sequence that is complimentary to three of the five core elements in the genedigit. An anti-genedigit can contain the same number, or a lesser number, of repeat sequences compared to the genedigit as long as the anti-genedigit is able to specifically bind to the genedigit. In this example, the anti-genedigit is a 24 base pair sequence that can bind to the 40 base pair genedigit in three different registers (see Figure 1C).

[0087] The anti-genedigit in this specific example is a 24 base pair sequence that only has the complexity of an eight base pair sequence. Complexity refers to the degree of repeated elements between two nucleic acids that are being hybridized together in a solution. When the nucleic acid molecules that are to be hybridized contain repeated core elements or homopolymeric regions, there are many possible pairing opportunities and so the hybridization proceeds quickly. When the nucleic acid molecules that are to be hybridized do not contain any repeated core elements then there is only one way that the two sequences can be hybridized and so the hybridization proceeds more slowly. Sequences that hybridize quickly are said to have a low complexity, while sequences that take longer to hybridize have a higher complexity. In this example, a 40 base pair genedigit sequence made up of five direct repeats of an 8 base pair core element, can be hybridized to a 24 base pair anti-genedigit containing three repeats of the 8 base pair core repeat in three different registers. Thus the anti-genedigit can hybridize to the 40 base pair genedigit through a 24 base pair sequence that only has a complexity of an 8 base pair sequence. The advantage of this method is that hybridization will proceed more quickly and efficiently.

[0088] In the specific example above, an 8 base pair core element was describe; however, a core element can be more or less than 8 base pairs. For example, a core element can be between 5 and 12 base pairs. A change in the core element of a repeat will correspondingly change complexity. For example, if a core element is between 5 and 12 base pairs, complexity will be between 5 and 12. In addition, an anti-genedigit can bind to a genedigit using all or part of its sequence. In the example above, a 24 base pair anti-genedigit can bind to the genedigit with less than 24 bases, for example, 15 to 23 base pairs.

[0089] The anti-genedigit serves as a connector between the genedigit and the label. The genedigit is bound by the anti-genedigit and the anti-genedigit is bound to a label directly or indirectly by being bound to a dendrimer that has a label(s) attached. The anti-genedigit also contains a linker sequence that allows attachment to a dendrimer. For example, an oligonucleotide tag can be synthesized at the stem of a dendrimer that is complementary to the linker sequence on the anti-genedigit. As described above, several types of linkers can be used. In this way the repeat sequences of the anti-genedigit are free to hybridize with their corresponding genedigit resulting in the specific attachment of a label to a genedigit.

[0090] Genedigits and anti-genedigits can be comprised of nucleic acids, including aptamers.

[0091] In connection with the invention, a target specific probe can be attached to one or more genedigits to form a "specifier." The genedigits can be directly linked or can be attached to the target specific region using an intervening or adapting sequence. As described above, the target specific area can be a nucleic acid, including an aptamer. The target specific area is designed to specifically bind an analyte in a mixture. In this way an analyte can be labeled with a unique label.

[0092] A specifier can contain one to several genedigits. For example, a specifier can contain four or five or more genedigits. The number of genedigits in a specifier will determine the number of unique labels that are available to bind to analytes in a mixture. Therefore, in order to label each analyte in a complex mixture, a large population of specifiers can be synthesized that contain several combinations of genedigits. Alternatively, to label one or a small number of

analytes in mixture, a specifier or specifiers can be synthesized that contain one or a few genedigits. In addition, a specifier can contain a common tag such as a biotin tag. These tags can facilitate synthesis and purification of specifiers.

[0093] As outlined herein, the invention relates to a population of uniquely labeled nucleic acid probes. The method for providing said population consists of synthesizing a population of target specific nucleic acid probes each having a different specifier; then synthesizing a corresponding population of anti-genedigits each having a unique label, and finally hybridizing the populations of target nucleic acid probes to the anti-genedigits, to produce a population where each of the target specific probes is uniquely labeled.

[0094] The invention further provides a method for detecting a nucleic acid analyte, by contacting a mixture of nucleic acid analytes with a population of target specific probes each attached to a unique label under conditions sufficient for hybridization of the probes to the target and measuring the resulting signal from one or more of the target specific probes hybridized to an analyte where the signal uniquely identifies the analyte.

[0095] The nucleic acid analyte can contain any type of nucleic acid, including for example, an RNA population or a population of cDNA copies. The invention provides for at least one target specific probe for each analyte in a mixture. The invention also provides for a target specific probe that contains a nucleic acid bound to a unique label. Furthermore, the invention provides two attached populations of nucleic acids, one population of nucleic acids containing a plurality of target specific nucleic acid probes, and a second population of nucleic acids containing a nucleic acid bound by a unique label. When the target specific probes are attached to unique labels, this allows for the unique identification of the target analytes.

[0096] The methods of the invention are advantageous because hybridizations can be performed in solution in a small volume (0.01 -2.0 μ l) which ensures high concentration of the nucleic acids which will drive the hybridization rate. Two different types of hybridizations are utilized by the methods of invention. The first hybridization is between a complex mixture of analytes and the specifiers and the second type of hybridization is between the specifiers and the labels.

[0097] In the first type of hybridization between a complex mixture of analytes and a population of specifiers, the specifiers are in excess. For example, the specifiers can be in 10 to about 100 fold excess over the analytes in the complex mixture. The kinetics of this reaction can be described by the following equation:

$$t_{1/2} = \frac{N \ln 2}{3.5 \times 10^5 \times L^{0.5} \times C_0}$$

where N is the complexity of the probe (the specifiers), L is the length, C_0 is the concentration of the probe, and $t_{1/2}$ is the time required for 50% completion of the reaction. Using this equation it is clear that higher concentrations of probe will result in shorter time required for 50% completion of the reaction. The sample in this hybridization can be either RNA or DNA. If the sample is not poly A RNA then it must first be tagged by some method, for example, platinum-digoxigenin. After the hybridization is complete, the duplexes containing the analyte and specifier are separated using the poly A, or the digoxigenin tag as a handle and washed from the excess specifier. The washed material is then used for the second hybridization.

[0098] In the second hybridization the isolated specifiers are mixed with labels. The labels are in excess over the specifiers and the kinetics follow the same kinetics as shown above for the first hybridization, and it is performed at an even faster rate because the complexity, N, is very low. The hybridized complexes containing specifiers and labels can be isolated using a biotin tag on the specifiers as a handle and processed further for detection.

[0099] The formation of macromolecule hybrids is dependent on several conditions known in the art, including temperature, salt concentration, and pH. Different conditions for the formation of nucleic acid hybrids are well known in the art and can be found in, for example, Hames and Higgins, Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford (1991). In addition, conditions for nucleic acid-protein and protein-protein interactions are well known in the art and can be found in, for example, Current Protocols in Molecular Biology (ed. Ausubel et al., Greene Publ., New York 1989) which is incorporated herein by reference.

[0100] The "specifier-label" complexes can be separated from one another, for example, by spreading them on a two-dimensional surface such as glass, or by splitting them in liquid droplets in a flow cytometer. In this example, visualization can be achieved either by scanning the 2-D surface or by flow cytometry. Co-localization of specific labels will determine the identity of the particular specifier, which will determine the identity of the particular analyte that the specifier bound initially.

[0101] The "specifier-label" complexes can be detected by various devices including but not limited to visual inspection, digital cameras, video cameras, photographic film, or the use of current instruments such as laser scanning devices, fluorometers, luminometers, photodiodes, quantum counters, plate readers, epifluorescence microscopes, scanning microscopes, confocal microscopes, or by other means for amplifying the signal such as a photomultiplier tube or other light detector capable of detecting the presence, location, intensity, excitation and emission spectra, fluorescence po-

larization, fluorescence lifetime and other physical properties of the fluorescent signal. Non-fluorescent signals can be detected using a Geiger counter, scintillation counters, chemiluminescence, enzyme assays, and autoradiography.

[0102] An example of a detection method that can be utilized by the invention is a CCD (charge-coupled device) camera. Use of this device requires that the specifier-label complexes are first spread on a glass slide. Since all complexes will have approximately the same total intensity this simplifies detection, because the camera can be adjusted for maximum sensitivity (maximum gain) and minimum dynamic range.

[0103] For example, if fluorescent dendrimers are used for labeling the sample can be excited at the corresponding absorption maxima and scanned at the corresponding emission maxima for each of the, for example, 8-10 fluorophores used. A benefit of CCD cameras is that they have very wide range of detection which allows for a choice of fluorophores with distant emission peaks. Thinned layer CCD cameras can detect from the soft X-ray to the near infrared spectrum.

[0104] Alternatively if nanoparticles labels are used the sample can be illuminated by angled white light and detection occurs at a few wavelengths. The number of wavelengths will depend on the quality of the camera used and its sensitivity and linearity. Cameras of good quality can reproducibly distinguish millions of colors with only the three standard filters (red, green, and blue).

[0105] In the case of unique labels that contain more than one label monomer, the overall signal generated by the unique label can be determined. For example, an overall signal for a unique label containing several fluorophores can be determined using a spectrometer. In addition to detecting an overall signal generated from the unique label, the invention provides for the determination of the spatial location of starting labels (which can be label monomers or combinations of label monomers) within a unique label. For example, the labels within a unique label can be separated from one another using a flow-stretch technique (Henegariu et al., *Biotechniques* 31:246-250 (2001)) a receding meniscus technique (Yokota et al., *Nuc. Acids Res.* 25:1064-1070 (1997)) or an electrostretching technique (Matsuura et al., *Nuc. Acids Res.* 29:E79 (2001)).

[0106] The use of flow-stretching, receding meniscus, or electro-stretching techniques allows for the separation of the labels within a unique label so that one can determine spatially where a particular label is positioned in the unique label (Example V, VI, VII and VIII). Therefore, unique labels that have the same combination of starting labels and the same overall signal can be differentiated from one another based on the location of those labels within the unique label. This ability to locate the position of a label within a unique label allows for the position of the starting label to be used as a distinguishing characteristic when generating a set of unique labels. Hence, a complex set of unique labels can be generated using the same combination of starting labels by varying the position of the labels within a unique label.

[0107] An observed signal can be modified using methods known in the art. For example, an observed signal can include subtraction of non-specific noise. An observed signal can also include, for example, treatment of the measured quantity by routine data analysis and statistical procedures which allow meaningful comparison and analysis of the observed values. Such procedures include, for example, normalization for direct comparison of values having different scales, and filtering for removal of aberrant or artifactual values.

[0108] In the first type of hybridization described above between a complex mixture of analytes and a population of specifiers, the specifiers were in excess. In contrast, hybridization on standard microarrays occurs under conditions where the analyte is in excess compared to the labeled probe. In addition, on microarrays only a small fraction of this probe will actually hybridize to the cognate target during the course of the hybridization, and so target coverage at the end of the experiment is usually less than 5%. The target coverage using the methods of the invention are theoretically 100% because the label is in excess compared to the target analyte.

[0109] One of the advantages of 100% target coverage is that this allows the target analytes to be directly counted since each analyte is bound by a label. Using the methods of the invention, molecular species can be directly counted one by one. Direct counting, or a digital output, is preferable to the indirect quantitation methods used for microarrays because the data does not need to go through several intermediary transformations. If the detection equipment directly counts the number of particles emitted it is said to have a digital output, however if the direct counts undergo several intermediary transformations then the data has an analog output. Quantitation data from microarrays is known to be subject to several distortions due to data extrapolation.

[0110] Another consequence of low target coverage in microarrays is that highly sensitive equipment is needed to detect the low amount of signal. However, since microarrays require a high dynamic range for signal detection, sensitivity is decreased because of the trade-off between sensitivity and dynamic range. The methods of the invention use a limited number of labels to create a large number of unique label combinations. This allows the detection method to operate in a narrow dynamic range. In contrast, microarray methods require a large dynamic range (4 orders of magnitude or more) to account for the large differences in abundance of the different molecular species. The low requirements in terms of dynamic range required by the methods of the invention will improve the sensitivity of the system since the trade-off between sensitivity and dynamic range is avoided.

[0111] The invention also provides a method of detecting a nucleic acid analyte as above where the anti-genedigit(s) each have a unique label.

[0112] The methods of the invention provide for detection of analytes in mixtures. The mixture can contain several

types of analytes or the mixture can contain just one type of analyte. In addition, the mixture could contain just a single copy of an analyte. If the target analyte has an unknown sequence or structure, a large population of target specific specifiers can be added to the mixture. This population can include specifiers with target specific regions of pre-determined sequence or structure or specifiers can be used with target specific regions of random sequence or structure. Alternatively,

if the target analyte has a known sequence or structure, a particular specifier containing a region that will specifically bind to that sequence or structure can be used either alone or in combination with other specifiers.

[0113] The methods of the invention are suited to nucleic acid analytes. A population of specifiers can be generated for any analyte where a target specific region can be found that specifically interacts with that analyte. In addition, the ability of nucleic acid aptamers to bind a wide variety of analytes allows these structures to be used in the target specific regions of specifiers. Several combinations are possible as long as the specifiers specifically bind to the analytes.

[0114] The labels and methods described herein can be used for diagnostic and therapeutic purposes. Analytes, or combinations of analytes, that are diagnostic of a disease can be detected and quantified from a sample derived from a subject. Many different analytes can be analyzed at one time from a single sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample. In addition, the methods of the invention can provide information that determines a course of treatment for a patient. For example, the amount of a particular marker for a tumor can be accurately quantified from even a small sample from a patient. For some disease like breast cancer overexpression of certain genes, such as Her2-neu, indicate a more aggressive course of treatment will be needed.

EXAMPLE I

Generation of Unique Labels Using Two Different Labels

[0115] In this example, ten unique labels are made from two different fluorescent labels. First, ten unique templates of a 220-base pair single-stranded DNA are synthesized. The templates consist of a pre-determined ratio of the following 20-base pair repeats:

5' (ACTCTCTCTCTCTCTCTC)n(GCTCTCTCTCTCTCTCTC)m 3' (SEQ ID NO:1) where
n=1,2,3,4,5,6,7,8,9,10, m=1,2,3,4,5,6,7,8,9,10, and n+m=11.

[0116] The second strand is synthesized using the primer GAGAGAGAGA (SEQ ID NO:2), Klenow polymerase, DNA ligase, dGTP, dATP, dUTP-fluorescein and dCTP-rhodamine. After the reaction is complete the product is treated with S1 nuclease to digest the DNA with gaps, and the remaining full length DNA is then purified. The labeled nucleotides will be incorporated into the DNA in a unique ratio determined by the ratio of the two repeats. The end result is ten uniquely labeled nucleic acids where the set ratio of fluorescein to rhodamine is 1:10, 2:9, 3:8, 4:7, 5:6, 6:5, 7:4, 8:3, 9:2, and 10:1.

[0117] A linker oligonucleotide is ligated to the labeled DNA and then this linker is used to attach the labeled DNA to the branches of a dendrimer. The dendrimer has an oligonucleotide tag of 5 bases at the stem to facilitate binding of an anti-genedigit (see Example II) and tags of 10-base pairs at the branches to facilitate binding of labeled DNA.

EXAMPLE II

Generation of a Labeled Specifier

[0118] The specifiers are synthesized by ligating together one target specific sequence (synthetic oligonucleotide, peptide-nucleic acid (PNA), PCR product, or linked-nucleic acid (LNA)), and several "genedigits" (see Figure 1A). In this example, each specifier contains a unique combination of 4 different genedigits. This results in the generation of 10,000 possible unique specifiers.

[0119] The genedigits are synthetic oligonucleotides that contain only two of the natural bases, plus two bases that not found in nature: isocytidine and isoguanine. Such base composition ensures that the genedigits will not non-specifically hybridize with analytes in a complex mixture. The sequence of each genedigit is composed of 5 repeats of an 8-base pair core sequence (see Figure 1B). Each core sequence unit differs from the others by at least two bases.

[0120] In order to make 10,000 unique specifiers, forty different genedigits are synthesized and split into 4 groups containing 10 genedigits each. The genedigits of each group have a 5-base pair tag on each end. One genedigit from each group is present in each specifier. The genedigits are ligated with the help of adapter 10mer oligonucleotides that are complimentary to the 5-base pair tags that are common for each group. In this way, for a specifier with 4 genedigits, there will be $10 \times 10 \times 10 \times 10 = 10,000$ possible combinations. All specifiers also contain a biotin tag.

[0121] The genedigit serves as an attachment point for a label and so the number of labels synthesized corresponds

to the number of genedigits. The genedigits are labeled through the use of anti-genedigits. An anti-genedigit sequence consisting of three 8-base pair repeats complimentary to the 8-base pair core repeat of the corresponding genedigit is ligated to the stem of a labeled dendrimer (from Example I).

[0122] The 24-base pair labeled anti-genedigit hybridizes to the 40-base pair genedigit sequence in the specifier in one of three different registers (see Figure 1C). Thus when the label hybridized to the specifier it will do so through a 24-base pair sequence that only has a complexity of an 8-base pair sequence.

EXAMPLE III

Gene Expression Analysis Using Specifiers

[0123] In order to determine differences in gene expression between astrocytes and LPS-activated astrocytes, RNA is isolated from both populations of astrocytes using cell lysis in guanidine isothiocyanine or phenol/chloroform. A population of specifiers is added to each RNA sample under conditions suitable for hybridization. The mRNA-specifier complexes are isolated with oligo dT beads and washed extensively to remove excess specifiers. The specifiers are eluted from the mRNA by digesting the mRNA with RNase A. The specifiers are then processed for labeling as described in Examples I and II and these labels are detected using a CCD camera. The number of specifiers corresponding to specific mRNAs from untreated astrocytes is then compared to the specifier pattern from LPS-treated astrocytes. Since the sequence of the target specific region of the specifier is known, this identifies the genes that are differentially expressed between the two samples.

EXAMPLE IV

Microbial Detection Using Specifiers

[0124] The invention can be used to detect strains of microorganisms with known sequence in biological samples. Total DNA is extracted from a blood sample from a patient with a suspected microbial infection. The total DNA is then tagged with digoxigenin. The DNA is denatured and hybridized with a population of specifiers containing target regions specific for a particular microorganism, or a panel of microorganisms, in a small volume (0.01 - 2.0 μ l). The sample DNA-specifier complexes are isolated using anti-digoxigenin antibodies, and washed extensively to remove excess specifiers. The specifiers are then processed for labeling and imaging as described above.

EXAMPLE V

Label Configurations

[0125] This example shows a DNA probe molecule (specifier) that contains labels that are separated by a distance sufficient such that the labels can be individually resolved (Figure 2). For example, in the case of fluorescent labels, this large spatial separation of labels allows for spectral identification of each label individually.

[0126] As shown in Figure 2, a DNA probe (specifier) can be generated that contains four distinct fluorescent labels. These labels are separated by a distance of about one micron. In this example, the labels are separated by about 2,000 to 3,000 bases of DNA, however other polymers could be used to separate the labels as well.

[0127] A diverse population of unique labels can be generated by varying the fluorescent label at each of the four locations. For example, a set of 10 distinct fluorophores can be used to generate 10,000 unique labels. Specifically, for a probe with four different labels, there will be $10 \times 10 \times 10 \times 10 = 10,000$ possible combinations. Similarly, for a probe that has five different labels, there will be 100,000 possible combinations.

EXAMPLE VI

Flow-stretch Separation of Labels

[0128] This example shows a DNA probe molecule (specifier) with one label monomer which has been stretched on a coverslip using a flow stretch technique. The flow stretch technique can also be used for a probe molecule with multiple label monomers along the probe molecule.

[0129] A double stranded DNA probe was constructed with a single stranded section at a desired location for label attachment. In this example, the single stranded section is at the end of the probe molecule. The label molecule contains about 300 Cy3 fluorophores and a section of DNA that is complementary to the single stranded area of the probe molecule. After hybridization, the probe molecule is stained with YOYO1 dye to verify that the label is attached to the

probe. The staining procedure is not necessary when multiple labels are attached. The probes with hybridized label were then stretched on a coverslip using a flow stretch technique (Henegariu et al., *supra*). An image, figure 3, was obtained using an inverted fluorescent microscope. Probes can also be stretched using a receding meniscus technique (Yokota et al., *supra*).

EXAMPLE VII

Electro-stretch Separation of Labels

[0130] This example shows that an electric field can be used to align probe (specifier) molecules that are attached by one end to a surface. In this manner, labels that are along the probe molecule are spatially separated.

[0131] A flow cell was constructed that allowed imaging of the probe molecules through a coverslip. The flow cell also allowed for the exchange of solutions and the application of an electric field. In order to prevent bubbling, a gel was used to separate the electrodes from the solution. With gel separation of the electrodes and solution, higher voltages are possible which improves the alignment of the probe molecules. The coverslip was coated with bovine serum albumin (BSA) to minimize non-specific binding. DNA was added to the coverslip and attached to the coverslip primarily by the ends of the DNA. Un-bound DNA was washed away and the remaining bound DNA was stained with YOYO1. Imaging was performed using an epi-fluorescent microscope. When the voltage was off, tethered DNA can be seen to move randomly (Figure 4A). However, when the voltage is on, the DNA becomes aligned (Figure 4B).

EXAMPLE VIII

Cytometer Detection of Labels

[0132] Another way to align DNA probes (specifiers) spatially is by the use of a constriction in the flow of a liquid containing the DNA in conjunction with an oscillating electric field (Asbury, C.L., and van den Engh, G., Biophys. J. 74: 1024-1030 (1998)). In this way, when the probe molecules pass through the constriction the labels are spatially separated (Figure 5). Confocal optics can allow for detection with enough spatial resolution to determine the sequence of the labels.

SEQUENCE LISTING

[0133]

<110> The Institute for Systems Biology
Dimitrov, Krassen
Dunaway, Dwayne

<120> Methods For Detection and Quantification
of Analytes In Complex Mixtures

<130> FP-IS 5320

<140> PCT/US02/21278

<141> 2002-07-03

<150> US 09/898,743

<151> 2001-07-03

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 220

<212> DNA

<213> Artificial Sequence

<220>

EP 1 448 581 B1

<223> synthetic primer

<400> 1

5
actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 120
actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 180
actctctctc tctctctctc gctctctctc tctctctctc 220

10

<210> 2

<211> 220

<212> DNA

<213> Artificial Sequence

15

<220>

<223> synthetic primer

20

<400> 2

actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 120
actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 180
25 gctctctctc tctctctctc gctctctctc tctctctctc 220

25

<210> 3

<211> 220

30

<212> DNA

<213> Artificial Sequence

<220>

35

<223> synthetic primer

<400> 3

actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 120
actctctctc tctctctctc actctctctc tctctctctc gctctctctc tctctctctc 180
40 gctctctctc tctctctctc gctctctctc tctctctctc 220

40

<210> 4

45

<211> 220

<212> DNA

<213> Artificial Sequence

<220>

50

<223> synthetic primer

<400> 4

55

EP 1 448 581 B1

5
 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 120
 actctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

<210> 5
 <211> 220
 <212> DNA
 <213> Artificial Sequence

10
 <220>

<223> synthetic primer

15
 <400> 5

20
 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 120
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

<210> 6
 <211> 220
 <212> DNA
 <213> Artificial Sequence

25
 <220>

30
 <223> synthetic primer

<400> 6

35
 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
 actctctctc tctctctctc actctctctc tctctctctc gctctctctc tctctctctc 120
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

<210> 7
 <211> 220
 <212> DNA
 <213> Artificial Sequence

40
 <220>

45
 <223> synthetic primer

<400> 7

50
 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
 actctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 120
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

55
 <210> 8
 <211> 220
 <212> DNA

EP 1 448 581 B1

<213> Artificial Sequence

<220>

5 <223> synthetic primer

<400> 8

10 actctctctc tctctctctc actctctctc tctctctctc actctctctc tctctctctc 60
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 120
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

15 <210> 9
 <211> 220
 <212> DNA
 <213> Artificial Sequence

20 <220>

 <223> synthetic primer

 <400> 9

25 actctctctc tctctctctc actctctctc tctctctctc gctctctctc tctctctctc 60
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 120
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

30

 <210> 10
 <211> 220
 <212> DNA
 <213> Artificial Sequence

35 <220>

 <223> synthetic primer

40 <400> 10

45 actctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 60
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 120
 gctctctctc tctctctctc gctctctctc tctctctctc gctctctctc tctctctctc 180
 gctctctctc tctctctctc gctctctctc tctctctctc 220

50 <210> 11
 <211> 10
 <212> DNA
 <213> Artificial Sequence

55 <220>

 <223> synthetic primer

 <400> 11
 gagagagaga 10

Claims

1. A diverse population of uniquely labeled probes, comprising thirty or more target-specific nucleic acid probes, each attached to a unique label bound to a nucleic acid, each nucleic acid probe being a nucleic acid molecule comprising a target-specific region and a region comprising genedigits, said genedigits being nucleic acid regions of predetermined nucleotide sequences that are specifically bound by complementary anti-genedigit nucleotide sequences, each said anti-genedigit nucleotide sequence being attached to a label monomer or a combination of label monomers, wherein the population is in solution, and wherein the label has a detectable signal that is fluorescence, luminescence, colorimetric, voltage, current, or a magnetic field.
2. The diverse population of claim 1, comprising thirty or more target-specific nucleic acid probes, wherein each nucleic acid probe comprises four or more genedigits, said genedigits being nucleic acid regions of predetermined nucleotide sequence that are specifically bound by complementary anti-genedigit nucleotide sequences, each said anti-genedigit nucleotide sequence being attached to a label monomer or a combination of label monomers, wherein the population is in solution, and wherein the label has a detectable signal that is fluorescence, luminescence, colorimetric, voltage, current, or a magnetic field.
3. The diverse population of any one of claims 1 to 2, wherein each probe is synthetic.
4. The diverse population of any one of claims 1 to 3, wherein each probe is DNA attached to a bridging nucleic acid.
5. The diverse population of any one of claims 1 to 4, wherein said label monomers are attached to dendrimers.
6. The diverse population of any one of claims 1 to 5 wherein the genedigits are DNA.
7. The diverse population of any one of claims 1 to 6, wherein the probes comprise tags.
8. The diverse population of claim 7, wherein the tags are biotin tags.
9. The diverse population of any one of claims 1 to 8, wherein a quantum dot is used as a label monomer.
10. The diverse population of any one of claims 1 to 8, wherein a fluorophore is used as a label monomer.
11. The diverse population of any one of claims 1 to 10, comprising 50, 100, 200, 500, 1,000, 2,000, 5,000, 1×10^4 , 3×10^4 , 1×10^5 or more different nucleic acid probes.
12. The diverse population of any one of claims 1 to 11, wherein the nucleic acid probes are separated from one another.
13. The diverse population of claim 12, wherein the nucleic acid probes are spread on a two-dimensional surface.
14. The diverse population of any one of claims 1 to 13, wherein one or more of the target-specific probes is hybridized to an analyte and wherein the signal from the hybridized target-specific probe uniquely identifies the analyte.
15. A method of detecting a nucleic acid analyte, comprising:
 - (a) contacting a mixture of nucleic acid analytes with the diverse population of uniquely labeled probes of any one of claims 1 to 11 under conditions sufficient for hybridization of the probes to target nucleic acid analytes, and
 - (b) measuring the resulting signal from one or more of the target specific probes hybridized to an analyte, where the signal uniquely identifies the analyte.
16. A method of quantifying a nucleic acid analyte, comprising:
 - (a) contacting a mixture of nucleic acid analytes with the diverse population of uniquely labeled probes of any one of claims 1 to 11 under conditions sufficient for hybridization of the probes to target nucleic acid analytes;
 - (b) measuring the resulting signal from one or more of the target specific probes hybridized to an analyte, where the signal uniquely identifies the analyte, thereby identifying the target nucleic acid analyte; and
 - (c) counting the individual molecules of the nucleic acid analyte in the sample, thereby quantifying the nucleic acid analyte.

17. The method according to any one of claims 15 and 16, wherein the position of the label monomer or combination of label monomers is used as a distinguishing characteristic.

18. The method according to any one of claims 15 to 17, wherein the nucleic acid probes are subjected to a flow-stretching technique, to a receding meniscus technique, to an electrostretching technique, or to constriction in the flow of a liquid containing said nucleic acid probes in conjunction with an oscillating electric field.

Patentansprüche

1. Eine verschiedenartige Population von einzigartig markierten Sonden, umfassend 30 oder mehr zielspezifische Nukleinsäuresonden, jeweils mit einer einzigartigen Markierung verbunden, die an eine Nukleinsäure gebunden ist, wobei jede Nukleinsäuresonde ein Nukleinsäuremolekül ist, welches eine zielspezifische Region und eine Genedigits umfassende Region umfasst, wobei die Genedigits Nukleinsäureregionen mit vorherbestimmten Nukleotidsequenzen sind, die spezifisch durch komplementäre Anti-Genedigit Nukleotidsequenzen gebunden werden, wobei die Anti-Genedigit Nukleotidsequenz mit einem Markierungs-Monomer oder einer Kombination von Markierungs-Monomeren verbunden ist, wobei die Population sich in Lösung befindet, und wobei die Markierung ein detektierbares Signal aufweist, das Fluoreszenz, Lumineszenz, kolorimetrisch, Spannung, Strom oder ein Magnetfeld ist.

2. Die verschiedenartige Population nach Anspruch 1, umfassend 30 oder mehr zielspezifische Nukleinsäuresonden, wobei jede Nukleinsäuresonde vier oder mehr Genedigits umfasst, wobei die Genedigits Nukleinsäureregionen mit vorherbestimmten Nukleotidsequenzen sind, die spezifisch durch komplementäre Anti-Genedigit Nukleotidsequenzen gebunden werden, wobei die Anti-Genedigit Nukleotidsequenz mit einem Markierungs-Monomer oder einer Kombination von Markierungs-Monomeren verbunden ist, wobei die Population sich in Lösung befindet, und wobei die Markierung ein detektierbares Signal aufweist, das Fluoreszenz, Lumineszenz, kolorimetrisch, Spannung, Strom oder ein Magnetfeld ist.

3. Die verschiedenartige Population nach einem der Ansprüche 1 oder 2, **dadurch gekennzeichnet, dass** jede Sonde synthetisch ist.

4. Die verschiedenartige Population nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** jede Sonde DNA ist, die mit einer überbrückenden Nukleinsäure verbunden ist.

5. Die verschiedenartige Population nach einem der Ansprüche 1 bis 4, **dadurch gekennzeichnet, dass** die Markierungs-Monomere mit Dendrimeren verbunden sind.

6. Die verschiedenartige Population nach einem der Ansprüche 1 bis 5, **dadurch gekennzeichnet, dass** die Genedigits DNA sind.

7. Die verschiedenartige Population nach einem der Ansprüche 1 bis 6, **dadurch gekennzeichnet, dass** die Sonden Kennzeichnungen umfassen.

8. Die verschiedenartige Population nach Anspruch 7, **dadurch gekennzeichnet, dass** die Kennzeichnungen Biotin-Kennzeichnungen sind.

9. Die verschiedenartige Population nach einem der Ansprüche 1 bis 8, **dadurch gekennzeichnet, dass** ein Quantenpunkt als Markierungs-Monomer verwendet wird.

10. Die verschiedenartige Population nach einem der Ansprüche 1 bis 8, **dadurch gekennzeichnet, dass** ein Fluorophor als Markierungs-Monomer verwendet wird.

11. Die verschiedenartige Population nach einem der Ansprüche 1 bis 10, umfassend 50, 100, 200, 500, 1000, 2000, 5000, 1×10^4 , 3×10^4 , 1×10^5 oder mehr unterschiedliche Nukleinsäuresonden.

12. Die verschiedenartige Population nach einem der Ansprüche 1 bis 11, **dadurch gekennzeichnet, dass** die Nukleinsäuresonden voneinander getrennt sind.

13. Die verschiedenartige Population nach Anspruch 12, **dadurch gekennzeichnet, dass** die Nukleinsäuresonden auf

einer zweidimensionalen Oberfläche ausgebreitet sind.

14. Die verschiedenartige Population nach einem der Ansprüche 1 bis 13, **dadurch gekennzeichnet, dass** eine oder mehrere der zielspezifischen Sonden mit einem Analyten hybridisiert ist, und wobei das Signal von der hybridisierten zielspezifischen Sonde den Analyten einzigartig identifiziert.

15. Ein Verfahren zur Detektion eines Nukleinsäureanalyten, Schritte umfassend, bei denen man:

(a) eine Mischung von Nukleinsäureanalyten mit der verschiedenartigen Population von einzigartig markierten Sonden nach einem der Ansprüche 1 bis 11 unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Sonden an Zielnukleinsäureanalyten ausreichend sind, und

(b) das sich ergebende Signal von einer oder mehreren der zielspezifischen Sonden, die mit einem Analyten hybridisiert ist, misst, wobei das Signal den Analyten einzigartig identifiziert.

16. Ein Verfahren zur Quantifizierung eines Nukleinsäureanalyten, Schritte umfassend, bei denen man:

(a) eine Mischung von Nukleinsäureanalyten mit der verschiedenartigen Population von einzigartig markierten Sonden nach einem der Ansprüche 1 bis 11 unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Sonden an Zielnukleinsäureanalyten ausreichend sind;

(b) das sich ergebende Signal von einer oder mehreren der zielspezifischen Sonden, die mit einem Analyten hybridisiert ist, misst, wobei das Signal den Analyten einzigartig identifiziert, wodurch der Zielnukleinsäureanalyt identifiziert wird, und

(c) die individuellen Moleküle des Nukleinsäureanalyten in der Probe zählt, wodurch der Nukleinsäureanalyt quantifiziert wird.

17. Das Verfahren nach einem der Ansprüche 15 und 16, **dadurch gekennzeichnet, dass** die Position des Markierungs-Monomers oder der Kombination von Markierungs-Monomeren als unterscheidendes Merkmal verwendet wird.

18. Das Verfahren nach einem der Ansprüche 15 bis 17, **dadurch gekennzeichnet, dass** die Nukleinsäuresonden eine Fluss-Dehnungs-Technik (flow-stretching technique) einer Technik des zurückgehenden Meniskus (receding meniscus technique), einer Elektro-Dehnungs-Technik (electrostretching technique) oder einer Einschränkung in dem Fluss einer die Nukleinsäuresonden enthaltenden Flüssigkeit in Verbindung mit einem schwingenden elektrischen Feld unterworfen werden.

Revendications

1. Population variée de sondes marquées de manière unique comprenant trente ou plus de trente sondes d'acides nucléiques spécifiques d'une cible, chacune étant fixée à un marqueur unique lié à un acide nucléique, chaque sonde d'acide nucléique étant une molécule d'acide nucléique comprenant une région spécifique d'une cible et une région comprenant des empreintes génétiques, lesdites empreintes génétiques étant des régions d'acides nucléiques de séquences nucléotidiques prédéterminées qui se lient spécifiquement à des séquences nucléotidiques complémentaires d'anti-empreinte génétique, chaque séquence nucléotidique d'anti-empreinte génétique étant fixée à un monomère marqueur ou à une combinaison de monomères marqueurs, dans laquelle la population est en solution et dans laquelle le marqueur possède un signal détectable qui est la fluorescence, la luminescence, la colorimétrie, une tension, un courant ou un champ magnétique.

2. Population variée selon la revendication 1, comprenant trente ou plus de trente sondes d'acides nucléiques spécifiques d'une cible, dans laquelle chaque sonde d'acide nucléique comprend quatre ou plus de quatre empreintes génétiques, lesdites empreintes génétiques étant des régions d'acides nucléiques possédant une séquence nucléotidique prédéterminée qui se lie spécifiquement à des séquences nucléotidiques complémentaires d'anti-empreinte génétique, chaque séquence nucléotidique d'anti-empreinte génétique étant fixée à un monomère marqueur ou à une combinaison de monomères marqueurs, dans laquelle la population est en solution et dans laquelle le marqueur possède un signal détectable qui est la fluorescence, la luminescence, la colorimétrie, une tension, un courant ou un champ magnétique.

3. Population variée selon l'une quelconque des revendications 1 à 2, dans laquelle chaque sonde est synthétique.

EP 1 448 581 B1

4. Population variée selon l'une quelconque des revendications 1 à 3, dans laquelle chaque sonde est un ADN fixé à un acide nucléique de pontage.
- 5 5. Population variée selon l'une quelconque des revendications 1 à 4, dans laquelle lesdits monomères marqueurs sont fixés à des dendrimères.
6. Population variée selon l'une quelconque des revendications 1 à 5, dans laquelle les empreintes génétiques sont des ADN.
- 10 7. Population variée selon l'une quelconque des revendications 1 à 6, dans laquelle les sondes comprennent des étiquettes.
8. Population variée selon la revendication 7, dans laquelle les étiquettes sont des étiquettes de biotine.
- 15 9. Population variée selon l'une quelconque des revendications 1 à 8, dans laquelle un point quantique est utilisé comme monomère marqueur.
10. Population variée selon l'une quelconque des revendications 1 à 8, dans laquelle un fluorophore est utilisé comme monomère marqueur.
- 20 11. Population variée selon l'une quelconque des revendications 1 à 10, comprenant 50, 100, 200, 500, 1000, 2000, 5000, 1×10^4 , 3×10^4 , 1×10^5 ou plus sondes d'acides nucléiques différentes.
- 25 12. Population variée selon l'une quelconque des revendications 1 à 11, dans laquelle les sondes d'acides nucléiques sont séparées les unes des autres.
13. Population variée selon la revendication 12, dans laquelle les sondes d'acides nucléiques sont étalées sur une surface bidimensionnelle.
- 30 14. Population variée selon l'une quelconque des revendications 1 à 13, dans laquelle une ou plusieurs des sondes spécifiques d'une cible s'hybrident à un analyte et dans laquelle le signal provenant de la sonde hybridée spécifique d'une cible identifie exclusivement l'analyte.
- 35 15. Procédé de détection d'un analyte d'acide nucléique, comprenant :
 - (a) le contact d'un mélange d'analytes d'acides nucléiques avec la population variée de sondes marquées de manière unique selon l'une quelconque des revendications 1 à 11 dans des conditions suffisantes pour l'hybridation des sondes aux analytes d'acides nucléiques cibles, et
 - (b) la mesure du signal résultant provenant d'une ou de plusieurs des sondes spécifiques d'une cible hybridées à un analyte, le signal identifiant exclusivement l'analyte.
- 40 16. Procédé de quantification d'un analyte d'acides nucléiques, comprenant :
 - (a) le contact d'un mélange d'analytes d'acides nucléiques avec la population variée de sondes marquées de manière unique selon l'une quelconque des revendications 1 à 11 dans des conditions suffisantes pour l'hybridation des sondes aux analytes d'acides nucléiques cibles ;
 - (b) la mesure du signal résultant provenant d'une ou de plusieurs des sondes spécifiques d'une cible hybridées à un analyte, le signal identifiant exclusivement l'analyte, ce qui permet ainsi d'identifier l'analyte d'acide nucléique cible ; et
 - (c) le comptage des molécules individuelles de l'analyte d'acide nucléique dans l'échantillon, pour ainsi quantifier l'analyte d'acide nucléique.
- 50 17. Procédé selon l'une quelconque des revendications 15 et 16, dans lequel la position du monomère marqueur ou de la combinaison de monomères marqueurs est utilisée comme caractéristique distinctive.
- 55 18. Procédé selon l'une quelconque des revendications 15 à 17, dans lequel les sondes d'acides nucléiques sont soumises à une technique d'étirage en écoulement, à une technique du ménisque descendant, à une technique d'électro-étirage, ou à un resserrement dans l'écoulement d'un liquide contenant lesdites sondes d'acides nucléiques

conjointement avec un champ électrique oscillant.

5

10

15

20

25

30

35

40

45

50

55

Figure 1

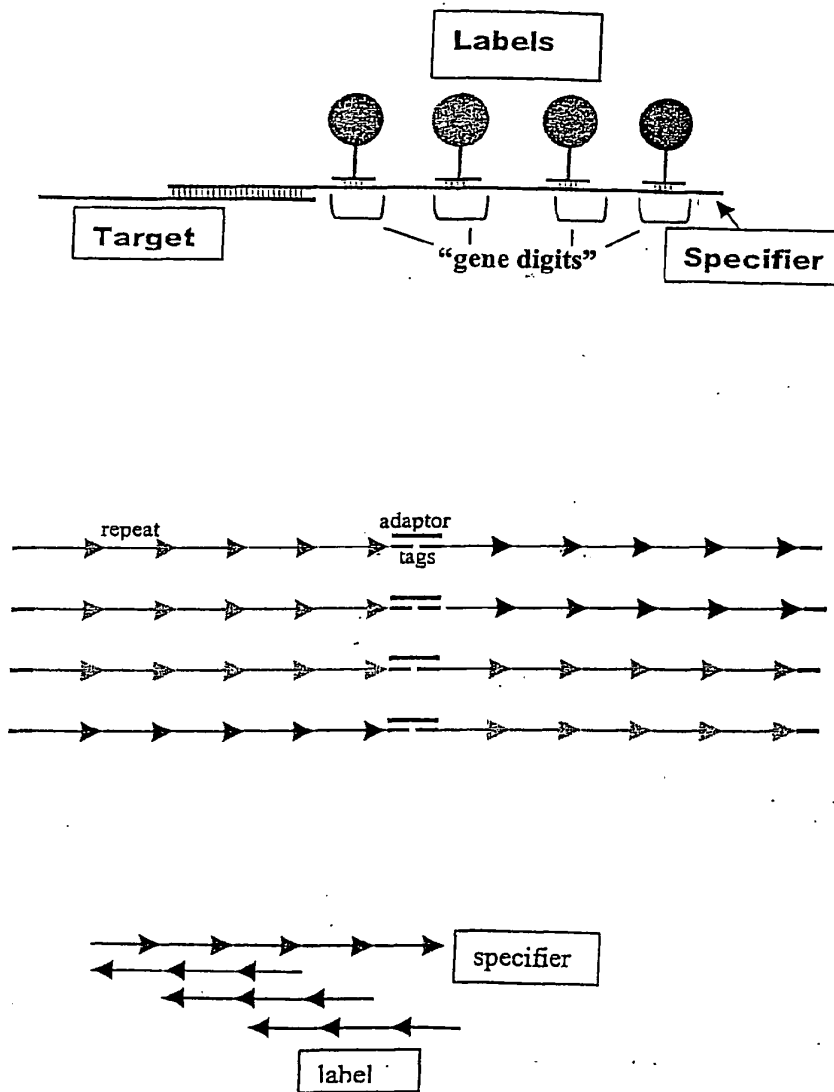


Figure 2

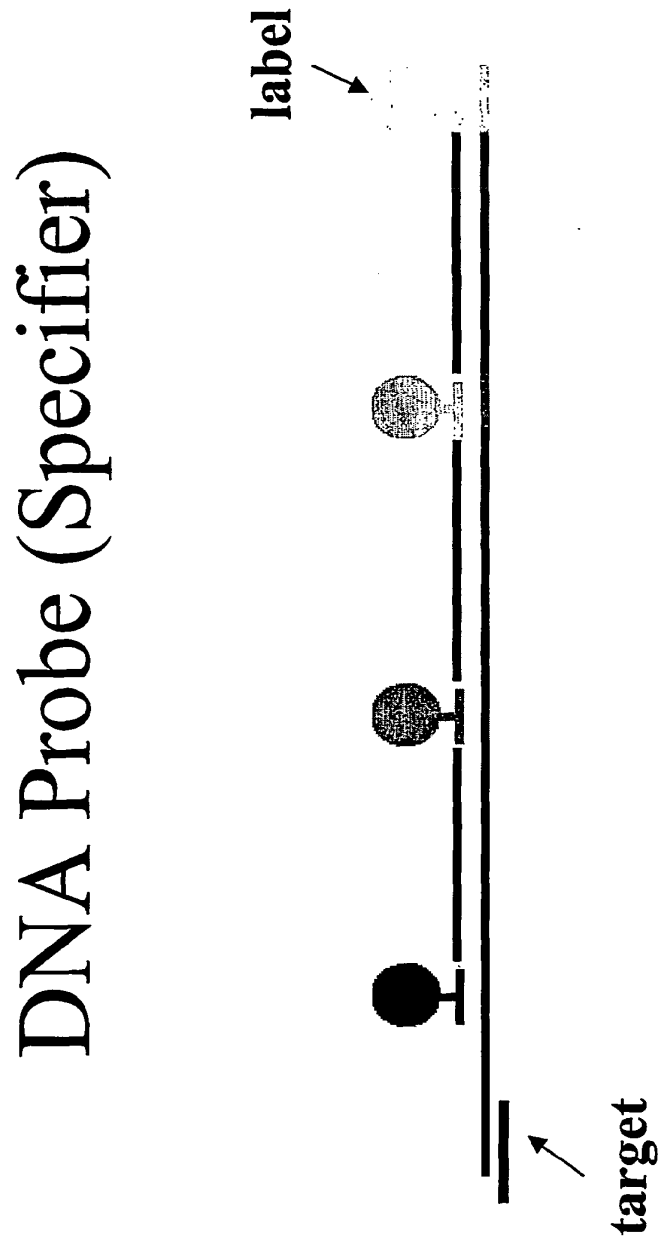


Figure 3

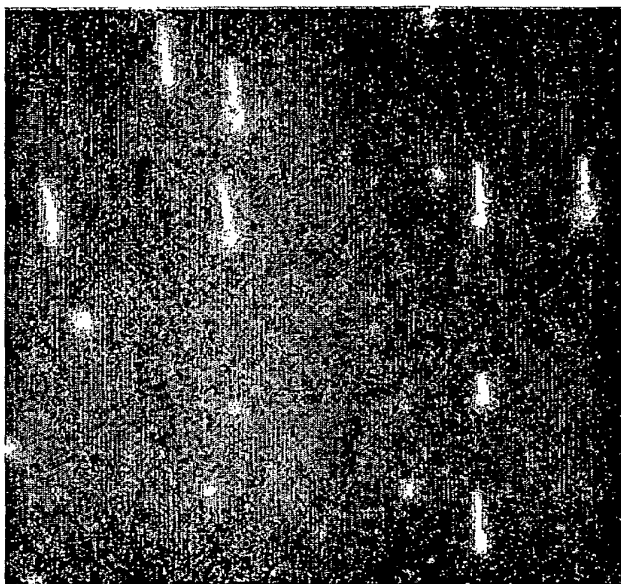
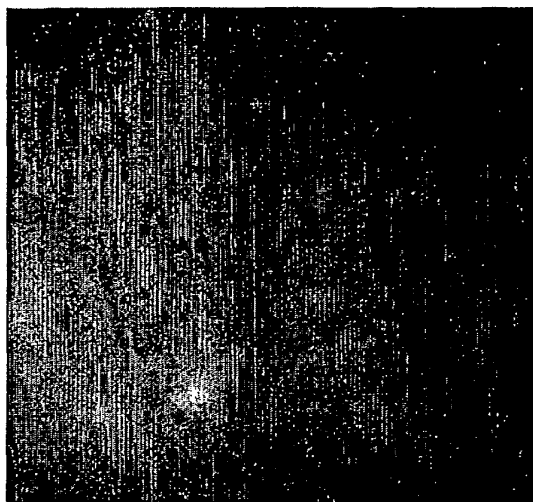
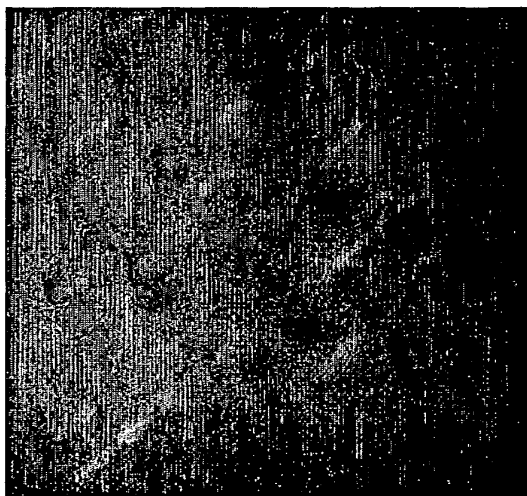


Figure 4

A



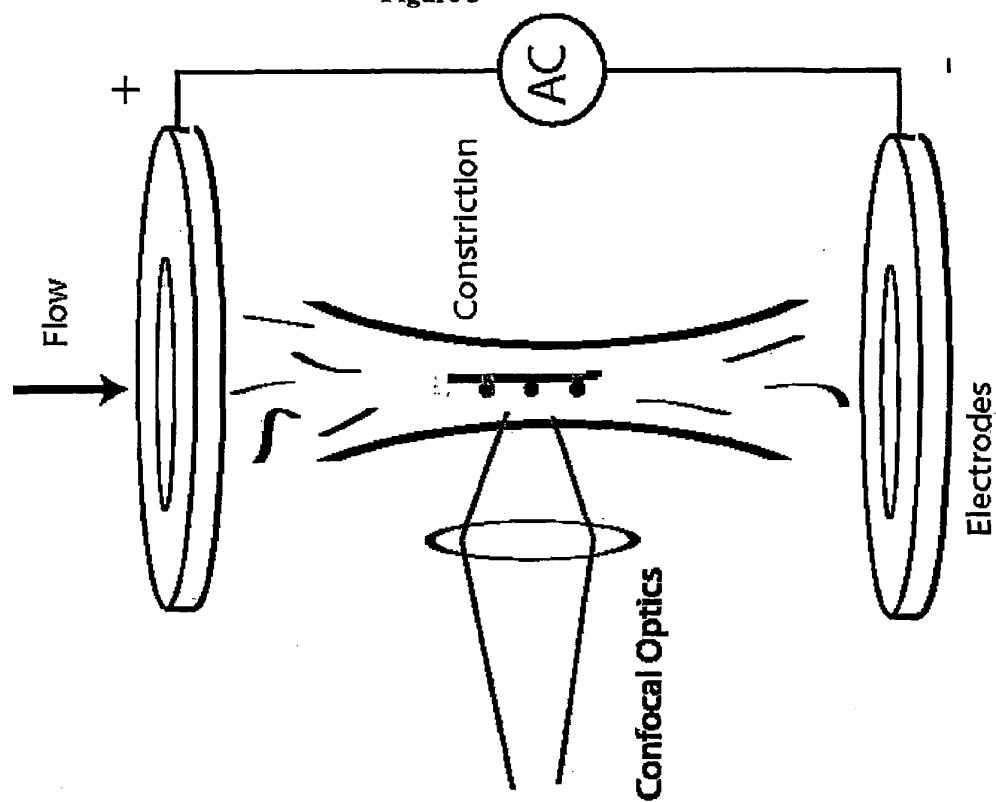
B



DNA alignment in AC field

EP 1 448 581 B1

Figure 5



REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5293050 A [0042]
- US 5354707 A, Chapple-Sokol [0042]
- US 0221278 W [0133]
- US 09898743 B [0133]

Non-patent literature cited in the description

- **COLLINS et al.** *Nucleic Acids Research*, 1997, vol. 25 (15), 2979-2984 [0007]
- **ANDERSON, M.L.M.** *Nucleic Acid Hybridization*. Springer-Verlag, 1999 [0027]
- **WANG ; GIESE.** *Anal. Chem.*, 1993, vol. 65, 3518 [0046]
- **JAYASENA, S.D.** *Clinical Chemistry*, 1999, vol. 45 (9), 1628-1650 [0073]
- **HAMES ; HIGGINS.** *Nucleic Acid Hybridisation: A Practical Approach*. IRL Press, 1991 [0099]
- *Current Protocols in Molecular Biology*. Greene Publ, 1989 [0099]
- **HENEGARIU et al.** *Biotechniques*, 2001, vol. 31, 246-250 [0105]
- **YOKOTA et al.** *Nuc. Acids Res.*, 1997, vol. 25, 1064-1070 [0105]
- **MATSUURA et al.** *Nuc. Acids Res.*, 2001, vol. 29, E79 [0105]
- **ASBURY, C.L. ; VAN DEN ENGH, G.** *Biophys. J.*, 1998, vol. 74, 1024-1030 [0132]



(11) **EP 1 629 108 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
31.12.2014 Bulletin 2015/01

(51) Int Cl.:
C12Q 1/68 (2006.01) C12Q 1/100 (2006.01)

(21) Application number: **04751175.3**

(86) International application number:
PCT/US2004/013652

(22) Date of filing: **29.04.2004**

(87) International publication number:
WO 2004/099434 (18.11.2004 Gazette 2004/47)

(54) **HOMOGENEOUS MULTIPLEX SCREENING ASSAYS AND KITS**

HOMOGENE MULTIPLEX-SCREENING-TESTS UND KITS

DOSAGES BIOLOGIQUES MULTIPLEXES HOMOGENES ET KITS ASSOCIES

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IT LI LU MC NL PL PT RO SE SI SK TR**

(56) References cited:
**WO-A-00/65094 WO-A-01/71044
WO-A-03/100101 WO-A1-03/020967
WO-A2-02/100157 US-A- 6 007 994
US-B1- 6 738 502**

(30) Priority: **30.04.2003 US 426556**

(43) Date of publication of application:
01.03.2006 Bulletin 2006/09

(73) Proprietor: **PHRI Properties, Inc.
Newark, NJ 07107-3001 (US)**

(72) Inventor: **KRAMER, Fred, R.
Riverdale, NY 10463 (US)**

(74) Representative: **Leoni, Richard Pietro et al
Avidity IP
Broers Building
Hauser Forum
21 J J Thomson Ave
Cambridge Cambridgeshire CB3 0FA (GB)**

- LEE L.G. ET AL.: 'Seven color homogeneous detection of six PCR products' BIOTECHNIQUES vol. 27, no. 2, 1999, pages 342 - 349, XP000851393
- HAN M. ET AL.: 'Quantum dot tagged microbeads for multiplexed optical coding of biomolecules' NATURE BIOTECHNOLOGY vol. 19, 2001, pages 631 - 635, XP001069134
- RIED T. ET AL.: 'Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy' PNAS vol. 89, 1992, pages 1388 - 1392, XP002073155
- TYAGI S. ET AL.: 'Wavelength-shifting molecular beacons' NATURE BIOTECHNOLOGY vol. 18, 2000, pages 1191 - 1196, XP002215066

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

TECHNICAL FIELD

[0001] This invention relates to homogenous *in vitro* multiplexed assays, including particularly homogeneous amplification assays, for screening biological samples for the presence of any of a large number of nucleic acid sequences.

BACKGROUND

[0002] A variety of assay techniques are available for testing a biological sample obtained from any of a variety of sources for the presence of a nucleic acid sequence that may indicate the presence, for example, of a particular bacterium, virus or other pathogen, including a particular strain or mutant. Assays are also available for testing such samples for the presence of a nucleic acid sequence of the subject's own genomic DNA that may indicate the presence, for example, of one or another disease-related genetic mutation. Assays may include oligonucleotide probes bearing detectable labels, for example, P³² or fluorophores. Nucleic acids, either DNA or RNA, in a sample may be probed directly. Alternatively, assays may include amplification of target sequences by any of several amplification techniques, for example, PCR, NASBA or TMA. Amplification assays may be monitored in real time utilizing intercalating dyes, for example SYBR green, or fluorescently labeled probes, such as 5' nuclease probes Livak, K.J. et al. (1995), Oligonucleotides with fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization, PCR Meth. Appl. 4: 357-362, dual FRET probes, Espy, M.J. et al. (2002), Detection of Vaccinia Virus, Herpes Simplex Virus, Varicella-Zoster Virus, and Bacillus anthracis by LightCycler Polymerase Chain Reaction after Autoclaving: Implications for Biosafety of Bioterrorism Agents, Mayo Clin. Proc. 77: 624-628, or molecular beacon probes, Tyagi, S. and Kramer, F.R. (1996), Molecular Beacons: Probes that Fluoresce upon Hybridization, Nature Biotechnol. 14: 303-308; Tyagi, S. et al. (1998), Multicolor Molecular Beacons for Allele Discrimination, Nature Biotechnol. 16: 49-53. Real-time multiplex assays utilizing PCR amplification have been demonstrated with TaqMan dual-labeled linear probes and the 5' nuclease detection process and, alternatively, with PCR amplification and molecular beacon probes, Tyagi, S. et al. (1998), *supra*; Vet, J.A. et al. (1999), Multiplex Detection of Four Pathogenic Retroviruses Using Molecular Beacons, Proc. Natl. Acad. Sci. USA 96: 6394-6399; El-Hajj, H. et al. (2001), Detection of Rifampin Resistance in Mycobacterium tuberculosis in a Single Tube with Molecular Beacons, J. Clin. Microbiol. 39: 4131-4137. Fluorescence-based multiplex assays currently are limited to about eight targets per sample by the need to minimize overlaps in emission spectra of fluorophores and, hence, are not expandable

for use as highly multiplexed screening assays.

[0003] Highly multiplexed assays rely on spatial segregation of targets for signal resolution. Spatial segregation enables the use of coding schemes involving combinations of differently colored fluorophores (combinatorial coding), combinations of different amounts of each fluorophore (ratio coding), and both. One example of an assay with spatial segregation is fluorescence *in situ* hybridization, or FISH, for chromosomal analysis. Speicher et al. (1996), Karyotyping Human Chromosomes by Combinatorial Multi-fluor FISH, Nature Genet. 12: 368-375, for example, report the use of 27 probes combinatorially labeled using a set of six different fluorophores for analysis of chromosome spreads. Similarly, segregation of transcription sites in cell nuclei has enabled combinatorial coding utilizing multiple, singly labeled probes per site, as well as ratio coding. Singer, R.H., International (PCT) patent application WO 00/65094; Levsky, J.M. et al. (2002), Single-Cell Gene Expression Profiling, Science 297: 836-840. Another spatial-segregation probe technique is the use of multiplex probe arrays, including arrays on DNA chips. Scheina, M. et al. (1995), Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray, Science 20: 467-470; Gingeras, T.R. et al. (1998), Simultaneous Genotyping and Species Identification Using Hybridization Pattern Recognition Analysis of Generic Mycobacterium DNA Arrays. Genet. Res. 8: 435-448; Han et al. (2001), Quantum-Dot-Tagged Microbeads for Multiplexed Optical Coding of Biomolecules, Nature Biotechnol. 19:631-635. Another segregation approach is the use of electrophoresis to separate ligated probe pairs of differing lengths. Tong, A.K. et al. (2001), Combinatorial Fluorescence Energy Transfer Tags for Multiplex Biological Assays, Nature Biotechnol. 19: 756-759. In that method, a variation of the oligonucleotide ligation assay ("OLA") for SNP detection, differently labeled probes are ligated to capture probes on a target, and the hybrids are immobilized, washed, released, separated from one another by electrophoresis, and read for each probes' fluorescent code, which is a combination of colors and ratios. Array methods and electrophoretic methods remain technically complex, requiring many separate steps, including amplification, hybridization, washing and analysis.

[0004] WO-A-03/020967 disclosed a method for determining at least three analytes, wherein one probe has a first label, a second probe has a second label, and a third probe has two sub-populations: one having the first label, the other having the second label. The analytes are detected according to the signal intensities determined for the first and second labels. The system may be extended to more analytes as exemplified by a six analyte system using three labels.

[0005] There is not currently available a homogeneous fluorescence hybridization assay that is suitable for use as a highly multiplexed screening assay, despite the need for such an assay. During suppressive treatment

of the HIV-1 virus with protease inhibitors, for example, any of about 30 mutations is likely to proliferate over time and to require change in treatment. Hirsch, M.S. et al. (1998), Antiretroviral Drug Resistance testing in Adults with HIV Infection, JAMA 279: 1984-1991. Lacking a highly multiplexed screening assay, current practice is to sequence the virus in response to a patient's increase in viral load. Sequencing is made difficult by the fact that the arising mutant is not the only allele present. During initial diagnosis of a patient with particular symptoms, for example fever, there is available no highly multiplexed homogeneous assay to screen for an early indication of one of numerous possible infectious agents that may be the cause of the patient's symptoms.

[0006] An aspect of this invention is highly multiplexed homogeneous assays for screening samples for the presence of at least six target nucleic acid sequences, utilizing conventional fluorescence detection equipment and techniques, and fluorescently labeled hybridization probes.

[0007] Another aspect of this invention is kits for carrying out particular screening assays according to this invention.

SUMMARY

[0008] Assays according to this invention are highly multiplexed assays suitable for screening. By "highly multiplexed," I mean that an assay is capable of detecting any of at least six, preferably at least 10, and in certain embodiments 30-60 and even more possible target sequences.

[0009] Assays according to this invention are homogeneous nucleic acid *in vitro* assays fluorescently labeled utilizing hybridization probes. By "homogeneous", I mean that nucleic acid targets are detected in solution without segregating targets spatially and without washing away unbound hybridization probes. Fluorescence signal is obtained from the reaction mixture itself. Certain preferred embodiments include target amplification by, for example PCR or NASBA, which enables detection of small amounts of pathogens in samples. When amplification is used, the reaction vessel, for example a microcentrifuge tube, is hermetically sealed to prevent contamination of the other samples being tested at the same time, samples not yet tested, equipment and workers.

[0010] Detection in assays according to this invention is by fluorescently labeled nucleic acid hybridization probes. Fluorescent labels, when stimulated emit radiation in the visible, ultraviolet or near infrared portions of the electromagnetic spectrum. Each probe is specific to one target among the multitude of possible targets being screened. If the goal and design of an assay are to distinguish among bacterial species, one will design a probe for a species, for example against a conserved region of *M. tuberculosis*, or design a probe that hybridizes to non-identical sequences of multiple strains of the same species. If the goal, on the other hand, is to distinguish among

strains or mutant alleles, one will design each probe against a variable region to provide the necessary distinction and make the probe strain-specific. Preferred probes for use in this invention are molecular beacon probes, which may be made mismatch tolerant or mismatch intolerant, depending on the need of a particular assay. As is known in the art, choice of target sequence and probe design are balanced to minimize both false negatives and false positives.

[0011] Multiplex capability is achieved by combinatorial coding, with or preferably without added ratio coding, utilizing a panel of distinguishable fluorophores. In a combinatorial code, each code element is identified by its unique pattern of different labels, in this case colors. The combinatorial code includes up to four different fluorophores per code element. The different fluorophores are from a panel of at least four distinguishable fluorophores, more preferably a panel of 4-8 distinguishable fluorophores. The combinatorial code may be duplex coding in which the code elements comprise two colors, optionally enlarged by additional code elements of one color. The combinatorial code may be triplex coding in which the code elements comprise three colors, optionally enlarged by additional code elements of two colors (triplex code combined with a duplex code), and optionally enlarged by code elements of one color. The combinatorial code may be quadruplex coding in which the code elements comprise four colors, optionally enlarged by additional code elements of three colors, two colors or one color, or some combination thereof.

[0012] Rather than labeling individual probe molecules with multiplex colors to give the probe the pattern assigned to it as a code element, each probe is subdivided into the required number of portions, and each portion is labeled to emit a different color signal. The portions are then mixed to create a probe that, when hybridized to its target, will include the unique color signal pattern of the code element.

[0013] For the preferred highly quenched probes such as molecular beacon probes, each portion will emit a single color. For less preferred probes labeled with different interacting fluorophores, such as 5'-nuclease probes, one option is to utilize a common short-wavelength fluorophore (which I call the "harvester" fluorophore), vary the long-wavelength fluorophores (which I call the "emitter" fluorophores), and detect the signal change in the latter. In that case, each portion will emit a signal in a different color. Another option is the reverse, namely, to use a common emitter fluorophore, vary the harvester fluorophores, and detect the signal change in the latter. In that case, each portion will also emit a signal in a different color, but for purposes of coding it is the harvester that becomes the detected emitter. However, a third option is to vary both fluorophores to permit detection of a ratio change in their emissions, so that the signal from one portion might be the ratio a/b , the ratio from a second portion might be the ratio c/d , and the ratio from a third portion might be e/f , where a , b , c , d , e and

f are distinguishable fluorophores. In this case the long-wavelength fluorophores are varied and can be considered to characterize the signal one obtains from any given probe portion. For ease of understanding, descriptions of combinatorial codes in this application and in the figures generally describe embodiments in which one detects for each probe portion the emission of a particular color, which can be characterized as the emitter fluorophore. It is to be understood that the teachings also apply to embodiments utilizing ratios, and persons skilled in the art will be able to apply the teachings to those embodiments.

[0014] If one begins with a panel of eight different fluorophores, for example, and chooses to subdivide each probe into four portions and to label each of the four portions with a different color, the codebook can include as many as 70 different combinations; if one begins with the same-sized panel of fluorophores but subdivides each probe into three portions and labels each of the three portions with a different color, the codebook can include as many as 56 different combinations; and if one begins with the same-sized panel of fluorophores but subdivides each probe into two portions and labels each of the two portions with a different color, the codebook can include as many as 28 different combinations. Four-color, three-color and two-color codes may be combined with each other and even with a one-color code, but my most preferred embodiments do not include such combinations. In ratio coding differing amounts of the same label are used for differentiation. Combinatorial codes according to this invention may be augmented with ratio coding, preferably integral ratio coding, that is, ratios of whole numbers such as 2:1, 3:1, 4:1 or 5:1, and vice versa. It is preferred that portions of molecular beacon probes be labeled with a non-fluorescent quencher and one signature-color fluorophore, most preferably without added ratio coding for reasons that will be explained.

[0015] As stated above, assays according to this invention are homogeneous detection assays. There is no segregation of targets. There is no separation of unbound probes from probes bound to their targets. Consequently, in assays according to this invention there must be a detectable signal change indicative of a probe hybridizing to its target. Preferred molecular beacon probes are highly quenched when free-floating at the detection temperature in the assay but change conformation and fluoresce when hybridized to their respective targets. Another type of probe whose fluorescence is restored by hybridization to its target and that is generally suitable for use in assays according to this invention is a "yin-yang" probe, a bimolecular probe labeled with a fluorophore on a target complementary strand and a quencher on a competing strand. Li, Q. et al. (2002), A New Class of Homogeneous Nucleic Acid Probes Based on Specific Displacement Hybridization, *Nucleic Acids Res.* 30: e5. Other probe-assay schemes result in a fluorescent signal change indicative of probes hybridizing to targets. In the 5'-nuclease detection process, end-labeled linear (or random-coil)

probes with two different fluorophores interact by fluorescence resonance energy transfer (FRET) when free-floating but are cleaved during primer extension in a polymerase chain reaction (PCR) amplification, resulting in a signal change (emission from the shorter wavelength fluorophore increases, emission from the longer wavelength fluorophore decreases, and the ratio of the two changes accordingly). Livak, K.J. et al. (1995), *supra*. Another system, so-called LightCycler probes, works in the opposite manner. Each probe is a pair of fluorescently labeled linear oligonucleotides that hybridize adjacently on a target and interact by FRET, thereby producing a signal change (emission from the shorter wavelength fluorophore decreases, emission from the longer wavelength fluorophore increases, and the ratio of the two changes accordingly). Espy, M.J. et al. (2002), *supra*. For FRET systems, I refer to the shorter wavelength fluorophore as a "harvester" and the longer wavelength fluorophore as an "emitter". However, because the background resulting from FRET quenching is considerably higher than contact-mediated quenching that is commonly used in molecular beacon probes and yin-yang probes (Tyagi et al. (1998), *supra*; Marras, S.A.E. et al. (2002), Efficiencies of Fluorescence Resonance Energy Transfer and Contact-Mediated Quenching in Oligonucleotide Probes, *Nucleic Acids Res.* 30: e122), FRET quenching is not preferred for screening assays according to this invention. Assays according to this invention preferably employ probes that are quenched by contact-mediated quenching, such as molecular beacon probes or yin-yang probes. My most preferred probes are molecular beacon probes that include a non-fluorescent quencher. These probes are most preferred, not only because of their low background signal, but also because of their ease of design for a multiplex system. Bonnet, G. et al. (1998), Thermodynamic Basis of the Enhanced Specificity of Structured DNA Probes, *Proc. Natl. Acad. Sci. USA* 96: 6171-6176.

[0016] Assays according to this invention include, but are not limited to, embodiments that employ target amplification. Several exponential amplification techniques are well known. Among them are the polymerase chain reaction (PCR), which includes thermal cycling, and isothermal amplification schemes such as nucleic acid sequence-based amplification (NASBA), strand-displacement amplification (SDA), transcription-mediated amplification (TMA), rolling-circle amplification (RCA), and ramification amplification methodology (RAM). Amplification assay embodiments may utilize end-point detection or real-time detection. It is preferred to perform amplification in a sealed environment, for example sealed tubes, in the presence of the hybridization probes to minimize the possibility of cross contamination between samples that results from opening sample tubes.

[0017] While not limited by sample source, assays according to this invention advantageously may utilize human or animal samples sometimes referred to as being normally sterile, such as blood, tissue or spinal fluid. Such

samples are advantageous as compared, for example, to sputum, stool or environmental samples, because they will contain fewer extraneous DNA sequences that could potentially interfere with amplification and detection. It is known that at least small amounts of many pathogens can be found in a patient's blood, even in septic syndromes resulting from bacterial infections. It is preferred to utilize normally sterile sources rather than non-sterile sources where a particular screening assay permits a choice.

[0018] Certain amplification assays according to this invention may advantageously take advantage of primers that amplify all or a number of nucleic acid targets being screened. When screening for mutations, it may be that numerous possible mutations occur in a particular variable nucleic acid sequence of interest, and one pair of primers flanking that region can be used to amplify all such possible mutations. Certain of these so-called "universal primers" bind to conserved regions of multiple species and are used for amplifying numerous targets, for example genes of various bacteria. See, for example, Kox, L.F.F. et al. (1995), PCR Assay Based on DNA Coding for 16S rRNA for Detection and Identification of Mycobacteria in Clinical Samples. *J. Clin. Microbiol.* 33: 3225-3233; Iwen, P.C. et al. (1995), Evaluation of Nucleic Acid-Based Test (PACE 2C) for Simultaneous detection of Chlamydia trachomatis and Neisseria gonorrhoeae in Endocervical Specimens, *J. Clin. Microbiol.* 33: 2587-2591; Yang, S. et al. (2002), Quantitative Multiprobe PCR Assay for Simultaneous Detection and Identification to Species Level of Bacterial Pathogens, *J. Clin. Microbiol.* 40: 3449-3454; Schonhuber, W. et al. (2001), Utilization of mRNA Sequences for Bacterial Identification, *BMC Microbiol.* 1: 2; Wong, R.S. and Chow, A.W. (2002), Identification of Enteric Pathogens by Heat Shock Protein 60kDa (HSP 60) Gene Sequences, *FEMS Microbiol. Lett.* 206: 107-113. One pair, more likely two or three pairs, of primers may suffice for particular screening assays.

[0019] This invention includes assay kits for performing particular multiplex screening assays according to this invention. Kits include at least the intended complement of detection probes. We refer to collections of oligonucleotides as "oligonucleotide sets." When the assay is an amplification assay, for example a PCR assay, the oligonucleotide sets preferably also include the primers needed for amplification. They may also include any amplifiable control sequence. Kits may include additional assay reagents such as enzymes, amplification buffer and, if desired, sample preparation reagents for isolation of nucleic acids; that is, up to all reagents needed for performing the assay, from sample preparation through detection.

[0020] Fluorescence-based screening assays of this invention are generally adaptable to detection instruments, which may include thermal cycling capability. Instruments may use a single-wavelength source or a multi-wavelength source, for example a white light combined

with selectable filters, multiple laser sources or multiple light-emitting diode sources. Detection instruments vary in their ability to distinguish emissions from multiple fluorophores, which needs to be taken into account when selecting fluorophore panels for a particular assay. In preferred embodiments of assays according to this invention, emissions from fluorophores for each probe are balanced in intensity such that intensities from the subdivided portions are closely spaced, which aids identification. Fluorophores vary in intensity, target sequences vary in accessibility for probe hybridization, and instruments may vary in their responses to different wavelengths. Relative amounts of portions may be varied to achieve balance. Where possible, it is preferred that all portions for all probes be balanced to further aid in identification. The results of these assays can be automatically decoded and interpreted by computer programs in real time, enabling the assays to be highly accurate, and to be carried out in an automated, high-throughput fashion in clinical diagnostic settings.

[0021] The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

[0022]

FIG 1 is a table showing the numbers of distinguishable combinations obtained using two or three colors per probe and panels of 4-8 distinguishable fluorophores.

FIG 2 illustrates fluorescence emissions of a real-time amplification assay in which a single target sequence is present in the sample being tested (out of many possible target sequences that the assay is designed to detect) and one probe (labeled with three different colors of equal fluorescence intensity) among the many different probes in the assay mixture binds to the amplified target sequence.

FIG 3 illustrates fluorescence emissions of a real-time amplification assay in which two different target sequences are present at different concentrations in the sample being tested and two probes (each labeled with a different combination of three different colors of equal fluorescence intensity, no color being shared by both probes) bind to the amplified target sequences.

FIG 4 illustrates fluorescence emissions of a real-time amplification assay in which two different target sequences are present at different concentrations in the sample being tested and two probes (each labeled with a different combination of three different

colors of equal fluorescence intensity, one color being shared by both probes) bind to the amplified target sequences.

FIG. 5 illustrates fluorescence emissions of a real-time amplification assay in which two different target sequences are present at the same concentration in a sample being tested, and two probes (labeled according to the first ternary color codebook shown in Fig. 6a) share two colors, leading to unambiguous interpretation of the results.

FIGS. 6a and 6b contain tables that show the effect of alternative probe coding schemes for detecting ten different possible targets in a sample, utilizing probes that are each labeled with three different colors of equal fluorescence intensity from a palette of five different fluorophores, on the results that would be obtained for every one of the 45 possible combinations of two different targets that could be present at the same concentration in a sample being tested.

FIG. 7 illustrates fluorescence emissions of a real-time amplification assay in which two different target sequences are present at the same concentration in a sample being tested, and two probes (labeled according to the first ternary color codebook shown in Fig. 6a) share one color, leading to an ambiguous result.

FIG 8 illustrates fluorescence emissions of a real-time amplification assay in which the same sample used in the assay shown in Fig. 7 is retested using probes labeled according to the second ternary color codebook shown in Fig. 6b, providing an example of how a second assay utilizing an alternative coding scheme can resolve ambiguous results.

DETAILED DESCRIPTION

[0023] FIG 1 is a table presenting certain combinatorial coding schemes useful in this invention. The top portion of the table presents the number of codes obtained by having each probe differently labeled with two different fluorophores from panels of 4, 5, 6, 7 or 8 fluorophores. If the panel includes eight fluorophores, there are 28 possible codes. The bottom portion of the table presents the number of codes obtained by having each probe labeled with three different fluorophores from the same panels of 5, 6, 7 or 8 fluorophores. If the panel includes eight fluorophores, there are 56 possibilities. Assays according to this invention utilize two or three (and sometimes four) differently labeled portions of each probe and a panel of fluorophores of sufficient size to permit at least six differently coded probes. This may be accomplished, for example, by combining the two portions of the table, that is, using both codes of two colors and codes of three

colors. For example, probes may be subdivided into three portions for use with a panel of five fluorophores, providing up to ten differently coded probes, as outlined in the bottom portion of the table, and additional probes may be subdivided into two portions for use with a panel of five fluorophores, providing up to ten additional differently coded probes, as outlined in the top portion of the table. In this manner, as many as 20 differently coded probes can be combined in the same assay mixture.

[0024] Assays according to this invention utilize one probe (which may be a single molecule probe, such as a molecular beacon probe or a TaqMan probe, or a bi-molecular probe, such as a yin-yang probe or a LightCycler probe pair) for each target. Coding is achieved, except for a code element represented by a single color, by subdividing probes into portions and labeling the portions according to a combinatorial code or combinatorial-plus-ratio code from a panel of at least four, preferably 5-8, distinguishable fluorophores. Utilizing our preferred molecular beacon probes for purposes of illustration, each probe portion is designed to have a different color signature. Certain of the probes may be wavelength-shifting probes, that is, probes in which one arm is labeled with a quencher and the other arm is labeled with a FRET pair comprising a "harvester" fluorophore to absorb energy from the stimulating light source and transfer that energy to an "emitter" fluorophore that fluoresces strongly at its own characteristic longer wavelength spectrum when the probe is in its open configuration. We utilize wavelength-shifting probes with instruments that employ a single wavelength source that is not highly absorbed by fluorophores that emit visible fluorescence at longer wavelengths, for example, Texas Red. See Tyagi, S. et al. (2000), Wavelength-Shifting Molecular Beacons, *Nature Biotechnol.* 18: 1191-1196; El-Hajj, H. et al. (2001), *supra*. It is also possible to utilize one fluorophore to quench another fluorophore by contact-mediated quenching. This creates two additional possibilities. First, a fluorophore that has relatively low emission in the system being used could be placed on each arm so that the fluorophores quench one another when the probe is closed but both fluoresce with the same color, increasing its intensity, when the probe is hybridized to its target and open. This technique can be used to add ratio coding to a combinatorial code. Second, one portion of a probe can carry two code elements (colors) by placing different fluorophores on the two arms, so that the fluorophores quench one another when the probe is closed but each fluoresces in its own characteristic emission spectrum when the probe is hybridized to its target and open. This can be used to combine two portions into one and thereby reduce the number of different portions needed for a two-color, three-color or four-color code element. It is equivalent to having separate portions. Because quenching by non-fluorescent quenchers is more efficient than quenching by fluorophores and because having two emitting fluorophores on the same molecule removes the option to balance their intensities by means of relative con-

centration, preferred embodiments of this invention utilize a non-fluorescent quencher on one arm of molecular beacon probes. Dabcyl is a preferred quencher, as is Black Hole Quencher No. 2.

[0025] "Codebooks" useful in this invention utilize four colors per probe, three colors per probe, two colors per probe, or combinations of four, three or two colors per probe, with or without single-color probes. Preferred codebooks utilize either only four colors per probe, only three colors per probe or only two colors per probe. The reason these are preferred is that the appearance of, for example, one color arising at a particular PCR cycle signals that an error has been made in assay preparation. If single-color probes are used, such an error is not obvious. In some embodiments of this invention, combinatorial coding may be supplemented with ratio coding to increase the number of code elements or possibilities. Preferred ratios are integral, from 5:1 to 1:5, preferably from 3:1 to 1:3. Ratio coding is most appropriate for screening assays in which it is highly likely that only one target will be found. The reason for this is that when the code is strictly combinatorial and two or more targets are present, changes in intensity levels are helpful for resolving ambiguities that may arise in analysis of the emission spectrum. Therefore, preferred assays that are not highly likely to find only one target utilize strictly combinatorial coding.

[0026] Various fluorophores can be included in panels of fluorophores. Certain fluorophores have relatively narrow emission spectra, for example, Vic, Alexa 488, and quantum dots (Gao et al. (2002), Quantum-Dot Nanocrystals for Ultrasensitive Biological Labeling and Multicolor Optical Encoding, J. Biomed. Opt. 7, 532-537), making them particularly useful for highly multiplexed assays utilizing panels of more than six fluorophores.

[0027] Oligonucleotide probes, including molecular beacon probes, are conveniently prepared by solid-phase DNA synthesis on, for example, an Applied Biosystems 394 DNA/RNA synthesizer. DNA nucleotides or RNA nucleotides may be used, including modified nucleotides such as 2'-O-methylribonucleotides, which yield strands resistant to cleavage. Our preferred method of molecular beacon synthesis starts with controlled-pore glass columns containing the quencher dabcyl for incorporation at the 3' end of the probe. These columns are available from Biosearch Technologies (Novato, California, USA). Our preferred method of incorporating fluorophores at the 5' end of the probe is to utilize thiolmodifier phosphoramidites or aminomodifier phosphoramidites so that iodoacetylated fluorophore derivatives can be coupled to the 5'-thiol groups and succinimidyl esters of fluorophores can be coupled to the 5'-amino groups. For tetrachlorofluorescein, our practice is to incorporate directly a tetrachlorofluorescein phosphoramidite. For wavelength-shifting probes a non-terminal fluorophore is required. We have used fluorescein phosphoramidites to incorporate internal fluorescein moieties for this purpose. Our molecular beacon probe synthesis includes

probe purification by gel exclusion chromatography through NAP-5 Sephadex columns, followed by HPLC through a reverse-phase column such as the C-18 column marketed by Waters. Finally, probes are precipitated with ethanol and dissolved in 100 μ l Tris-EDTA buffer. Molecular beacon probes should have a signal-to-noise ratio of at least 25, preferably closer to 100, keeping in mind that for a triplet code, for example, each fluorophore will be common to 30-60% of the probes.

[0028] As indicated above, the probes are subdivided into two, three or four portions for labeling and use. It is advantageous that the various portions of a probe, and where possible all portions in the assay, produce roughly the same signal intensity. Accordingly, our preferred assays and kits have portions "balanced" for that purpose. By "balanced" I mean that the fluorescence intensities of the signals from the different portions are generally within 20% of one another, most preferably within 5%. The fluorescence intensity produced by a particular probe portion, for example a molecular beacon probe portion, will depend on the intrinsic properties of its emitter fluorophore (and, if it is a wavelength-shifting molecular beacon probe, also its harvester fluorophore), the nature of the target, as some targets accept probes more readily than others, and quite possibly on the instrument utilized for detection. By testing a probe against its target in the assay on the instrument to be used, the amounts of the portions relative to one another can be adjusted to produce balanced signals. I prefer further that all portions of all of the combinatorially labeled probes be balanced. After the portions of each probe are balanced, the amounts of the probes relative to one another can be adjusted to produce balanced signals. It is possible to adjust signal readings by adjusting the detection instrument rather than the amounts of the portions of each probe to account for differences in the fluorophores and the instrument. For example, if a probe with three portions labeled a, b, c produces relative intensities with equal volumes of each portion of 1.0, 1.5 and 3.0, respectively, the instrument could be programmed to multiply readings from portion "a" by three, from portion "b" by two, and from portion "c" by one. Then only the relative amounts of the probes would have to be balanced. My preference, however, is to adjust the portion amounts rather than the instrument. Utilizing balanced emissions makes interpretation by a human observer significantly easier, as one can see immediately whether or not three colors, for example, indicate a particular code element.

[0029] Assays according to this invention are highly multiplexed. While not limited as to sample source, the assays are particularly applicable for use with samples from human or other animal body sites that are considered sterile, such as blood, tissue and cerebral spinal fluid. Non-sterile specimens, such as sputum or stool, are likely to have nucleic acids from a variety of sources and are more difficult to screen for that reason. Pathogens generally infect the blood stream to at least a small degree. However, nucleic acid amplification assays re-

quire minimal amounts of target sequence in the sample. For example, Kane *et al.* compared blood culturing to PCR assays utilizing "universal" bacterial primers to diagnose bacterial infection in critically ill surgical patients. Kane, T.D. *et al.* (1998), The Detection of Microbial DNA in Blood: a Sensitive Method for Diagnosing Bacteremia and/or Bacterial Translocation in Surgical Patients. *Ann. Surg.* 227: 1-9. Whereas blood cultures were positive for only 14% of the samples, the PCR assays were positive for 64%.

[0030] For embodiments employing amplification, such as PCR or NASBA, primers are required for the various possible targets. Primer design is known in the art. Primers can be tested with other primers and then with probes to check for undesired interactions. It is important that primers, and to a lesser extent probes, not interact in multiplex reactions. A type of high-specificity primer useful in this invention is disclosed in published international patent application WO 00/71562. Such primers form hairpin stems that significantly reduce undesired hybridizations. We have decreased primer interactions by utilizing such high-specificity primers, taking care to ensure that the melting temperature (T_m) of the stem is approximately 6°C above the primer annealing temperature utilized in the assay. We have constructed two pairs of universal primers that together amplify a segment of the 16S ribosomal RNA gene of many bacterial species. One pair includes 5'-TGACGACAACCATGCACC-3' (SEQ. ID NO. 1) and 5'-ATGTGGTTTAATTCGAAGCAA-3' (SEQ. ID NO. 2) to amplify target regions in *B. anthracis* and related gram-positive bacteria. The other pair includes 5'-GTGGACTTAGATACCCCTGGTAGTCCAC-3' (SEQ. ID NO. 3) and 5'-GCGTTCATCGAATTAA-3' (SEQ. ID NO. 4), which amplify target regions in several important gram-positive bacteria and gram-negative bacteria. The underlined sequences of the third identified probe are hairpin-forming sequences added to reduce undesired hybridizations by this primer.

[0031] Amplification assays with real-time detection can be performed on a variety of instruments that utilize different techniques for exciting fluorophores and detecting emissions. Excitation can be by lasers, white light coupled with filters, or light-emitting diodes. In addition, the assay instrument may utilize a synchronous scanning mode, in which narrow excitation and emission wavelengths are continuously varied, while maintaining a fixed wavelength difference between them. Lee *et al.* (1999), Seven-Color, Homogeneous Detection of Six PCR Products, *Biotechniques* 27, 342-349. The choice of instrumentation may affect the number of different fluorophores that can be included in a panel.

[0032] Screening assays according to this invention, including but not limited to real-time amplification assays, may be positive for a single target, in which case the target is identifiable by its unique signature, preferably its unique color signature but possibly by its color-plus-intensity signature. On occasion two or even more than

two targets may be present in a sample. Deciphering the spectrum in such a case may take advantage of the fact that, although the absolute fluorescence of a probe is dependent on target concentration, the fluorescence ratios of two probe portions hybridizing to the same target will be independent of target concentration. Taking the case of strictly combinatorial coding utilizing three-color codes wherein the intensities of the portions have been balanced, assume that two targets are present. If the targets share no color, there will result six colors. However, because equal concentrations of the two targets in the sample are highly unlikely, direct probing of the sample is likely to result in three colors of one intensity and three colors of a second intensity, thereby identifying the two targets. If the assay, on the other hand, is a real-time amplification assay, the times at which signals rise above background, for example the threshold cycle (C_T) of a PCR reaction, are likely to differ, leading to two groups of three colors, each group arising at a different time, thereby identifying the two targets.

[0033] If the codes assigned to the probes for the two targets share a color, in an assay based on directly probing the nucleic acids in a sample, the fluorescence intensity for that color will not match the fluorescence intensity for the other colors. Rather it will be a combination from two targets and will stand out from the level of the other colors, thereby identifying it as the color in common. If the assay, on the other hand, is a real-time amplification assay, three colors will come up together as signal from the more abundant target. When the second threshold for the less abundant target is reached, two more colors will come up together and the intensity curve for the common color will change slope to a higher value, thereby identifying it as the color in common.

[0034] Taking the case where ratio coding is also used, assume one color is common to two targets, but its intensity ratio for the more abundant target is 2:1 as compared to the other portions for that target. In a real-time amplification assay three colors will rise above background as the first threshold is reached, but one color (the common color) will have a higher slope than the other two colors. Nonetheless, when the second threshold is reached, the intensity curve of the common color will change slope to a higher value, signifying its commonality.

[0035] When two different target species are present in the sample at the same concentration, the colors encoding each probe will appear at the same time. In some instances of this situation it may not be possible to unambiguously identify each code. In such an instance a repeat assay utilizing an alternative coding scheme for the targets, should in almost all cases resolve the ambiguity.

Example 1

[0036] In this example, a real-time PCR amplification assay is carried out on DNA isolated from a normally

sterile blood sample suspected of containing an infectious bacterium. The sets of primers that are present in the reaction mixture enable the amplification of a segment of the 16s ribosomal RNA gene that is present in all of the bacterial species that are suspected of possibly being present in the blood sample. A multiplicity of molecular beacon probes is present in the assay, one for each of the suspected bacterial target species. Each different molecular beacon possesses a universal quencher moiety (dabcyl) covalently linked to its 3' end. Each different species-specific molecular beacon probe is labeled combinatorially according to this invention with a unique set of three differently colored fluorophores. These fluorophores are covalently linked to the 5' end of each oligonucleotide. For every species-specific probe that is present in the assay mixture, the abundance of each of the three differently colored oligonucleotides that together make up the species-specific probe is adjusted prior to performing the assay so that the fluorescence intensity of each of the three colors of that probe will be approximately the same when measured in the assay instrument used to carry out the assay. Fig. 2 is a plot of results that will be obtained in this assay. The "x" axis represents the number of amplification cycles performed by the time that each measurement of fluorescence is made. This measurement is made automatically by the assay instrument in a manner that permits the fluorescence intensity of each of the differently colored fluorophores that are used to label the probes to be measured. The "y" axis represents the fluorescence intensity due to a particular colored fluorophore. Measurements are taken during the annealing phase of each amplification cycle. Measurements are taken in real time during each amplification cycle throughout the course of the amplification. The fluorescence intensity of each of the differently colored fluorophores used to label the probes is plotted as a function of the number of amplification cycles completed.

[0037] In the experiment shown in Fig. 2, a signal appears that is sufficiently intense to be detected above the background fluorescence after a particular number of cycles of amplification have been carried out. The amplification cycle at which this occurs is known as the "threshold cycle", generally abbreviated as "C_T". The signal that appears after the threshold cycle consists of three different colors (a, b, and c). This signal rises in a linear fashion as more cycles of amplification are completed. The set of three colors that appear together in combination identify which bacterial species is present in the blood sample. The abundance of the species in the sample is determined from the threshold cycle at which the signal arises. The threshold cycle is inversely proportional to the logarithm of the number of target molecules in the sample

Example 2

[0038] In this example, an assay is carried out in the same manner and for the same purpose as the assay described in Example 1. The results that will be obtained

in this experiment are shown in Fig. 3. Two different signals occur, one (consisting of colors a, b, and c) appears early in the reaction, and the other (consisting of colors d, e, and f) appears later in the reaction. This result indicates that two different bacterial species are present in the sample, one (identified by colors a, b, and c) is relatively abundant, and the other (identified by colors d, e, and f) is relatively rare and therefore requires more cycles of amplification before there is sufficient amplified product for the fluorescence signal due to the hybridization of the species-specific probe to be seen above the background fluorescence. The unique set of colors in each of the two signals unambiguously identifies each of the two bacterial species that are present in the sample.

Example 3

[0039] In this example, an assay is carried out in the same manner and for the same purpose as the assay described in Example 1. The results that will be obtained in this experiment are shown in Fig. 4. Two different signals occur, one (consisting of colors a, b, and c) appears early in the reaction, and the other (consisting of colors d and e) appears later in the reaction. In addition, the slope of the curve showing the fluorescence intensity of color "c" increases after the threshold cycle at which the later signal (consisting of colors d and e) occurs. This result indicates that two different bacterial species are present in the sample, one (identified by colors a, b, and c) is relatively abundant, and the other (identified by colors c, d, and e) is relatively rare and appears later. The increase in the slope of the curve showing the fluorescence intensity of color "c" is due to the contributions of "c" fluorescence from probes binding to amplified segments of the DNA from the rare bacterial species being added the "c" fluorescence from probes binding to amplified segments of the DNA from the abundant bacterial species. Even though the three-color codes for each of the two bacterial species that are present in the sample have one color in common, the occurrence of the two species at different concentrations in the sample enables the signals from each species to be distinguished from one another. In general, when two (or even three) species are simultaneously present in a clinical sample, each will occur at a different concentration, and the coded signal due to the presence of each of the bacterial species can be distinguished from the signals due to the other species in the sample.

Example 4

[0040] This example illustrates how unambiguous results can be obtained even though two different species are present in a sample at approximately the same concentration. The assay described in this example is carried out in the same manner and for the same purpose as the assay described in Example 1. In this experiment, the probes present in the assay mixture are designed to de-

tect ten different bacterial species (designated species 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9). Three-color codes are used to label each of the ten probes, from a palette of five differently colored fluorophores (designated a, b, c, d, and e). For every species-specific probe that is present in the assay mixture, the abundance of each of the three differently colored oligonucleotides that together make up the species-specific probe is adjusted prior to performing the assay so that the fluorescence intensity of each of the three colors of that probe will be approximately the same when measured in the assay instrument used to carry out the assay. For this assay, the following coding scheme (which I call the "first ternary color codebook") is used:

probe for species 0	abc
probe for species 1	abd
probe for species 2	abe
probe for species 3	acd
probe for species 4	ace
probe for species 5	ade
probe for species 6	bcd
probe for species 7	bce
probe for species 8	bde
probe for species 9	cde

[0041] The results that will be obtained in the experiment carried out with probes labeled according to this codebook are shown in Fig. 5. A signal consisting of four different colors (a, b, c, and d) arises at the same threshold cycle. The slope of the curves for two of the colors (a and b) is twice as high as the slope of the curves for the other two colors (c and d). This result indicates that two different bacterial species are present in the sample at approximately the same concentration, and that the three-color codes of their respective probes have two colors in common. Since the slope of the curves for colors "a" and "b" is approximately twice as high as the slope of the curves for colors "c" and "e", colors "a" and "b" are components of the codes for both bacterial species that are present in the sample. The conclusion to be drawn from these results is that one of the probes that bound to amplified target sequences was encoded "abc" and the other probe that bound to amplified target sequences was encoded "abd". By comparing these results with the coding scheme in the first ternary color codebook (shown above), "abc" is seen to be the code for bacterial species "0" and "abd" is seen to be the code for bacterial species "1". Thus, the results unambiguously indicate that the sample contains species "0" and species "1" in approximately equal concentration.

Example 5

[0042] This example illustrates how ambiguous results can sometimes be obtained when two different species

are present in a sample at approximately the same concentration. The assay described in this example is carried out in the same manner and for the same purpose as the assay described in Example 4. The probes used in this assay utilize the first ternary color codebook (shown in Example 4).

[0043] Fig. 6a contains a two-part table (based on the first ternary color codebook) that shows the results that should occur when two different species are present in a sample at the same concentration for each of the 45 combinations of two species that are possible. In 30 of these combinations four colors arise, two at a higher rate of fluorescence intensity increase (identified by underlined letters) and two at a lower rate of fluorescence intensity increase (identified by letters that are not underlined). Each of these 30 combinations is unique within the table, therefore providing an unambiguous result.

[0044] However, from an examination of Fig. 6a it can be seen that in 15 of the 45 possible combinations of two different species that can be present in a sample at the same concentration five colors arise, one at a higher rate of fluorescence intensity increase (identified by an underlined letter) and four at a lower rate of fluorescence intensity increase (identified by letters that are not underlined). These combinations, however, are not unique. For example, the code "abcde" occurs three times in the upper table: once for the combination of species "2" and "9", a second time for the combination of species "4" and "8", and a third time for the combination of species "5" and "7". An example of an experiment that yields such an ambiguous outcome is shown in Fig. 7. The results of this assay show five colors arising at the same threshold cycle (a, b, c, d, and e), one color of which (e) increases in intensity at twice the rate as the other four colors.

Example 7

[0045] The ambiguities inherent in results such as those shown in Fig. 7 can be resolved by repeating the assay with the same probes encoded by an alternative coding scheme. For example, the assay giving results shown in Fig. 8 is carried out on the same sample used to obtain the results shown in Fig. 7, except that the probes are encoded by the following alternative coding scheme (which I call the "second ternary color codebook"):

probe for species 0	cde
probe for species 1	bde
probe for species 2	bce
probe for species 3	bcd
probe for species 4	ade
probe for species 5	ace
probe for species 6	acd
probe for species 7	abe
probe for species 8	abd

(continued)

probe for species 9 abc

[0046] Fig 8 shows the results that will be obtained from an assay on the same sample used in the experiment shown in Fig. 7, except that the probes are encoded according to the second ternary color codebook. A signal consisting of four different colors (a, b, c, and e) arises at the same threshold cycle. The slope of the curves for two of the colors (b and c) is twice as high as the slope of the curves for the other two colors (a and e). Since the slope of the curves for colors "b" and "c" are approximately twice as high as the slope of the curves for colors "a" and "e", colors "b" and "c" are present in the codes for both bacterial species that are present in the sample. The conclusion drawn from these results is that one of the probes that bound to amplified target sequences is encoded "abc" and the other probe that bound to amplified target sequences is encoded "bce". By comparing these results with the coding scheme in the second ternary color codebook (shown above), "bce" is seen to be the code for bacterial species "2" and "abc" is seen to be the code for bacterial species "9". Thus, the results shown in Fig. 8 resolve the ambiguity posed by the results shown in Fig. 7, unambiguously indicating that the sample contains species "2" and species "9" in approximately equal concentration.

[0047] Fig. 6b shows the results that should occur for all 45 different combinations of two different bacterial species in a sample at approximately the same concentration when the probes are encoded according to the second ternary color codebook. A comparison of the two-part table in Fig. 6a with the two-part table in Fig. 6b shows that all 45 possible combinations of two different bacterial species can be unambiguously identified by carrying out a second assay with alternatively coded probes when it is necessary to resolve ambiguities arising from the first assay.

[0048] A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

Claims

1. A multiplexed homogeneous *in vitro* assay to screen a sample for the presence of at least six different nucleic acid targets comprising:

a) contacting the sample with a set of at least six combinatorially coded fluorescent hybridization probes, each probe specific for a different nucleic acid target, wherein

i) each probe in said set is labeled with a

combinatorial code comprising at least two different emitter fluorophores selected from a panel of at least four spectrally differentiable fluorophores,

ii) each target specific probe is subdivided into portions, each portion comprising a population of molecules, and wherein the number of portions is equal to the number of emitter fluorophores used to label each probe,

iii) each portion is labeled with a different emitter fluorophore and wherein molecules of a portion are labeled with the same single emitter fluorophore,

iv) said portions are mixed to create a single uniquely-labeled probe specific for one of said nucleic acid targets, and

v) hybridization of each probe portion with its corresponding nucleic acid target leads to a detectable fluorescent signal change indicative of its labeling;

b) stimulating the fluorophores present with the sample; and

c) detecting changes in emissions signals from said emitter fluorophores without spatially segregating the nucleic acid targets from one another.

2. The assay according to claim 1 further comprising exponentially amplifying said nucleic acid targets, optionally wherein changes in emission signals are detected in real time during amplification.

3. The assay according to claim 2 wherein each probe comprises two, three or four differently labeled portions.

4. The assay according to any one of claims 1 to 3 wherein the relative amounts of the portions of each probe are adjusted relative to one another to provide balanced signal changes.

5. The assay according to any one of claims 1 to 4 wherein said set of probes additionally includes at least one probe that is ratio coded.

6. The assay according to any one of claims 1 to 5 wherein said combinatorially coded probes are quenched probes whose fluorescence is restored by hybridization to their targets, optionally wherein said combinatorially coded probes are molecular beacon probes having a non-fluorescent quencher on one arm and an emitter fluorophore on the other arm and preferably the fluorescence intensities of the portions of each probe are balanced relative to one another.

7. The assay according to claim 6 further comprising exponentially amplifying said nucleic acid targets in the presence of said probes and detecting increases in fluorescence from emitter fluorophores in real time.

8. The assay according to claim 7 wherein each probe comprises two, three or four differently labeled portions from a panel of four to eight differentiable fluorophores, wherein the relative amounts of the portions of each probe are adjusted relative to one another to provide balanced fluorescence emissions, optionally wherein the relative amounts of all portions of all probes are adjusted relative to one another to provide balanced fluorescence emissions.

9. A kit of reagents for use in a multiplexed *in vitro* assay to screen a sample for the presence of at least six different nucleic acid targets comprising a set of at least six combinatorially coded fluorescent hybridization probes, each probe specific for a different nucleic acid target, wherein

- i) each probe in said set is labeled with a combinatorial code comprising at least two different emitter fluorophores selected from a panel of at least four spectrally differentiable fluorophores,
- ii) each target specific probe is subdivided into portions, each portion comprising a population of molecules, and wherein the number of portions is equal to the number of emitter fluorophores used to label each probe,
- iii) each portion is labeled with a different emitter fluorophore and wherein molecules of a portion are labeled with the same single emitter fluorophore,
- iv) said portions are mixed to create a single uniquely-labeled probe specific for one of said nucleic acid targets, and
- v) hybridization of each probe portion with its corresponding nucleic acid target leads to a detectable fluorescent signal change indicative of its labeling.

10. The kit according to claim 9 further comprising:

- (a) primers for exponential amplification of said nucleic acid targets;
- (b) amplification reagents for amplifying said nucleic acid targets utilizing said primers; or
- (c) at least one probe that is ratio coded.

11. The kit according to claim 9 or claim 10 wherein the combinatorially coded probes are quenched probes whose fluorescence is restored by hybridization to their targets, optionally wherein said probes are molecular beacon probes having a non-fluorescent quencher on one arm and an emitter fluorophore on

the other arm.

12. The kit according to claim 11 wherein the relative amounts of the portions of each probe are adjusted relative to one another to provide balanced fluorescence emissions, optionally wherein the relative amounts of all portions of all probes are adjusted relative to one another to provide balanced fluorescence emissions.

Patentansprüche

1. Homogener Mehrfach-*in-vitro*-Assay zur Durchmusterung einer Probe auf die Anwesenheit von mindestens sechs Nukleinsäure-Targets, umfassend:

a) Inkontaktbringen der Probe mit einem Satz von mindestens sechs kombinatorisch kodierten fluoreszierenden Hybridisierungssonden, wobei jede Sonde spezifisch ist für ein anderes Nukleinsäure-Target, wobei

- i) jede Sonde in dem Satz mit einem kombinatorischen Code markiert ist, der mindestens zwei verschiedene Emitterfluorophore umfasst, die ausgewählt sind aus einem Panel von mindestens vier spektral unterschiedlichen Fluorophoren,
- ii) jede targetspezifische Sonde in Abschnitte unterteilt ist, wobei jeder Abschnitt eine Population von Molekülen umfasst, und wobei die Anzahl der Abschnitte gleich der Anzahl von Emitterfluorophoren ist, die verwendet werden, um die einzelnen Sonden zu markieren,
- iii) jeder Abschnitt mit einem anderen Emitterfluorophor markiert ist, und wobei Moleküle eines Abschnitts mit dem gleichen, einzigen Emitterfluorophor markiert sind,
- iv) die Abschnitte gemischt sind, um eine einzige, eindeutig markierte Sonde zu erzeugen, die spezifisch ist für eines von den Nukleinsäure-Targets, und
- v) die Hybridisierung jedes Sondenabschnitts mit seinem entsprechenden Nukleinsäure-Target zu einer nachweisbaren Fluoreszenzsignaländerung führt, die für seine Markierung bezeichnend ist;

b) Stimulieren der vorhandenen Fluorophore mit der Probe; und
c) Nachweisen von Änderungen in Emissionssignalen von den Emitterfluorophoren, ohne die Nukleinsäure-Targets räumlich voneinander zu trennen.

2. Assay nach Anspruch 1, ferner das exponentielle

- Amplifizieren der Nukleinsäure-Targets umfassend, wobei Änderungen in Emissionssignalen während der Amplifizierung optional in Echtzeit erkannt werden.
3. Assay nach Anspruch 2, wobei jede Sonde zwei, drei oder vier unterschiedlich markierte Abschnitte umfasst.
4. Assay nach einem der Ansprüche 1 bis 3, wobei die relativen Mengen der Abschnitte jeder Sonde in Bezug aufeinander angepasst sind, um ausgewogene Signaländerungen zu ermöglichen.
5. Assay nach einem der Ansprüche 1 bis 4, wobei der Satz von Sonden außerdem mindestens eine Sonde umfasst, die verhältniscodiert ist.
6. Assay nach einem der Ansprüche 1 bis 5, wobei die kombinatorisch codierten Sonden gequenchte Sonden sind, deren Fluoreszenz durch Hybridisierung an ihre Targets wiederhergestellt wird, wobei die kombinatorisch codierten Sonden optional Molecular-Beacon-Sonden sind, die einen nicht-fluoreszierenden Quencher an einem Zweig und einen Emitterfluorophor am anderen Zweig aufweisen, und wobei die Fluoreszenzintensitäten der Abschnitte jeder Sonde in Bezug aufeinander vorzugsweise ausgeglichen sind.
7. Assay nach Anspruch 6, ferner das exponentielle Amplifizieren der Nukleinsäure-Targets in Anwesenheit der Sonden und das Nachweisen von Zunahmen der Fluoreszenz von Emitterfluorophoren in Echtzeit umfassend.
8. Assay nach Anspruch 7, wobei jede Sonde zwei, drei oder vier verschieden markierte Abschnitte aus einem Panel von vier bis acht unterscheidbaren Fluorophoren umfasst, wobei die relativen Mengen jeder Sonde in Bezug aufeinander angepasst sind, um ausgewogene Fluoreszenzemissionen zu ermöglichen, wobei die relativen Mengen aller Abschnitte aller Sonden in Bezug aufeinander optional angepasst sind, um ausgewogene Fluoreszenzemissionen zu ermöglichen.
9. Kit von Reagenzien zur Verwendung in einem Mehrfach-*in-vitro*-Assay zur Durchmusterung einer Probe auf die Anwesenheit von mindestens sechs verschiedenen Nukleinsäure-Targets, einen Satz von mindestens sechs kombinatorisch codierten fluoreszierenden Hybridisierungssonden umfassend, wobei jede Sonde spezifisch ist für ein anderes Nukleinsäure-Target, wobei
- i) jede Sonde in dem Satz mit einem kombinatorischen Code markiert ist, der mindestens
- zwei verschiedene Emitterfluorophore umfasst, die ausgewählt sind aus einem Panel von mindestens vier spektral unterschiedlichen Fluorophoren,
- ii) jede targetspezifische Sonde in Abschnitte unterteilt ist, wobei jeder Abschnitt eine Population von Molekülen umfasst, und wobei die Anzahl der Abschnitte gleich der Anzahl von Emitterfluorophoren ist, die verwendet werden, um die einzelnen Sonden zu markieren,
- iii) jeder Abschnitt mit einem anderen Emitterfluorophor markiert ist, und wobei Moleküle eines Abschnitts mit dem gleichen, einzigen Emitterfluorophor markiert sind,
- iv) die Abschnitte gemischt sind, um eine einzige, eindeutig markierte Sonde zu erzeugen, die spezifisch ist für eines von den Nukleinsäure-Targets, und
- v) die Hybridisierung jedes Sondenabschnitts mit seinem entsprechenden Nukleinsäure-Target zu einer nachweisbaren Fluoreszenzsignaländerung führt, die für seine Markierung bezeichnend ist;
10. Kit nach Anspruch 9, ferner umfassend:
- (a) Primer für eine exponentielle Amplifizierung der Nukleinsäure-Targets,
- (b) Amplifikationsreagenzien zum Amplifizieren der Nukleinsäure-Targets unter Verwendung der Primer; oder
- (c) mindestens eine Sonde, die verhältniscodiert ist.
11. Kit nach Anspruch 9 oder Anspruch 10, wobei die kombinatorisch codierten Sonden gequenchte Sonden sind, deren Fluoreszenz durch Hybridisierung an ihre Targets wiederhergestellt wird, wobei die Sonden optional Molecular-Beacon-Sonden sind, die einen nicht-fluoreszierenden Quencher an einem Zweig und einen Emitterfluorophor am anderen Zweig aufweisen.
12. Kit nach Anspruch 11, wobei die relativen Mengen der Abschnitte jeder Sonde in Bezug aufeinander angepasst sind, um ausgewogene Fluoreszenzemissionen zu ermöglichen, wobei die relativen Mengen aller Abschnitte aller Sonden in Bezug aufeinander optional angepasst sind, um ausgewogene Fluoreszenzemissionen zu ermöglichen.

Revendications

1. Essai *in vitro* homogène multiplexe servant à balayer un échantillon pour détecter la présence d'au moins six cibles d'acide nucléique différentes, comprenant :

- a) la mise en contact de l'échantillon avec au moins six échantillons d'hybridation fluorescent à codage combinatoire, chaque échantillon étant spécifique à une cible d'acide nucléique différente, dans lequel
- i) chaque échantillon dudit ensemble est étiqueté avec un code combinatoire comprenant au moins deux fluorophores émetteurs différents sélectionnés dans un groupe d'au moins quatre fluorophores différenciables au niveau spectral,
 - ii) chaque échantillon spécifique à une cible est subdivisé en sections, chaque section comprenant un peuplement de molécules, et dans lequel le nombre de sections est égal au nombre de fluorophores émetteurs utilisés pour étiqueter chaque échantillon,
 - iii) chaque section est étiquetée avec un fluorophore émetteur différent et dans lequel les molécules d'une section sont étiquetées avec le même unique fluorophore émetteur,
 - iv) lesdites sections sont mélangées pour créer un seul échantillon à étiquetage unique spécifique à une desdites cibles d'acide nucléique ; et
 - v) l'hybridation de chaque section d'échantillon avec sa cible d'acide nucléique correspondante entraîne un changement de signal fluorescent détectable indiquant son étiquetage ;
- b) la stimulation des fluorophores présents avec l'échantillon ; et
- c) la détection de changements dans des signaux d'émissions desdits émetteurs fluorophores sans isoler spatialement les cibles d'acide nucléique les unes des autres.
2. Essai selon la revendication 1, comprenant en outre l'amplification exponentielle desdites cibles d'acide nucléique, dans lequel, en option, les changements des signaux d'émissions sont détectés en temps réel pendant l'amplification.
3. Essai selon la revendication 2, dans lequel chaque échantillon comprend deux, trois ou quatre sections étiquetées différemment.
4. Essai selon l'une quelconque des revendications 1 à 3, dans lequel les quantités relatives des sections de chaque échantillon sont ajustées les unes par rapport aux autres pour permettre des changements de signaux équilibrés.
5. Essai selon l'une quelconque des revendications 1 à 4, dans lequel ledit ensemble d'échantillons inclut
- au moins un échantillon qui est codé par ratio.
6. Essai selon l'une quelconque des revendications 1 à 5, dans lequel lesdits échantillons à codage combinatoire sont des échantillons trempés dont la fluorescence est restaurée par hybridation vers leurs cibles, dans lequel, en option, lesdits échantillons à codage combinatoire sont des échantillons à balise moléculaire dotés d'une balise non fluorescente sur un bras et d'un fluorophore émetteur sur l'autre bras et les intensités de fluorescence des sections de chaque échantillon sont de préférence équilibrées entre elles.
7. Essai selon la revendication 6, comprenant en outre l'amplification exponentielle desdites cibles d'acide nucléique en présence desdits échantillons et la détection d'augmentations de fluorescence provenant des fluorophores émetteurs en temps réel.
8. Essai selon la revendication 7, dans lequel chaque échantillon comprend deux, trois ou quatre fluorophores étiquetés différemment sur un groupe de quatre à huit fluorophores différenciables, dans lequel les quantités relatives de chaque échantillon sont ajustées les unes par rapport aux autres, dans lequel, en option, les quantités relatives de toutes les sections de tous les échantillons sont ajustées les unes par rapport aux autres pour permettre des émissions à fluorescence équilibrée.
9. Ensemble de réactifs destinés à un usage dans un essai *in vitro* homogène multiplexe servant à balayer un échantillon pour détecter la présence d'au moins six cibles d'acide nucléique différentes, comprenant un ensemble d'au moins six échantillons d'hybridation fluorescents à codage combinatoire, chaque échantillon étant spécifique à une cible d'acide nucléique différente, dans lequel
- i) chaque échantillon dudit ensemble est étiqueté avec un code combinatoire comprenant au moins deux fluorophores émetteurs sélectionnés dans un groupe d'au moins quatre fluorophores différenciables au niveau spectral,
 - ii) chaque échantillon spécifique à une cible est subdivisé en sections, chaque section comprenant un peuplement de molécules, et dans lequel le nombre de sections est égal au nombre de fluorophores émetteurs utilisés pour étiqueter chaque échantillon,
 - iii) chaque section est étiquetée avec un fluorophore émetteur différent et dans lequel les molécules d'une section sont étiquetées avec le même unique fluorophore émetteur,
 - iv) lesdites sections sont mélangées pour créer un seul échantillon à étiquetage unique spécifique à une desdites cibles d'acide nucléique ; et

v) l'hybridation de chaque section d'échantillon avec sa cible d'acide nucléique correspondante entraîne un changement de signal fluorescent détectable indiquant son étiquetage.

5

10. Ensemble selon la revendication 9, comprenant en outre :

(a) des primaires pour l'amplification exponentielle desdites cibles d'acide nucléique ;

10

(b) des réactifs d'amplification pour amplifier lesdites cibles d'acide nucléique en utilisant lesdits primaires ;

(c) au moins un échantillon qui est codé par ratio.

15

11. Ensemble selon la revendication 9 ou la revendication 10, dans lequel les échantillons à codage combiné sont des échantillons trempés dont la fluorescence est restaurée par hybridation vers leurs cibles, dans lequel, en option, lesdits échantillons à codage combiné sont des échantillons à balise moléculaire dotés d'une balise fluorescente sur un bras et d'un fluorophore émetteur sur l'autre bras.

20

12. Ensemble selon la revendication 11, dans lequel les quantités relatives des sections de chaque échantillon sont ajustées les unes par rapport aux autres pour permettre des émissions à fluorescence équilibrée, dans lequel, en option, les quantités relatives de toutes les sections de tous les échantillons sont ajustées les unes par rapport aux autres pour permettre des émissions à fluorescence équilibrée.

25

30

35

40

45

50

55

15

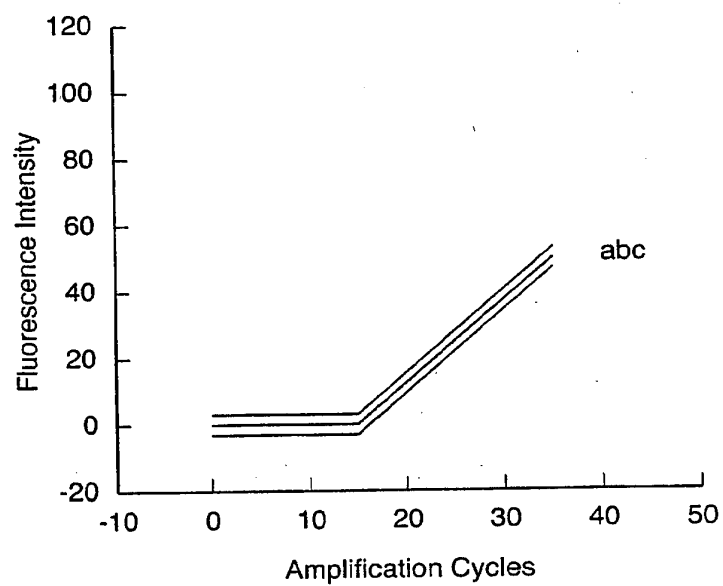


FIGURE 2

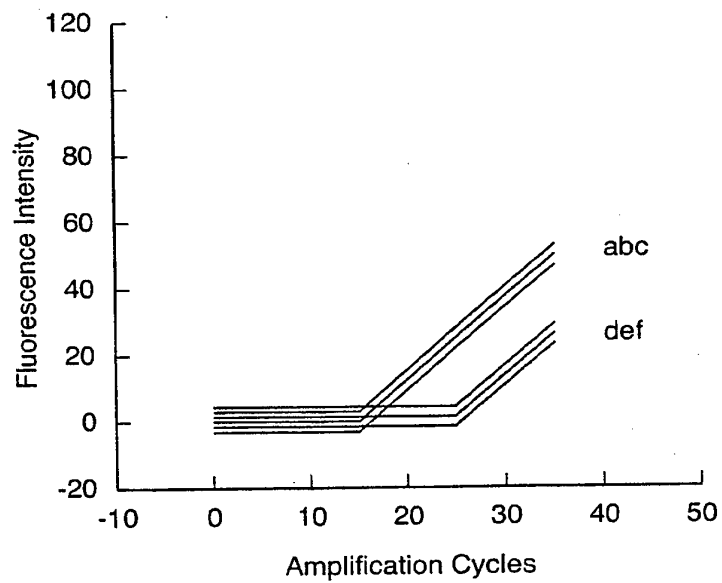


FIGURE 3

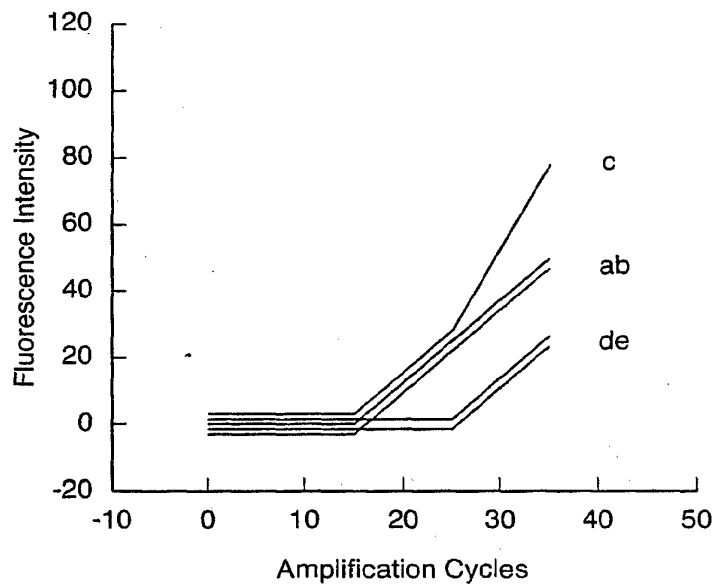


FIGURE 4

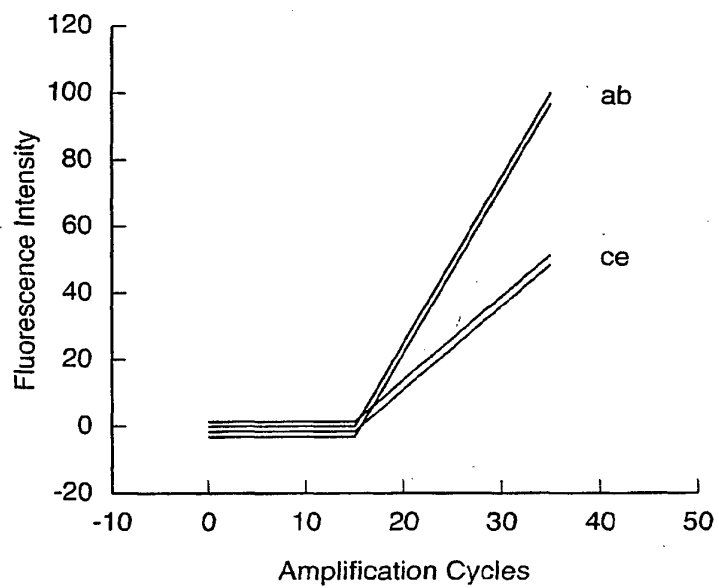


FIGURE 5

FIRST TERNARY COLOR CODEBOOK

	0(abc)	1(abd)	2(abe)	3(acd)	4(ace)
0(abc)		0+1(<u>a</u> bcd)	0+2(<u>a</u> bce)	0+3(<u>a</u> bcd)	0+4(<u>a</u> bce)
1(abd)	0+1(<u>a</u> bcd)		1+2(<u>a</u> bde)	1+3(<u>a</u> bcd)	1+4(<u>a</u> bcd)e)
2(abe)	0+2(<u>a</u> bce)	1+2(<u>a</u> bde)		2+3(<u>a</u> bcd)e)	2+4(<u>a</u> bce)
3(acd)	0+3(<u>a</u> bcd)	1+3(<u>a</u> bcd)	2+3(<u>a</u> bcd)e)		3+4(<u>a</u> cde)
4(ace)	0+4(<u>a</u> bce)	1+4(<u>a</u> bcd)e)	2+4(<u>a</u> bce)	3+4(<u>a</u> cde)	
5(ade)	0+5(<u>a</u> bcd)e)	1+5(<u>a</u> bde)	2+5(<u>a</u> bde)	3+5(<u>a</u> cde)	4+5(<u>a</u> cde)
6(bcd)	0+6(<u>a</u> bcd)	1+6(<u>a</u> bcd)	2+6(<u>a</u> bcd)e)	3+6(<u>a</u> bcd)	4+6(<u>a</u> bcd)e)
7(bce)	0+7(<u>a</u> bce)	1+7(<u>a</u> bcd)e)	2+7(<u>a</u> bce)	3+7(<u>a</u> bcd)e)	4+7(<u>a</u> bce)
8(bde)	0+8(<u>a</u> bcd)e)	1+8(<u>a</u> bde)	2+8(<u>a</u> bde)	3+8(<u>a</u> bcd)e)	4+8(<u>a</u> bcd)e)
9(cde)	0+9(<u>a</u> bcd)e)	1+9(<u>a</u> bcd)e)	2+9(<u>a</u> bcd)e)	3+9(<u>a</u> cde)	4+9(<u>a</u> cde)

	5(ade)	6(bcd)	7(bce)	8(bde)	9(cde)
0(abc)	0+5(<u>a</u> bcd)e)	0+6(<u>a</u> bcd)	0+7(<u>a</u> bce)	0+8(<u>a</u> bcd)e)	0+9(<u>a</u> bcd)e)
1(abd)	1+5(<u>a</u> bde)	1+6(<u>a</u> bcd)	1+7(<u>a</u> bcd)e)	1+8(<u>a</u> bde)	1+9(<u>a</u> bcd)e)
2(abe)	2+5(<u>a</u> bde)	2+6(<u>a</u> bcd)e)	2+7(<u>a</u> bce)	2+8(<u>a</u> bde)	2+9(<u>a</u> bcd)e)
3(acd)	3+5(<u>a</u> cde)	3+6(<u>a</u> bcd)	3+7(<u>a</u> bcd)e)	3+8(<u>a</u> bcd)e)	3+9(<u>a</u> cde)
4(ace)	4+5(<u>a</u> cde)	4+6(<u>a</u> bcd)e)	4+7(<u>a</u> bce)	4+8(<u>a</u> bcd)e)	4+9(<u>a</u> cde)
5(ade)		5+6(<u>a</u> bcd)e)	5+7(<u>a</u> bcd)e)	5+8(<u>a</u> bde)	5+9(<u>a</u> cde)
6(bcd)	5+6(<u>a</u> bcd)e)		6+7(<u>a</u> bcd)e)	6+8(<u>a</u> bcd)e)	6+9(<u>a</u> bcd)e)
7(bce)	5+7(<u>a</u> bcd)e)	6+7(<u>a</u> bcd)e)		7+8(<u>a</u> bcd)e)	7+9(<u>a</u> bcd)e)
8(bde)	5+8(<u>a</u> bde)	6+8(<u>a</u> bcd)e)	7+8(<u>a</u> bcd)e)		8+9(<u>a</u> bcd)e)
9(cde)	5+9(<u>a</u> cde)	6+9(<u>a</u> bcd)e)	7+9(<u>a</u> bcd)e)	8+9(<u>a</u> bcd)e)	

FIGURE 6a

SECOND TERNARY COLOR CODEBOOK

	9(abc)	8(abd)	7(abe)	6(acd)	5(ace)
9(abc)		8+9(<u>abcd</u>)	7+9(<u>abce</u>)	6+9(<u>abcd</u>)	5+9(<u>abce</u>)
8(abd)	8+9(<u>abcd</u>)		7+8(<u>abde</u>)	6+8(<u>abcd</u>)	5+8(<u>abcde</u>)
7(abe)	7+9(<u>abce</u>)	7+8(<u>abde</u>)		6+7(<u>abcde</u>)	5+7(<u>abce</u>)
6(acd)	6+9(<u>abcd</u>)	6+8(<u>abcd</u>)	6+7(<u>abcde</u>)		5+6(<u>acde</u>)
5(ace)	5+9(<u>abce</u>)	5+8(<u>abcde</u>)	5+7(<u>abce</u>)	5+6(<u>acde</u>)	
4(ade)	4+9(<u>abcde</u>)	4+8(<u>abde</u>)	4+7(<u>abde</u>)	4+6(<u>acde</u>)	4+5(<u>acde</u>)
3(bcd)	3+9(<u>abcd</u>)	3+8(<u>abcd</u>)	3+7(<u>abcde</u>)	3+6(<u>abcd</u>)	3+5(<u>abcde</u>)
2(bce)	2+9(<u>abce</u>)	2+8(<u>abcde</u>)	2+7(<u>abce</u>)	2+6(<u>abcde</u>)	2+5(<u>abce</u>)
1(bde)	1+9(<u>abcde</u>)	1+8(<u>abde</u>)	1+7(<u>abde</u>)	1+6(<u>abcde</u>)	1+5(<u>abcde</u>)
0(cde)	0+9(<u>abcde</u>)	0+8(<u>abcde</u>)	0+7(<u>abcde</u>)	0+6(<u>acde</u>)	0+5(<u>acde</u>)

	4(ade)	3(bcd)	2(bce)	1(bde)	0(cde)
9(abc)	4+9(<u>abcde</u>)	3+9(<u>abcd</u>)	2+9(<u>abce</u>)	1+9(<u>abcde</u>)	0+9(<u>abcde</u>)
8(abd)	4+8(<u>abde</u>)	3+8(<u>abcd</u>)	2+8(<u>abcde</u>)	1+8(<u>abde</u>)	0+8(<u>abcde</u>)
7(abe)	4+7(<u>abde</u>)	3+7(<u>abcde</u>)	2+7(<u>abce</u>)	1+7(<u>abde</u>)	0+7(<u>abcde</u>)
6(acd)	4+6(<u>acde</u>)	3+6(<u>abcd</u>)	2+6(<u>abcde</u>)	1+5(<u>abcde</u>)	0+6(<u>acde</u>)
5(ace)	4+5(<u>acde</u>)	3+5(<u>abcde</u>)	2+5(<u>abce</u>)	1+5(<u>abcde</u>)	0+5(<u>acde</u>)
4(ade)		3+4(<u>abcde</u>)	2+4(<u>abcde</u>)	1+4(<u>abde</u>)	0+4(<u>acde</u>)
3(bcd)	3+4(<u>abcde</u>)		2+3(<u>bcde</u>)	1+3(<u>bcde</u>)	0+3(<u>bcde</u>)
2(bce)	2+4(<u>abcde</u>)	2+3(<u>bcde</u>)		1+2(<u>bcde</u>)	0+2(<u>bcde</u>)
1(bde)	1+4(<u>abde</u>)	1+3(<u>bcde</u>)	1+2(<u>bcde</u>)		0+1(<u>bcde</u>)
0(cde)	0+4(<u>acde</u>)	0+3(<u>bcde</u>)	0+2(<u>bcde</u>)	0+1(<u>bcde</u>)	

FIGURE 6b

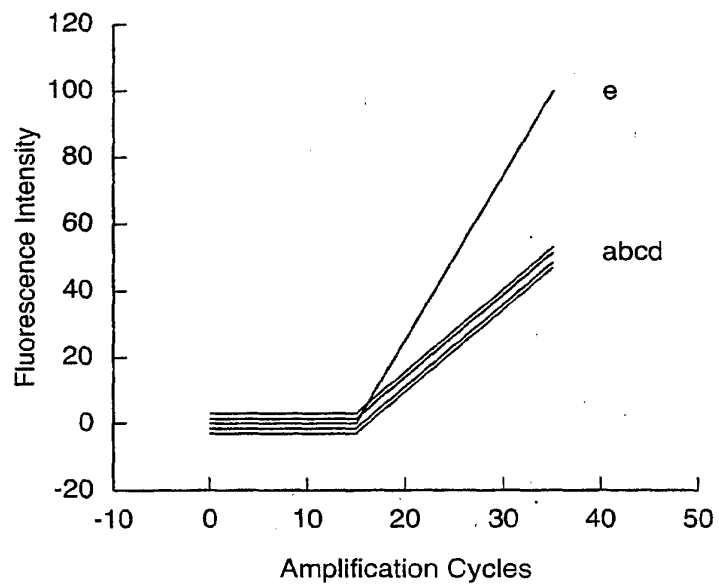


FIGURE 7

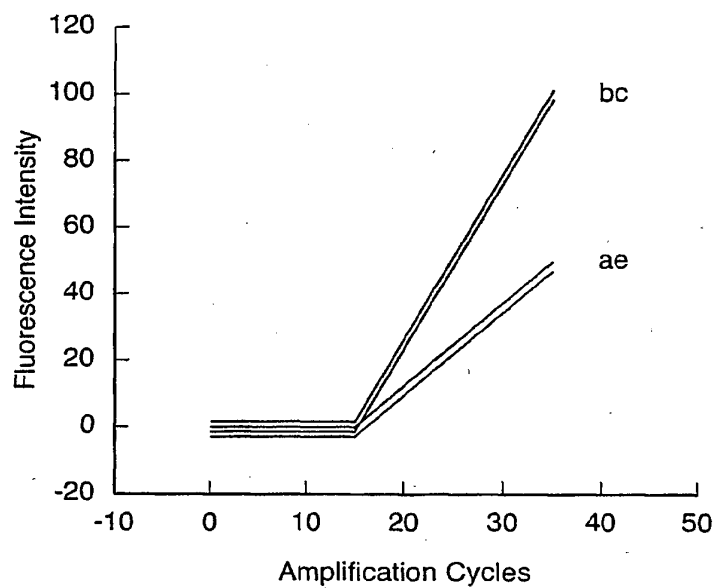


FIGURE 8

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 0065094 A, Singer, R.H. [0003]
- WO 03020967 A [0004]
- WO 0071562 A [0030]

Non-patent literature cited in the description

- **LIVAK, K.J. et al.** Oligonucleotides with fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization. *PCR Meth. Appl.*, 1995, vol. 4, 357-362 [0002]
- **ESPY, M.J. et al.** Detection of Vaccinia Virus, Herpes Simplex Virus, Varicella-Zoster Virus, and Bacillus anthracis by LightCycler Polymerase Chain Reaction after Autoclaving: Implications for Biosafety of Bio-terrorism Agents. *Mayo Clin. Proc.*, 2002, vol. 77, 624-628 [0002]
- **TYAGI, S. ; KRAMER, F.R.** Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nature Biotechnol.*, 1996, vol. 14, 303-308 [0002]
- **TYAGI, S. et al.** Multicolor Molecular Beacons for Allele Discrimination. *Nature Biotechnol.*, 1998, vol. 16, 49-53 [0002]
- **VET, J.A. et al.** Multiplex Detection of Four Pathogenic Retroviruses Using Molecular Beacons. *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, 6394-6399 [0002]
- **EL-HAJJ, H. et al.** Detection of Rifampin Resistance in Mycobacterium tuberculosis in a Single Tube with Molecular Beacons. *J. Clin. Microbiol.*, 2001, vol. 39, 4131-4137 [0002]
- **SPEICHER et al.** Karyotyping Human Chromosomes by Combinatorial Multi-fluor FISH. *Nature Genet.*, 1996, vol. 12, 368-375 [0003]
- **LEVSKY, J.M. et al.** Single-Cell Gene Expression Profiling. *Science*, 2002, vol. 297, 836-840 [0003]
- **SCHENA, M et al.** Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, 1995, vol. 20, 467-470 [0003]
- **GINGERAS, T.R. et al.** Simultaneous Genotyping and Species Identification Using Hybridization Pattern Recognition Analysis of Generic Mycobacterium DNA Arrays. *Genet. Res.*, 1998, vol. 8, 435-448 [0003]
- **HAN et al.** Quantum-Dot-Tagged Microbeads for Multiplexed Optical Coding of Biomolecules. *Nature Biotechnol.*, 2001, vol. 19, 631-635 [0003]
- **TONG, A.K. et al.** Combinatorial Fluorescence Energy Transfer Tags for Multiplex Biological Assays. *Nature Biotechnol.*, 2001, vol. 19, 756-759 [0003]
- **HIRSCH, M.S. et al.** Antiretroviral Drug Resistance testing in Adults with HIV Infection. *JAMA*, 1998, vol. 279, 1984-1991 [0005]
- **LI, Q. et al.** A New Class of Homogeneous Nucleic Acid Probes Based on Specific Displacement Hybridization. *Nucleic Acids Res.*, 2002, vol. 30, e5 [0015]
- **MARRAS, S.A.E. et al.** Efficiencies of Fluorescence Resonance Energy Transfer and Contact-Mediated Quenching in Oligonucleotide Probes. *Nucleic Acids Res.*, 2002, vol. 30, e122 [0015]
- **BONNET, G. et al.** Thermodynamic Basis of the Enhanced Specificity of Structured DNA Probes. *Proc. Natl. Acad. Sci. USA*, 1998, vol. 96, 6171-6176 [0015]
- **KOX, L.F.F. et al.** PCR Assay Based on DNA Coding for 16S rRNA for Detection and Identification of Mycobacteria in Clinical Samples. *J. Clin. Microbiol.*, 1995, vol. 33, 3225-3233 [0018]
- **IWEN, P.C. et al.** Evaluation of Nucleic Acid-Based Test (PACE 2C) for Simultaneous detection of Chlamydia trachomatis and Neisseria gonorrhoeae in Endocervical Specimens. *J. Clin. Microbiol.*, 1995, vol. 33, 2587-2591 [0018]
- **YANG, S. et al.** Quantitative Multiprobe PCR Assay for Simultaneous Detection and Identification to Species Level of Bacterial Pathogens. *J. Clin. Microbiol.*, 2002, vol. 40, 3449-3454 [0018]
- **SCHONHUBER, W. et al.** Utilization of mRNA Sequences for Bacterial Identification. *BMC Microbiol.*, 2001, vol. 1, 2 [0018]
- **WONG, R.S. ; CHOW, A.W.** Identification of Enteric Pathogens by Heat Shock Protein 60kDa (HSP 60) Gene Sequences. *FEMS Microbiol. Lett.*, 2002, vol. 206, 107-113 [0018]
- **TYAGI, S. et al.** Wavelength-Shifting Molecular Beacons. *Nature Biotechnol.*, 2000, vol. 18, 1191-1196 [0024]
- **GAO et al.** Quantum-Dot Nanocrystals for Ultrasensitive Biological Labeling and Multicolor Optical Encoding. *J. Biomed. Opt.*, 2002, vol. 7, 532-537 [0026]
- **KANE, T.D. et al.** The Detection of Microbial DNA in Blood: a Sensitive Method for Diagnosing Bacteremia and/or Bacterial Translocation in Surgical Patients. *Ann. Surg.*, 1998, vol. 227, 1-9 [0029]

EP 1 629 108 B1

- **LEE et al.** Seven-Color, Homogeneous Detection of Six PCR Products. *Biotechniques*, 1999, vol. 27, 342-349 **[0031]**



(11) **EP 1 963 531 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
21.09.2011 Bulletin 2011/38

(51) Int Cl.:
C12Q 1/68^(2006.01) C12P 19/34^(2006.01)

(21) Application number: **06848157.1**

(86) International application number:
PCT/US2006/049274

(22) Date of filing: **22.12.2006**

(87) International publication number:
WO 2007/076128 (05.07.2007 Gazette 2007/27)

(54) **NANOREPORTERS AND METHODS OF MANUFACTURING AND USE THEREOF**

NANOREPORTER UND VERFAHREN ZU DEREN HERSTELLUNG UND VERWENDUNG

NANORAPORTEURS ET PROCEDES DE PRODUCTION ET D'UTILISATION DE CEUX-CI

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI
SK TR**

- **WEBSTER, Philippa, J.**
Seattle, WA 98103 (US)
- **DIMITROV, Krassen, M.**
Seattle, WA 98115 (US)

(30) Priority: **23.12.2005 US 753758 P**
08.09.2006 US 843528 P

(74) Representative: **Crump, Julian Richard John et al**
Mintz Levin Cohn Ferris Glovsky & Popeo
Intellectual Property LLP
Alder Castle
10 Noble Street
London EC2V 7JX (GB)

(43) Date of publication of application:
03.09.2008 Bulletin 2008/36

(73) Proprietors:

- **Nanostring Technologies, Inc.**
Seattle, WA 98119 (US)
- **Institute for Systems Biology**
Seattle WA 98103-8904 (US)

(56) References cited:
WO-A-03/003810 WO-A-2005/071401
US-A1- 2002 039 732 US-A1- 2002 137 036
US-A1- 2004 091 864 US-B1- 6 268 147

(72) Inventors:

- **GEISS, Gary, K.**
Seattle, WA 98117 (US)
- **FERREE, Sean, M.**
Seattle, WA 98117 (US)

- **GEISS GARY K ET AL: "Direct multiplexed
measurement of gene expression with color-
coded probe pairs." NATURE BIOTECHNOLOGY**
MAR 2008, vol. 26, no. 3, March 2008 (2008-03),
pages 317-325, XP002505107 ISSN: 1546-1696

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

1. FIELD OF THE INVENTION

5 **[0001]** The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to coded, labeled reporter molecules, referred to herein as labeled "nanoreporters," that are capable of binding individual target molecules. Through the nanoreporters' label codes, the binding of the nanoreporters to target molecules results in the identification of the target molecules. Methods of making and using such nanoreporters are also provided. The nanoreporters can be used in diagnostic, prognostic, quality control and screening applications.

2. BACKGROUND OF THE INVENTION

15 **[0002]** This invention relates generally to the field of detection, identification, and quantification of target molecules in mixtures.

[0003] Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression are at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in normal cells compared to diseased cells can aid the discovery of new drug and diagnostic targets.

20 **[0004]** Nucleic acids can be detected and quantified based on their specific polynucleotide sequences. The basic principle underlying existing methods of detection and quantification is the hybridization of a labeled complementary probe sequence to a target sequence of interest in a sample. The formation of a duplex indicates the presence of the target sequence in the sample and the degree of duplex formation, as measured by the amount of label incorporated in it, is proportional to the amount of the target sequence.

25 **[0005]** This technique, called Molecular hybridization, has been a useful tool for identifying and analyzing specific nucleic acid sequences in complex mixtures. This technique has been used in diagnostics, for example, to detect nucleic acid sequences of various microbes in biological samples. In addition, hybridization techniques have been used to map genetic differences or polymorphisms between individuals. Furthermore, these techniques have been used to monitor changes in gene expression in different populations of cells or in cells treated with different agents.

30 **[0006]** WO 2005/071401 discloses a method for the simultaneous analysis, in a homogeneous format, for detection of multiple nucleic acid target sequences. The invention utilizes capture probes conjugate to substrates that are distinguishable based on unique spectral signature in combination with reporter probes having a separately detectable label. In particular, the substrates are subsets of microspheres, created by impregnating a combination of two fluorescent dyes in different proportions to produce a unique spectral signature for each subset. In the presence of a target sequence, detectable complexes are formed that are distinguishable for each combination of target, capture probe (carrying bead subsets containing a single spectral signature) and reporter probe (containing a single signal produced by a single fluorescent label).

35 **[0007]** US 6,268,147 relates to a method for analyzing genomic DNA and expressed sequences using auxiliary oligonucleotides, preannealed to the single-stranded target nucleic acid to form a partially duplex target molecule. Unlabeled nucleic acid analyte (the "target sequence") is denatured and annealed or hybridized with a molar excess of two or more oligonucleotide probes, at least one of which is labeled, and which bind to target sequences in one or more regions of known sequence, to form a partially duplex structure in which at least two oligonucleotide probes bind to the target sequence in tandem, forming a duplex region in which binding of at least one probe is stabilized by uninterrupted contiguous base stacking with the tandemly hybridizing probe. The probes are simply labeled with detectable labels or tags.

45 **[0008]** WO 03/003810 relates to single probes, present in populations. In particular, this document discloses a diverse population of uniquely labeled probes, containing about thirty or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid, and methods of using these probes to detect a nucleic acid analyte.

50 **[0009]** In the past, only a few genes could be detected in a complex sample at one time. Within the past decade, several technologies have made it possible to monitor the expression level of a large number of transcripts within a cell at any one time (see, e.g., Schena et al., 1995, Science 270: 467-470; Lockhart et al., 1996, Nature Biotechnology 14: 1675-1680; Blanchard et al., 1996, Nature Biotechnology 14:1649). In organisms for which most or all of the genome is known, it is possible to analyze the transcripts of large numbers of the genes within the cell. Most of these technologies employ, DNA microarrays, devices that consist of thousands of immobilized DNA sequences present on a miniaturized surface that have made this process more efficient. Using a microarray, it is possible in a single experiment to detect the presence or absence of thousands of genes in a biological sample. This allows researchers to simultaneously perform several diagnostic tests on one sample, or to observe expression level changes in thousands of genes in one experiment.

Generally, microarrays are prepared by binding DNA sequences to a surface such as a nylon membrane or glass slide at precisely defined locations on a grid. Then nucleic acids in a biological sample are labelled and hybridized to the array. The labeled sample DNA marks the exact position on the array where hybridization occurs, allowing automatic detection.

Unfortunately, despite the miniaturization of array formats, this method still requires significant amounts of the biological sample. However, in several cases, such as biopsies of diseased tissues or samples of a discrete cell type, the biological sample is in limited supply. In addition, the kinetics of hybridization on the surface of a microarray is less efficient than hybridization in small amounts of aqueous solution. Moreover, while methods exist to estimate the amount of nucleic acid present in a sample based on microarray hybridization result, microarray technology thus far does not allow for detection of target molecules on an individual level, nor are there microarray-based methods for directly quantifying the amount of target molecule in a given sample.

Thus, there exists a need for accurate and sensitive detection, identification and quantification of target molecules in complex mixtures. Discussion or citation of a reference herein shall not be construed as an admission that such reference is prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention relates to methods for the generation of a diverse populations of uniquely-labeled molecules, preferably synthetic molecules, referred to herein as nanoreporters, that can be used for the detection, identification, and direct quantification of a wide variety of target molecules. The methods are advantageous in that they generate large numbers of distinctly labeled reporter molecules, each capable of detecting a single target molecule, starting from just a small number of different types of label monomers.

In certain embodiments, the present invention provides a dual nanoreporter, or a "probe pair," comprising two components: a first probe and a second probe. The first probe is a complex comprising: (a) a first molecule, or a scaffold, comprising: (i) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal and a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached (directly, or indirectly) one or more label monomers that emit light constituting a second signal; (ii) a first target-specific sequence attached to the first molecule; and (iii) an affinity tag attached to said first molecule. The second probe is a second molecule comprising (i) a second target-specific sequence; (ii) optionally, a third label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a third signal; and (iii) an affinity tag attached to said second molecule; wherein the first target-specific sequence and the second target-specific sequence bind to different regions of the same target molecule wherein the target molecule is a naturally occurring molecule or a DNA thereof or a complement of said cDNA; wherein when said probe pair is bound to its target molecules, the identity of the first and second signals and their locations relative to each other constitutes at least part of a code that identifies the target molecule, wherein the first, second and third label attachment regions are predetermined nucleotide sequences, wherein, the first, second, and third signals are spatially distinguishable, and wherein a first DNA or RNA molecule is hybridized to the first label attachment region, to which said first DNA or RNA molecule is bound said one or more label monomers that emit light constituting said first signal; and wherein a second DNA or RNA molecule is hybridized to the second label attachment region, to which said second DNA or RNA molecule is bound said one or more label monomers that emit light constituting said second signal.

In alternative embodiments, the present invention provides a dual nanoreporters, or a "probe pair," comprising two components: a first probe and a second probe. The first probe is a complex comprising: (i) a first molecule comprising: a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal and a second label attachment region which is non-overlapping with the first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a second signal; (ii) a first target-specific sequence attached to the first and (iii) an affinity tag attached to said first molecule. The second probe is a second molecule comprising (i) a second target-specific sequence; (ii) optionally, a third label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a third signal; and (iii) an affinity tag attached to said second molecule; wherein the first signals, the second signal and the third signal are spatially or spectrally distinguishable; wherein the first target-specific sequence and the second target-specific sequence bind to different regions of the same target molecule wherein the target molecule is a naturally occurring molecule or a cDNA thereof or a complement of said DNA, wherein when said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitutes at least part of a code that identifies the target molecule, wherein the first, second, and third label attachment regions are predetermined nucleotide sequences, and wherein a first DNA or RNA molecule is hybridized to the first label attachment region, to which said first DNA or RNA molecule is bound said one or more label monomers that emit light constituting said first signal; and wherein a second

DNA or RNA molecule is hybridized to the second label attachment region, to which said second DNA or RNA molecule is bound to said one or more label monomers that emit light constituting said second signal.

[0014] The code can consist of the first and second signals and their locations relative to each other, or can consist of the first and second signals and additional signals similarly generated from additional label attachment regions (on

the first or second molecule) and the locations of the signals relative to each other.

[0015] The target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.

[0016] In a first aspect of the dual nanoreporters of the invention in which the target attachment regions are nucleotide sequences, a first DNA molecule is hybridized to the first label attachment region, to which first DNA molecule are bound (directly or indirectly) said one or more label monomers that emit light constituting said first signal; and wherein a second DNA molecule is hybridized to the second label attachment region, to which second DNA molecule are bound (directly or indirectly) said one or more label monomers that emit light constituting said second signal.

[0017] In a second aspect of the dual nanoreporters of the invention in which the target attachment regions are nucleotide sequences, a first RNA molecule is hybridized to the first label attachment region, to which first RNA molecule are bound (directly or indirectly) said one or more label monomers that emit light constituting said first signal; and wherein a second DNA molecule is hybridized to the second label attachment region, to which second DNA molecule are bound (directly or indirectly) said one or more label monomers that emit light constituting said second signal.

[0018] In a third aspect of the dual nanoreporters of the invention in which the target attachment regions are nucleotide sequences, a plurality of first DNA molecules are hybridized to the first label attachment region, to which DNA molecules are bound (directly or indirectly) said one or more label monomers that emit light constituting said first signal; and wherein a plurality of second DNA molecules are hybridized to the second label attachment region, to which second DNA molecules are bound (directly or indirectly) said one or more label monomers that emit light constituting said second signal.

[0019] In a fourth aspect of the dual nanoreporters of the invention in which the target attachment regions are nucleotide sequences, a plurality of first RNA molecules are hybridized to the first label attachment region, to which RNA molecules are bound (directly or indirectly) said one or more label monomers that emit light constituting said first signal; and wherein a plurality of second RNA molecules are hybridized to the second label attachment region, to which second RNA molecules are bound (directly or indirectly) said one or more label monomers that emit light constituting said second signal.

[0020] In certain embodiments of the dual nanoreporters of the invention, the second probe is a nucleic acid complex comprising: (a) the second nucleic acid (scaffold) molecule, wherein the second nucleic acid molecule comprises a third label attachment region to which is hybridized a third RNA molecule, to which third RNA molecule are bound (directly or indirectly) one or more label monomers that emit light constituting a third signal; and (b) the second target-specific sequence covalently attached to the second nucleic acid molecule, wherein the code comprises the identity of the first, second and third signals and their locations relative to each other. In certain embodiments, the label monomers attached to the first label attachment regions emit light at the same wavelength, which light constitutes said first signal, and the label monomers attached to the second label attachment region emit light at the same wavelength, which light constitutes the second signal, and the label monomers attached to the third label attachment region emit light at the same wavelength, which light constitutes the third signal. In other embodiments, at least one of the first signal, the second signal and the third signal comprises light at a plurality of different wavelengths. The first, second and third signals may be spectrally distinguishable. Alternatively, the first and third signals emit at the same wavelength or wavelengths.

[0021] The first probe and the second probe comprise an affinity tag. Also, in specific embodiments, the second probe comprises a third label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a third signal. Preferably, the first and second molecules are nucleic acid molecules; more preferably, the label attachment regions and target-specific sequences are predetermined nucleotide sequences. In certain embodiments of the dual nanoreporters of the invention, the one or more label monomers attached to one or both of the label attachment regions are covalently bound to nucleic acids hybridized to their respective attachment regions, or are covalently bound to nucleic acids that are bound to the respective label attachment regions indirectly through one or more bridging nucleic acids.

[0022] In certain embodiments of the dual nanoreporters of the invention, the second target-specific sequences are unlabeled with any one or more of said label monomers; however, at least one target-specific sequence in a nanoreporter is attached to a scaffold that is directly or indirectly attached to one or more label monomers.

[0023] In other embodiments of the dual nanoreporters of the invention, the second molecule further comprises a fourth label attachment region to which is attached one or more label monomers that emit light constituting a fourth signal. Preferably, when such a dual nanoreporter is bound to its target molecule, the code comprises the identity of the first signal, second signal, third signal and fourth signals and their locations relative to each other.

[0024] However, the second nucleic acid need not comprise a third label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a third signal. Such probes are referred to herein as "ghost probes." Ghost probes contain a target-specific sequence, which improves the hybridization kinetics of a nanoreporter to its target molecule, and, optionally, an affinity tag that can be used to immobilize and stretch the

dual nanoreporter. In nanoreporter embodiments employing a ghost probe, the first molecule (to which the label monomers are attached and is sometimes referred to herein as a "reporter probe") is preferably a nucleic acid molecule of about 2,000 to about 10,000 bases in length, more preferably about 5,000 to about 8,000 bases in length, and the second molecule (the "ghost probe") is preferably a nucleic acid molecule of about 40 to about 250 bases in length, more preferably about 50 to 100 bases in length. In a specific embodiment, both the first molecule and the second molecule are DNA molecules.

[0025] As used herein, the use of the terms "about" and "approximately" before a number or range means that number or range plus or minus 5%.

[0026] An affinity tag can be indirectly attached to the first probe or second probe. For example an affinity tag can be covalently attached to an oligonucleotide hybridized to the first or second probe, or it can be covalently attached to an oligonucleotide hybridized to a "patch" that is hybridized to the scaffold of the first and/or second probe. Alternatively, the affinity tag is covalently attached to the first nucleic acid molecule and/or to the second, nucleic acid molecule of the first and second probes.

[0027] The signals emitted by the label monomers attached to a given label attachment region can be the same, or different. Thus, in one embodiment, the label monomers attached to a specific (e.g., first or second) label attachment region emit light at the same wavelength, which light constitutes the signals corresponding to that label attachment region. Alternatively, in another embodiment, at least one of the first signal and the second signal comprises light at a plurality of different wavelengths.

[0028] Preferably, in certain embodiments, signals emitted by label monomers attached to adjacent labels attachment regions are spectrally distinguishable. Thus, in specific embodiments, the first and second signals are spectrally distinguishable.

[0029] We also disclose a probe, which can be a singular nanoreporter or one or both of the two components of a dual nanoreporter, said probe being a nucleic acid complex comprising (a) a nucleic acid molecule comprising (i) a first label attachment region to which is hybridized a first RNA molecule, to which first RNA molecule are bound (directly or indirectly) one or more label monomers that emit light constituting a first signal; (ii) a second label attachment region to which is hybridized a second RNA molecule, to which second RNA molecule are bound (directly or indirectly) one or more label monomers that emit light constituting a second signal; and (b) a target-specific sequence covalently attached to the nucleic acid molecule, wherein the first and second RNA molecules preferably are each at least 20 nucleotides, wherein the target-specific sequence binds to a target molecule, which target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA, and wherein when said probe is bound to its target molecules, the identity of the first and second signals and their locations relative to each other constitutes at least part of a code that identifies the target molecule. The first and second RNA molecules preferably are each at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90 or at least 100 nucleotides.

[0030] Additionally, we disclose a probe, said probe being a nucleic acid complex comprising (a) a synthetic nucleic acid molecule comprising (i) a first label attachment region to which is hybridized a first RNA molecule, to which first RNA molecule are attached one or more label monomers that emit light constituting a first signal; (ii) a second label attachment region to which is hybridized a second RNA molecule, to which second RNA molecule are attached one or more label monomers that emit light constituting a second signal; and (b) a target-specific sequence covalently attached to the synthetic nucleic acid molecule, wherein the first and second RNA molecules are each at least 50 nucleotides, wherein the target-specific sequence binds to a target molecule, which target molecule is a naturally occurring molecule or a DNA of a naturally occurring molecule or the complement of said cDNA, and wherein when said probe is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitutes at least part of a code that identifies the target molecule.

[0031] Also disclosed herein is a probe, said probe being a nucleic acid complex comprising: (a) a nucleic acid molecule comprising (i) a first label attachment region to which is hybridized a first RNA molecule, to which first RNA molecule are covalently attached one or more label monomers that emit light constituting a first signal; (ii) a second label attachment region to which is hybridized a second RNA molecule, to which second RNA molecule are covalently attached one or more label monomers that emit light constituting a second signal; and (b) a target-specific sequence covalently attached to the nucleic acid molecule, wherein the first and second RNA molecules are each at least 50 nucleotides, wherein the target-specific sequence binds to a target molecule, which target molecule is a naturally occurring molecule or a DNA of a naturally occurring molecule or the complement of said cDNA, and wherein when said probe is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitutes at least part of a code that identifies the target molecule.

[0032] In certain aspects of the present invention, the label monomers are attached to the label attachment regions of the nanoreporters of the invention by way of one or more RNA patches, or RNA molecules. Such RNA patches are preferably about (+/- 5%) 100 to about 3,000 nucleotides each, more preferably about 500 to about 1,500 nucleotides each.

[0033] The affinity tag can be attached to a patch and, in a dual nanoreporter, the affinity tag can be attached to one or more patches on one or both components of the probe pair.

[0034] The affinity tag can be attached by being covalently attached to an oligonucleotide hybridized to one or more patches on a nanoreporter. Alternatively, the affinity tag can be attached to the nucleic acid molecule that is the scaffold of a nanoreporter.

[0035] Also disclosed herein are nanoreporter-target molecule complexes. Optionally, one or more components of the nanoreporter complexes are covalently attached to an affinity tag for purification and/or for immobilization (either before or after stretching, as described in Section 5,12 below). Alternatively, the target molecule itself is attached to an affinity tag. For example, where the target molecule is a nucleic acid, the affinity tag can be biotin that is incorporated into the nucleic acid as biotin-modified nucleotides. The target molecule can be immobilized through the affinity tag prior to or after formation of the nanoreporter-target molecule complex. Where the affinity tag is a biotin moiety, the target molecule can be immobilized on a solid surface that is coated with avidin or streptavidin.

[0036] One or all of the components (e.g., scaffold, target-specific sequence) of a nanoreporter can be (or have the sequence of) a naturally-occurring molecule. However, the fully assembled and labeled nanoreporter is generally a synthetic molecule, for example a chimeric molecule made by joining naturally-occurring sequences (e.g., a viral or plasmid-based scaffold and a mammalian target-specific sequence) and/or manmade sequences.

[0037] The nanoreporters of the invention, when complexed to their target molecules, are preferably imaged when immobilized and stretched. The nanoreporter and complexes comprising the nanoreporters and target molecules can be immobilized by any method known in the art. Preferably, the nanoreporter is attached to an affinity tag (e.g., biotin or digoxigenin) that can be used to tether the nanoreporter to a surface containing a ligand for the affinity tag (e.g., streptavidin or an anti-digoxigenin antibody, respectively). The immobilized nanoreporter can be stretched by any method known in the art, including but not limited to flow-stretching, a receding meniscus technique, electrostretching, constriction in the flow of a liquid containing the nanoreporter in conjunction with an oscillating electric field.

[0038] We also disclose herein populations of one or more nanoreporters, or probes, each probe comprising a synthetic molecule, said synthetic molecule comprising, sequentially in the following order: (a) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal; (b) a second label attachment region to which is attached one or more label monomers that emit light constituting a second signal; (c) a third label attachment region to which is attached one or more label monomers that emit light constituting a third signal; and (d) a target-specific sequence that binds to a target molecule; wherein the first and second signal are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable; wherein the first and second signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the second and third signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the first and third signals are spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein each probe comprises a target-specific sequence that binds to a different target molecule; and wherein the identities of the first, second and third signals and the location of the first and third signal relative to each other constitute at least part of a code that identifies each target molecule.

[0039] Preferably, the target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said DNA.

[0040] Also preferably, the label attachment regions and target-specific sequences are predetermined nucleotide sequences.

[0041] In certain aspects of the invention, the probe pairs of the invention are purified.

[0042] In specific embodiments, the label monomers attached to the first label attachment region emit light at the same wavelength, which light constitutes said first signal; wherein the label monomers attached to the second label attachment region emit light at the same wavelength, which light constitutes the second signal; and the label monomers attached to the third label attachment region emit light at the same wavelength, which light constitutes the third signal. Thus, in such embodiments, the label monomers attached to a particular label attachment region emit light at the same wavelength (s),

[0043] In other specific embodiments, at least one of the first signal, second signal, and third signals comprises light at a plurality of different wavelengths. In such embodiments, two or more monomers attached to at least one label attachment region emit light at different wavelengths.

[0044] We also disclose herein a population of one or more nanoreporter-target molecule complexes, each complex comprising a synthetic molecule bound to a target molecule, said synthetic molecule comprising, sequentially in the following order: (a) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal; (b) a second label attachment region to which is attached one or more label monomers that emit light constituting a second signal; (c) a third label attachment region to which is attached one or more label monomers that emit light constituting a third signal; and (d) a target-specific sequence that binds to said target molecule; wherein the first and second signals are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable; wherein the first and second signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the second and third signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the first and third signals are

spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein each target-specific sequence (or, in a dual nanoreporter, pair of target-specific sequences) binds to a different said target molecule; and wherein the identities of the first, second and third signals and the locations of the first and third signal relative to each other constitute at least part of a code that identifies each target molecule.

[0045] Optionally, the populations is immobilized to a solid surface, and is preferably stretched. For example the population can be subjected to flow-stretching, a receding meniscus technique, electrostretching, or constriction in the flow of a liquid containing said complexes in conjunction with an oscillating electric field.

[0046] The present invention yet further provides methods of detecting a target molecule, said methods comprising: (a) contacting a biomolecular sample with a probe pair according to the invention under conditions that permit binding of the first target - specific sequence and the second target-specific sequence to the target molecule and (b) detecting the code that identifies a target molecule.

[0047] Preferably, the probe and the target molecule are nucleic acids. In certain specific embodiments, the label monomers attached to the first label attachment region emit light at the same wavelength, the label monomers attached to the second label attachment region emit light at the same wavelength, and the label monomers attached to the third label attachment region emit light at the same wavelength. In such embodiment, each of the first signal, second signal, and third signal consists of light at a single wavelength.

[0048] In other embodiments, at least one of the first signal, second signal, and third signal comprises light at a plurality of different wavelengths.

[0049] We also disclose an artificial nucleic acid molecule comprising one or more synthetic regions (hereinafter, "Region"), each Region being at least 50 nucleotides and comprising a regularly repeated base, said regularly repeated base having a periodicity of four to twenty five, such that said type of base occurs at regular intervals of every fourth to every twenty fifth position in said region. In a specific embodiment, not more than 5% of said base in said Region appears at other than said regular intervals. The artificial nucleic acid can be a single-stranded nucleic acid (e.g., useful as a nanoreporter scaffold) or at least partially (or fully) double-stranded (e.g., a nanoreporter scaffold to which patches are hybridized). The artificial nucleic acid can be RNA, DNA, or a DNA/RNA hybrid. Preferably the regularly repeated base is a thymidine or a uracil. In specific embodiments, the periodicity of the regularly repeated base in said one or more Regions is about 6 to about 15; alternatively, the periodicity of the regularly repeated base in said one or more Regions is about 8 to about 12.

[0050] Preferably, the artificial nucleic acid comprises a plurality of Regions, for example about 2 to about 50 Regions, or about 3 to about 10 Regions.

[0051] The plurality of Regions can be immediately adjacent to one another, or separated by one or more spacer regions. Spacer regions can include, for example, one or more restriction endonuclease recognition sequences (or the complement thereof) and/or one or more RNA polymerase promoter sequence (or the complement thereof).

[0052] Optionally, the plurality of Regions differ from one another in sequence.

[0053] The artificial nucleic acid may further comprises an origin of replication.

[0054] Preferably, the GC content of the one or more Regions is about 25% to about 75%, more preferably about 40 to about 60%, and most preferably is about 50% (+/- 5%).

[0055] Each of said one or more Region in the artificial nucleic acids of the invention is preferably about 50 to about 50,000 bases in length, and is more preferably about 200 to about 2,000 bases in length.

[0056] The artificial nucleic acids disclosed herein are Preferably about 50 to about 100,000 bases in length, and are more preferably about 200 to about 10,000 or about 500 to about 20,000 bases in length.

[0057] The artificial nucleic acids disclosed herein may comprise one or a plurality of Regions, which are useful as label attachment regions. Where an artificial nucleic acid comprise a plurality of Regions, the type of regularly repeated base can be the same in the plurality of Regions, or different. Also, the periodicity of the regularly repeated base can be the same in the plurality of Regions, or different. We therefore disclose an artificial nucleic acid comprising a plurality of Regions, in which both the type and the periodicity of the regularly repeated base are the same in the plurality of Regions.

[0058] The artificial nucleic acids disclosed herein may further comprise a target-specific sequence.

[0059] Preferably, the Regions do not comprise a recognition site for at least one six-cutter restriction endonucleotides, such as one of the following restriction endonucleases: BamHI, EcoRI, PstI, SmaI, XbaI, NotI, SacI, EcoRV, HindIII, ClaI, SalI, XhoI, ApaI or KpnI. In a specific embodiment, the Regions do not comprise any recognition site for 6-cutter enzymes, or at least 6-cutter enzymes that are used for cloning within the artificial nucleic acid, for example 6-cutter enzymes present in the polylinker of a plasmid or in regions flanking the Regions.

[0060] The artificial nucleic acids or at least the Regions therein, preferably are not predicted by the MFOLD program to contain hairpin loops.

[0061] The artificial nucleic acids or at least the Regions therein, preferably do not have direct or inverted repeats that are greater than 12 bases in length. Alternatively, the artificial nucleic acids and/or Regions do not have direct or inverted repeats that are greater than about 11, about 10 or about 9 bases in length,

[0062] The artificial nucleic acids preferably comprises a promoter sequence (such as T7, T3 or SP6) or its complement.

[0063] The artificial nucleic acids may further comprise a phage encapsidation sequence.

[0064] The Regions in the artificial nucleic acids may be flanked by restriction endonuclease recognition sites.

[0065] The artificial nucleic acids can be used as nanoreporter scaffolds, in which the Regions are useful as labels attachment regions. Accordingly, we disclose an artificial nucleic acid comprising one or more Regions, the regularly repeated base of said one or more Regions being attached to at least one light-emitting label monomer. At least one regular-repeated base can be covalently attached to the label monomer, or it can be attached to the label monomer via hybridization to another nucleic acid to which the label monomer is covalently attached. Optionally, at least two different Regions are attached to a different light-emitting labeling monomer.

[0066] The artificial nucleic acids may further comprise an affinity tag and/or an origin of replication.

[0067] Optionally, an artificial nucleic acid is a partially or fully double-stranded molecules in which one strand is DNA to which is hybridized a plurality of RNAs, each RNA comprising a single region in which the regularly-repeated base is aminoallyl-modified uracil. Optionally, at least 50% of occurrences of said aminoallyl-modified uracil is covalently attached to at least one light-emitting label monomer; in other embodiments, the percentage of occurrences is at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or at least 98%. This percentage of occurrences can be measured by any means known in the art.

[0068] As discussed elsewhere, the nanoreporters of the invention may comprise nucleic acid scaffold that are attached to patches and, optionally, flaps. Details of such structures are discussed in Section 5.3, *infra*, and exemplary patch and flap structures are depicted in Figure 7.

[0069] The present invention further provides methods of detecting a plurality of target molecules, said methods comprising: (a) contacting a biomolecular sample that may contain the target molecule with a population of probe pairs according to the invention under conditions that permit binding of the first target-specific sequence and the second target-specific sequence of each probe pair to their respective target molecule, wherein each probe pair in said population when bound to its respective target molecule is associated with a distinguishable code; and (b) detecting the codes that identify the plurality of the target molecules.

[0070] In certain aspects, the methods nanoreporter is a synthetic molecule comprising: (i) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal, all said first signals being the same; (ii) a second label attachment region to which is attached one or more label monomers that emit light constituting a second signal, all said second signals being the same; (iii) a third label attachment region to which is attached one or more label monomers that emit light constituting a third signal, all said third signals being the same; wherein each attachment region comprises a plurality of patch pairs; wherein the first and second signals are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable, and determining whether or not the nanoreporter code is present and/or produced comprises (b) detecting said first, second and third signals under conditions; wherein the first and second signals are not spatially resolvable; wherein the second and third signals are not spatially resolvable; wherein the first and third signals are spatially resolvable under said conditions; and wherein the identities of the first, second and third signals and the locations of the first and third signal relative to each other constitute at least part of a code that identifies a target molecule. Thereby, the target molecule is detected.

4. BRIEF DESCRIPTION OF THE FIGURES

[0071] FIGURE 1A-1F: Figure 1A illustrates a dual nanoreporter with a 16-position nanoreporter code, using two 8-position nanoreporter components. Figure 1B illustrates a dual nanoreporter with a 9-position nanoreporter code, using one 8-position nanoreporter component and one single-position nanoreporter component. Figure 1C illustrates a dual nanoreporter with an 8-position nanoreporter code, using one ghost probe and one 8-position nanoreporter component. Figure 1D illustrates a single nanoreporter with an 8-position nanoreporter code. In Figures 1A-1D, the star shape (depicted with an arrow) is illustrative and discussed in Section 5.3, *infra*, and exemplary patch and flap structures are depicted in Figure 7.

[0072] The present invention further provides methods of detecting a target molecule, said methods comprising: (a) contacting a biomolecular sample that may contain the target molecule with any of the nanoreporter structures described herein under conditions that permit binding of the synthetic molecule to the target molecule, (b) subjecting the conditions that permit formation and/or detection of the nanoreporter code and (c) determining whether or not the nanoreporter code is produced or present, thereby detecting said target molecule.

[0073] In certain aspects of the invention, the methods nanoreporter is a synthetic molecule comprising: (i) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal, all said first signals being the same; (ii) a second label attachment region to which is attached one or more label monomers that emit light constituting a second signal, all said second signals being the same; (iii) a third label attachment region to which is attached one or more label monomers that emit light constituting a third signal, all said third signals being the same; wherein each attachment region comprises a plurality of patch pairs; wherein the first and second signals are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable, and

determining whether or not the nanoreporter code is present and/or produced comprises (b) detecting said first, second and third signals under conditions: wherein the first and second signals are not spatially resolvable; wherein the second and third signals are not spatially resolvable; wherein the first and third signals are spatially resolvable under said conditions; and wherein the identities of the first, second and third signals and the locations of the first and third signal relative to each other constitute at least part of a code that identifies a target molecule. Thereby, the target molecule is detected.

4. BRIEF DESCRIPTION OF THE FIGURES

[0074]

FIGURE 1A-1F: Figure 1A illustrates a dual nanoreporter with a 16-position nanoreporter code, using two 8-position nanoreporter components. Figure 1B illustrates a dual nanoreporter with a 9-position nanoreporter code, using one 8-position nanoreporter component and one single-position nanoreporter component. Figure 1C illustrates a dual nanoreporter with an 8-position nanoreporter code, using one ghost probe and one 8-position nanoreporter component. Figure 1D illustrates a single nanoreporter with an 8-position nanoreporter code. In Figures 1A-1D, the star shape (depicted with an arrow) is illustrative of an affinity tag, which can be used to purify the nanoreporter or immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging. The numbered region in Figure 1A-1D refer to separate label attachment regions. All except for position 12 of Figure 1A are labeled with one of four types of label monomers, depicted as grey, white, hatched or stripe "sun" diagrams. Position 12 of Figure 1A is an unlabeled "dark spot." Figures 1E and 1F represent variations on the nanoreporters of Figures 1B and 1D, respectively, in which the target molecule to which the nanoreporters are bound comprises biotin moieties (shown as small asterisks), for example biotin-modified nucleotides randomly incorporated into a target nucleic acid. The nanoreporters themselves further optionally comprise an affinity tag (not shown).

FIGURE 2A-2C: Figure 2A shows an illustration of a label unit of a nanoreporter, containing a scaffold with patch units and corresponding split flaps disposed along its length. Figure 2B illustrates the components of a single patch pair and its corresponding flap, containing: 1: a portion of a nanoreporter scaffold (e.g., M13 single-stranded DNA); 2: A patch pair; 3: a split flap pair; and 4: labeled oligonucleotides, each with a label monomer incorporated, hybridized to the split flap. Figure 2C shows a nanoreporter with 4 "spots," each spot designed to contain 9 patch pairs of 60-65 nucleotides, each attached to a split flap pair of 95-100 nucleotides. Each split flap pair had binding site for 12 oligonucleotides each attached to a single label monomer. Each spot therefore had binding sites for 108 label monomers.

FIGURE 3: A nanoreporter in which the patches are RNA segments can be used with (Figure 3A) and without registers (Figure 3B). Both Figure 3A and 3B depict a (1) nanoreporter scaffold (heavy black line) to which are attached (2) 8 RNA segments (heavy grey lines 1-8), (3) a target-specific sequence (dotted line "T") and (4) an oligonucleotide (checkered line "O") that is partly complementary to the nanoreporter scaffold and partly complementary to the target-specific sequence. This oligonucleotide is referred to as a "ligator" oligonucleotide. In Figure 3A, only one register, i.e., every alternate RNA segment is labeled. The second register positions serve as "spacers," making it possible to generate a nanoreporter code in which consecutive positions in the code are the same "color," or spectrally indistinguishable. In Figure 3B, both registers, i.e., adjacent RNA segments with no intervening spacers, are labeled, with no nearest neighbor of the same "color."

FIGURE 4: Is an image of a dual nanoreporter hybridized to a target molecule. Here, both registers are labeled. The nanoreporters are labeled with three different colors, Alexa 488, Cy3 and Alexa 647 (labeled 1, 2 and 3, respectively). The left brackets show one probe of the dual nanoreporter and the right brackets show the other probe of the dual nanoreporter. Colors 1, 2 and 3 were each acquired in different channels and the first and second registers, seen as rows of spots, were shifted up by several pixels to be able to show each register individually.

FIGURE 5A-5D: This figure illustrates the various components of the dual nanoreporters shown in Figure 4. Figure 5A illustrates one color (here, Alexa 488, depicted in the left column as open circles), which is spectrally distinguishable from Cy3 (shown in Figure 5B, depicted in the left column as vertically striped circles) and Alexa 647 (shown in Figure 5C as diagonally striped circles). The images obtained from each were superimposed to generate Figure 5D. FIGURE 6A-6E: Figure 6A is a schematic illustration of the experiment shown in Figures 6B and 6C. In this case, the star represents biotin that was used to attach the complex by one end to the surface prior to stretching. Figures 6B and 6C show images from experiments in which S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA (Figure 6B) or S2 target RNA (Figure 6C) were hybridized. Figure 6E shows a close-up of a nanoreporter complexes from Figure 6B, each containing S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA. Figure 6D shows an image of a negative control experiment, in which S2-A ghost probe, S2-B labeled nanoreporter and no S2 target RNA were hybridized.

FIGURE 7A-G. Figures 7A, 7B, 7C and 7D depict different permutations of patches on a nanoreporter scaffold,

Figures 7E and 7F depict different permutations of split flaps on a nanoreporter scaffold, optionally hybridized to one or more oligonucleotides, as in Figure 7G. In Figure 7A-G, α refers to a 5' or 3' molecule or end of a molecule, and β refers to a corresponding 3' or 5' molecule or end of a molecule.

FIGURE 8: Figure 8 depicts a scheme in which single-stranded M13 phage is linearized for use as a nanoreporter scaffold. The circular M13 phage is annealed to a fivefold excess of BamHI cutter oligonucleotide (hatched lines) (1), and the resulting partially double-stranded M13 digested with the restriction endonuclease BamHI (2), resulting in a linearized M13 in which BamHI cutter oligonucleotide is still attached (3). This M13-oligonucleotide complex is heated in the presence of an excess oligonucleotide complementary to the BamHI cutter oligonucleotide (an "anti-BamHI oligonucleotide") (grey lines) (4). The BamHI cutter oligonucleotide anneals to the excess of anti-BamHI oligonucleotide, and the M13 molecule is purified from the oligonucleotide, for example by using size exclusion columns, to yield M13 scaffold.

FIGURE 9A-9B: Shows a labeled nanoreporter with an affinity tag at each end, A1 and A2. In Figure 9, the labeled nanoreporter is immobilized through the binding of A1 to an immobilized affinity partner. In the absence of an affinity binding partner for A2, the A2 end of the nanoreporter remains in solution (Figure 9A), but in the presence of an affinity binding partner (A2'), the A2 end of the nanoreporter is also immobilized (Figure 9B). Upon immobilization, the nanoreporter can be stretched, or "elongated" as depicted in Figure 9B, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code.

FIGURE 10A-10C: Figure 10A shows a labeled nanoreporter containing a single affinity tag, A1. Another affinity tag, A2, can be attached to the nanoreporter by direct binding of the nanoreporter to a molecule containing A2 (e.g., if the nanoreporter is or comprises a nucleic acid, it can hybridize directly with another nucleic acid to which A2 is attached), as depicted in Figure 10B. Alternatively, the second affinity tag, A2, can be attached to the labeled nanoreporter via a bridging molecule, such as the bridging nucleic acid ("X") depicted in Figure 10C.

FIGURE 11A-11B: Shows a labeled (nucleic acid-based) nanoreporter with an affinity tag, A1, at one end. In Figure 11, the labeled nanoreporter is immobilized through the binding of A1 to an immobilized affinity partner. The other end of the nanoreporter is in solution (Figure 11A), but can be immobilized by hybridization to a complementary oligonucleotide which contains another affinity tag (A2) used to immobilize the nanoreporter (Figure 11B). A and A2 can be the same, for example biotin, for immobilization on an avidin- or streptavidin-coated surface. Upon immobilization of A1, the nanoreporter can be stretched, or "elongated" as depicted in Figure 11, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code. Optionally, while the nanoreporter is in an elongated state, A2 is introduced and binds the end of the nanoreporter that is complementary to A2 down to the surface.

FIGURE 12A-12B. Figure 12A provides an illustration of a nanoreporter comprising an immobilized first portion F1; and Figure 12B provides an illustration of a nanoreporter extended in an electrical field and comprising immobilized first portion F1 and immobilized second portion F2, wherein F2 is immobilized via a complex with molecule F3.

FIGURE 13A-3C. Figure 13A provides an illustration of a three-member complex for immobilization of an extended nanoreporter; FIG. 13B provides an illustration of a two-member complex for immobilization of an extended nanoreporter; and FIG. 13C provides an illustration of an incomplete complex for immobilization of an extended nanoreporter.

FIGURE 14A-14D. Figure 14A provides an illustration of a nanoreporter comprising an immobilized first portion F1; FIG. 14B provides an illustration of an extended nanoreporter immobilized at first portion F1 and at a second portion via complexes with F2; FIG. 14C provides an illustration of a nanoreporter comprising a first portion immobilized to an avidin surface via biotin; and FIG. 14D provides an illustration of an extended nanoreporter immobilized at a first portion and at a second portion via selective binding of biotin to an avidin surface.

FIGURE 15A-15C. Figure 15A illustrates immobilization of one terminus of a DNA molecule in a microfluidic device; Figure 15B illustrates extension of the DNA in an electric field; and Figure 15C illustrates selective immobilization of a second terminus of the extended DNA molecule.

FIGURE 16 provides an image of extended nanoreporters selectively immobilized by the methods of the present invention.

FIGURE 17 depicts the relationship between the number of label attachment regions to the calculated entanglement threshold for nanoreporters for label attachment region sizes of 900 and 1100 bp

FIGURE 18 is a scatter plot showing normalized and average log2 signal values from each positive sample (n=3) for all 509 genes whose expression was measured in a nanoreporter multiplex assay as described in Example 9 (Section 14) below.

5. DETAILED DESCRIPTION OF THE INVENTION

[0075] Disclosed herein are nanoreporters, and their manufacture and use. A fully assembled and labeled nanoreporter comprises two main portions, a target-specific sequence that is capable of binding to a target molecule, and a labeled

region which emits a "code" of signals (the "nanoreporter code") associated with the target-specific sequence. Upon binding of the nanoreporter to the target molecule, the nanoreporter code identifies the target molecule to which the nanoreporter is bound.

[0076] Nanoreporters are modular structures. Generally, a nanoreporter is a molecular entity containing three basic elements: a scaffold containing two or more label attachment regions, one or more patches attached to the scaffold, and a target-specific sequence, also attached to the scaffold. The elements of a nanoreporter can be found in a single molecular entity (a "singular" nanoreporter), or two distinct molecular entities (a "dual" nanoreporter). Each molecular entity may be composed of one molecule or more than one molecule attached to one another by covalent or non-covalent means. Generally, each component of a dual nanoreporter has a target-specific sequence that binds to a different site on the same target molecule. This allows for smaller nanoreporter components with more efficient kinetics of binding of the nanoreporter to the target molecule and better signal:noise ratios resulting from the greater binding specificity.

[0077] The patches attached to a nanoreporter scaffold serve to attach label monomers to a nanoreporter scaffold. Patches may be directly labeled, for example by covalent incorporation of one or more label monomers into nucleic acid patches. Alternatively, patches may be attached to flaps, which maybe labeled directly, for example by covalent incorporation of one or more label monomers into a nucleic acid flap, or indirectly, for example by hybridization of a nucleic acid flap to an oligonucleotide which is covalently attached to one or more label monomers. Where the label monomers attached to a label attachment region are not directly incorporated into a patch or flap, the patch or flap serves as a "bridge" between the label monomer and the label attachment region, and may be referred to as a "bridging molecule," e.g., a bridging nucleic acid.

[0078] Additionally, nanoreporters may have affinity tags for purification and/or for immobilization (for example to a solid surface). Nanoreporters, or nanoreporter-target molecule complexes, are preferably purified in two or more affinity selection steps. For example, in a dual nanoreporter, one probe can comprise a first affinity tag and the other probe can comprise a second (different) affinity tag. The probes are mixed with target molecules, and complexes comprising the two probes of the dual nanoreporter are separated from unbound materials (e.g., the target or the individual probes of the nanoreporter) by affinity purification against one or both individual affinity tags. In the first step, the fixture can be bound to an affinity reagent for the first affinity tag, so that only probes comprising the first affinity tag and the desired complexes are purified. The bound materials are released from the first affinity reagent and optionally bound to an affinity reagent for the second affinity tag, allowing the separation of complexes from probes comprising the first affinity tag. At this point only full complexes would be bound. The complexes are finally released from the affinity reagents for the second affinity tag and then preferably stretched and imaged. The affinity reagent can be any solid surface coated with a binding partner for the affinity tag, such as a column, bead (e.g., latex or magnetic bead) or slide coated with the binding partner. Immobilizing and stretching nanoreporters using affinity reagents is fully described in US2010-0261026 A1.

[0079] Nanoreporter and nanoreporter-target complexes which are or comprise nucleic acids may be affinity-purified or immobilized using a nucleic acid, such as an oligonucleotide, that is complementary to at least part of the nanoreporter or target. In a specific application where the target includes a poly A or poly dA stretch, the nanoreporter-target complex can be purified or immobilized by an affinity reagent coated with a poly dT oligonucleotide.

[0080] The sequence of signals emitted by the label monomers associated with the various label attachment regions of the scaffold of a given nanoreporter allows for the unique identification of the nanoreporter. A nanoreporter having a unique identity or unique spectral signature is associated with a target-specific sequence that recognizes a specific target molecule or a portion thereof. When a nanoreporter is exposed to a mixture containing the target molecule under conditions that permit binding of the target-specific sequence(s) of the nanoreporter to the target molecule, the target-specific sequence(s) preferentially bind(s) to the target molecule. Detection of the spectral code associated with the nanoreporter allows detection of the presence of the target molecule in the mixture (qualitative analysis). Counting all the label monomers associated with a given spectral code or signature allows the counting of all the molecules in the mixture associated with the target-specific sequence coupled to the nanoreporter (quantitative analysis). Nanoreporters are thus useful for the diagnosis or prognosis of different biological states (e.g., disease vs. healthy) by quantitative analysis of known biological markers. Moreover, the exquisite sensitivity of single molecule detection and quantification provided by the nanoreporters of the invention allows for the identification of new diagnostic and prognostic markers, including those whose fluctuations among the different biological states is too slight to detect a correlation with a particular biological state using traditional Molecular methods. The sensitivity of nanoreporter-based Molecular detection permits detailed pharmacokinetic analysis of therapeutic and diagnostic agents in small biological samples.

[0081] Many nanoreporters, referred to as singular nanoreporters, are composed of one molecular entity, as depicted in Figure 1D. However, to increase the specificity of a nanoreporter and/or to improve the kinetics of its binding to a target molecule, a preferred nanoreporter is a dual nanoreporter composed of two molecular entities, each containing a different target-specific sequence that binds to a different region of the same target molecule. Various embodiments of dual nanoreporters are depicted in Figures 1A-1C, in a dual nanoreporter, at least one of the two molecular entities is labeled. The other molecular entity is not necessarily labeled. Such unlabeled components of dual nanoreporters are

referred to herein as "ghost probes" (see Figure 1C) and often have affinity tags attached, which are useful to immobilize and/or stretch the complex containing the dual nanoreporter and the target molecule to allow visualization and/or imaging of the complex.

[0082] Because of their modular structures, nanoreporters may be assembled and labeled in a variety of different ways. For example, a nanoreporter scaffold can be attached to a target-specific sequence (for example by hybridization and, optionally, ligation), and the structure comprising the scaffold and target-specific sequence attached to one or more patches and, where desired, flaps. Alternatively, the nanoreporter scaffold can first be attached to one or more patches (and, optionally, flaps), and the scaffold/patch structure then attached to a target specific sequence. Thus, unless stated otherwise, a discussion or listing of steps in nanoreporter assembly does not imply that a specific route of assembly must be followed.

[0083] Nanoreporter assembly and use is exemplified herein largely by way of description of a variety of nucleic acid-based nanoreporters; however, one of skill in the art would recognize that the methods described herein are applicable to an amino acid-based (or hybrid nucleic acid/amino acid-based) nanoreporter. Illustrative embodiments of partially and fully assembled nanoreporters are listed below.

[0084] At its simplest, a scaffold has at least two label attachment regions capable of being labeled and resolved. The scaffold can be any Molecular entity that allows the formation of label attachment regions on the scaffold that can be separately labeled and resolved. The number of labels attachment regions to be formed on a scaffold is based on the length and nature of the scaffold, the means of labeling the nanoreporter, as well as the type of label monomers emitting a signal to be attached to the label attachment regions of the scaffold. A nanoreporter may have a scaffold including two or more label attachment regions. Suitable scaffold structures include DNA- based scaffolds.

[0085] We also disclose labeled nanoreporter wherein one or more label attachment regions are attached to corresponding label monomers, each label monomer emitting a signal. For example a labeled nanoreporter according to the invention is obtained when at least two label monomers are attached to two corresponding label attachment regions of the scaffold such that these labeled label attachment regions, or "spots," are distinguishable. Label monomers emitting a signal associated with different label attachment regions of the scaffold can emit signals that are spectrally indistinguishable under the detections conditions ("like" signals), or can emit signals that are spectrally distinguishable, at least under the detection conditions (e.g., when the nanoreporter is immobilized, stretched and observed under a microscope).

[0086] We also disclose a nanoreporter wherein two or more label monomers are attached to a label attachment region. The signal emitted by the label monomers associated with said label attachment region produces an aggregate signal that is detected. The aggregate signal produced may be made up of like signals or made up of at least two spectrally distinguishable signals.

[0087] Additionally we disclose a nanoreporter wherein at least two label monomers emitting like signals are attached to two corresponding label attachment regions of the scaffold and said two label monomers are spatially distinguishable. We also disclose a nanoreporter wherein at least two label monomers emitting two distinguishable signals are attached to two neighboring label attachment regions, for example two adjacent label attachment regions, whereby said at least two label monomers are spectrally distinguishable.

[0088] We disclose a nanoreporter wherein two spots emitting like signals are separated by a spacer region, whereby interposing the spacer region allows resolution or better resolution of said like signals emitted by labels monomers attached to said two spots. Optionally, the spacer regions have a length determined by the resolution of an instrument employed in detecting the nanoreporter.

[0089] Also disclosed herein is a nanoreporter with one or more "double spots." Each double spot contains two or more (e.g., three, four or five) adjacent spots that emit like signals without being separated by a spacer region. Double spots can be identified by their sizes.

[0090] A label monomer emitting a signal may be attached covalently or non-covalently (e.g., via hybridization) to a patch that is attached to the label attachment region. The label monomers may also be attached covalently or non-covalently (e.g., via hybridization) to a flap attached to a patch that is in turn attached to the scaffold. The flap can be formed by one molecule or two or more molecules ("flap pieces") that form a split flap.

[0091] We also disclose a nanoreporter associated with spectral code determined by the sequence of signals emitted by the label monomers attached (e.g., indirectly via a patch) to label attachment regions on the scaffold of the nanoreporter, whereby detection of the spectral code allow identification of the nanoreporter.

[0092] Additionally, we disclose a nanoreporter further comprising an affinity tag attached to the nanoreporter scaffold, such that attachment of the affinity tag to a support allows scaffold stretching and resolution of signals emitted by label monomers corresponding to different label attachment regions on the scaffold. Nanoreporter stretching may involve any stretching means known in the art including but not limited to, means involving physical, hydrodynamic or electrical means.

[0093] We further disclose a nanoreporter further comprising flaps attached to label attachment regions of the scaffold, wherein a flap attached to a label attachment region of the scaffold attaches the label monomer corresponding to said label attachment region, thereby indirectly attaching label monomers to corresponding label attachment regions on said scaffold. Optionally, each label monomer comprises a signals emitting portion and an oligonucleotide portion of a pre-

determined sequence, and the flaps comprise repeats of a flap sequence complementary to the oligonucleotide portion of a corresponding label, whereby one or more label monomers attach to a corresponding label attachment region through hybridization of said oligonucleotide portions of said label monomers to said repeats of said flap sequence thereby producing a labeled nanoreporter.

5 [0094] A nanoreporter can further include a target-specific sequence coupled to the scaffold. The target-specific sequence is selected to allow the nanoreporter to recognize, bind or attach to a target molecule. The nanoreporters disclosed herein are suitable for identification of target molecules of all types. For example, appropriate target-specific sequences can be coupled to the scaffold of the nanoreporter to allow detection of a target molecule. Preferably the target molecule is DNA (including cDNA), RNA (including mRNA and cRNA), a peptide, a polypeptide, or a protein.

10 [0095] One embodiment of the invention provides increased flexibility in target molecule detection with label monomers. In this embodiment, a dual nanoreporter (or probe pair) comprising two different molecular entities, each with a separate target-specific region, at least one of which is labeled, bind to the same target molecule. Thus, the target-specific sequences of the two components of the dual nanoreporter bind to different portions of a selected target molecule, whereby detection of the spectral code associated with the dual nanoreporter provides detection of the selected target molecule in a biomolecular sample contacted with said dual nanoreporter.

15 [0096] The invention also provides a method of detecting the presence of a specific target molecule in a biomolecular sample comprising: (i) contacting said sample with a dual nanoreporter (or probe pair) according to the invention under conditions that allow binding of the target-specific sequences in the dual nanoreporter to the target molecule and (ii) detecting the code that identifies the target molecule. Depending on the nanoreporter architecture, the dual nanoreporter may be labeled before or after binding to the target molecule.

20 [0097] In certain embodiments, the methods of detection are performed in multiplex assays, whereby a plurality of target molecules are detected in the same assay (a single reaction mixture). In a preferred embodiment, the assay is a hybridization assay in which the plurality of target molecules are detected simultaneously. In certain embodiments, the plurality of target molecules detected in the same assay is at least 5 different target molecules, at least 10 different target molecules, at least 20 different target molecules, at least 50 different target molecules, at least 75 different target molecules, at least 100 different target molecules, at least 200 different target molecules, at least 500 different target molecules, or at least 750 different target molecules, or at least 1000 different target molecules. In other embodiments, the plurality of target molecules detected in the same assay is up to 50 different target molecules, up to 100 different target molecules, up to 150 different target molecules, up to 200 different target molecules, up to 300 different target molecules, up to 500 different target molecules, up to 750 different target molecules, up to 1000 different target molecules, up to 2000 different target molecules, or up to 5000 different target molecules. In yet other embodiments, the plurality of target molecules detected is any range in between the foregoing numbers of different target molecules, such as, but not limited to, from 20 to 50 different target molecules, from 50 to 200 different target molecules, from 100 to 1000 different target molecules, from 500 to 5000 different target molecules, and so on and so forth.

35 [0098] In certain embodiments, the invention is directed to detecting different splice forms of the same RNA. The different splice forms can be detected using a plurality of nanoreporter probes, each with a different target-specific sequence complementary to a different exon of the same gene.

[0099] Structural stability of a nanoreporter can be increased through ligation of the patches and, optionally, ligation of the split flaps and/or the labeled oligonucleotides hybridized to the split flaps.

40 [0100] In addition to the qualitative analytical capabilities provided by the nanoreporters of the invention and the analytical techniques based thereon, the nanoreporters of the invention are uniquely suitable for conducting quantitative analyses. By providing a one to one binding between the nanoreporters (whether singular or dual nanoreporters) of the invention and their target molecules in a biomolecular sample, all or a representative portion of the target molecules present in the sample can be identified and counted. This individual counting of the various molecular species provides an accurate and direct method for determining the absolute or relative concentration of the target molecule in the biomolecular sample. Moreover, the ability to address each molecule in a mixture individually leverages benefits of miniaturization including high sensitivity, minimal sample quantity requirements, high reaction rates which are afforded by solution phase kinetics in a small volume, and ultimately very low reagent costs.

45 [0101] As will be appreciated from the description and examples provided below, the present invention provides numerous advantages. For example, the complex modularity in forming nanoreporters according to the invention allows for systematic creation of libraries of unique nanoreporters having a very high degree of diversity (*e.g.*, millions of uniquely recognizable nanoreporters). This modularity allows flexibility in customizing nanoreporter populations to specific applications which in turn provides significant manufacturing efficiencies. Another advantage that will be appreciated through the following description stems from the flexibility in assembling the nanoreporters of the invention. That is, due to their modular structure, the nanoreporters of the invention can be assembled prior to shipment to a point of use or assembled at the point of use.

5.1 NANOREPORTER NOMENCLATURE

- [0102] NANOREPORTER:** The term "nanoreporter" refers to a molecular entity that has (i) a molecule ("scaffold") containing at least two label attachment regions; (ii) at least one patch attached to at least one label attachment region; and (iii) a target-specific sequence. As described in detail below, nanoreporters can be singular nanoreporters (all components being in a single molecular entity) or dual nanoreporters (all the components being in two separate molecular entities). Nanoreporters are preferably synthetic, *i.e.*, non-naturally-occurring molecules, for example are chimeric molecules made by joining two or more manmade and/or naturally occurring sequences that normally exist on more than one molecule (*e.g.*, plasmid, chromosome, viral genome, protein, etc.).
- [0103] LABELED NANOREPORTER:** A labeled nanoreporter is a nanoreporter in which at least one patch of the nanoreporter is attached to one or more label monomers that generate(s) a signal that forms at least part of the nanoreporter code.
- [0104] LABEL UNIT:** The term "label unit" refers to the non-target-specific portions of a labeled nanoreporter.
- [0105] PROBE:** This refers to a molecule that has a target-specific sequence. In the context of a singular nanoreporter, the term "probe" refers to the nanoreporter itself; in the context of a dual nanoreporter, the term "probe" refers to one or both of the two components of the nanoreporter.
- [0106] PROBE PAIR:** This refers to a dual nanoreporter.
- [0107] PATCH:** The term "patch" refers to a molecular entity attached to the label attachment region of the nanoreporter scaffold, generally for the purpose of labeling the nanoreporter. The patch can have one or more label monomers either directly (covalently or noncovalently) or indirectly attached to it, either prior to or after its attachment to the scaffold.
- [0108] FLAP:** The term "flap" as used herein refers to a molecular entity attached to a patch or patch pair attached to a label attachment region. The flap is one or more molecule containing label monomers or capable of binding one or more molecules containing label monomers. By providing indirect labeling of the regions, the flaps provide more flexibility in controlling the number of signal emitting monomers associated with a region as well as the nature of those monomers. Flaps may be formed by a single molecular piece or several molecular pieces (*e.g.*, two pieces) forming a "split flap" (see, *e.g.*, Figure 7)
- [0109] TARGET-SPECIFIC SEQUENCE:** The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter scaffold. The target molecule is preferably (but not necessarily) a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.
- [0110] GHOST PROBE:** A molecule comprising a target-specific sequence, but which is not labeled with a label monomer that emits a signal that contributes to the nanoreporter code.
- [0111] REPORTER PROBE:** A molecule comprising a target-specific sequence that is labeled with at least one label monomer that emits a signal that contributes to the nanoreporter code. A singular nanoreporter is a reporter probe, as is a labeled component of a dual nanoreporter.
- [0112] F-HOOK and G-HOOK:** In the context of a dual nanoreporter, F- and G-hooks are each an affinity tag that is capable of being selectively bound to one of the probes. In preferred embodiments, the F-hook and G-hook are biotinylated oligonucleotides that are hybridizable to respective complementary sequences present in (*e.g.*, via ligation) or attached to (*e.g.*, via hybridization) the respective nanoreporter probes in a dual nanoreporter. Thus, the F-hooks and G-hooks can be used for purification, immobilization and stretching of the nanoreporter. Generally, where a dual nanoreporter contains one reporter probe and one ghost probe, the G-hook becomes attached to the reporter probe and the F-hook becomes attached to the ghost probe. F-hooks and G-hooks can be biotinylated on either end or internally. They can also be amine-modified to allow for attachment to a solid substrate for affinity purification.
- [0113] F-TAG and G-TAG:** Tandemly-repeated sequences of about 10 to about 25 nucleotides that are complementary to the F-hook and G-hook, respectively. G-tags and F-tags are attached to the nanoreporter probes. Generally, an F-tag is present in or attached to a ghost probe via a ligator sequence and a G-tag is present in or attached the reporter probe scaffold via a ligator sequence.
- [0114] SPOT:** A spot, in the context of nanoreporter detection, is the aggregate signal detected from the label monomers attached to a single label attachment site on a nanoreporter, and which, depending on the size of the label attachment region and the nature (*e.g.*, primary emission wavelength) of the label monomer, may appear as a single point source of light when visualized under a microscope. Spots from a nanoreporter may be overlapping or non-overlapping. The nanoreporter code that identifies that target molecule can comprise any permutation of the length of a spot, its position relative to other spots, and/or the nature (*e.g.*, primary emission wavelength(s)) of its signal. Generally, for each probe or probe pair of the invention, adjacent label attachment regions are non-overlapping, and/or the spots from adjacent label attachment regions are spatially and/or spectrally distinguishable, at least under the detection conditions (*e.g.*, when the nanoreporter is immobilized, stretched and observed under a microscope, as described herein).
- [0115]** Occasionally, reference is made to a spot "size" as a certain number of bases or nucleotides. As would be readily understood by one of skill in the art, this refers to the number of bases or nucleotide in the corresponding label

attachment region.

[0116] NANOREPORTER CODE: The order and nature (e.g., primary emission wavelength(s), optionally also length) of spots from a nanoreporter serve as a nanoreporter code that identifies the target molecule capable of being bound by the nanoreporter through the nanoreporter's target specific sequence(s). When the nanoreporter is bound to a target molecule, the nanoreporter code also identifies the target molecule. Optionally, the length of a spot can be a component of the nanoreporter code.

[0117] DARK SPOT: The term "dark spot" refers to a lack of signal, or "spot," from a label attachment site on a nanoreporter. Dark spots can be incorporated into the nanoreporter code to add more coding permutations and generate greater nanoreporter diversity in a nanoreporter population.

[0118] REGISTER: The term "register" refers to a set of alternating label attachment regions.

5.2 THE NANOREPORTER SCAFFOLD

[0119] The nanoreporter scaffold can be any molecular entity, more preferably a nucleic acid molecule, containing label attachment regions to which label monomers can be directly or indirectly attached. In one embodiment, the nanoreporter scaffold is a protein scaffold; in a preferred embodiment, the nanoreporter scaffold is a nucleic acid scaffold in which the label attachment regions are single-stranded regions to which other nucleic acids, such as oligonucleotide patches, RNA patches, or DNA patches, can attach by hybridization. In specific embodiments, the nanoreporter scaffold is a nucleic acid molecule.

[0120] There are no particular limitations on the types of scaffolds that are suitable for forming nanoreporters. A scaffold can essentially have any structure including, for example, single stranded linear scaffold, double stranded linear scaffold, single stranded circular scaffold or double stranded circular scaffold. Examples of scaffold structures include, for example, a scaffold made of one molecular entity such as polypeptides, nucleic acids or carbohydrates. A scaffold may also include a combination of structures, for example, a scaffold may be made of one or more polypeptide stretches coupled to one or more carbohydrate stretches.

[0121] Suitable molecular entities for scaffolds according to the invention include polymeric structures particularly nucleic acid based polymeric structures such as DNA. DNA based structures offer numerous advantages in the context of the present invention due at least in part to the vast universe of existing techniques and methodologies that allow manipulation of DNA constructs.

[0122] As indicated above, the scaffold may be single stranded or double stranded. Double stranded scaffold can be either conventional double stranded DNA or a double strand that is composed of a linear single stranded stretch of nucleic acid with patch units or flat-patches attached.

[0123] A scaffold can have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21-100 label attachment regions or more.

[0124] The label attachment regions of a nanoreporter scaffold will vary in size depending on the method of labeling. In various embodiments, a label attachment region can have a length anywhere from 10 nm to 10,000 nm, but is more preferably from 50 nm to 5,000 nm, and is more preferably from 100 nm to 1,000 nm. In various embodiments, the label attachment region is from about 100 nm to about 500 nm, from about 150 nm to about 450 nm, from about 200 nm to about 400 nm, or from 250 to about 350 nm. In a preferred embodiment, the label attachment region corresponds closely to the size of a diffraction-limited spot, *i.e.*, the smallest spot that can be detected with standard optics, which is about 300 nm.

[0125] Where the scaffold is a nucleic acid, 1 nm corresponds to approximately 3 nucleotides; thus, an approximately 300 nm-label attachment region corresponds to approximately 900 bases. In other preferred embodiments, the label attachment region is from about 300 nucleotides to about 1.5kb, from about 450 nucleotides to about 1.35 kb, from about 0.6 kb to about 1.2 kb, or from 0.75 kb to about 1.05 kb.

[0126] An illustrative example of a molecular entity for a nanoreporter scaffold according to the invention is M13 DNA, which is single-stranded. In one embodiment, the nanoreporter scaffold is circular at least partially single stranded DNA, such as circular M13. In a more preferred embodiment, the nanoreporter scaffold is linear at least partially single stranded DNA, such as linear M13. In a specific embodiment, the M13 single-stranded DNA obtained by operating a cut at the BamHI site of circular M13 DNA. It should be noted that within the context of the present invention, linear DNA provides additional advantages compared to circular DNA. One advantage of using linear DNA in forming a scaffold according to the invention relates to the significantly reduced torsional stress associated with linear DNA. The added torsional stress associated with circular DNA may interfere with the structural integrity of the scaffold upon the addition to the scaffold of other components of the nanoreporter, such as patch units. Severe torsional stress may lead to the breaking of the structure of the scaffold.

[0127] It should be noted however that the nanoreporters where only a few, short label attachment sites are labeled, circular DNA may be suitable.

5.2.1 NOVEL SYNTHETIC NANOREPORTER SCAFFOLD

SEQUENCES

5 **[0128]** The present invention employs nanoreporter scaffold that are artificial nucleic acid molecules (DNA5 RNA, or DNA/RNA hybrids) designed to have features that optimize labeling and detection of the nanoreporter. In these aspects of the invention, a nanoreporter scaffold is an artificial nucleic acids comprising one or more synthetic sequences from 50 to 50,000 bases long. Accordingly, the nanoreporter scaffold, which is preferably a DNA, is designed to have one or more Regions, useful as label attachment regions, comprising a regular pattern of a particular base (the "regularly-repeated base"). In such regions, the regularly-repeated base occurs with a periodicity of every nth residue, where n is any number, and preferably from 4 to 25.

10 **[0129]** Preferably, not more than 25% of the regularly-repeated base in a Region appears at other than said regular intervals. For example, if in a Region of 100 nucleotides there are 12 thymidine bases, and thymidine is the regularly-repeated base, in this aspect of the invention not more than 25% of these, i.e., 3 thymidine bases, appear outside the regular pattern of thymidines. In specific embodiments, not more than 20%, not more than 15%, not more than 10%, not more than 9%, not more than 8%, not more than 7%, not more than 6%, not more than 5%, not more than 4%, not more than 3%, not more than 2% or not more than 1% of said base appears at other than said regular intervals in said region.

15 **[0130]** The regularly-repeated base in the Regions in a nanoreporter scaffold, or its complementary regularly-repeated base in an annealed patch (or segment) can be used to attach label monomers, preferably light emitting label monomers, to the nanoreporter in a regular, evenly spaced pattern for better distribution of the nanoreporter signal. Preferably, where a Region is labeled, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or at least 98% of occurrences of the regularly-repeated base is attached to at least one light-emitting label monomer, either by covalent attachment of a label monomer to a base, or by hybridization to a nucleic acid in which the complements of the regularly-repeated base are so-labeled.

20 **[0131]** This percentage of occurrences can be measured by any means known in the art. In one method, the amount of nucleic acid produced in a labeling reaction is purified (for example, RNA can be purified using a Qiagen RNeasy kit) and subjected to UV spectrophotometry. The absorbance ("A") at the appropriate wavelengths is measured for each of the nucleic acid (260 nm) and the label monomer whose occurrence is to be measured (e.g., 495 nm for Alexa Fluor 488; 590 nm for Alexa Fluor 594; 650 for Alexa Fluor 647; and 550 nm for Cy3). The absorbance of the nucleic acid is corrected by adjusting the value of the absorbance at 260 nm ("A₂₆₀") to remove the "noise" contribution from the label monomer by subtracting the absorbance at the peak wavelength for the label monomer (A_{LM}) minus the correction factor for that label monomer. Where the nucleic acid is RNA, the number of label monomers per one thousand nucleotides is calculated according to the formula:

$$\frac{\text{no. of label monomers}}{1000 \text{ nucleotides}} = \frac{A_{260}}{A_{LM}} \times \frac{9010}{EC_{LM}} \times 1000$$

40 where EC_{LM} is the extinction coefficient for the label monomer. From this formula, the percentage of occurrences of the regularly-repeated base that are attached to a light-emitting label monomer can be calculated.

45 **[0132]** Generally, the preferred regularly-repeating base in a label attachment region is thymidine, so that the region can be labeled by hybridization to one or more complementary patches (e.g., RNA segments) in which the regularly-repeated base is uridine. This permits the use of amino-allyl-modified UTPs, which are readily commercially available, as label monomer attachment sites, in an otherwise random sequence. Preferably, in addition to the regular periodicity of the Regions, the regions (and the nucleic acid comprising them) contain minimal secondary structure. The overall GC-content is preferably maintained close to 50%, and is preferably consistent over relatively short stretches to make local T_m's similar.

50 **[0133]** The artificial nucleic acids of the invention, or at least the Regions therein, preferably do not have direct or inverted repeats that are greater than 12 bases in length. In other embodiments, the artificial nucleic acids and/or Regions do not have direct or inverted repeats that are greater than about 11, about 10 or about 9 bases in length.

55 **[0134]** In an exemplary Region in which the regularly-repeated nucleotide is a thymidine and a GC content of approximately 50%, excess adenines would make up the loss in abundance of T's. To generate the selected sequence, random sequences with fixed patterns of T's ranging from every 4th base to every 25th base are created and screened to minimize the presence of inverted and direct repeats.

[0135] Sequences are also screened preferably to avoid common six-base-cutter restriction enzyme recognition sites. Selected sequences are additionally subjected to predicted secondary structure analysis, and those with the least sec-

ondary structure are chosen for further evaluation. Any program known in the art can be used to predict secondary structure, such as the MFOLD program (Zuker, 2003, Nucleic Acids Res. 31 (13):3406-15; Mathews et al., 1999, J. Mol. Biol. 288:911-940).

[0136] An appropriate sequence is divided into label attachment regions ranging from 50 bases to 2 kilobases long (could be longer). Each label attachment region is a unique sequence, but contains a consistent number and spacing of T's in relation to the other label attachment regions in a given reporter sequence. These label attachment regions can interspersed with other regions whose sequence does not matter. The synthetic label attachment regions in a nanoreporter scaffold can be of different lengths and/or have different regularly-repeated bases. An optimized start sequence for transcription by RNA polymerase T7, T3, or SP6 (beginning at position +1 of the transcript) can be added to the 5' end of each label attachment region. Restriction sites are optionally added at the boundaries of each label attachment region to allow specific addition or deletion of individual label attachment regions to the sequence using conventional cloning techniques. The number of synthetic label attachment regions in a nanoreporter preferably ranges from 1 to 50. In yet other embodiments, the number of synthetic label attachment regions in a nanoreporter ranges from 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 synthetic label attachment regions to 15, 20, 30, 40, or 50 synthetic label attachment regions, or any range in between.

[0137] An example of such a novel synthetic label attachment region is given below. In this sequence, shown 5' to 3', the T's are placed in every 8th position and the region is bounded by a 5' Sac I restriction site and a 3' Kpn I restriction site. An optimized transcript start site for T7 polymerase (GGGAGA) is included at the 5' end of the region, downstream of the 5' restriction site. The complement of this sequence, when generated as a single-stranded forms the scaffold for the RNA molecule transcribed from this label attachment region.

```

GAGCTCGGGAGATGGCGAGCTGGAAGCATCAGAAAGTAGGAAGATGACAAAAT
25 AGGCCCATAGAAGCATGAAGAACTGAACGCATGAGACAATAGGAAGCTACGCC
ACTAGGGACCTGAGAAGCTGAGCGGCTCAGCGGGTCCGAGCGTCAAAAAATAA
AAGAGTGAAACAATAGACGAATGACGCGGTAAAACCATCCAGAAGTAAACGGG
30 TACAAACATACAGAGATAGCCACCTGGACCAATAGGCACGTACAAACGTACAA
GCCTGGCGCGATGAGGCAATCCACACGTGCAGAGCTGGAACAATGGAAAGATG
CAAGAATAAACCGATACCGGGATCGAGGGCTCAGCGAATAAAGCAGTCAACAA
CTGGAAAGATCCACACATACCGGCGTAACCGAGTCCAAACATACAGACCTGCA
35 AGACTCGCGACATGGGACGGTAAAACCATCCGACCGTAAACCGGTAAACAGGT
AGCCGGGTAAAAACATAGCAGGGTGGAGACCTCAGAACGTAAAGACGTCCAAG
GGTCGCCGGATAGCGAACTACGCGCATCCCCCAATGGGCCAATCAACAGATAA
40 ACGAGTAGAAAAGTCAGAAAATAAGAAACTAACGAAATACGAGGGTCCAAGG
ATGCAAGACTGAGGCCCTAAGGAGATAAGGAAATAGGCCGATGCAGACCTGAA
ACGATGCACCGATCCGACGGTAAAAGACTAGACACGTAGCCGGATCAGGGCCT
45 GGGAGGCTGGAACCGTGAGCACATAGCAAAGTCGCAGCGTCGGCAGATGCGCC
GGTAAAAAAGTAGAGGCATGACCGGATGGGCAAATAGCGACGTACAGCAGTGA
AGCACTAAAAGCATCCAAGGGTAGGAGACTAGGCGCCTCGACGGGTAGGTACC

```

[0138] The synthetic nucleic acids employed in the present invention can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotide designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the label attachment region and the annealed patches or segments, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the synthetic nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcyto-

sine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D- mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2- thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3- N-2-carboxypropyl)uracil, (acp3)w, and 2,6-diaminopurine.

[0139] Alternatively, the synthetic nucleic acid can be produced biologically using a vector into which a nucleic acid has been subcloned.

[0140] In various embodiments, the synthetic nucleic acid molecules which may be employed in the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al., 1996, Bioorganic & Medicinal Chemistry 4(1):5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al., 1996, Bioorganic & Medicinal Chemistry 4(1): 5-23; Perry-O'Keefe et al, 1996, Proc. Natl. Acad. Sci. USA 93: 14670- 675.

[0141] In an exemplary embodiment, the selected novel synthetic sequence can be constructed synthetically as double-stranded DNA by a commercial gene synthesis company and cloned in an oriented fashion into a "phagemid", a plasmid vector containing an M13 or f1 phage intergenic (IG) region which contains the cis-acting sequences necessary for DNA replication and phage encapsidation, such as pUC119. The appropriate orientation of the cloned insert relative to the phage origin of replication allows for the generation of a single- stranded DNA scaffold which is the reverse complement of the RNA molecules generated by *in vitro* transcription for each label attachment region.

[0142] In order to generate the single-stranded DNA scaffold of the novel reporter, the phagemid is transformed into an E. coli strain containing an F' episome. Subsequent infection of the transformed bacteria with a helper phage such as the M1 3 mutant K07 results in the secretion of the phagemid carrying the novel reporter sequence as a single-stranded, packaged phage from which the circular, single-stranded DNA is prepared using a standard protocol. This DNA is linearized and the vector portion is excised by annealing short, complementary oligonucleotides to either end of the novel reporter sequence to generate double-stranded restriction sites, followed by treatment with the appropriate restriction enzymes.

[0143] To make the RNA molecules (patches or "segments") for each label attachment region, polymerase chain reaction ("PCR") primers are designed to generate a double-stranded template beginning with an RNA polymerase promoter (T7, T3, or SP6) directly upstream (5') of the transcription start site and ending following the 3' restriction enzyme site. Using this template, *in vitro* transcription of RNA molecules is performed in the presence of amino-allyl modified regularly-repeated base in the RNA (e.g., UTP) and unmodified other bases (e.g., ATP, CTP and GTP). This leads to an RNA product in which every regularly-repeated base (e.g., U) is modified to allow covalent coupling of a label monomer at that position in the RNA molecule.

[0144] Coupling of light-emitting label monomers to the RNA molecules and annealing of the labeled RNA molecules to the scaffold are carried out as described below.

[0145] Some design considerations for the de novo sequence are listed in Table 1 below.

Feature Of Synthetic Scaffold	Advantages
Novel synthetic sequence	Can be of any length and be designed to incorporate any desired sequence feature including but not limited to those listed in this table.
Minimal secondary structure (select against inverted repeats)	Allows for consistent transcription of full-length RNA molecules. Allows for consistent annealing of RNA molecules to scaffold at predictable temperatures. Minimizes self-annealing and/or cross-annealing between RNA molecules or scaffolds.
Minimal repeated sequences	Avoids mis-annealing between RNA molecules and inappropriate regions of the scaffold.
Unique restriction sites at borders of label attachment regions	Allows addition and deletion of individual label attachment regions using standard molecular cloning techniques.

(continued)

Feature Of Synthetic Scaffold	Advantages
Defined, even spacing of T's and transcription with amino-allyl-modified UTP (no unmodified UTP)	Controls number of coupling sites for monomers in each label attachment region, allowing for consistent brightness of individual labeled RNA molecules. Controls distance between monomers: spacing can be optimized to avoid steric hindrance and fluorescence quenching.
Optimized start sequence for transcription by RNA polymerase T7, T3, or SP6	Promotes efficient in vitro transcription of each label attachment region.

5.3 PATCHES

[0146] Label monomers that emit signals which constitute all or part of the nanoreporter code are attached to label attachment region(s) of the nanoreporter scaffold through a structure referred to herein as a "patch." The label monomers can be directly (e.g., covalently or noncovalently) attached to a patch, or indirectly attached to a patch (e.g., through hybridization).

[0147] Nucleic acid patches can be anywhere from 25 nucleotides to several kilobases (e.g., 5kb) in length, and are preferably 50 nucleotides to 2kb in length. In specific embodiments, nucleic acid patches are approximately 25 to 250, 50 to 200, 50 to 150, or 50 to 100 nucleotides in length. In other embodiments, nucleic acid patches are approximately 500 to 2,000, 500 to 1,500, 500 to 1,000, 750 to 1,250, or 750 to 1,000 nucleotides in length. Nucleic acid patches can be RNA patches or DNA patches.

[0148] A label monomer can be covalently attached to a patch before or after the patch is attached to the label attachment region of a nanoreporter scaffold. For example, where the patch is a nucleic acid molecule, the label can be covalently attached by incorporation of a nucleotide containing a label monomer into the nucleic acid during its synthesis but before it is attached, e.g., via hybridization, to the label attachment region of the scaffold. Alternatively, during the synthesis of a nucleic acid patch, a nucleotide containing a label monomer acceptor group can be included, and the label monomer added to the nucleic acid patch after its synthesis, either before or after it is attached to the label attachment region of the scaffold. Alternatively, the label monomer can be indirectly attached to the patch, for example by hybridization of the patch to a "flap" that serves as a basis for attachment of the label monomer to the nanoreporter.

[0149] Thus, where a patch is a nucleic acid, it can range anywhere from 20 nucleotides to more than 5 kb in length, depending on the method of assembly of the nanoreporter.

[0150] For example, where a patch has covalently incorporated into it one or more label monomers that emit signals that are part of the nanoreporter code in the context of the labeled nanoreporter, the patch is preferably about 100 to about 10,000 bases, more preferably 200 to about 2000 bases, and yet more preferably 700 to about 1200 nucleotides in length, and is generally referred to herein as a "segment," a "dark" segment being the patch prior to the incorporation of the label monomer (but, in a preferred embodiment, containing label monomer acceptor sites, such as amino allyl nucleotides), and a "colored" segment being one containing the desired label monomer or label monomers. The T_m of a segment when hybridized to its label attachment region preferably is $>80^\circ\text{C}$, more preferably $>90^\circ\text{C}$, in 825 mM Na^+ (5X SSC).

[0151] Where a patch merely serves as a template for flap attachment to the nanoreporter, then it is preferably smaller in size, for example about 25-250 nucleotides in length, and is most preferably about 50-100 nucleotides in length. Such patches are referred to herein as "oligonucleotide patches." As detailed in Section below, an oligonucleotide is preferably partially complementary in sequence to a scaffold, such that when it is annealed to the scaffold, an overhang is generated that is complementary to all or portion of a flap.

[0152] The terms "segment" and "oligonucleotide patch" are used herein merely for convenience of description, however, there is no size cutoff to distinguish a "segment" from an "oligonucleotide patch." The purpose of both types of structures is to maximize the labeling --and thus signal intensity-- from the nanoreporter, thereby allowing for single target molecule detection by a nanoreporter.

[0153] In certain aspects, the present invention employs a synthetic molecule, whose configuration is illustrated by reference to **Figure 7A**, comprising a strand of a nucleic acid (**scaffold**) and a plurality of patch pairs hybridized to said strand, wherein each patch pair comprises an "A" patch and a "B" patch, wherein, for each patch pair, (a) each "A" patch is an oligonucleotide comprising a first region (**1P**) and a second region (**2P**), said first region being (i) at the alpha end of said "A" patch, and (ii) hybridized to a first portion of said strand, said second region being (i) at the beta end of said "A" patch; (b) each "B" patch is an oligonucleotide comprising a third region (**3P**) and a fourth region (**4P**), said third region being (i) at the alpha end of said "B" patch, and (ii) hybridized to said second region of said "A" patch, said fourth

region being (i) at the beta end of said "B" patch and (ii) hybridized to a second portion of said strand, said second portion of said strand being to the beta end of said first portion of said strand, wherein said second region or said third region further comprises at its beta end or alpha end, respectively, a hybridizable region that is not hybridized to said "B" patch or "A" patch, respectively.

[0154] In the synthetic molecule of **Figure 7A**, the second region may further comprise at its beta end a hybridizable region that is not hybridized to said "B" patch, as depicted in **Figure 7B**, or the third region further comprises at its alpha end a hybridizable region that is not hybridized to said "A" patch, as depicted in **Figure 7C**.

[0155] The present invention further employs a synthetic molecule, whose configuration is illustrated by reference to **Figure 7D**, comprising a strand of a nucleic acid (**scaffold**) and a plurality of patch pairs hybridized to said strand, wherein each patch pair comprises an "A" patch and a "B" patch, wherein, for each patch pair, (a) each "A" patch is an oligonucleotide comprising a first region (**1P**) and a second region (**2P**), said first region being (i) at the alpha end of said "A" patch, and (ii) hybridized to a first portion of said strand, said second region being (ii) at the beta end of said "A" patch; (b) each "B" patch is an oligonucleotide comprising a third region (**3P**) and a fourth region (**4P**), said third region being (i) at the alpha end of said "B" patch, and (ii) hybridized to said second region of said "A" patch, said fourth region being (i) at the beta end of said "B" patch and (ii) hybridized to a second portion of said strand, said second portion of said strand being to the first of said first portion of said strand, wherein said second region further comprises at its beta end a first hybridizable region that is not hybridized to said "B" patch, and wherein said third region further comprises at its alpha end a second hybridizable region that is not hybridized to said "A" patch.

[0156] In the synthetic molecule of **Figure 7B**, each patch pair can be attached to a flap pair, as depicted in **Figure 7F**, wherein each flap pair comprises an "A" flap and a "B" flap, wherein, for each flap pair, (a) each "A" flap is an oligonucleotide comprising a first flap region (**1F**) and a second flap region (**2F**); said first flap region being at the alpha end of said "A" flap; said second flap region (i) being at the beta end of said "A" flap and (ii) comprising at its beta end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "B" flap; and (b) each "B" flap is an oligonucleotide comprising a third flap region (**3F**), a fourth flap region (**4F**), and a fifth flap region (**5F**); said third flap region being (i) being at the alpha end of said "B" flap and (ii) comprising at its alpha end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "A" flap; said fourth flap region (i) being between the third flap region and the fifth flap region and (ii) hybridized to said first flap region of said "A" flap; said fifth flap region being (i) at the beta end of said "B" flap, and (ii) hybridized to said hybridizable region of said second region of said "A" patch.

[0157] In the synthetic molecule of **Figure 7C**, each patch pair can be attached to a flap pair, as depicted in **Figure 7E**, wherein each flap pair comprises an "A" flap and a "B" flap, wherein, for each flap pair, (a) each "A" flap is an oligonucleotide comprising an first flap region (**1F**), a second flap region (**2F**), and a third flap region (**3F**); said "A" flap region being (i) at the alpha end of said "A" flap and (ii) hybridized to said hybridizable region of said third region of said "B" patch; said second flap region being between the first flap region and the third flap region; said third flap region (i) being at the beta end of said "A" flap and (ii) comprising at its beta end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "B" flap, and (b) each "B" flap is an oligonucleotide comprising a fourth flap region (**4F**) and a fifth flap region (**5F**); said fourth flap region being (i) being at the alpha end of said "B" flap and (ii) comprising at its alpha end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "A" flap; said fifth flap region being (i) at the beta end of said "B" flap, and (ii) hybridized to said second flap region of said "A" flap.

[0158] In the synthetic molecule of **Figures 7D** and **7E**, the split flaps can be attached one (*e.g.*, (10)), or more (*e.g.*, (20) and (30)) oligonucleotides, as depicted in **Figure 7G**. Thus, the one or more oligonucleotides can be attached to the all or a portion of the "A" flap individually (*e.g.*, (10)), the "B" flap individually (*e.g.*, (30)), or span all or a portion of each of the "A" flap and "B" flap (*e.g.*, (20)). Such oligonucleotides are preferably covalently bound to one or more label monomers.

[0159] The hybridizable regions of said synthetic molecules may be hybridized to a plurality of oligonucleotides, each bound, preferably covalently bound, to at least one label monomer, more preferably to at least five label monomers. In certain embodiments, all the oligonucleotides attached to a single patch pair comprise the same label monomers, *e.g.*, comprise label monomers that emit light at the same wavelength(s); in specific embodiments, all the oligonucleotides attached to at least two, or at least four, adjacent patch pairs preferably comprise the same label monomers. One or more of the oligonucleotides may be bound to at least one affinity tag.

[0160] In certain preferred embodiments, the label monomers are fluorophores or quantum dots.

[0161] In the synthetic molecule described above, alpha can refers to either 5' or 3', and the corresponding beta to either 3' or 5', respectively.

[0162] The region of complementary in each patch pair, or between a given patch and corresponding flap, is preferably about 20 to 5,000 nucleotides. In certain embodiments, the region of complementary is about 20 to 100 nucleotides, or about 5 to 50 nucleotides.

[0163] In the synthetic molecules described above, each flap is preferably about 50 to 5,000 nucleotides in length. In certain embodiments, each flap is about 50 to 150 nucleotides.

[0164] The synthetic molecules described above may further comprise a target-specific region which binds to a target

molecule. The target-specific region can be attached to the beta or alpha end of said strand.

[0165] In certain embodiments, the synthetic molecule described above comprise at least ten patch pairs, or at least fifty patch pairs.

[0166] In the synthetic molecules described above, the strand, or scaffold, can be a linearized vector, such as linearized M 13.

[0167] The synthetic molecule described above may further comprise (a) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal; (b) a second label attachment region, which is non-overlapping with the first label attachment region, to which is attached one or more label monomers that emit light constituting a second signal; (c) a third label attachment region, which is non-overlapping with the first and second label attachment regions, to which is attached one or more label monomers that emit light constituting a third signal; wherein each attachment region comprises a plurality of patch pairs; wherein the first and second signals are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable; wherein the first and second signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the second and third signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the first and third signals are spatially resolvable under conditions that can be used to detect said first, second and third signals; and wherein the identities of the first, second and third signals and the locations of the first and third signal relative to each other constitute at least part of a code that identifies the target molecule.

5.4 LABEL MONOMERS

[0168] The nanoreporters of the present invention can be labeled with any of a variety of label monomers, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other monomer known in the art that can be detected directly (e.g., by light emission) or indirectly (e.g., by binding of a fluorescently-labeled antibody). At least the first and second label attachments regions in the nanoreporter are labeled with one or more label monomers, and the signals emitted by the label monomers attached to the label attachment regions of a nanoreporter constitute a code that identifies the target to which the target-specific region of the nanoreporter binds. In certain embodiments, the lack of a given signal from the label attachment region (i.e., a "dark" spot) can also constitute part of the nanoreporter code. An example of a dark spot is depicted at position 12 of the nanoreporter in Figure 1A.

[0169] Radioisotopes are an example of label monomers that can be utilized by the invention. Several radioisotopes can be used as label monomers for labeling nucleotides or proteins, including, for example, ^{32}P , ^{33}P , ^{35}S , ^3H , and ^{125}I . These radioisotopes have different half-lives, types of decay, and levels of energy which can be tailored to match the needs of a particular experiment. For example, ^3H is a low energy emitter which results in low background levels, however this low energy also results in long time periods for autoradiography. Radioactively labeled ribonucleotides, deoxyribonucleotides and amino acids are commercially available. Nucleotides are available that are radioactively labeled at the first, or α , phosphate group, or the third, or γ , phosphate group. For example, both [α - ^{32}P] dATP and [γ - ^{32}P] dATP are commercially available. In addition, different specific activities for radioactively labeled nucleotides are also available commercially and can be tailored for different experiments.

[0170] Another example of label monomers that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, and Texas Red. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum. Also, different formulations of the same fluorophore have been produced for different applications. For example, fluorescein, can be used in its isothiocyanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or as isomeric dichlorotriazine forms of fluorescein (DTAF). These monomers are chemically distinct, but all emit light with a peak between 515-520 nm, thereby generating a similar signal. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein. In addition, different formulations or related dyes have been developed around other fluorophores that emit light in other parts of the spectrum.

[0171] Non-radioactive and non-fluorescent label monomers are also available. For example, biotin can be attached directly to nucleotides and detected by specific and high affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin labeled nucleotides can also similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

[0172] Very small particles, termed nanoparticles, also can be used as label monomers to label nucleic acids. These particles range from 1-1000 nm in size and include diverse chemical structures such as gold and silver particles and quantum dots.

[0173] When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120 nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle. Four to five different particles in close proximity will each scatter monochromatic light which when superimposed will give a specific, unique color. The particles are being manufactured by companies such as Genicon Sciences. Derivatized silver or gold particles can be attached to a broad array of molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivitized to allow attachment to a nucleotide.

[0174] Another type of nanoparticle that can be used as a label monomer are quantum dots. Quantum dots are fluorescing crystals 1-5 nm in diameter that are excitable by a large range of wavelengths of light. These crystals emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties.

[0175] Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to desired size and/or emission wavelengths. Use of quantum dots for labeling particles, in the context of the present invention, is new, but is old in the art of semiconductors. Two examples of earlier references in which quantum dots are embedded within intrinsic silicon epitaxial layers of semiconductor light emitting/detecting devices are United States Patent Nos. 5,293,050 and 5,354,707 to Chapple Sokol, et al.

[0176] In specific embodiments, one or more of the label attachments regions in the nanoreporter is labeled with one or more light-emitting dyes, each label attachment region containing, directly or indirectly, one or more label monomers. The light emitted by the dyes can be visible light or invisible light, such as ultraviolet or infra red light. In exemplary embodiments, the dye is a fluorescence resonance energy transfer (FRET) dye; a xanthene dye, such as fluorescein and rhodamine; a dye that has an amino group in the alpha or beta position (such as a naphthylamine dye, 1-dimethyl-aminonaphthyl-5-sulfonate, 1-anilino-8-naphthalene sulfonate and 2-p-touidiny-6-naphthalene sulfonate); a dye that has 3-phenyl-7-isocyanatocoumarin; an acridine, such as 9-isothiocyanatoacridine and acridine orange; a pyrene, a benoxadiazole and a stilbene; a dye that has 3-(s-carboxypentyl)-3'-ethyl-5,5'-dimethyloxycarbocyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); 6-carboxy-X-rhodamine (ROX); 6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein (JOE); ALEXA Fluor™; Cy2; Texas Red and Rhodamine Red; 6-carboxy-2',4,7,7'-tetrachlorofluorescein (TET); 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (HEX); 5-carboxy-2',4',5',7'-tetrachlorofluorescein (ZOE); NAN; NED; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; and Cy7.5; Alexa Fluor 350; Alexa Fluor 488; Alexa Fluor 532; Alexa Fluor 546; Alexa Fluor 568; Alexa Fluor 594; or Alexa Fluor 647.

[0177] The label monomers can be incorporated into a nanoreporter at different stages of its assembly, or into a component (e.g., a "flap" or of the nanoreporter prior to its assembly into the nanoreporter.

[0178] A label monomer can be directly attached to a nucleotide using methods well known in the art. Nucleotides can also be chemically modified or derivitized in order to attach a label monomer. For example, a fluorescent monomer such as a fluorescein molecule can be attached to dUTP (deoxyuridine-triphosphate) using a four-atom aminoalkynyl group. Each label monomer is attached to a nucleotide making a label monomer: nucleotide complex.

[0179] This label monomer: nucleotide complex can be incorporated into nucleic acids (for example, a DNA patch or a detection oligonucleotide) in a variety of ways. For example, a label monomer: nucleotide complex can be incorporated at only one location within a nucleic acid or at two or more locations within a nucleic acid.

[0180] Amine-reactive and thiol-reactive fluorophores are available and used for labeling nucleotides and biomolecules. Generally, nucleotides are fluorescently labeled during chemical synthesis, for example, incorporation of amines or thiols during nucleotide synthesis permit addition of fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum.

[0181] A nucleotide can be attached to a label monomer first and then be incorporated into a nucleic acid. Alternatively, an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid. For example aminoalkyl- ("AA-") modified UTP nucleotides can be incorporated into the RNA product during transcription. In various embodiments, 20% or more of UTP nucleotides in a transcription reaction to generate RNA patches are AA modified. In various embodiments, about 20% to 100%, 20% to 80%, 30 to 80%, 40 to 60% or 50% to 75% of UTPs in a transcription reaction are AA-modified, in a preferred embodiment, approximately 50% of UTPs in a transcription reaction are AA-modified.

[0182] In addition, for example, different types of label monomer: nucleotide complexes can be incorporated into a single acid nucleic acid, where one component of the nanoreporter code comprises more than one type of signal.

[0183] Fluorescent dyes that can be bound directly to nucleotides can also be utilized as label monomers. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These fluorescently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-

ddGTP and ROX-ddUTP, are commercially available.

[0184] Other types of label monomers that may be used to label a nanoreporter are quantum dots. Due to their very small size the quantum dots can be coupled into oligonucleotides directly without affecting the solubility or use of the oligonucleotide. In a preferred embodiment, only one oligonucleotide molecule is coupled to each nanoparticle. To synthesize an oligonucleotide-nanoparticle complex in a 1:1 ratio by conventional batch chemistry, both the oligonucleotide and the nanoparticle require a single reactive group of different kinds that can be reacted with each other. For example, if an oligonucleotide has an amino group and a nanoparticle has an aldehyde group, these groups can react to form a Schiff base. An oligonucleotide can be derivitized to attach a single amino or other functional group using chemistry well known in the art. However, when a nanoparticle is derivitized, it is covered with a chemical reagent which results in coating the entire surface of the nanoparticle with several functional groups.

[0185] The invention provides a method of coupling one oligonucleotide to one nanoparticle by chemically coupling the oligonucleotide on a solid surface such as the glass support used for the oligonucleotide synthesis.

[0186] For example, commercially available resins for oligonucleotide synthesis such as long chain alkylamino controlled pore glass (Icaa CPG) can be used.

[0187] Alternatively, a flat surface such as a derivitized microscope slide can be used. The surface density of the nascent oligonucleotide chains should be lower than the diameter of the nanoparticle. This can be achieved by either choosing a glass support with low surface density of the reactive groups, or by using diluted reagent for the first step of the oligonucleotide synthesis so that the surface is not saturated. Another point of consideration when using the standard glass matrices for oligonucleotide synthesis is to use a pore diameter higher than the nanoparticle diameter to ensure the flow of the reagents. For example, an oligonucleotide can be synthesized on a diluted basis relative to the solid support, for example one tenth of a normal synthesis, to ensure good spacing of the oligonucleotides on the glass support. After the oligonucleotide is synthesized with a reactive functional group, for example, an amino group, derivitized nanoparticles are passed over the glass support to react with the oligonucleotides. A sufficiently large pore size of the glass support can be chosen to prevent slogging with nanoparticles. For example, a pore size of about 200 nm can be used. After the reaction is complete, un-reacted groups on the nanoparticle can be blocked and the complexes can be uncoupled from the glass support.

5.5 THE NANOREPORTERCODE

5.5.1 DUAL NANOREPORTERS

[0188] A nanoreporter whose components exist in two molecular entities is referred to as a dual nanoreporter (or probe pair). In a dual nanoreporter according to the present invention, each component contains a target-specific sequence, which improves the specificity of and binding kinetics of the nanoreporter to its target. The two different target-specific sequences are designed or selected such that each recognizes a different portion of a target molecule.

[0189] Figures 1A- 1C illustrates embodiments of the invention involving dual nanoreporters. In Figures 1A and 1B, each of the two component of the nanoreporter is labeled, such that the nanoreporter's spectral code is formed only when the two components of the nanoreporter come together upon binding of the dual nanoreporter to its target molecule. However, in a dual nanoreporter, it is not necessary that both components are labeled. For example, as depicted in Figure 1C, one component of a dual nanoreporter is labeled with the nanoreporter code; and the other component attached to an affinity tag (arrow) that is useful to immobilize the nanoreporter for stretching an visualization.

5.5.2 REGISTERS

[0190] The term "register" refers to a set of alternating (every other) label attachment regions. Registers are useful where it is desirable to label adjacent label attachment regions without a spacer region, and where the signal emanating from adjacent label attachment regions cannot be spatially resolved using the desired method of detection. Thus, the signals detected with use of a register is that form by the alternating, rather than adjacent, label attachment regions. Signals detected from a plurality of registers (e.g., that together are all the label attachment regions) can be combined to form a nanoregister code. Generally when using registers, adjacent label attachment regions are labeled with spectrally distinguishable label monomers.

[0191] Examples of registers are depicted in Figures 3 and 5. For example, in Figure 3A-3B, there are 8 label attachment regions 1-8. Alternating label attachment regions 1, 3, 5 and 7 form one register, and label attachment regions 2, 4, 6 and 8 form another register. In Figure 3A, only one of the registers (1, 3, 5 and 7) is labeled; in Figure 3B, both registers are labeled.

5.6 AFFINITY TAGS

[0192] A variety of affinity tags known in the art may be used to purify and/or immobilize nanoreporters.

[0193] Where an affinity tag is used to immobilize a nanoreporter for the purpose of detection or imaging, it may be referred to herein as an "anchor." In a preferred embodiment, a biotin anchor is attached to the nanoreporter, allowing immobilization of the nanoreporter on a streptavidin coated slide.

[0194] An affinity tag that can be used to used for attachment to beads or other matrixes for a variety of useful applications including but not limited to purification.

[0195] Non-limiting examples of suitable affinity tags are provided below. It should be understood that most affinity tags could serve dual purposes: both as anchors for immobilization of the nanoreporters and tags for purification of the nanoreporters (whether fully or only partially assembled) or their components.

[0196] In certain embodiments, the affinity tag is a protein monomer. Examples of protein monomers include, but are not limited to, the immunoglobulin constant regions (see Petty, 1996, Metal-chelate affinity chromatography, in Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel et al., Greene Publish. Assoc. & Wiley Interscience), glutathione S-transferase (GST; Smith, 1993, Methods Mol. Cell Bio. 4:220-229), the *E. coli* maltose binding protein (Guan et al., 1987, Gene 67:21-30), and various cellulose binding domains (U.S. Pat. Nos. 5,496,934; 5,202,247; 5,137,819; Tomme et al., 1994, Protein Eng. 7:117-123), etc. Other affinity tags are recognized by specific binding partners and thus facilitate isolation and immobilization by affinity binding to the binding partner, which can be immobilized onto a solid support. For example, the affinity tag can be an epitope, and the binding partner an antibody. Examples of such epitopes include, but are not limited to, the FLAG epitope, the myc epitope at amino acids 408-439, the influenza virus hemagglutinin (HA) epitope, or digoxigenin ("DIG"). In other embodiments, the affinity tag is a protein or amino acid sequence that is recognized by another protein or amino acid, for example the avidin/streptavidin and biotin.

[0197] In certain aspects of the invention, the affinity tag is a nucleotide sequence. A large variety of sequences of about 8 to about 30 bases, more preferably of about 10 to about 20 bases, can be used for purification and immobilization of nanoreporters, and the sequence can be tandemly repeated (e.g., from 1 to 10 tandem repeats). Such a sequence is preferably not widely represented (that is, present in fewer than 5% of the genes, more preferably, present in fewer than 3% of the genes, and, most preferably, present in fewer than 1% of the genes) in the sample being assayed (for example, where the nanoreporter is used for detection of human cellular RNA, the sequence is preferably not widely represented in the human genome); have little or no secondary structure or self-complementarity either internally or with copies of itself when multimerized (that is, all secondary structures of the multimerized tag preferably have a T_m less than 25°C at 1 M NaCl); have no significant identity or complementarity with scaffold or segment sequences (that is, the T_m of complementary sequences is preferably less than 25°C at 0.2 M NaCl); and have a T_m of about 35-65°C, more preferably about 40-50°C, in 50 mM Na⁺.

[0198] In certain embodiments, different sequences are used as purification and immobilization tags. In this case, for example, the purification tag can be as described above, but the immobilization tag can be in the range of 10 to 100 bases, with a T_m up to 95°C in 50 mM Na⁺. An alternative embodiment would be to have the purification tag nested within the immobilization tag (e.g., the affinity tag would comprise a 25-base sequence of which 15 bases are used as a purification tag and the entire 25 bases are used as the immobilization tag).

[0199] In certain instances, the affinity tag can be used for labeling a nanoreporter in addition to purifying or immobilizing the nanoreporter.

[0200] As will be appreciated by those skilled in the art, many methods can be used to obtain the coding region of the affinity tags, including but not limited to, DNA cloning, DNA amplification, and synthetic methods. Some of the affinity tags and reagents for their detection and isolation are available commercially.

5.7 TARGET-SPECIFIC SEQUENCES

[0201] The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter scaffold.

[0202] The target specific sequence is generally an amino acid sequence (i.e., a polypeptide or peptide sequence) or a nucleic acid sequence.

[0203] In specific embodiments, where the target-specific sequence is an amino acid sequence, the target-specific sequence is an antibody fragment, such as an antibody Fab' fragment, a single chain Fv antibody.

[0204] The target-specific sequence is preferably a nucleic acid sequence, and is most preferably within an oligonucleotide that is either covalently attached (e.g., by ligation) or noncovalently attached (e.g., by hybridization) to the nanoreporter scaffold. A target-specific nucleic acid sequence is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific embodiments, the target-specific sequence is approximately 10 to 500, 20 to 400, 30 to 300, 40 to 200, or 50 to 100 nucleotides in length. In other embodiments, the target-specific sequence is approximately 30 to 70, 40 to 80, 50 to 90, or 60 to 100, 30 to 120, 40 to 140, or 50 to 150 nucleotides in length.

[0205] A target-specific nucleotide sequence preferably has a Tm of about 65-90°C for each probe in 825 mM Na⁺ (5X SSC), most preferably about 78-83°C.

[0206] In certain preferred embodiments, the target specific sequence of each probe of a dual nanoreporter is about 35 to 100 nucleotides (for a total target sequence of about 70 to 200 nucleotides, covered by 2 probes), most preferably about 40 to 50 nucleotides for each probe (for a total of about 80 to 100 nucleotides).

5.8 TARGET MOLECULES

[0207] The term "target molecule" is the molecule detected or measured by binding of a labeled nanoreporter whose target-specific sequence(s) recognize (are specific binding partners thereto). Preferably, a target molecule can be, but is not limited to, any of the following: DNA, cDNA, RNA, mRNA, peptide, a polypeptide/protein (e.g., a bacterial or viral protein or an antibody), a lipid, a carbohydrate, a glycoprotein, a glycolipid, a small molecule, an organic monomer, or a drug. Generally, a target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.

[0208] A target molecule can be part of a biomolecular sample that contains other components or can be the sole or major component of the sample. A target molecule can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The target molecule can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also, the target molecule can have either a known or unknown structure or sequence.

[0209] In certain specific embodiments, that target molecule is not a chromosome. In other specific embodiments, the target molecule is no greater than 1,000 kb (or 1 mb) in size, no greater than 500 kb in size, no greater than 250 kb in size, no greater than 175 kb in size, no greater than 100 kb in size, no greater than 50 kb in size, no greater than 20 kb in size, or no greater than 10 kb in size. In yet other specific embodiments, the target molecule is isolated from its cellular milieu.

[0210] In specific, non-limiting embodiments, the target molecule is one of the following antibodies or an antigen recognized by one of the following antibodies: anti-estrogen receptor antibody, an anti-progesterone receptor antibody, an anti-p53 antibody, an anti-Her-2/neu antibody, an anti-EGFR antibody, an anti-cathepsin D antibody, an anti-Bcl-2 antibody, an anti-E-cadherin antibody, an anti-CA125 antibody, an anti-CA15-3 antibody, an anti-CA19-9 antibody, an anti-c-erbB-2 antibody, an anti-P-glycoprotein antibody, an anti-CEA antibody, an anti-retinoblastoma protein antibody, an anti-ras oncoprotein antibody, an anti-Lewis X antibody, an anti-Ki-67 antibody, an anti-PCNA antibody, an anti-CD3 antibody, an anti-CD4 antibody, an anti-CD5 antibody, an anti-CD7 antibody, an anti-CD8 antibody, an anti-CD9/p24 antibody, an anti-CD 10 antibody, an anti-CD11c antibody, an anti-CD13 antibody, an anti-CD14 antibody, an anti-CD 15 antibody, an anti-CD19 antibody, an anti-CD20 antibody, an anti-CD22 antibody, an anti-CD23 antibody, an anti-CD30 antibody, an anti-CD31 antibody, an anti-CD33 antibody, an anti-CD34 antibody, an anti-CD35 antibody, an anti-CD38 antibody, an anti-CD41 antibody, an anti-LCA/CD45 antibody, an anti-CD45RO antibody, an anti-CD45RA antibody, an anti-CD39 antibody, an anti-CD100 antibody, an anti-CD95/Fas antibody, an anti-CD99 antibody, an anti-CD106 antibody, an anti-ubiquitin antibody, an anti-CD71 antibody, an anti-c-myc antibody, an anti-cytokeratins antibody, an anti-vimentins antibody, an anti-HPV proteins antibody, an anti-kappa light chains antibody, an anti-lambda light chain antibody, an anti-melanosome antibody, an anti-prostate specific antigen antibody, an anti-S-100 antibody, an anti-tau antigen antibody, an anti-fibrin antibody, an anti-keratins antibody, an anti-Tn-antigen antibody receptor protein, a lymphokine, an enzyme, a hormone, a growth factor, or a nucleic acid binding protein, a ligand for a cell adhesion receptor; a ligand for a signal transduction receptor; a hormone; a molecule that binds to a death domain family molecule; an antigen; a viral particle, a viral coating protein or fragment thereof, a toxic polypeptide selected from the group consisting of: (a) ricin, (b) Pseudomonas exotoxin (PE); (c) bryodin; (d) gelonin; (e) α -sarcin; (f) aspergillin; (g) restrictocin; (h) angiogenin; (i) saporin; (j) abrin; (k) pokeweed antiviral protein (PAP); and (1) a functional fragment of any of (a)-(k); a cytokine, or a soluble cytokine selected from the group consisting of erythropoietin, interleukins, interferons, fibroblast growth factors, transforming growth factors, tumor necrosis factors, colony stimulating factors and epidermal growth factor, Class I MHC antigens, class II MHC antigens, internalizing cell-surface receptors and/or viral receptors.

[0211] In specific, non-limiting embodiments, the target molecule is an antigen such as alpha fetoprotein, alpha-1 antitrypsin, α -2 macroglobulin, adiponectin, apolipoprotein-A-1, apolipoprotein-CIII, apolipoprotein-H, BDNF, β -2 microglobulin, C reactive protein, calcitonin, cancer antigen 19-9, cancer antigen 125, CEA, CD 40, CD 40 ligand, complement 3, GK-MB, EGF, ENA-78, endothelin-1, enrage, eotaxin, erythropoietin, Factor VII, FABP, ferritin, FGF-basic, fibrinogen, G-CSF, GST, GM-CSF, growth hormone, haptoglobin, ICAM-1, IFN-gamma, IgA, IgE, IGF-1, IgM, IL-1 α , IL-1 β , IL-1ra, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12 p40, IL-12 p70, IL-13, IL-15, IL-16, insulin, leptin, lipoprotein (a), lymphotactin, MCP-1, MDC, MIP-1 α , MIP-1 β , MMP-2, MMP-3, MMP-9, myeloperoxidase, myoglobin, PAI-1, PAP, PAPP-A, SGOT, SHBG, PSA (free), RANTES, serum amyloid P, stem cell factor, TBG, thrombopoietin, TIMP-1, tissue factor, TNF- α , TNF- β , TNF RII, TSH, VCAM-1, VEGF, or vWF.

[0212] In some embodiments, the target molecule is an autoimmune related molecule such as ASCA, β -2 glycoprotein,

Clq, centromere Prot. B, collagen type 1, collagen type 2, collagen type 4, collagen type 6, Cyto P450, ds DNA, histone, histone H1, histone H2A, histone H2B, histone H3, histone H4, HSC-70, HSP-32, HSP-65, HSP-71, HSP-90 α , HSP-90 β , insulin, JO-I, mitochondrial, myeloperoxidase, pancreatic islet cells, PCNA, PM-I, PR3, ribosomal P, RNP-A, RNP-C5 RNP, Sel-70, Smith, SSA, SSB, T3, T4, thyroglobulin, tTG, (celiac disease), or thyroid microsomal.

[0213] In some embodiments, the target molecule is a component isolated from an infectious agent, such as Cholera Toxin, Cholera Toxin β , Campylobacter jejuni, cytomegalovirus, Diphtheria toxin, Epstein-Barr NA, Epstein-Barr EA, Epstein-Barr VCA, Helicobacter pylori, HBV core, HBV envelope, HBV surface (Ad), HBV surface (Ay), HCV core, HCV NS3, HCV NS4, HCV NS5, hepatitis A, hepatitis D, HEV or f2. 3KD, HEV or f2 6 KD, HEV or f3 3KD, HIV-1 p24, HIV-1 gp41, HIV-1 gp120, HPV, HSV-1/2, HSV-1 gD, HSV-2 gD, HTLV-1/2, influenza A, influenza A H3N2, influenza B, Leishmania donovani, Lyme disease, mumps, M. pneumonia, M. tuberculosis, parainfluenza 1, parainfluenza 2, parainfluenza 3, polio virus, RSV, Rubella, Rubeola, Streptolysin O, Tetanus Toxin, T. pallidum 15kD, T. pallidum p47, T. cruzi, Toxoplasma, Varicella roster.

5.9 NANOREPORTER POPULATIONS

[0214] Disclosed herein are nanoreporter or nanoreporter label unit populations, for example nanoreporter or nanoreporter label unit libraries, that contain at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, or at least 1,000 unique nanoreporters or nanoreporter label units, respectively. As used herein, "unique" when used in reference to a nanoreporter or nanoreporter label units within a population is intended to mean a nanoreporter or label unit that has a code that distinguishes it from other nanoreporters or label units in the same population.

[0215] In particular we disclose nanoreporter populations with at least 5,000, at least 10,000, at least 20,000 or at least 50,000 unique nanoreporters or nanoreporter label units.

[0216] The nanoreporters in a population of nanoreporters can be singular nanoreporters, dual nanoreporters, or a combination thereof. The nanoreporters can be labeled or unlabeled.

[0217] The size of a nanoreporter population and the nature of the target-specific sequences of the nanoreporters within it will depend on the intended use of the nanoreporter. Nanoreporter populations can be made in which the target-specific sequences correspond to markers of a given cell type, including a diseased cell type. Nanoreporter populations may be generated in which the target-specific sequences represent at least 0.1%, at least 0.25%, at least 0.5%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the different type of transcripts in a cell. Optionally, nanoreporter populations are generated in which the target-specific sequences represent at least 0.1%, at least 0.25%, at least 0.5%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the different genes in a cell. Optionally, nanoreporter populations are generated in which at least some of the target-specific sequences represent rare transcripts in a cell or tissue. Such nanoreporter populations preferably represent at least 5 rare transcripts. In specific embodiments, such nanoreporter populations represent at least 10, at least 20, at least 30, at least 40 or at least 50 rare transcripts.

[0218] The cell or tissue may be a mammalian cell or tissue, and more Preferably is a human cell or tissue.

[0219] Optionally, the nanoreporter population is a diagnostic or prognostic nanoreporter populations. For example, a diagnostic nanoreporter population can be generated that is useful for screening blood products, in which the target-specific sequences bind to the nucleic acids of contaminating viruses such the hepatitis B, hepatitis C, and the human immunodeficiency virus. Alternatively, the diagnostic nanoreporter population may contain target-specific sequences corresponding to cellular disease markers, such as tumor antigens. Prognostic nanoreporter populations generally include target-specific markers that represent different stages of a given disease such as cancer. By selecting appropriate target-specific sequences, a nanoreporter population can be used both to diagnose and prognose disease.

5.10 BIOMOLECULAR SAMPLES

[0220] The probe pairs of the invention can be used to detect target molecule in any biomolecular sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: cells (including both primary cells and cultured cell lines), cell lysates or extracts (including but not limited to RNA extracts; purified mRNA), tissues and tissue extracts (including but not limited to RNA extracts; purified mRNA); bodily fluids (including, but not limited to, blood, urine, serum, lymph, bile, cerebrospinal fluid, interstitial fluid, aqueous or vitreous humor, colostrum, sputum, amniotic fluid, saliva, nasal and vaginal secretions, perspiration and semen, a transudate, an exudate (e.g., fluid obtained from an abscess or any other site of infection or inflammation) or fluid obtained from a joint (e.g., a normal joint or a joint affected by disease such as rheumatoid arthritis, osteoarthritis, gout or septic arthritis) of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred;

environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples including extracellular fluids, extracellular supernatants from cell cultures, inclusion bodies in bacteria, cellular compartments, cellular periplasm, mitochondria compartment, etc.

[0221] The biomolecular samples can be indirectly derived from biological specimens. For example, where the target molecule of interest is a cellular transcript, e.g., a messenger RNA, the biomolecular sample of the invention can be a sample containing cDNA produced by a reverse transcription of messenger RNA. In another example, the biomolecular sample of the invention is generated by subjecting a biological specimen to fractionation, e.g., size fractionation or membrane fractionation.

[0222] The biomolecular samples of the invention may be either "native," i.e., not subject to manipulation or treatment, or "treated," which can include any number of treatments, including exposure to candidate agents including drugs, genetic engineering (e.g. the addition or deletion of a gene), etc.

5.11 REPARATION OF LABEL MONOMERS

[0223] In addition to detecting an overall signal generated from a labeled nanoreporter, the invention provides for the determination of the spatial location of signals emanating from the label monomers (i.e., spots) on a nanoreporter, each spot representing the aggregate signal from label monomers attached to a given label attachment region. A spot may contain signals of the same wavelength or of different wavelengths. Thus, the nature of the spots on a nanoreporter and their location constitutes the nanoreporter code.

[0224] Any of a variety of means can be used to "stretch" the nanoreporter to separate the individual spots. For example, a nanoreporter can be stretched using a flowstretch technique (Henegariu et al, 2001, *Biotechniques* 31: 246-50), a receding meniscus technique (Yokota et al., 1997, *Nuc. Acids Res.* 25:1064-1070) or an electrostretching technique (Matsuura et al, 2001, *Nuc. Acids Res.* 29: E79).

[0225] The use of now-stretching, receding meniscus, or electro-stretching techniques allows for the separation of the label attachment regions within a nanoreporter so that one can determine spatially where a particular signal is positioned in the nanoreporter. Therefore, unique nanoreporters that have the same combination of label monomers and the same overall signal can be differentiated from one another based on the location of those label monomers within the nanoreporter.

[0226] This ability to locate the position of a label attachment region or spot within a nanoreporter allows for the position of the signal(s) emitted by the label monomers in each label attachment region to be used as a distinguishing characteristic when generating a set of unique nanoreporters. Hence, a complex set of nanoreporters can be generated using the same combination of starting label monomers by varying the positions of the label monomers within a nanoreporter.

[0227] Prior to stretching a nanoreporter, it is preferable to immobilize the nanoreporter to a solid surface using an affinity tag, as described in Section 5.6 above.

[0228] In certain aspects of the invention, one end of a nanoreporter is immobilized, either through specific or non-specific binding to a solid surface, the nanoreporter is stretched, and then the other end of the reporter is immobilized, also either through specific or non-specific binding to a solid surface. Accordingly, the nanoreporter is "frozen" in its stretched, or extended, state, to facilitate resolution of the nanoreporters code by detecting and/or imaging the signals emitted by the label monomers attached to a nanoreporter and their locations relative to one another. These aspects of the invention are described below in Section 5.12.

5.12 IMMOBILIZATION OF STRETCHED NANOREPORTERS

[0229] We disclose herein methods and compositions that facilitate the identification of primary structures of a variety of nanoreporters. In particular we disclose methods for the selective immobilization of nanoreporters in an extended state. Within the present invention, a nanoreporter can be selectively immobilized while fully extended under whatever force is used for the extension. In addition, the methods facilitate the selective immobilization of extended nanoreporters that are oriented with respect to each other. In other words, a plurality of nanoreporter can readily be immobilized in the same orientation with respect to each other.

[0230] Firstly we disclose methods for selectively immobilizing a nanoreporter in an extended state. For the methods, generally, a first portion of the nanoreporter is immobilized by any technique known to those of skill in the art. Indeed, the technique for immobilizing the first portion of the nanoreporter is not critical. In certain methods, the first portion of the nanoreporter can be immobilized selectively or non-selectively. In certain methods the first portion is immobilized by one or more covalent bonds. In certain methods, the first portion is immobilized by one or more non-covalent bonds. Exemplary immobilized first portions are described in the sections below.

[0231] With an immobilized first portion, the nanoreporter can be extended by any technique for extending a nanoreporter apparent to those of skill in the art. In certain methods, the technique for extending the nanoreporter is not critical for the methods of the invention. In certain methods, the technique for extending the nanoreporter appropriate for the

class of nanoreporter according to the judgment of one of skill in the art. In certain methods, the nanoreporter is extended by application of a force capable of extending the nanoreporter. The force can be any force apparent to one of skill in the art for extending the nanoreporter. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the nanoreporter are described in the sections below.

[0232] The nanoreporter is in an extended state if it would be recognized as extended by one of skill in the art. In certain methods, the nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter. In certain methods, the nanoreporter is in an extended state when its average hydrodynamic radius is more than double the average hydrodynamic radius of the nanoreporter in its native state as recognized by those of skill in the art.

[0233] The methods generally comprise the step of selectively immobilizing a second portion of the nanoreporter while it is in an extended state. This can result in an immobilized nanoreporter that is extended between the first and the second portion. Remarkably, since the nanoreporter is selectively immobilized while extended, that extension can be preserved in the immobilized nanoreporter. Generally, the first portion and the second portion of the nanoreporter are not the same.

[0234] The selective immobilization can be according to any technique for selective immobilization of a portion of a nanoreporter apparent to those of skill in the art. The selective immobilization can be through, for example, the formation of one or more covalent bonds or one or more non-covalent bonds, or both. Particular examples of selective immobilization techniques are described in the sections below. In particular methods, one or more binding pairs are used to immobilize the second portion of the nanoreporter.

[0235] The second portion can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any substrate judged to be useful for immobilization known to those of skill in the art. In certain methods, the second portion can be immobilized to another molecule. Further useful substrates include surfaces, membranes, beads, porous materials, electrodes, arrays and any other substrate apparent to those of skill in the art.

[0236] Secondly we disclose a composition comprising a selectively immobilized, extended nanoreporter. The compositions generally comprise a substrate and an extended nanoreporter selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the nanoreporter are immobilized onto the substrate, and the nanoreporter is in an extended state between the two portions. In certain compositions, at least one portion of the nanoreporter is selectively immobilized onto the substrate. In certain compositions, two or more portions of the nanoreporter are selectively immobilized onto the substrate. The nanoreporter can be extended and/or immobilized by any technique apparent to those of skill, including particularly the methods disclosed herein.

[0237] Thirdly we disclose methods for selectively immobilizing a nanoreporter in an oriented state. The nanoreporter can be any nanoreporter described above. In certain methods, the nanoreporter can be flexible, or in certain embodiments the nanoreporter can be rigid or semirigid. For these methods, generally, a first portion of the nanoreporter is immobilized as described above. With an immobilized first portion, the nanoreporter can be oriented by any technique for extending a nanoreporter apparent to those of skill in the art. In certain methods, the technique for orienting the nanoreporter is not critical. In certain methods, the technique for orienting the nanoreporter appropriate for the class of nanoreporter according to the judgment of one of skill in the art. In certain methods, the nanoreporter is oriented by application of force capable of orienting the nanoreporter. The force can be any force apparent to one of skill in the art for orienting the nanoreporter. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the nanoreporter are described in the subsections below.

[0238] The nanoreporter is in an oriented state if it would be recognized as oriented by one of skill in the art. For example, the nanoreporter is in an oriented state when it is in the field of a force capable of orienting the nanoreporter. Alternatively, the nanoreporter is in an oriented state when its termini are arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the nanoreporter. Additionally, a plurality of nanoreporters is in an oriented state when the termini of the nanoreporters are arranged in parallel, as recognized by those of skill in the art.

[0239] The methods generally comprise the step of selectively immobilizing a second portion of the nanoreporter while it is in an oriented state. This can result in an immobilized nanoreporter that is oriented between the first and the second portion. Remarkably, since the nanoreporter is selectively immobilized while extended, that orientation can be preserved in the immobilized nanoreporter. The selective immobilization can according to the methods described above.

[0240] Also disclosed herein is a composition comprising a selectively immobilized, oriented nanoreporter. The compositions generally comprise a substrate and an oriented nanoreporter selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the nanoreporter are immobilized onto the substrate, and the nanoreporter is in an oriented state between the two portions. In certain compositions, at least one portion of the nanoreporter is selectively immobilized onto the substrate. In certain compositions, both portions of the nanoreporter are selectively immobilized onto the substrate. The nanoreporter can be oriented and/or immobilized by any technique apparent to those of skill,

including particularly the methods taught herein.

[0241] The methods and compositions can be used for any purpose apparent to those of skill in the art. For instance, the immobilized and extended and/or oriented nanoreporter can be used as a label for a substrate on which the nanoreporter is immobilized. The primary sequence of the immobilized and extended and/or oriented nanoreporter can be identified by any technique apparent to those of skill. Advantageously, immobilization of the extended and/or oriented nanoreporter can facilitate such techniques. Optionally, the immobilized and extended and/or oriented nanoreporter can be used to guide the manufacture of nanopaths, for example to create nanowires or nanocircuits. Further uses for the immobilized and extended and/or oriented nanoreporters are described in the sections below.

[0242] All terms used herein have their ordinary meanings to those of skill in the art unless indicated otherwise. The following terms shall have the following meanings.

[0243] As used herein, the term "binding pair" refers to first and second molecules or moieties that are capable of selectively binding to each other, i.e. binding to each other with greater affinity than to other components in a composition. The binding between the members of the binding pair can be covalent or non-covalent. In certain embodiments, the binding is noncovalent. Exemplary binding pairs included immunological binding pairs (e.g., any haptenic or antigenic compound in combination with a corresponding antibody or binding portion or fragment thereof, for example digoxigenin and anti-digoxigenin, fluorescein and anti-fluorescein, dinitrophenol and anti-dinitrophenol, bromodeoxyuridine and anti-bromodeoxyuridine, mouse immunoglobulin and goat anti-mouse immunoglobulin) and nonimmunological binding pairs (e.g., biotin-avidin, biotin-streptavidin, hormone- hormone binding protein, receptor-receptor ligand (e.g., acetylcholine receptor-acetylcholine or an analog thereof), IgG-protein A, lectin-carbohydrate, enzyme-enzyme cofactor, enzyme-enzyme inhibitor, complementary polynucleotide pairs capable of forming nucleic acid duplexes, and the like). For instance, immunoreactive binding members may include antigens, haptens, aptamers, antibodies (primary or secondary), and complexes thereof, including those formed by recombinant DNA methods or peptide synthesis. An antibody may be a monoclonal or polyclonal antibody, a recombinant protein or a mixture(s) or fragment(s) thereof as well as a mixture of an antibody and other binding members. Other common binding pairs include but are not limited to, biotin and avidin (or derivatives thereof), biotin and streptavidin, carbohydrates and lectins, complementary nucleotide sequences (including probe and capture nucleic acid sequences), complementary peptide sequences including those formed by recombinant methods, effector and receptor molecules, hormone and hormone binding protein, enzyme cofactors and enzymes, enzyme inhibitors and enzymes, and so forth.

[0244] "Selective binding" refers to the any preferential binding of a pair of molecules or moieties for each other with respect to other molecules or moieties in a composition that would be recognized by one of skill in the art. Optionally, a pair of molecules or moieties selectively binds when they preferentially bind each other compared to other molecules or moieties. Selective binding can include affinity or avidity, or both, of one molecule or moiety for another molecule or moiety. Optionally, selective binding requires a dissociation constant (K_D) of less than about 1×10^{-5} M or less than about 1×10^{-6} M, 1×10^{-7} M, 1×10^{-8} M, 1×10^{-9} M, or 1×10^{-10} M. In contrast, non-selective binding optionally has significantly less affinity, for example, a K_D greater than 1×10^{-3} M.

[0245] "Extended state" refers to a nanoreporter in a state that would be recognized as extended by one of skill in the art. A nanoreporter may be in an extended state when it is extended relative to its native conformation in solution. A nanoreporter may be in an extended state when it is in the field of a force capable of extending the nanoreporter. An extended state of a nanoreporter can be determined quantitatively, such that, those of skill in the art will recognize R as the end-to-end vector of the nanoreporter, i.e. the distance between two termini of the nanoreporter, and $\langle R \rangle$ as the average end-to-end vector such that 95% of R will be within $2\langle R \rangle$ in a solution deemed appropriate to one of skill in the art. Exemplary solutions include, for example, a dilute solution of the nanoreporter in water or in a pH buffer. Optionally, a nanoreporter is in an extended state when R is greater than $2.0\langle R \rangle$.

[0246] "Oriented state" refers to a nanoreporter in a state that would be recognized as oriented by one of skill in the art. A nanoreporter may be in an oriented state when it is oriented relative to its native conformation in solution. The nanoreporter may be oriented when it is arranged in parallel with the field of a force capable of orienting the nanoreporter. The nanoreporter may be oriented when it is one of a plurality of nanoreporters that are arranged in parallel, as recognized by those of skill in the art.

5.12.1 METHODS OF SELECTIVE IMMOBILIZATION

[0247] As described above, we disclose methods for the selective immobilization of a nanoreporter in an extended state. The nanoreporter, once selectively immobilized, can be used for any purpose apparent to those of skill in the art.

[0248] The nanoreporter may be any polymer known to those of skill in the art. For instance, the nanoreporter can be a polysaccharide, a polypeptide or a polynucleotide. Useful polynucleotides include ribonucleic acids, deoxyribonucleic acids and other polynucleotides known to those of skill in the art.

[0249] The nanoreporter can be of any size that is sufficient to allow extension and immobilization of the nanoreporter according to the methods of the invention. In certain embodiments when the nanoreporter is a polynucleotide, the

nanoreporter can have a length of greater than 500 bp, greater than 750 bp, greater than 1 kb, greater than 1.5 kb, greater than 2.0 kb, greater than 2.5 kb, greater than 3.0 kb, greater than 4.0 kb or greater than 5.0 kb. In certain embodiments, when the nanoreporter is a polypeptide, the nanoreporter can have a size of greater than 50 amino acids, greater than 100 amino acids, greater than 200 amino acids, greater than 300 amino acids, greater than 400 amino acids, greater than 500 amino acids, greater than 750 amino acids, greater than 1000 amino acids, greater than 1500 amino acids, greater than 2000 amino acids, greater than 2500 amino acids, greater than 3000 amino acids, greater than 4000 amino acids or greater than 5000 amino acids. In certain embodiments, when the nanoreporter is a polysaccharide, the nanoreporter can have a size of greater than 50 saccharides, greater than 100 saccharides, greater than 200 saccharides, greater than 300 saccharides, greater than 400 saccharides, greater than 500 saccharides, greater than 750 saccharides, greater than 1000 saccharides, greater than 1500 saccharides, greater than 2000 saccharides, greater than 2500 saccharides, greater than 3000 saccharides, greater than 4000 saccharides or greater than 5000 saccharides.

[0250] The nanoreporter can be a native nanoreporter as understood by those of skill in the art, or the nanoreporter can be a non-native nanoreporter. When the nanoreporter is a polypeptide, the nanoreporter can comprise only naturally occurring amino acids, or the nanoreporter can comprise naturally occurring amino acids and non-naturally occurring amino acids. The other amino acids can be any amino acids, or derivatives or analogs thereof, known to those of skill in the art. When the nanoreporter is a polynucleotide, the polynucleotide can comprise only naturally occurring nucleotides, or the polynucleotide can comprise naturally occurring nucleotides and non-naturally occurring nucleotides. When the nanoreporter is a polysaccharide, the polysaccharide can comprise only naturally occurring saccharides, or the polysaccharide can comprise naturally occurring saccharides and non-naturally occurring saccharides. The polymers can comprise only non-natural monomers. The nanoreporter can comprise a plurality of classes of monomers, such as amino acids, nucleotides and/or saccharides.

[0251] Optionally, the nanoreporter comprises only one primary, covalently linked chain of monomers. For instance, when the nanoreporter is a polypeptide, in certain embodiments, the nanoreporter comprises only one primary amino acid chain. When the nanoreporter is a polynucleotide, the nanoreporter may be single stranded. The nanoreporter may comprise two primary, covalently linked chains of monomers. For instance, when the nanoreporter is a polypeptide, the nanoreporter may comprise two primary amino acid chains. When the nanoreporter is a polynucleotide, the nanoreporter may comprise two polynucleotide strands; the nanoreporter can be double stranded, in part or in whole. Further, the nanoreporter may comprise three or more primary, covalently linked chains of monomers. For instance, when the nanoreporter is a polypeptide, the nanoreporter optionally comprises three primary amino acid chains. When the nanoreporter is a polynucleotide, the nanoreporter optionally comprises three polynucleotide strands. For instance, the nanoreporter can comprise three strands F1, X and F2 where a portion of strand X is complementary to strand F1 and a portion of strand X is complementary to strand F2. An example is illustrated in Figure 13A. Optionally, the nanoreporter comprises more than three primary, covalently linked chains of monomers.

[0252] Advantageously, a nanoreporter can comprise one or more label that facilitate the detection, imaging or identification of the nanoreporter by techniques known to those of skill in the art. The label can be any detectable moiety known to those of skill in the art. Exemplary labels for nanoreporters include detectable isotopes, radioisotopes, fluors, dyes, enzymes, ligands, receptors, antigens, antibodies, lectins, carbohydrates, nucleotide sequences, and any other detectable label apparent to those of skill in the art.

[0253] A polynucleotide is optionally a polymer of natural (e.g. A, G, C, T, U) or synthetic nucleobases, or a combination of both. The backbone of the polynucleotide can be composed entirely of "native" phosphodiester linkages, or it may contain one or more modified linkages, such as one or more phosphorothioate, phosphorodithioate, phosphoramidate or other modified linkages. As a specific example, a polynucleotide may be a peptide nucleic acid (PNA), which contains amide interlinkages. Additional examples of synthetic bases and backbones that can be used in conjunction with the invention, as well as methods for their synthesis can be found, for example, in U.S. Patent No. 6,001,983; Uhlman & Peyman, 1990, Chemical Review 90(4):544 584; Goodchild, 1990, Bioconjugate Chem. 1(3):165 186; Egholm et al., 1992, J. Am. Chem. Soc. 114:1895 1897; Gryaznov et al., J. Am. Chem. Soc. 116:3143 3144, as well as the references cited in all of the above. Common synthetic nucleobases of which polynucleotides may be composed include 3-methyluracil, 5,6-dihydrouacil, 4 thiouracil, 5 bromouracil, 5-thorouracil, 5-iodouracil, 6-dimethyl aminopurine, 6-methyl aminopurine, 2-aminopurine, 2,6-diamino purine, 6-amino-8-bromopurine, inosine, 5-methylcytosine, 7-deazaadenine, and 7-deazaguanosine. Additional non-limiting examples of synthetic nucleobases of which the target nucleic acid may be composed can be found in Fasman, CRC Practical Handbook of Biochemistry and Molecular Biology, 1985, pp. 385-392; Beilstein's Handbuch der Organischen Chemie, Springer Verlag, Berlin and Chemical Abstracts, all of which provide references to publications describing the structures, properties and preparation of such nucleobases.

[0254] The nanoreporter can be prepared according to any technique apparent to those of skill in the art. Advantageously, nanoreporters according to the invention can comprise labels and/or members of binding pairs, as described in the sections below, that can be used to facilitate preparation and/or purification of the nanoreporter. In addition, certain nanoreporters of the invention are capable of forming complexes with molecules that comprise members of binding

pairs, as described below. These complexes can be used to facilitate preparation and/or purification of the nanoreporter or complex.

5.12.2 IMMOBILIZATION OF FIRST PORTION

[0255] In the methods of the invention, a first portion of the nanoreporter is immobilized. Generally, the first portion is immobilized if it would be recognized as immobilized by one of skill in the art. The first portion can be immobilized by any technique apparent to those of skill in the art. In certain embodiments, the technique for immobilization of the first portion of the nanoreporter is not critical for the methods of the invention.

[0256] The first portion of the nanoreporter can be at any location in the nanoreporter. In certain embodiments, the first portion is at a terminus of the nanoreporter. For the purposes of the invention and this disclosure in general, a portion of a nanoreporter can be "at a terminus" when it is less than five, four, three, two, one or zero monomers from a terminus of the nanoreporter. Of course, although many nanoreporters have two termini, the methods disclosed herein are applicable to nanoreporters having more than two termini and to nanoreporters having one or zero termini, e.g. circular nanoreporters. The first portion is not necessarily at a terminus of the nanoreporter.

[0257] The nanoreporter can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any moiety to which the nanoreporter can be immobilized without limitation. Optionally, the substrate is a surface, membrane, bead, porous material, electrode or array.

[0258] In some examples, the first portion of the nanoreporter can be immobilized non-selectively. In further examples, the first portion of the nanoreporter can be immobilized selectively. In advantageous examples, after the first portion of the nanoreporter is immobilized, some portion of the nanoreporter should be free to move sufficiently so that the nanoreporter can be extended in the following steps of the method. In particular, in certain embodiments, when the first portion of the nanoreporter is immobilized non-selectively, it is important that the entire nanoreporter not be immobilized non-selectively to an extent that prevents extension of any portion of the nanoreporter.

[0259] The immobilization can be by any interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain cases, the immobilization can be via electrostatic interaction with an electrode. In further cases, the immobilization is via electrostatic interaction with a substrate other than the electrode.

[0260] Optionally, the first portion of the nanoreporter comprises a first member of a binding pair. The first member of the binding pair can be covalently bound to the first portion of the nanoreporter, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. Usefully, the substrate onto which the first portion of the nanoreporter is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound. Figure 12A illustrates a nanoreporter that comprises a moiety F1 that is capable of selectively binding a moiety of the substrate. Moiety F1 can be, for example, biotin, capable of binding, for example, a substrate coated with avidin.

[0261] Optionally, the first portion of the nanoreporter can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins, and antibodies. The first portion of the nanoreporter would comprise a binding moiety capable of binding with the binding moiety of the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, surfaces comprising epoxy, aldehyde, gold, hydrazide, sulfhydryl, NHS-ester, amine, thiol, carboxylate, maleimide, hydroxymethyl phosphine, imidoester, isocyanate, hydroxyl, pentafluorophenyl-ester, psoralen, pyridyl disulfide or vinyl sulfone, or mixtures thereof. Such surfaces can be obtained from commercial sources or prepared according to standard techniques.

[0262] Advantageously, the first portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. Optionally, the nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. Similarly, a polypeptide nanoreporter can comprise a biotinylated amino acid residue. The substrate comprising avidin can be any substrate comprising avidin known to those of skill in the art. Useful substrates comprising avidin are commercially available including TB0200 (Accelr8), SAD6, SAD20, SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), streptavidin slide (catalog #MPC 000, Xenopore) and STREPT AVIDIN slide (catalog #439003, Greiner Bio-one).

[0263] The first portion of the nanoreporter may comprise a nucleotide sequence that is capable of selectively binding a nucleotide sequence on the substrate.

[0264] The first portion of the nanoreporter can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are commercially available including Optiarray-biotin (Accelr8), BD6, BD20, BD100, BD.500 and BD2000 (Xantec).

[0265] The first portion of the nanoreporter may be capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, a first

portion of the nanoreporter can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. Figure 13A illustrates a nanoreporter that is capable of selectively binding a second molecule X that is capable of selectively binding a third molecule that comprises F1. F1 is capable of selectively binding a moiety on a substrate. Figure 13B illustrates a nanoreporter that is capable of selectively binding a second molecule that comprises F1, and F1 is capable of selectively binding a moiety on a substrate.

[0266] The first portion of the nanoreporter can comprise a member of a binding pair that is capable of reacting with a member of binding pair on the substrate to form one or more covalent bonds. Exemplary useful substrates comprising reactive groups include those that comprise a reactive moiety selected from the group consisting of succinamides, amines, aldehydes, epoxies and thiols. The first portion of the nanoreporter would comprise a reactive moiety capable of reacting with the reactive moiety of the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, OptArray-DNA NHS group (Accler8), Nexterion Slide AL (Schott) and Nexterion Slide E (Schott).

[0267] The first portion of the nanoreporter can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the first portion of the nanoreporter could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as N-((2-pyridyl)dithio)ethyl-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

[0268] The first portion of the nanoreporter can be immobilized to the substrate via other binding pairs apparent to those of skill in the art.

5.12.3 EXTENSION OF THE NANOREPORTER

[0269] In certain methods of the invention, the nanoreporters comprising the probe pair are in an extended state. Generally, any nanoreporter is in an extended state if it would be recognized as such by one of skill in the art.

[0270] In certain embodiments, the nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter under conditions suitable for extending the nanoreporter. Such forces and conditions should be apparent to those of skill in the art. For instance, many nanoreporters can be extended by hydrodynamic force or by gravity, and many charged nanoreporters can be extended by electromagnetic force. In certain embodiments, the force can be applied to the nanoreporter indirectly. For instance, the nanoreporter can comprise or can be linked, covalently or noncovalently, to a moiety capable of being moved by a force. In certain embodiments, the nanoreporter can be linked to a moiety

[0271] In certain embodiments, the force is an electromagnetic force. For instance, when the nanoreporter is charged, such as a polynucleotide, the nanoreporter can be extended in an electric or magnetic field. The field should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter in an electric or magnetic field are described in Matsuura et al., 2002, *JBiomol Struct Dyn.* 20(3):429-36; Ferree & Blanch, 2003, *Biophys J.* 85(4):2539-46; Stigter & Bustamante, 1998, *Biophys J.* 1998 75(3):1197-210; Matsuura et al., 2001, *Nucleic Acids Res.* 29(16); Ferree & Blanch, 2004, *Biophys J.* 87(1):468-75; the contents of which are hereby incorporated by reference in their entirety.

[0272] In certain embodiments, the force is a hydrodynamic force. For instance, many nanoreporters, including polysaccharides, polypeptides, and polynucleotides, can be extended in the field of a moving fluid. The hydrodynamic force should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter in hydrodynamic field are described in Bensimon et al., 1994, *Science* 265: 2096-2098; Henegariu et al., 2001, *BioTechniques* 31: 246-250; Kraus et al., 1997, *Human Genetics* 99:374 - 380; Michalet et al., 1997, *Science* 277:1518-1523; Yokota et al., 1997, *Nucleic Acids Res.* 25(5):1064-70; Otobe et al., 2001, *Nucleic Acids Research* 29:109; Zimmerman & Cox, 1994, *Nucleic Acids Res.* 22(3):492-7, and U.S. Patent Nos. 6,548,255; 6,344,319; 6,303,296; 6,265,153; 6,225,055; 6,054,327; and 5,840,862, the contents of which are hereby incorporated by reference in their entirety.

[0273] In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force to extend the nanoreporter. In certain embodiments, The force should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter with gravity are described in Michalet et al, 1997, *Science* 277:1518-1523; Yokota et al., 1997, *Nucleic Acids Res.* 25(5); 1064-70; Kraus et al., 1997, *Human Genetics* 99:374 - 380.

[0274] In particular embodiments, the force is applied through a moving meniscus. Those of skill in the art will recognize that a moving meniscus can apply various forces to a nanoreporter including hydrodynamic force, surface tension and any other force recognized by those of skill in the art. The meniscus can be moved by any technique apparent to those of skill in the art including evaporation and gravity. Exemplary techniques for extending a nanoreporter with a moving meniscus are described in, for example, U.S. Patent Nos. 6,548,255; 6,344,319; 6,303,296; 6,265,153; 6,225,055; 6,054,327; and 5,840,862.

[0275] In particular, embodiments, the nanoreporter can be extended by an optical trap or optical tweezers. For instance, the nanoreporter can comprise or can be linked, covalently or noncovalently, to a particle capable of being trapped or moved by an appropriate source of optical force. Useful techniques for moving articles with optical traps or optical tweezers are described in Ashkin et al., 1986, Optics Letters 11 :288-290; Ashkin et al., 1987, Science 235: 1517-1520; Ashkin et al., Nature 330:769-771; Perkins et al., 1994, Science 264:822-826; Simmons et al., 1996, Biophysical Journal 70:1813-1822; Block et al., 1990, Nature 348:348 - 352; and Grier, 2003, Nature 424: 810-816.

[0276] In certain embodiments, the nanoreporter can be extended by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain nanoreporters are extended by a combination of an electric field and hydrodynamic force.

[0277] The nanoreporter is extended when it would be recognized as extended by one of skill in the art according to standard criteria for extension of a nanoreporter. In certain embodiments, the nanoreporter is extended when it loses most of its tertiary structural features as recognized by those of skill in the art. In certain embodiments, the nanoreporter is extended when it loses most of its secondary structural features as recognized by those of skill in the art. In certain embodiments, the nanoreporter is extended when its primary structural features are detectable in sequence when imaged according to standard techniques. Exemplary imaging techniques are described in the examples below.

[0278] In certain embodiments, an extended state of a nanoreporter can be recognized by comparing its hydrodynamic radius to its average hydrodynamic radius when free in dilute solution. For instance, in certain embodiments, a nanoreporter, or portion thereof, is extended when its hydrodynamic radius is more than about double its average hydrodynamic radius in dilute solution. More quantitatively, R represents the hydrodynamic radius of the nanoreporter, or portion thereof, and $\langle R \rangle$ represents the average hydrodynamic radius of the nanoreporter, or portion thereof, in dilute solution. The average $\langle R \rangle$ should be calculated such that R for the nanoreporter, or portion thereof, when unbound in dilute solution is less than $2\langle R \rangle$ 95% of the time. In certain embodiments, a nanoreporter, or portion thereof, is in an extended state when R is greater than $1.5\langle R \rangle$, greater than $1.6\langle R \rangle$, greater than $1.7\langle R \rangle$, greater than $1.8\langle R \rangle$, greater than $1.9\langle R \rangle$, greater than $2.0\langle R \rangle$, greater than $2.1\langle R \rangle$, greater than $2.2\langle R \rangle$, greater than $2.3\langle R \rangle$, greater than $2.4\langle R \rangle$, greater than $2.5\langle R \rangle$ or greater than $3.0\langle R \rangle$. In particular embodiments, a nanoreporter, or portion thereof, is in an extended state when R is greater than $2.0\langle R \rangle$.

5.12.4 ORIENTATION OF THE NANOREPORTER

[0279] Optionally, the nanoreporter is in an oriented state. Generally, any nanoreporter is in an oriented state if it would be recognized as such by one of skill in the art.

[0280] For example, the nanoreporter is in an oriented state when it is in the field of a force capable of orienting the nanoreporter under conditions suitable for orienting the nanoreporter. Such forces and conditions should be apparent to those of skill in the art.

[0281] The force may be an electromagnetic force. For instance, when the nanoreporter is charged, such as a polynucleotide, the nanoreporter can be oriented in an electric or magnetic field. The field should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter in an electric or magnetic field are described above.

[0282] Alternatively, the force is a hydrodynamic force. For instance, many nanoreporters, including polysaccharides, polypeptide, and polynucleotides, can be oriented in the field of a moving fluid. The hydrodynamic force should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter in hydrodynamic field are described above.

[0283] Alternatively, the force is gravity. Advantageously, the force of gravity can be combined with, for example, hydrodynamic force to orient the nanoreporter. In certain examples, the force should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter with gravity are described above.

[0284] Optionally, the nanoreporter can be oriented by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain nanoreporters are oriented by a combination of an electric field and hydrodynamic force.

[0285] The nanoreporter is oriented when it would be recognized as oriented by one of skill in the art according to standard criteria for orientation of a nanoreporter. For example, the nanoreporter is oriented when it is arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the nanoreporter. In certain examples, the nanoreporter is oriented when it is one of a plurality of nanoreporters that are arranged in parallel, as recognized by those of skill in the art. For instance, a plurality of nanoreporters can be oriented when the vector from a first terminus to a second terminus of a nanoreporter is parallel, as recognized by those of skill in the art, to the vectors between corresponding termini of other nanoreporters in the plurality.

5.12.5 SELECTIVE IMMOBILIZATION OF SECOND PORTION OF NANOREPORTER

[0286] As discussed above, in the methods disclosed herein, a second portion of the nanoreporter is selectively immobilized. The second portion of the nanoreporter can be any portion of the nanoreporter that is not identical to the first portion of the nanoreporter. Optionally, the second portion of the nanoreporter does not overlap any part of the first portion of the nanoreporter.

[0287] We disclose methods that comprise the single step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended or oriented state, and while a first portion of the nanoreporter is immobilized. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension or orientation of the nanoreporter are described in detail in the sections above.

[0288] We also disclose methods that comprise the step of extending a nanoreporter, while a first portion of the nanoreporter is immobilized, and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension of the nanoreporter are described in detail in the sections above.

[0289] We also disclose methods that comprise the step of immobilizing a first portion of a nanoreporter, the step of extending the nanoreporter while the first portion is immobilized and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension of the nanoreporter are described in detail above. In certain embodiments, the present invention provides methods that comprise the step of orienting a nanoreporter, while a first portion of the nanoreporter is immobilized, and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an oriented state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for orienting the nanoreporter are described in detail in the sections above.

[0290] We also disclose methods that comprise the step of immobilizing a first portion of a nanoreporter, the step of orienting the nanoreporter while the first portion is immobilized and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an oriented state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for orienting the nanoreporter are described in detail above.

[0291] The selective immobilization of the second portion of the nanoreporter can follow any technique for selective immobilization of a nanoreporter apparent to those of skill in the art. Significantly, in advantageous embodiments of the invention, the second portion of the nanoreporter is not immobilized non-selectively. Selective immobilization can allow the nanoreporter to be immobilized while in a fully extended state or nearly fully extended state. Selective immobilization can also allow the nanoreporter to be immobilized in an oriented manner. In other words, the first portion and second portion of the nanoreporter can be immobilized along the direction of the field or fields used to extend the nanoreporter, with the first portion preceding the second portion in the field. When a plurality of nanoreporters are immobilized, the can be uniformly oriented along the field.

[0292] The second portion of the nanoreporter can be at any location in the nanoreporter. Optionally, the second portion is at a terminus of the nanoreporter. Optionally, the second portion is not at a terminus of the nanoreporter. Optionally, the first portion, described in the sections above, is at one terminus of the nanoreporter, and the second portion is at another terminus of the nanoreporter.

[0293] As discussed above, the second portion of the nanoreporter is immobilized selectively. The immobilization can be by any selective interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. The immobilization can be via electrostatic interaction with an electrode. The immobilization may be via electrostatic interaction with a substrate other than the electrode.

[0294] If the first portion and the second portion of the nanoreporter are selectively immobilized to the same substrate, the techniques of selective immobilization should of course be compatible with the substrate. In particular, the techniques of immobilization are the same. For instance, on a substrate coated with avidin, both the first and second portion of the nanoreporter can be immobilized selectively via biotin - avidin interactions. However, as will be apparent to those of skill in the art, the same interaction need not be used at both the first and second portions for immobilization on the same substrate. For instance, the substrate can comprise multiple moieties capable of selective binding, or the first portion can be immobilized non-selectively, or other techniques apparent to those of skill in the art.

[0295] Optionally, the second portion of the nanoreporter comprises a first member of a binding pair. The second member of the binding pair can be covalently bound to the second portion of the nanoreporter, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. Usefully, the substrate onto which the second portion of the nanoreporter is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound.

[0296] The second portion of the nanoreporter can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates,

nucleic acids, receptors, lectins, and antibodies such as those described in the sections above.

[0297] Advantageously, the second portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. The nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. Similarly, a polypeptide nanoreporter can comprise a biotinylated amino acid residue. Useful substrates comprising avidin are described in the sections above.

[0298] Optionally, the second portion of the nanoreporter can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are described in the sections above.

[0299] Optionally, the second portion of the nanoreporter can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Exemplary useful substrates comprising reactive groups are described in the sections above.

[0300] The second portion of the nanoreporter can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the second portion of the nanoreporter could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

[0301] Alternatively, the second portion of the nanoreporter can be immobilized to the substrate via other binding pairs described in the sections above.

[0302] Alternatively, the second portion of the nanoreporter is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, the second portion of the nanoreporter can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. Figure 12B illustrates a nanoreporter of selectively binding a second molecule that comprises F3 that is, in turn, capable of selectively binding a moiety on a substrate. The interaction between the second portion of the nanoreporter and the molecule that comprises F3 can be mediated, for example, by an antigen-antibody interaction.

[0303] FIGS. 14A and 14B illustrate the selective immobilization of a nanoreporter according to methods of the present invention. In Figure 14A, a first portion of the nanoreporter comprises binding moiety F1 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F1 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. The nanoreporter of Figure 14A is extended by a force as described in the sections above. In Figure 14B, the force is an electrical potential. While extended, the nanoreporter is contacted with molecules comprising binding moiety F2 that is capable of selectively binding a moiety on the illustrate substrate S. Binding moiety F2 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. Significantly, up to three molecules comprising F2 are capable of selectively binding to a second portion of the nanoreporter to selectively immobilize it in its extended state. As illustrated, the molecules comprise a second binding moiety that electively binds a repeated binding moiety of the nanoreporter. The binding moieties can be, for instance, complementary nucleic acid sequences, as illustrated in Figure 14B. The resulting nanoreporter is selectively immobilized in an extended state and should remain extended even when the force is removed. The selectively immobilized, extended nanoreporter can be used for any purpose apparent to those of skill in the art.

5.12.6 IMMOBILIZATION OF TWO PORTIONS OF AN EXTENDED OR ORIENTED NANOREPORTER

[0304] Also disclosed herein are methods for selective immobilization of a first portion and a second portion of a nanoreporter that is in an extended or oriented state. Significantly, according to these methods, the nanoreporter need not be immobilized prior to application of a force capable of extending or orienting the nanoreporter.

[0305] In these methods, the nanoreporter is extended or oriented, or both, by a force capable of extending or orienting the nanoreporter. Such forces are described in detail in the sections above. The force is a force capable of extending or orienting the nanoreporter while maintaining the nanoreporter in one location, i.e. a force capable of extending or orienting without substantially moving the nanoreporter. Exemplary forces include oscillating electromagnetic fields and oscillating hydrodynamic fields. In a particular method, the force is an oscillating electrical field. Exemplary techniques for extending or orienting a nanoreporter in an oscillating electric field are described in Asbury et al, 2002, *Electrophoresis* 23(16):2658-66; Kabata et al, 1993, *Science* 262(5139):1561-3; and Asbury and van den Engh, 1998, *Biophys J.* 74: 1024-30.

[0306] In the methods, the nanoreporter is immobilized at a first portion and at a second portion while extended or oriented. Both the first portion and the second portion can be immobilized non-selectively, both can be immobilized selectively, or one can be immobilized selectively and the other non-selectively. Techniques for immobilization of the first portion and second portion are described in detail in the sections above.

5.12.7 SUBSTRATE FOR IMMOBILIZATION

[0307] In the methods discussed herein, the substrate for immobilization can be any substrate capable of selectively binding the nanoreporter apparent to those of skill in the art. Further, we disclose compositions comprising a selectively immobilized nanoreporter in an extended state. The compositions comprise a substrate, as described herein, having immobilized thereto a nanoreporter in an extended state. The nanoreporter can be, of course, immobilized according to a method of the invention.

[0308] The only requirement of the substrate is that it be capable of selectively binding the second portion of the nanoreporter as described above. Thus, the substrate can be a filter or a membrane, such as a nitrocellulose or nylon, glass, a polymer such as polyacrylamide, a gel such as agarose, dextran, cellulose, polystyrene, latex, or any other material known to those of skill in the art to which capture compounds can be immobilized. The substrate can be composed of a porous material such as acrylic, styrene methyl methacrylate copolymer and ethylene/acrylic acid.

[0309] The substrate can take on any form so long as the form does not prevent selective immobilization of the second portion of the nanoreporter. For instance, the substrate can have the form of a disk, slab, strip, bead, submicron particle, coated magnetic bead, gel pad, microtiter well, slide, membrane, frit or other form known to those of skill in the art. The substrate is optionally disposed within a housing, such as a chromatography column, spin column, syringe barrel, pipette, pipette tip, 96 or 384 well plate, microchannel, capillary, etc., that aids the flow of liquid over or through the substrate.

[0310] The nanoreporter can be immobilized on a single substrate or on a plurality of substrates. For instance, the first and second portions of nanoreporter may be immobilized on the same substrate, as recognized by those of skill in the art. The first portion of the nanoreporter can be immobilized on a first substrate while the second portion of the nanoreporter can be immobilized on a second substrate, distinct from the first.

[0311] The substrate can be prepared according to any method apparent to those of skill in the art. For a review of the myriad techniques that can be used to activate exemplary substrates of the invention with a sufficient density of reactive groups, see, the Wiley Encyclopedia of Packaging Technology, 2d Ed., Brody & Marsh, Ed., "Surface Treatment," pp. 867-874, John Wiley & Sons (1997), and the references cited therein. Chemical methods suitable for generating amino groups on silicon oxide substrates are described in Atkinson & Smith, "Solid Phase Synthesis of Oligodeoxyribonucleotides by the Phosphite Triester Method," In: Oligonucleotide Synthesis: A Practical Approach, M J Gait, Ed., 1984, IRL Press, Oxford, particularly at pp. 45-49 (and the references cited therein); chemical methods suitable for generating hydroxyl groups on silicon oxide substrates are described in Pease et al., 1994, Proc. Natl. Acad. Sci. USA 91:5022-5026 (and the references cited therein); chemical methods for generating functional groups on polymers such as polystyrene, polyamides and grafted polystyrenes are described in Lloyd Williams et al., 1997, Chemical Approaches to the Synthesis of Peptides and Proteins, Chapter 2, CRC Press, Boca Raton, FL (and references cited therein).

[0312] Exemplary useful substrates include surfaces coated with streptavidin, e.g. Accelr8 TB0200. Further useful substrates include surfaces coated with N-hydroxysuccinamide that are capable of reacting with a portion of a nanoreporter that comprises an amine. One such surface is OptArray-DNA (Accelr8). Additional useful surfaces are coated with aldehyde (e.g. Nexterion Slide AL., Schott) and surfaces coated with epoxy (e.g. Nexterion Slide E, Schott). Another useful surface is a biotinylated BSA coated surface useful for selective immobilization of a portion of a nanoreporter that comprises avidin or streptavidin.

5.12.8 METHODS OF USING SELECTIVELY IMMOBILIZED, EXTENDED OR ORIENTED NANOKEPORTERS

[0313] The selectively immobilized, elongated nanoreporters can be used to create macromolecular barcodes for the purposes of separation and sequential detection of labels. These labels spaced along the molecule provide a unique code that can be read when the nanoreporter is extended and immobilized. Extension and selective immobilization can facilitate the decoding of the macromolecular barcode.

[0314] The selectively immobilized, elongated nanoreporters can further be used for can be used in any context where detection or imaging of a nanoreporter might be useful. They can be used for diagnostic, prognostic therapeutic and screening purposes. For instance, they can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease. They can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample. The compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state. In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient.

5.12.9 KITS COMPRISING SELECTIVELY IMMOBILIZED, EXTENDED OR ORIENTED NANOREPORTERS

[0315] We also disclose kits comprising one or more components of the invention. The kits can comprises, for example, a substrate according to the invention and one or more extended or oriented, or both, nanoreporters selectively immobilized on the substrate. The kits can be used for any purpose apparent to those of skill in the art, including, those described above.

[0316] Furthermore, we disclose kits useful for the extension and selective immobilization of nanoreporters. The kits can comprise a substrate for immobilization and one or more binding partners to facilitate extension or immobilization of a nanoreporter. The binding partners could comprise a moiety useful for extension of the nanoreporter in an appropriate force. In certain embodiments, the binding partners could facilitate immobilization or selective immobilization of the nanoreporter to the surface. The kit could comprise a nanoreporter for extension and immobilization. The kit could comprise a device capable of extending the nanoreporter.

5.13 DETECTION OF NANOREPORTERS

[0317] Nanoreporters are detected by any means available in the art that is capable of detecting the specific signals on a given nanoreporter. Where the nanoreporter is fluorescently labeled, suitable consideration of appropriate excitation sources may be investigated. Possible sources may include but are not limited to are lamp, xenon lamp, lasers, light emitting diodes or some combination thereof. The appropriate excitation source is used in conjunction with an appropriate optical detection system, for example an inverted fluorescent microscope, an epi-fluorescent microscope or a confocal microscope. Preferably, a microscope is used that can allow for detection with enough spatial resolution to determine the sequence of the spots on the nanoreporter.

5.13.1 MICROSCOPY AND OBJECTIVE LENS SELECTION.

[0318] The major consideration regarding the microscope objective lens is with the optical resolution, which is determined by its numerical aperture (NA). Generally, the larger the NA, the better the optical resolution. The required NA is preferably at least 1.07 based on the relationship of $[\Delta] = 0.617 \lambda / \text{NA}$ ($[\Delta]$ = optical resolution and $[\lambda]$ = wavelength). The amount of light that is collected by an objective is determined by $\text{NA}^2 / \text{Mag}^2$ (Mag = magnification of the objective). Therefore, in order to collect as much light as possible, objectives with high NA and low magnifications should be selected.

5.13.2 CCD CAMERA SELECTION AND IMAGE CAPTURE TECHNIQUES.

[0319] When selecting a CCD camera, the first consideration is the pixel size, which partially determines the resolution of the imaging system. Optimally the optical resolution should not be compromised by the CCD camera. For example, if the optical resolution is 210-300 nm, which corresponds to 12.6-18 μm on a CCD chip after a 60 x magnification, in order to resolve and maintain the optical resolution there should be at least two pixels to sample each spot. Or the pixel size of the CCD chip should be at most 6.3- 9 μm .

[0320] The second consideration is detection sensitivity which can be determined by many factors that include but are not limited to pixel size, quantum efficiency, readout noise and dark noise. To achieve high sensitivity, select a qualitative camera with big pixel size (which can give big collection area), high quantum efficiency and low noise. An exemplary camera with these criteria is the Orca-Ag camera from Hamamatsu Inc. The chip size is 1344 x 1024 pixels; when using the 60x objective, the field of view is 144 x 110 μm .

5.14 COMPUTER SYSTEMS

[0321] We Also disclose computer systems that may be used to computerize nanoreporter image collection, nanoreporter identification and/or decoding of the nanoreporter code. Specifically, we disclose various computer systems comprising a processor and a memory coupled to the processor and encoding one or more programs. The computer systems can be connected to the microscopes employed in imaging the nanoreporter, allowing imaging, identification and decoding the nanoreporter, as well as storing the nanoreporter image and associated information, by a single apparatus. The one or more programs encoded by the memory cause the processor to perform the methods of the invention.

[0322] We also disclose computer program products for use in conjunction with a computer system (e.g., one of the above-described computer systems of the invention) having a processor and a memory connected to the processor. The computer program products of the invention comprise a computer readable storage medium having a computer

program mechanism encoded or embedded thereon. The computer program mechanism can be loaded into the memory of the computer and cause the processor to execute the steps of the methods of the invention.

[0323] The methods described in the previous subsections can preferably be implemented by use of following computer systems, and according to the following methods. An exemplary computer system suitable for implementation of the methods of this invention comprises internal components and being linked to external components. The internal components of this computer system include a processor element interconnected with main memory. For example, the computer system can be an Intel Pentium-based processor of 200 MHz or greater clock rate and with 32 MB or more of main memory.

[0324] The external components include mass storage. This mass storage can be one or more hard disks which are typically packaged together with the processor and memory. Such hard disks are typically of 1 GB or greater storage capacity. Other external components include user interface device, which can be a monitor and a keyboard, together with pointing device, which can be a "mouse", or other graphical input devices (not illustrated). Typically, the computer system is also linked to a network link, which can be part of an Ethernet link to other local computer systems, remote computer systems, or wide area communication networks, such as the internet. This network link allows the computer system to share data and processing tasks with other computer systems.

[0325] Loaded into memory during operation of this system are several software components, which are both standard in the art and special to the instant disclosure. These software components collectively cause the computer system to function according to the methods discussed herein. The software components are typically stored on mass storage. A first software component is an operating system, which is responsible for managing the computer system and its network interconnections. This operating system can be, for example, of the Microsoft Windows(R) family, such as Windows 95, Windows 2000, or Windows XP or, alternatively, a Macintosh operating system, a Linux operating system or a Unix operating system. A second software component may include common languages and functions conveniently present in the system to assist programs implementing the methods specific to this invention. Languages that can be used to program the analytic methods of the invention include, for example, C, C++, JAVA, and, less preferably, FORTRAN, PASCAL5 and BASIC. Another software component of the present invention comprises the analytic methods of this invention as programmed in a procedural language or symbolic package.

[0326] In an exemplary implementation, a nanoreporter code (i.e., a correlation between the order and nature of spots on a nanoreporter and the identity of a target molecule to which such a nanoreporter binds) is first loaded in the computer system. Next the user causes execution of analysis software which performs the steps of determining the presence and, optionally, quantity of nanoreporters with a given nanoreporter code.

[0327] The analytical systems also include computer program products that contain one or more of the above-described software components such that the software components may be loaded into the memory of a computer system. Specifically, a computer program product of the invention includes a computer readable storage medium having one or more computer program mechanisms embedded or encoded thereon in a computer readable format. The computer program mechanism encoded, e.g., one or more of the analytical software components described above which can be loaded into the memory of computer system and cause the processor of the computer system to execute the analytical methods of the present invention.

[0328] The computer program mechanism or mechanism are preferably stored or encoded on a computer readable storage medium. Exemplary computer readable storage media are discussed above and include, but are not limited to: a hard drive, which may be, e.g., an external or an internal hard drive of a computer system of the invention, or a removable hard drive; a floppy disk; a CD-ROM; or a tape such as a DAT tape. Other computer readable storage media will also be apparent to those skilled in the art that can be used in the computer program mechanisms disclosed herein.

[0329] We also disclose herein databases useful for practicing the methods of the present invention. The databases may include reference nanoreporter codes for a large variety of target molecules. Preferably, such a database will be in an electronic form that can be loaded into a computer system. Such electronic forms include databases loaded into the main memory of a computer system used to implement the methods of this invention, or in the main memory of other computers linked by network connection, or embedded or encoded on mass storage media, or on removable storage media such as a CD-ROM or floppy disk.

5.15 APPLICATIONS OF NANOREPORTER TECHNOLOGY

[0330] The compositions and methods of the invention are for use in diagnostic, prognostic therapeutic and screening purposes. The present invention provides the advantage that many different target molecules can be analyzed at one time from a single biomolecular sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample.

5.15.1 DIAGNOSTIC/PROGNOSTIC METHODS

[0331] The present methods can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease.

[0332] For example, a blood sample can be assayed according to any of the methods described herein to determine the presence and/or quantity of markers of a cancerous cell type in the sample, thereby diagnosing or staging the cancer.

[0333] Alternatively, the methods described herein can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample.

[0334] Thus, the target molecules detected using the compositions and methods of the invention can be either patient markers (such as a cancer marker) or markers of infection with a foreign agent, such as bacterial or viral markers.

[0335] Because of the quantitative nature of nanoreporters, the compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state.

[0336] In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient. For example, the amount of a particular marker for a tumor can be accurately quantified from even a small sample from a patient. For certain diseases like breast cancer, overexpression of certain genes, such as Her2-neu, indicate a more aggressive course of treatment will be needed.

5.15.2 ANALYSIS OF PATHOLOGY SAMPLES

[0337] RNA extracted from formaldehyde- or paraformaldehyde-fixed paraffin-embedded tissue samples is typically poor in quality (fragmented) and low in yield. This makes gene expression analysis of low-expressing genes in histology samples or archival pathology tissues extremely difficult and often completely infeasible. The nanoreporter technology can fill this unmet need by allowing the analysis of very small quantities of low-quality total RNA.

[0338] To use nanoreporter technology in such an application, total RNA can be extracted from formaldehyde- or paraformaldehyde-fixed paraffin-embedded tissue samples (or similar) using commercially available kits such as RecoverAll Total Nucleic Acid Isolation Kit (Ambion) following manufacturer's protocols. RNA in such samples is frequently degraded to small fragments (200 to 500 nucleotides in length), and many paraffin-embedded histology samples only yield tens of nanograms of total RNA. Small amounts (5 to 100 ng) of this fragmented total RNA can be used directly as target material in a nanoreporter hybridization following the assay conditions described herein. As described in Example 6 in Section 11 below, nanoreporter analysis of approximately 3.3 ng cellular RNA permitted detection of transcripts present at approximately 0.5 copy/cell.

5.15.3 SCREENING METHODS

[0339] The methods of the present invention may be for use, inter alia, in determining the effect of a perturbation, including chemical compounds, mutations, temperature changes, growth hormones, growth factors, disease, or a change in culture conditions, on various target molecules, thereby identifying target molecules whose presence, absence or levels are indicative of a particular biological states. Optionally, the methods are for use to elucidate and discover components and pathways of disease states. For example, the comparison of quantities of target molecules present in a disease tissue with "normal" tissue allows the elucidation of important target molecules involved in the disease, thereby identifying targets for the discovery/screening of new drug candidates that can be used to treat disease.

5.16 KITS

[0340] We disclose kits comprising one or more components of the invention. The kits can contain pre-labeled nanoreporters, or unlabeled nanoreporter with one or more components for labeling the nanoreporters. Moreover, the nanoreporters provided in a kit may or may not have target-specific sequences pre-attached. Optionally, the target sequences are provided in the kit unattached to the nanoreporter scaffold.

[0341] The kit can include other reagents as well, for example, buffers for performing hybridization reactions, linkers, restriction endonucleases, and DNA ligases,

[0342] The kit also will include instructions for using the components of the kit, and/or for making and/or using the labeled nanoreporters.

6. EXAMPLE 1: NANOREPORTER MANUFACTURING AND PROTOCOL

[0343] Herein is a step-by-step example of a method construction of a nanoreporter from various components.

[0344] It can be appreciated that various components can be constructed or added either at the same time, before or after other components. For example, annealing patch units or flaps to a scaffold can be done simultaneously or one after the other.

[0345] In this example the starting material is a circular M13mp18 viral vector. Using a single Linear strand M13Mp18, patch units are annealed to it to form a double stranded scaffold. Next, flaps are added then a target-specific sequence is ligated. Meanwhile purification steps aid to filter out excess, unattached patch units and flaps. Construction of labeled nucleic acids (patches and/or flaps and/or other labeled oligonucleotides) that bind the nanoreporter are also described.

[0346] Upon attachment (e.g., via hybridization) of a target molecule, the nanoreporter is attached to a surface and stretched. Finally the nanoreporters are imaged by a camera.

[0347] Nanoreporters were generated and successfully employed to detect target molecules using methods substantially as described in this example. An example of target detection using such this method is shown in Figure 4.

6.1 SCAFFOLD CONSTRUCTION

[0348] The oligonucleotide scaffold sequence selected was analyzed using Vector NTI® software. First, a single stranded nucleic acid was made from linearizing a circular M13mp18 single stranded DNA, which was commercially purchased from New England Biolabs. The circular M13mp18 was digested with BamH1 enzyme to linearize it. Materials used consisted of M13mp18 vector (250 ng/μl), Patch_1L_BamH1.02 (10 μM dilution of a 100 μM stock), 10X BamH1 Buffer, BamH1 enzyme. Protocol for making 0.8 pmol total of linear M13mp18 involve the following steps. 1) preheat heating block to 37 °C; 2) in a 0.65ml ependorff tube combine 40 μl of 250 ng/μl M13mp18 vector, 2 μl of 10 μM Patch_1L_BamH1.02, and 5 μl of 10X BamH1 Buffer; 3) place the ependorff tube in the 37°C heating block with foil over the top. Incubate the tube at 37°C for 15 minutes to allow the patch unit to hybridize to the M13mp18 scaffold; 4) after 15 minutes add 2 μl of BamH1 enzyme and let the reaction digest at 37 °C for 30 minutes, after which add an additional 2 μl of BamH1 enzyme and let the reaction continue to digest for another 30 minutes at 37°C (final volume of BamH1 enzyme is 8%); and 5) aliquot 10 μl into 0.65ml ependorff tubes and store in freezer (final concentration of linear M13mp18 is 200 ng/μl).

6.2 PATCH UNIT PREPARATION OF THE BASE PATCH POOLS (BPP).

[0349] Second, patch units are prepared in pools. Patch oligonucleotide sequences were selected for optimal length and desired homology/non-homology to M13mp18 strand and the human genomic sequence. Patches were commercially manufactured oligonucleotides (purchased from Integrated DNA technologies) either 60 or 65 nucleotide bases in length. 50 nucleotide bases of each patch oligonucleotide are complementary to the M13mp18 single stranded DNA, 10 nucleotide bases are complementary to an adjacent patch, and 5 nucleotides base pairs are complementary to a corresponding flap. The 10 nucleotide base match between patches forms a stem structure which stabilizes the structure and helps lift the flaps off the covered scaffold so they are more available to bind labeled oligonucleotides. Synthetic binding sites, the 5 nucleotide bases, on the patches for binding to the flaps make leveraging the power of a modular system possible.

[0350] The base patch pools contain nine patch units all corresponding to a specific letter grouping and position on the nanoreporter. For this example, there are four different fluorescent dyes (color) labeled A, B, C, and D and 8 different positions or regions where labeled nucleic acids can bind on a nanoreporter. For example, BPP A3 corresponds to all of the A patch units at position 3 (patch units 19-27) on the nanoreporter.

[0351] The nanoreporter positions are as follows:

- Position 1: Patch units 1-9 (A or C)
- Position 2: Patch units 10-18 (B or D)
- Position 3: Patch units 19-27 (A or C)
- Position 4: Patch units 28-36 (B or D)
- Position 5: Patch units 37-45 (A or C)
- Position 6: Patch units 46-54 (B or D)
- Position 7: Patch units 55-63 (A or C)
- Position 8: Patch units 64-72 (B or D)

[0352] Materials: right and left patches, pre-annealed to each other (each oligonucleotide is at a concentration of 10 μM). Materials for making 100 pmol of BPP 1: (In position 1, patch coordinate 1L is used for the BamHI digest -this patch is not included in BPP 1): 10 μl each pre-annealed (10 μM/each) patch unit (coordinates 2-9), 5 μl [20 μM] Patch_1R (A or C). Final concentration of each patch is 1.18 pmol/ μl. Materials for making 100 pmol of BPP 2-8: 10 μl each pre-annealed (10 μM/each) appropriate patch unit. There are 9 patch units added to each, or 90 μl total. Final concentration of each patch is 1.11 pmol/ μl.

EP 1 963 531 B1

[0353] Below is a table of all the patch unit pools made for this example, with 8 positions or regions for dye-labeled nucleic acids to bind on the nanoreporter. Positions 1, 3, 5, and 7 can bind to nucleic acid labeled with dye A or dye C, and a positions 2, 4, 6, and 8 can bind to nucleic acid labeled with dye B or dye D.

5 Table 2 of resulting Basic Patch Pools (correspond to labels on tubes)

10	BPP-A1 [Pre-Paired, Color=A, Coordinates 1-9] Patch_(1-9)R.A Patch_(2-9)L
15	BPP-B2 [Pre-Paired, Color=B, Coordinates 10-18] Patch_(10-18)R.B Patch (10-18)L
20	BPP-A3 [Pre-Paired, Color=A, Coordinates 19-27] Patch_(19-27)R.A Patch (19-27)L
25	BPP-B4 [Pre-Paired, Color=B, Coordinates 28-36] Patch_(28-36)R.B Patch_(28-36)L
30	BPP-A5 [Pre-Paired, Color=A, Coordinates 37-45] Patch_(37-45)R.A Patch (37-45)L
35	BPP-B6 [Pre-Paired, Color=B, Coordinates 46-54] Patch_(46-54)R.B Patch (46-54)L
40	BPP-A7 [Pre-Paired, Color=A, Coordinates 55-63] Patch_(55-63)R.A Patch_(55-63)L
45	BPP-B8 [Pre-Paired, Color=B, Coordinates 64-72] Patch_(64-72)R.B Patch (64-72)L
50	BPP-C1 [Pre-Paired, Color=C, Coordinates 1-9] Patch_(1-9)R.C Patch (2-9)L
55	BPP-D2 [Pre-Paired, Color=D, Coordinates 10-18] Patch_(10-18)R.D Patch (10-18)L
	BPP-C3 [Pre-Paired, Color=C, Coordinates 19-27] Patch_(19-27)R.C Patch (19-27)L
	BPP-D4 [Pre-Paired, Color=D, Coordinates 28-36] Patch_(28-36)R.D Patch (28-36)L
	BPP-C5 [Pre-Paired, Color=C, Coordinates 37-45] Patch_(37-45)R.C Patch (37-45)L
	BPP-D6 [Pre-Paired, Color=D, Coordinates 46-54] Patch_(46-54)R.D Patch (46-54)L
	BPP-C7 [Pre-Paired, Color=C, Coordinates 55-63]

(continued)

Patch_(55-63)R.C
Patch_(55-63)L
BPP-D8 [Pre-Paired, Color=D, Coordinates 64-72]
Patch_(64-72)R.D
Patch_(64-72)L

6.3 MATERIALS AND PREPARATION FOR ANNEALING THE SINGLE STRANDED OLIGONUCLEOTIDE WITH PATCH UNITS FOR A DOUBLE STRANDED SCAFFOLD.

[0354] Third, patch units are prepared to be annealed to the single stranded linear M13mp18, covering the strand in order to make a double stranded oligonucleotide scaffold. Conditions for annealing 60 and 65 nucleotide base patches to the M13mp18 need to occur at high salt concentrations so that binding will be very specific and patches will not anneal to an incorrect coordinate on the M13mp18 strand. For the annealing step, each patch unit is added at a 2:1 to 4:1 ratio with the single stranded M13mp18 sequence at 0.5 pmol total volume. Excess patches are removed before annealing flaps.

[0355] Materials used consisted of 20x SSC, linear M13mp18 (BamH1 digested at 0.08 pmol/ μ l or 200 ng/ μ l), appropriate base patch pools (BPP) (need 8 total at 1.11 pmol/ μ l - see above) and digital heat block set at 45°C. Annealing reaction make up is as follows. General guidelines: 2X each patch unit per M13mp18 molecule, pre-ligated flaps/patches (in position 1 or 8) added for purification later, and 5X SSC. Example (0.5 pmol of scaffold with F8 hook flaps) reaction consists of: 7.1 μ l BamH1 Digested M13mp18 strand at 0.071 μ M, 0.9 μ l each new Base Patch Pools at 1.11 μ M for first 7 positions: A1, B2, A3, B4, C5, B6 and A7:

1.7 μ l A1 BPP (Pre-Annealed, 12/15; at 1.18 μ M/each patch)
 1.8 μ l B2 BPP (Pre-Annealed, 12/15; at 1.11 μ M/each patch)
 1.8 μ l A3 BPP (Pre-Annealed, 12/15; at 1.11 μ M/each patch)
 1.8 μ l B4 BPP (Pre-Annealed, 12/15; at 1.11 μ M/each patch)
 1.8 μ l C5 BPP (Pre-Annealed, 12/15; at 1.11 μ M/each patch)
 1.8 μ l B6 BPP (Pre-Annealed, 12/15; at 1.11 μ M/each patch)
 1.8 μ l A7 BPP (Pre-Annealed, 12/15; at 1.11 μ M/each patch),
 2.4 μ l BPP-D8 (pool of the first seven patch units - coordinates 64, 65, 66, 67, 68, 69, and 70 at position 8 - "D" specificity) with purification tags- F8 (FHF, which anneal to patch coordinates 7 1 L, 71R, 72L, 72R, 73L making full split-flap/patch units that have "F" specificity for use as biotin linkers, at position F8) at 0.83 μ M, and 7.3 μ l 20X SSC. The final reaction volume will be 29.3 μ l at 0.027 pmol/ μ l.

[0356] Anti-Bam oligonucleotide is also added to anneal to region in M13 that is complementary to the (missing) 1L patch unit and to prevent recircularization of the M13 scaffold during ligation.

6.4 ANNEALING PATCH UNITS TO SINGLE STRANDED M13MP18 TO FORM A DOUBLE STRANDED SCAFFOLD.

[0357] The fourth step involves the protocol to anneal the patch units to the single stranded linear M13mp18, covering the strand in order to make a double stranded oligonucleotide scaffold, is performed in the following steps: 1) preheat heating block to 42°C, heat above reaction solution to 45°C in small PCR (or strip) tube(s) with foil over top for 15 minutes, turn heat block to 65°C and incubate for an additional 1 hour and 45 minutes and remove tubes, place on ice or freeze.

6.5 PURIFICATION OF NANOREPORTER SCAFFOLD USING BIOTIN AND MAGNETIC BEADS WITH STREPTAVIDIN.

[0358] The fifth step occurs before attaching the flaps, where excess patch units that have not annealed to the M13mp18 strand are separated from the double stranded oligonucleotide scaffold. A purification tag with a 5 nucleotide base homologous region to some of the patch units' complementary 5 nucleotide base overhang is annealed to 'hook' the scaffold. Biotinylated oligonucleotides are annealed to the 'purification tag' and magnetic beads with streptavidin are used to capture the scaffold using the biotinylated oligonucleotides. Excess patch units are removed with the supernatant. The scaffold melts off of the magnetic beads into solution for recovery.

6.6 ANNEAL THE D-BIOTIN CATCHERS TO THE PURIFICATION TAGS

[0359] Anneal the D-Biotin catchers to the purification tags on the nanoreporter (making 2X to amount of D8-flap positions available in solution, which is 2X to M13, or 4X final): 0.5 pmol X 25 hook oligonucleotide positions (5 multiplied by 5), 4X makes 50 pmols translates to 0.50 μ l of 100 pmol/ μ l D-biotin, add 0.5 μ l (D, E, F) - Biotin (at 100 μ M) to sample, mix and incubate at room temperature for 30 minutes.

6.7 PURIFICATION PROTOCOL TO WASH OFF UNATTACHED PATCH UNITS FROM DOUBLE STRANDED SCAFFOLD.

[0360] Anneal F-hook oligonucleotides in a 25 fold excess to nanoreporters in 5X SSC for 30 min at room temperature. Pipet 200 μ l DynaBead MyOne Streptavidin™ bead solution into 1.5 ml tubes, place on magnet and remove supernatant. Wash twice with 5X SSC by resuspending and clearing with magnet as in step above. Add 80 μ l of sample in 5X SSC (80 fmoles of sample in this example). Resuspend well, by placing on vortex for 15 minutes. Clear solution with magnet and transfer supernatant to fresh tubes for later gel analysis. While on magnet, wash pellets (do not resuspend) with 80 μ l TE by pipeting over pellet three times with the same 80 μ l volume originally added. Remove wash, place in freshly "wash" tubes for analysis. Heat up TE buffer to 45°C, add 80 μ l to each pellet and resuspend. Place tubes on 45°C heat block for 15 minutes, pipetting up/down once to insure beads remain suspended. Immediately clear product with magnet while warm and save. The majority of purified nanoreporters should be present in this product eluted at 45°C.

6.8 ANNEALING AND LIGATION OF FLAPS TO SCAFFOLD.

[0361] The sixth step involves split flap oligonucleotides which are annealed to the scaffold to make a 'covered scaffold.' Purification with magnetic beads is performed afterwards to remove excess split flaps. Ligation of the covered scaffold is done using T4 ligase to increase the stability of the structure. Only one type of flap is needed per fluorescent dye. Flaps are either 95 or 100 bases in length and have regions complementary to the patches, to labeled oligonucleotides and to each other. Each flap has 15 base repeating sequences for binding to labeled oligonucleotides. The repeat sequences are based on Lambda sequences that have been analyzed to remove any palindromes and hairpin structures.

[0362] Conditions for annealing the flaps are as follows. The sequence on the flaps that corresponds to the patch is 5 nucleotide base pairs long, and therefore the flaps anneal specifically to the patches even at high salt concentrations. The ratio of flaps to patches is 2:1. In order to increase stability at high temperatures, ligation of patches to each other and the flap to the patches may be carried out in the same reaction.

[0363] 1) Quantify the purified scaffold using a spectrometer at A260 nm. Calculate the volume needed for appropriate amount of nanoreporter to prepare. For this example we used 110 ng or 0.023 pmol, reading at A260 nm shows 7.7 ng/ μ l, or 14.3 μ l for 110 ng. 2) Setup ligation reaction as follows (volume will vary, depending on the purification and scale). Currently using 1.5X flaps to patches, calculate accordingly. For this example, there are four different fluorescent dyes (color) labeled A, B, C, and D and 8 different positions or regions where dye-labeled nucleic acids can bind on a nanoreporter. The number of positions for each color (in this case 1-4) multiply by 9 multiply by 1.5 moles of scaffold = moles of flaps to use.

[0364] For the nanoreporter with fluorescent dye in the sequence/positions [ABABCBAD] :

ABABCBAD =

A:40.5 x .023 = .93 pmol; vol: .93 μ l of SF (split flap)-AL at 1 μ M

.93 μ l of SF-AR at 1 μ M

B:40.5 x .023 = .93 pmol; vol: .93 μ l of SF-BL at 1 μ M

.93 μ l of SF-BR at 1 μ M

C:13.5 x .023 = .31 pmol; vol: .31 μ l of SF-CL at 1 μ M

.31 μ l of SF-CR at 1 μ M

D:13.5 x .023 = .31 pmol; vol: .31 μ l of SF-DL at 1 μ M

.31 μ l of SF-DR at 1 μ M

[0365] Ligation reaction (25 μ l total) consists of : Split Flaps (see above; 4.96 μ l, or $\sim$ 5 μ l total), 14.3 μ l of MODB-Scaffold at 0.0016 pmol/ μ l, 2.5 μ l 10X T4 ligation Buffer, 2.2 μ l NanoPure H₂O and 1 μ l T4 ligase. Incubate tubes 5 minutes at 45°C. Move to 37°C water bath, inc. for 5 minutes. Add 1 μ l T4 ligase to samples. Incubate for additional 1 hour at 37°C. Freeze immediately, or heat at 75°C for 5 minutes to kill T4 ligase.

6.9 LIGATION OF TARGET-SPECIFIC SEQUENCES TO NANOREPORTERS

[0366] The seventh step involves ligation of a target-specific sequence to the nanoreporter. A DNA target-specific sequence is designed to be complementary to the target molecule, which can be RNA (*e.g.*, mRNA) or DNA (*e.g.*, cDNA or genomic DNA). The target-specific sequence can be from 35, 60 or 70 nucleotide bases in length. The target-specific sequence can be ligated to the scaffold using a single stranded overhanging region on the covered scaffold. The scaffold with a single type of target-specific sequence can be manufactured separately and then mixed to form libraries.

6.10 NANOREPORTER CONSTRUCTION

[0367] Addition of oligonucleotides to a nanoreporter can be done at any point during the construction of a nanoreporter. In certain aspects of the present invention, a labeled oligonucleotide is 15 nucleotide bases long. On the 5' end, a single fluorophore dye is attached. Oligonucleotides with a particular fluorophore dye will generally have the same sequence. These labeled oligonucleotides bind to the repeat sequences of the split flaps. Fluorophores best suited for this example include but are not limited to Alexa 488, cy3, Alexa 594, and Alexa 647. The 15 nucleotide base length holds the fluorophores far enough apart so that they cannot quench each other and ensure that the labeled nucleic acids will be stable (will not melt off complementary strand) at conditions in the visualization process. Labeled oligonucleotides are stable at 40°C. This short length also allows for packing a large number of fluorescent dyes onto the flaps. In certain aspects of the invention, labeled oligonucleotides are introduced during the target sample processing.

6.11 ATTACHMENT OF NANOREPORTERS TO TARGET MOLECULES

[0368] Nanoreporters can be attached to target molecules using any means known to one of skill in the art. In an exemplary embodiment, dual nanoreporters are hybridized to target molecules by mixing 250 pmols each of both the first probe and the second probe with 125 pmols of target. The total volume is adjusted to 4 μ l and a final concentration of buffer of 5X SSC. This mixture is incubated in a covered PCR tube overnight at 42 degrees to allow hybridization to occur.

6.12 SURFACE ATTACHMENT

[0369] Once the nanoreporters are attached to both target molecule and corresponding labeled nucleic acids, *i.e.*, nucleic acids attached to label monomers, they are attached to a surface and stretched in resolve the order of signals emitted by the label monomers and thus identify the target molecule. In this example, the nanoreporters are stretched to spatially resolve their fluorescent dye codes which correspond to a particular target molecule. The nanoreporters are stretched by attaching one end to a surface (in this example - a coverslip, see preparations below). Two methods for surface attachment may be used: A) streptavidin coated slides from Accelr8 Corporation with the nanoreporters being biotinylated and B) biotin coated slides with the nanoreporters having streptavidin. In buffer, the nanoreporters are brought into contact with the active surface and allowed to incubate for a period of time. The reaction is performed in flow cells which were made from PDMS molded in etched silicon wafers to make the channels. Metal tubing is used to core wells at the ends of the channels for buffer and sample insertion. Channel dimensions are 0.5 mm or 1 mm wide and 54 μ m high. Once the sample has been loaded into the flow cell lane and incubated, the nanoreporters should be attached. Nanoreporters can be stretched either by applying a voltage or by removing the liquid with a receding meniscus leaving the strings stretched and dry.

6.13 PREPARATION OF SURFACE AND ASSEMBLY OF DEVICE

[0370] The binding surfaces (Accelr8 brand Streptavidin-OptiChem, coated coverslips) are shipped in units of 5 surfaces per slide container, and each container is enclosed with a package of silica dessicant in a foil pouch. The pouches are stored at -20 °C until use.

[0371] To prepare the surface for binding, a pouch is first pulled from the freezer and allowed to come to room temperature over several minutes. If previously unopened, the pouch is then sliced along one edge to form a slit, and the container of surfaces is removed. Upon removal of the required surface, the container is replaced in the pouch with its dessicant, the slit is sealed closed with a strip of packaging tape, and the pouch is replaced in the freezer.

[0372] The surface is then lightly rinsed with a stream of Nanopure water (Barnstead Nanopure Diamond) and soaked

for 10 minutes in 0.2 μm -filtered 1X PBS in a clean, slotted Coplin jar. After soaking, the surface is dipped in Nanopure water and dried by blowing filtered nitrogen across the surface edge.

[0373] The PDMS device used to mate with the surface and provide localization of the sample is cleaned just before use by applying cellophane tape to the PDMS surface and then peeling away dust or other particles which may have become attached during storage. The binding side of the Accelr8 surface is laid face-up, and the clean PDMS structure is centered, channel side down, on the surface. PDMS adheres readily to coated glass, and no further attachment mechanism is necessary.

6.14 SAMPLE BINDING AND WASHING

[0374] The sample is bound to the surface by first applying a 5 μL drop of the sample (currently diluted in 100 mM sodium borate buffer, pH 9.8) in one well of the chosen lane. The drop should just touch the point at which the channel joins the well (some sample may wick into the channel at this point). The channel is filled, and binding is equalized throughout the channel, by pulling the droplet through the channel to the opposite well using a very weak vacuum (< 2 kPa). The process is repeated for the other samples in their respective lanes. Excess fluid is then removed from the wells, the wells are taped to reduce evaporation, and the device is incubated at room temperature in the dark for 20 minutes.

[0375] After binding, the tape is removed, and the top well of each lane is filled with 100 μL of the borate buffer described above. About 20 μL of that buffer is pulled through the channels to the other wells using the vacuum, and the process is repeated once. All borate buffer is then removed from all wells, and the top well is filled with 1X TAE, pH 8.3. About 50 μL TAE is pulled through the channel, then all TAE is removed and the well is refilled. The process is repeated three times, for a total of about 150 μL of TAE rinse. Finally, all wells are filled with 100 μL 1X TAE.

6.15 ELECTROSTRETCHING

[0376] The bottom of the coverslip/PDMS device is spotted with immersion oil and placed on the microscope. Electrodes are inserted into the wells on opposite ends of the first PDMS channel (negative electrode in top well, positive in bottom). The first image of the channel will be taken close to the bottom well; the microscope stage is adjusted so that the area of interest is in focus.

[0377] Voltage (200 V) is then applied across the channel. Voltage is supplied by a DC power supply (Agilent E3630A) and amplified 100X through a home-built amplifier. After the current is applied, focus is readjusted, and the imaging process begins.

[0378] The electrostretching and imaging process is then repeated with the remaining channels. Image the bindings.

6.16 LIGHT SOURCE FOR THE FLUORESCENT DYES ON THE NANOREPORTER

[0379] In using an arc lamp as a light source, then best fluorophore selection is the brightest types without leading to fluorescent overlap such as Alexa 488, Cy3, and Alexa 594. Weaker fluorescent dyes such as Alexa 647 and Cy5.5 may also be used.

6.17 FILTERS TO IMAGE THE FLUORESCENT DYES ON THE NANOREPORTER

[0380] For the selected fluorophores Alexa 488, Cy3, Alexa 594 and Alexa 647 there maybe an overlap between the Cy3 and Alexa 594. However, custom ordering an emission filter with a bandwidth of 572-600 nm minimizes the overlap.

6.18 MICROSCOPE AND OBJECTIVE LENS TO IMAGE THE NANOREPORTERS

[0381] The microscope model used was the Nikon Eclipse TE2000E from Nikon Incorporation using the inverted fluorescence imaging station which has 6 filter cassettes that allow the selection of fluorescent emission from multiple fluorescent dye candidates. For the selected dyes, the optical resolution required is about 400 nm for all the wavelengths (500-700 nm). The selected objective lens is the Nikon Plan Apo TIRF lens which has a NA of 1.45 and magnification of 60. The optical resolution is ~ 210 -300 nm for different wavelengths.

7. EXAMPLE 2: PATCH/FLAP NANOREPORTER MANUFACTURING PROTOCOL

[0382] This example demonstrates another way of making a nanoreporter which consists of a single stranded linear M13mp18 viral DNA, oligonucleotide patch units and long flaps.

[0383] Nanoreporter label units were successfully generated using methods substantially as described in this example.

EP 1 963 531 B1

[0384] Pre-phosphorylated patch units and flaps are added together with the M13mp18 DNA vector and ligated together. After the ligation of the flaps to the patch units which are ligated to the M13mp18 DNA, the BamH1 enzyme is introduced to linearize the vector.

[0385] Prepare a batch of nanoreporters starting with 5 µg of M13mp18 as a scaffold. The hybridization may be scaled up accordingly to the desired amount. This process will take about 1-2 days to complete.

Materials:

Qty	Item	Vendor
20	250 ug/µl M13mp18 viral ssDNA	New England Biolabs
27 µl	0.74 pmol/µl Oligonucleotide Patch Unit Mix	IDT
8 µl	Long Flap Oligonucleotide A 100 pmol/µl	IDT
8 µl	Long Flap Oligonucleotide B 100 pmol/µl	IDT
0.5 µl	Flap patch Oligos at 100 pmol/µl from plates #529916 and #610591	IDT
31 µl	T4 Ligase 10x buffer	Fermentas
19 µl	T4 Ligase	Fermentas
15 µl	Optikase 10x buffer	USB
4.2 µl	100 mM ATP	ANY
5 µl	Optikase Enzyme 10 units/µl	USB
1 µl	BamH1 oligonucleotide 10 pmol/µl	IDT
20 µl	BamH1 10x buffer	Fermentas
3 µl	BamH1 Enzyme 10 units/µl	Fermentas

[0386] Preheat water bath to 37°C and 55°C before beginning protocol. Make sure buffers are all well mixed and thawed before using. A work plate should be available and labeled with the ordered oligos from IDT in plates #529916 and #610591. Take these two plates out and thaw at room temperature for 0.5-1 hours and spin down contents before removing the tape that covers the wells. Four separate reactions will be set up in 1.5 ml eppendorf tubes using specific oligonucleotides from these plates. To begin label these four separate tubes with roman numerals on their caps. Columns 5 and 6 A through H are for reaction I, Columns 7 and 8 A through H are for reaction ii are all found in plate #529916. Columns 1 and 2 are for reaction iv, and Columns 3 and 4 are for reaction iii.

[0387] Flap Ligations (Step A): Label four separate 1.5 ml tubes with roman numerals i through iv (mentioned above). Add the reagents below accordingly to each 50 µl reaction containing: 5 µl 10x ligase buffer, 0.5 µl/oligonucleotide from designated wells from plates #529916 and #610591, 4 µl Long Flap Oligo/reaction (A or B) for reactions I, ii and iv. 3 µl of LF for area iii, 29 H₂O for reactions I,ii and iv. 32 µl H₂O for reaction iii, and 4 µl T4 ligase. Preanneal oligos in this mix without the ligase at 37 °C for half an hour. Add ligase as last reagent and allow to ligate at room temperature for at least four hours. Product concentration is 1 pmol/flap/µl.

[0388] Flap Ligation Phosphorylation (Step B) Label four separate 1.5 ml tubes with roman numerals again, one through four with a P inside a circle to designate that the products are phosphorylated. Add the following reagents to the corresponding tube: 10 µl/Flap ligation. reaction (take 10 µl/flap ligation reaction above), 2.5 µl Optikase buffer, 0.5 µl 100 mM ATP, 11.5 µl H₂O, and 0.5 µl Optikase enzyme. Incubate at 37°C for 1 hour. Product concentration 0.4 pmol/flap/µl.

[0389] Oligonucleotide Patch Unit Phosphorylation (Step C) 27 µl Oligonucleotide Patch Unit mix 0.74 pmol/µl, 5 µl 10x buffer, 1 µl 100 mM ATP, 3 µl Optikase enzyme, and 14 µl H₂O. Once reagents are all together gently mix the solution by flicking the tube a few times and spin down. Incubate at 37 °C for 1 hour.

[0390] Hybridization to M13mp18 scaffold (Step D) In a new 1.5 ml tube add the following reagents: 20 µl M13mp18 at 250 ng/µl, 27 µl Phosphorylated Oligonucleotide Patch Units 0.4 pmol/µl (Step C), 12.5 µl/ Phosph. Flap Ligation (Step B) preheat at 55 °C for 5 minutes and put on ice, 11 µl 10x ligase buffer and heat entire mixture at 55 °C for 1 minute. Hybridize mixture at 37 °C for at least 4 hours.

[0391] Ligation (Step E) Spin down eppendorf contents. Add 1.2 µl 100 mM ATP and 3 µl T4 ligase. Gently mix contents by flicking the tube, then spin down.

[0392] BamH1 Digest (Step F): 1 µl of 10 pmol BamH1 oligo, 20 µl 10x BamH1 buffer and hybride at 37 ~ 1 hour. Adjust volume to 200 µl. Add 3 µl BamH1 enzyme. Incubate at 37 °C for 1 hour.

EP 1 963 531 B1

[0393] First step: start by adding 20 μ l of M13mp18 (NEB 250 μ g/ml) to a clean 1.7 ml eppendorf tube. Take 5 μ l of Phosphorylated Flap ligation reaction and preheat it at 70 for 2 minutes and immediately put on ice. Add the 5 μ l of each Phosphorylated Flap Ligation reaction (1 pmol/flap/ μ l) to the tube and gently mix by pipetting a few times. Incubate the eppendorf tube at 37 °C for 1 hour.

5 [0394] Second step: put 13.5 μ l Oligonucleotide Patch Unit Mix (0.74 pmol/ μ l) and 1 μ l of Acrydite Mix (10 pmol/ μ l) in a new eppendorf 1.7 ml eppendorf tube. Add 5 μ l 10x Optikinase buffer, 1 μ l 100 mM ATP and 27.5 μ l H₂O. Mix gently by pipetting the solution. Add 2 μ l Optikinase enzyme, gently mix by pipetting and incubate at 37°C for 1 hr.

[0395] Third step: take the phosphorylated oligos rxn and add it entirely to the contents of the M13mp18+Flaps Hybridization. The reaction is mixed gently by pipetting and it is allowed to incubate at 30 °C for 1 hour. After the hybridization is complete adjust the ATP by adding 1 μ l (100 ATP) to the reaction.

10 [0396] Fourth step: spin down contents in eppendorf tube and add 4 μ l T4 Ligase enzyme (5 units/ μ l), mix gently by pipetting. Incubate at room temperature for at least four hours. Add 1 μ l BamH 1 oligonucleotide (10 pmol/ μ l) to hybridize at room temperature while ligation is taking place.

[0397] Fifth step: digest ligation reaction by adding 4 μ l BamH1 enzyme (5 units/ μ l), mix gently by pipetting and incubate at 37°C for 1 hour. Once the incubation period is over. Take an aliquot of 500 ng for QC.

15 [0398] Sixth step: treat with Psoralen, UV or DMPA light for 15 minutes.

[0399] Calculations include:

20
$$5 \mu\text{g of M13} = 20 \mu\text{l stock from New England Biolabs} = 2 \text{ pmols}$$

$$\text{Oligonucleotide mix: 180-34 flap areas} - 10 \text{ Acrydite modified Oligos} = 0.74$$

25
$$\text{pmol/oligo}$$

$$10 \text{ pmols/oligonucleotide} = 13.5 \mu\text{l} = 1350 \text{ pmols}$$

30 [0400] Optikinase 1 unit converts 1 nmol of phosphate to ends - use excess. 4 μ l of Optikinase was used.

[0401] SEQ ID NO: 1 = M13mp18.

8. EXAMPLE 3: PROTOCOL FOR PRODUCTION OF RNA NANOREPORTERS

35 [0402] Nanoreporters were generated and successfully employed to detect target molecules using methods substantially as described in this example. An example of target detection using such this method is shown in Figure 6.

8.1 SCAFFOLD PRODUCTION

40 [0403] Single-stranded circular M13mp18 DNA (USB Corporation) is annealed to a 10-fold molar excess of an oligonucleotide complementary to the Bam HI recognition site (Bam Cutter oligo) and cut with Bam HI restriction enzyme to yield a linear single-stranded DNA backbone. An oligonucleotide complementary to the Bam Cutter oligonucleotide (anti-Bam oligonucleotide) is subsequently added in 50-fold excess to the Bam Cutter oligonucleotide to sequester free Bam Cutter oligonucleotide and thus prevent recircularization of the M13 during later steps.

45 [0404] The linear M13 molecule serves as a scaffold onto which RNA patches, or RNA segments, with incorporated fluorophores can be annealed.

8.2 PCR TO FORM DOUBLE-STRANDED POSITIONS ON THE M13 SCAFFOLD

50 [0405] Ten sets of oligonucleotide primer pairs were designed to create 10 different regions along the M13 scaffold. Each pair contains one primer which has a T7 RNA polymerase promoter at the 5' end. Regions 2-7 are designed to be 900 bases (approximately 300 nm) long, as this is the approximate size of a diffraction-limited spot (the smallest spot that can be achieved with standard optics). Regions 1 and 8 have both long and short versions: the long versions cover the whole 900-base region, while the short versions cover only a portion of the 900-base region to allow a target-specific sequence to be ligated. Thus a target-specific sequence can be attached to either end. The ends can also be used for attachment of anchors or tags.

55 [0406] PCR is performed using Taq polymerase and 0.5 ng of double-stranded M13mp18 (USB Corporation) as a

template. Reactions are cleaned up using a Qiaquick purification kit from Qiagen. Each PCR reaction yields a double-stranded fragment corresponding to one specific segment as illustrated below. These fragments are used as templates for the *in vitro* transcription of the RNA segments.

8.3 *IN VITRO* TRANSCRIPTION TO PRODUCE DARK RNA SEGMENTS

[0407] Using the PCR products described above as double-stranded templates, RNA segments are generated using an *in vitro* transcription kit from Ambion (Megascript T7 kit). The products of the transcription reactions are purified (including treatment with DNase I to remove template) using a RNeasy Kit from Qiagen.

8.4 *IN VITRO* TRANSCRIPTION TO PRODUCE RNA SEGMENTS MODIFIED WITH AMINOALLYL GROUPS

[0408] Using the PCR products described above as double-stranded templates, RNA segments for later dye-coupling are generated using an *in vitro* transcription kit from Ambion (MessageAmp aRNA kit). Aminoallyl-modified UTP nucleotides are incorporated into the RNA segments during transcription. The products of the transcription reactions are purified (including treatment with DNase I to remove template) using a RNeasy Kit from Qiagen.

8.5 DYE COUPLING OF AMINOALLYL RNA SEGMENTS TO PRODUCE COLORED RNA SEGMENTS

[0409] 20-100 µg of aminoallyl-modified RNA segment is coupled with NHS-ester dyes using Ambion Aminoallyl Labeling Kit. Dyes used include Alexa 488, Alexa 594 and Alexa 647 (Invitrogen/Molecular Probes) as well as Cy3 (Amersham).

[0410] Each segment is made separately in 4 colors so that each position on the scaffold can be filled with a segment in any of the four colors; thus different colors can be added at different positions to create many unique color combinations.

[0411] In this particular embodiment, adjacent segments must be of different colors or there may be dark segments interspersed so that each segment is detected as an individual 'spot'. Dark segments may be used as part of the nanoreporter code.

8.6 ASSEMBLY OF THE LABEL MOLECULE

[0412] Segments for each position are annealed in a 2:1 ratio of segment to M13 scaffold in 1X SSPE buffer at 70°C for 2 hours.

[0413] An assembled nanoreporter with labeled RNA segments is depicted in Figure 3A-3B. Figure 3A depicts a nanoreporter in which only alternate "spots" (1, 3, 5 and 7) are labeled, and Figure 3B depicts a nanoreporter in which every spot is labeled.

9. EXAMPLE 4: DETECTION OF TARGET (S2) RNA AND DNA MOLECULES USING AN RNA NANOREPORTER/ GHOST PROBE COMBINATION

9.1 SYNTHESIS OF PROBE AND TARGET OLIGONUCLEOTIDES

[0414] S2 DNA target oligonucleotide was synthesized and purified by polyacrylamide gel electrophoresis (Integrated DNA Technologies). S2 RNA target molecules were generated by *in vitro* transcription of PCR products corresponding to region of cloned SARS coronavirus gene (Invitrogen) using an Ambion Megascript™ kit per manufacturer's instructions. The S2 ghost probe (Figure 6A (i)) was complementary to a specific 50-base region of the S2 target sequence (S2-a) and was synthesized with a biotin-TEG monomer at the 5' end and purified by high performance liquid chromatography (Integrated DNA Technologies). A second oligonucleotide with 50 bps complementary to the S2 target (S2-b) plus 9 bp of a additional sequence used for ligation to the M13 scaffold (59bp total) was synthesized and purified by HPLC (Integrated DNA Technologies). Note that S2-a and S2-b target regions were not overlapping.

9.2 NANOREPORTER SYNTHESIS

[0415] Oligonucleotide S2-b was ligated to the 5' end of linearized M13 [Figure 6A (iii)], and the resulting product was purified away from residual unligated oligonucleotide by size-exclusion filtration through a YM100 filter (Millipore) per manufacturer's instructions. Amino-allyl-modified RNA segments complementary to M13 is positions 2, 4, 6, and 8 (SEQ ID NOs:) (Figure 1C) were generated from *in vitro*-transcription of DNA templates (PCR products) via the Ambion Megascript™ kit per manufacturer's instructions. The segments were then coupled to NHS-ester-modified Alexa 647 dye (Invitrogen) per Ambion's instructions (amino allyl MessageAmp™ II aRNA kit). RNA segments corresponding to

positions 1, 3, 5, and 7 of the M13 scaffold (Figure 1C) were generated as unmodified *in vitro*-transcribed RNAs from DNA templates as described above. Assembly of the nanoreporter was carried out by annealing 10 fmol/ μ l of each of the eight segments to 5 fmol/ μ l of the M13-S1-*b* scaffold for 2 hours at 70°C in 1X SSPE buffer (150 mM sodium chloride, 10 mM sodium phosphate, 1mM EDTA). The final product was a nanoreporter with 4 segments labeled with A647 (red) interspersed with dark segments.

9.3 HYBRIDIZATION CONDITIONS

[0416] Hybridization of nanoreporters and ghost probes to target were carried out under the following conditions: 5X SSPE (750 mM sodium chloride, 50 mM sodium phosphate, 5 mM disodium EDTA), 40 pM ghost probe (attachment oligonucleotide S2-a), 40 pM Nanoreporter S2-b, 100 ng/ μ l sheared salmon sperm DNA, 5X Denhardt's solution and 0.1% Tween. Final target concentrations were 20 pM S2 DNA target (Figure 6B) and 1pM S2 RNA target (Figure 6C). No target was added to the negative control (Figure 6D). The hybridization reaction was incubated at 65°C for at least 16h.

[0417] Hybridization reactions were diluted 1:2 with 100 mM Borate buffer solution (pH 9.8) and introduced into a flow cell channel and bound to a streptavidin-coated coverslip forming the bottom of the channel (Streptavidin-OptiChem coverslips from Accel8). Attachment to the slide by one end of the nanoreporter/target/ghost probe complex was achieved via interaction of the biotinylated ghost probe with the streptavidin surface. After rinsing the channel with additional borate buffer to remove excess reporters not bound to the surface, the buffer was exchanged with 1X TAE (40 mM Tris-acetate, 1 mM EDTA) and a current of 200V was applied to stretch out the nanoreporter/target complexes during image capture.

[0418] Images were obtained using a Leica DMI 6000B microscope with a 63X oil immersion objective (1.4 NA), Xcite-120 light source (Exfo), customized filter sets (Chroma Technologies), an Orca-ER CCD camera (Hamamatsu) and Metamorph data acquisition software (Molecular Devices).

[0419] As predicted, when the correct target molecule S2 hybridizes [Figure 6A (ii)] to both ghost probe [Figure 6A (i), S2-a] and S2-b-target-specific nanoreporter [Figure 6A (iii)], the ghost probe/target/nanoreporter complex forms a single species that attaches to the slide and was visualized as 4 spots when exposed to 647 nm wavelength light (Figure 6B, 6C, and 6E). The amount of binding was dependent on the target concentration. There was no significant binding in absence of S2 target sequence (Figure 6D).

10. EXAMPLE 5: NANOREPORTER COMPRISING A MONOVALENT OR BIVALENT ANTIBODY FRAGMENT

[0420] Where a target molecule is a protein or polypeptide, a nanoreporter can be generated in which the nanoreporter scaffold is a nucleic acid and the target-specific sequence is a monovalent or bivalent antibody fragment.

[0421] Using routine methods, an antibody that recognizes a target molecule of interest is optimally digested with pepsin to generate F(ab')₂ fragments. The two parts of the antibody or the two F(ab')₂ fragments generated by the pepsin digestion are separated by mild reduction, for example with 2-mercaptoethylamine. This reduction separates either the antibody or the two F(ab')₂ fragments into two monovalent fragments with two sulfhydryl groups that can be functionalized.

[0422] A heterobifunctional crosslinking reagent (e.g., m-Maleimidobenzoyl-N-hydroxysuccinimide ester from Pierce Biotechnology Inc.) is used to attach a maleimide to an oligonucleotide with an amine modification (which can be ordered from many sources, such as Integrated DNA Technologies). The NHS on the cross-linking reagent is reacted with the amine on the oligonucleotides to produce a maleimide-conjugated oligonucleotide.

[0423] This maleimide conjugated oligonucleotide is then reacted with one of the sulfhydryl groups on the antibody fragment. Due to steric limitations, it is preferable that only one oligonucleotide be attached to each fragment.

[0424] This monovalent or bivalent antibody fragment attached to an oligonucleotide can then be hybridized to a complementary sequence on a nanoreporter scaffold, to generate a reporter probe in which the target-specific sequence is an antibody sequence. Such a reporter probe can be used alone to detect the target molecule, or in conjunction with a ghost probe or another reporter probe whose target-specific sequence is a monovalent or bivalent antibody or antibody fragment that binds to a different portion of the same target molecule.

11. EXAMPLE 6: HYBRIDIZATION OF 25 CELLULAR GENES TO 100 NG OF PLACENTAL TOTAL RNA USING NANOSTRING REPORTER SYSTEM

[0425] Detection and quantitation of 25 endogenous cellular genes was carried out in a single multiplexed hybridization reaction. In addition, three non-human control sequences were spiked into each reaction that corresponded to approximately 10, 100 and 300 copies per cell, respectively. A negative control hybridization was also performed in the absence of cellular RNA.

11.1 HYBRIDIZATION REACTION

[0426] Each sample was hybridized in triplicate. Final concentrations of the hybridization reagents were as follows: 1.12nM total Nanoreporters (28 individual Nanoreporters at 40 pM each), 1.12nM total ghost probe (28 individual ghost probes), 5X SSPE (pH 7.5), 5X Denhardt's reagent, 100 ng/μl sheared salmon sperm DNA, 0.1% tween 20, 150 fM S3 spike DNA, 50 fM S4 spike, and 5 fM S6 spike. The final concentration of total placental RNA was 33 ng/μl. No total placental RNA was added to the negative control hybridizations. The final volume of the reaction was 30 μl. Reagents were mixed and incubated at 65°C in thermocycler block with heated lid for 20 hours.

Master mix	(1 Reaction)	(6 reactions)
1.8X hybridization mix*	16.7 μl	100 μl
25 endogenous gene reporters (0.6 nM each)	2 μl	12 μl
25 endogenous gene ghost probes (0.6 nM each)	2 μl	12 μl
Control reporters (0.6 nM each)	2 μl	12 μl
Control ghost probes (0.6 nM each)	2 μl	12 μl
10X control target mix	3 μl	18 μl
H ₂ O	1.3 μl	8 μl
Total	29 μl	174 μl
*Hybridization mix (9X SSPE, 9X Denhardt's reagent, 180 ng salmon sperm DNA, 0.18% tween 20)		

Reactions	1	2	3	4	5	6
Master mix	29 μl	29 μl	29 μl	29 μl	29 μl	29 μl
100 ng/μl placental RNA	1 μl	1 μl	1 μl	0 μl	0 μl	0 μl
H ₂ O	0 μl	0 μl	0 μl	1 μl	1 μl	1 μl
Total Rxn volume	30 μl	30 μl	30 μl	30 μl	30 μl	30 μl
Incubate reactions in thermocycler with heated lid overnight (18hours).						

11.2 POST-HYBRIDIZATION PURIFICATION

[0427] Hybridization reactions were purified to remove unhybridized reporters using an oligonucleotide complimentary to ghost probe attached to magnetic beads (F-bead). Hybridization reactions were diluted 5 fold in 0.1 % tween 20 to bring the final salt concentration to 1X SSPE and the solution added to 30 μl of F-beads (prewashed 2 times in 150 μl of 1X SSPE/0.1% tween 20). Hybridized complexes were allowed to bind to the beads at room temperature for 15 minutes with continuous rotation, washed once in 150 μl of 0.5X SSPE, and eluted in 25 μl of 0.1X SSPE for 15 minutes at 45°C.

11.3 BINDING, STRETCHING, AND IMMOBILIZATION

[0428] The samples were prepared for binding by addition of 1 μl of 1/1000 dilution of 0.1 uM Tetraspec™ fluorescent microspheres (product # T7279, Molecular Probes) and 3 μl of 1 M bis-tris propane (pH 9.0). Samples were loaded into a Nanostring fluidic device for attachment to Accer18 Optichem® slide coated with streptavidin (product #TB0200). After loading, slide surface was washed once with 1X TAE and prepared for electrostretching by addition of 40 μl of TAE to each well. Attached complexes were stretch by applying 200V across the fluidic channel. After 1 minute the samples were immobilized in the stretched position by adding 60 μl of 500 mM of G-hook oligo solution to the well containing the negatively charged electrode while continuing to apply voltage for 5 minutes. After immobilization the TAE solution is removed and replaced with anti-photobleaching reagent for imaging.

11.4 IMAGING

[0429] Slides were imaged on Nikon Eclipse TE2000E equipped with a metal halide light source (X-cite 120, Exfo Corporation) and a 60X oil immersion lens (1.4 NA Plan Apo VC, Nikon). For each field of view, 4 images at different excitation wavelengths (480, 545, 580 and 622) were acquired with an Orca Ag CCD camera (Hamamatsu) under control

of Metamorph software (Universal Imaging Corporation). Images were processed with custom image processing software.

11.5 DATA ANALYSIS

[0430] Raw data were extracted from processed images using custom software. Data were normalized to the average counts for control spikes in each sample. To determine if a gene was "detected" by the system, the counts obtained for each gene from hybridizations containing RNA were compared to counts obtained in hybridizations without RNA using a Student's ttest. Genes with p values <0.05 were determined to be detected. After background subtraction, the concentrations of cellular mRNA were estimated from the linear regression of the spike controls. These concentrations were converted to copies per cell using the following assumptions: 1 cell contains 10 pg total RNA; each cell contains 300,000 mRNA molecules; final volume of the reaction is 30 μ l.

11.6 RESULTS AND CONCLUSION

[0431] Table 3 below shows the results of the data analysis described above. These results show that using the nanoreporter technology described herein, it was possible to detect transcripts, such as CASP3, that are present at a concentration of less than 1 transcript/cell. Thus, the nanoreporter technology provides an exquisitely sensitive means of detecting and quantifying gene expression.

Table 3: Transcript Concentration and Abundances

Gene	Avg counts*	error (counts)	concentration (fM)	error (conc.)	calculated copies/cell	error (copy/cell)	Detected/ Not Detected (p<0.05)
GM2A	149	17	3.39	0.39	6.12	0.07	D
ATF4	68	2	1.55	0.06	2.80	0.01	D
CTNNB1	792	50	17.95	1.19	32.44	0.22	D
IRF1	221	20	5.01	0.47	9.05	0.09	D
STAT5A	120	11	2.72	0.25	4.91	0.05	D
CREG1	409	17	9.28	0.44	16.76	0.08	D
CASP3	13	1	0.30	0.03	0.54	0.00	D
CCL20	2	1	0.04	0.03	0.07	0.01	ND
NMI	115	2	2.61	0.07	4.72	0.01	D
XBP1	719	46	16.30	1.10	29.45	0.20	D
PCGF4	75	18	1.70	0.40	3.08	0.07	D
IFI27	747	41	16.94	1.00	30.61	0.18	D
TAF7	185	11	4.19	0.26	7.57	0.05	D
OAS3	74	9	1.68	0.20	3.03	0.04	D
C2	850	49	19.28	1.19	34.83	0.21	D
IL6	8	3	0.19	0.07	0.34	0.01	D
MyD88	94	6	2.13	0.14	3.85	0.03	D
HIF1A	130	7	2.95	0.17	5.33	0.03	D
APOA2	-1	2	-0.01	-0.05	-0.03	-0.01	ND
KISS	6825	130	154.79	4.52	279.65	0.82	D
ELK3	55	4	1.25	0.09	2.27	0.02	D
CBF2	72	3	1.64	0.07	2.96	0.01	D

(continued)

Gene	Avg counts*	error (counts)	concentration (fM)	error (conc.)	calculated copies/cell	error (copy/cell)	Detected/ Not Detected (p<0.05)
IF130	625	47	14.16	1.10	25.59	0.20	D
RELB	35	5	0.78	0.11	1.42	0.02	D
CTCF	103	3	2.35	0.09	4.24	0.02	D
*Normalized and background subtracted.							

[0432] The hybridization methods described herein have been performed in single multiplexed reactions containing up to 120 different reporters with similar hybridization efficiencies and results.

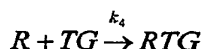
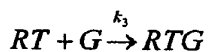
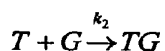
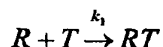
12. EXAMPLE 7: CONSIDERATIONS REGARDING NANOREPORTER HYBRIDIZATION KINETICS

12.1 BACKGROUND

[0433] Solution hybridizations with a large excess of probe over target follow pseudo-first order kinetics. In this regime the speed of the reaction depends only on the probe concentration and not on the target concentration. For a two-probe, one-target strategy to provide accurate information on the concentration of a target in solution, the probes should both be present in excess of the target. The possible concentration range is preferably therefore bounded on the lower end by the concentration of the target. However, the useful concentration range for the nanoreporter technology described herein is practically bounded on the lower end by the amount of time needed to perform the hybridization.

12.2 HYBRIDIZATION KINETICS

[0434] In preferred embodiments, target detection and quantification assays are performed in which the target (T) must hybridize to both a reporter probe (R) and a ghost probe (G) to be detected (for example by affinity selection and detection of complexes comprising only (R) and (G), which in turn only form complexes in the presence of (T)). Assuming that these reactions are irreversible, there are four possible elementary reactions that occur.



[0435] Because RT and TG are intermediate complexes of two out of the three species, these four reactions can be simplified to



[0436] However, to quantitatively calculate the rate of production of RTG (the reporter-target-ghost probe complex), all four reactions must be considered. The differential equations describing the system are:

$$\begin{aligned}
 \frac{dC_G}{dt} &= -k_2 C_G C_T - k_3 C_G C_{RT} & \frac{dC_R}{dt} &= -k_1 C_R C_T - k_4 C_R C_{TG} & \frac{dC_T}{dt} &= -k_2 C_G C_T - k_2 C_R C_T \\
 \frac{dC_{TG}}{dt} &= k_2 C_G C_T - k_4 C_R C_{TG} & \frac{dC_{RT}}{dt} &= k_1 C_R C_T - k_3 C_G C_{RT} \\
 \frac{dC_{RTG}}{dt} &= k_4 C_R C_{TG} + k_3 C_G C_{RT}
 \end{aligned}$$

where C_R , C_T , C_G , C_{RT} , C_{TG} , and C_{RTG} are the concentrations of the various species, and k_1 – k_4 are the kinetic constants for the four elementary reactions. Values for these kinetic constants when the probes and targets are complementary single-stranded molecules (i.e. when there is no purification tag on the ghost probe and no reporter) can be calculated from data available in the literature (Wetmur, J. Annu. Rev. Biophys. Bioeng. 1976.5:337-361).

$$k = k_N \frac{\sqrt{L}}{N} \frac{\alpha_{salt}}{\alpha_{ref}}$$

[0437] In the above equation, k_N is the nucleation rate constant, L is the nucleic acid length (in base pairs), N is the nucleic acid complexity (equal to L for non-repetitive sequences) and α_{salt} and α_{ref} are corrections for salt concentration (Britten et al., 1974, Methods in Enzymology 29E:363-406). In the nanoreporter systems described herein, the kinetic constants will depend on the sizes of the attached ghost probe tags and reporter probe. Without being bound by any theory, it is the inventors' belief that the kinetic constants will have the same dependence on length that an elementary reaction has on the diffusion constants of the reactants.

$$k = k_N \frac{\sqrt{L}}{N} \frac{\alpha_{salt}}{\alpha_{ref}} \frac{D_1 + D_2}{2D_{50}}$$

[0438] In the above equation D_1 and D_2 are the diffusion constants of the two reacting species (see the reactions above) and D_{50} is the diffusion constant of a 50-mer single-stranded DNA molecule. Assuming a 100-base single-stranded target, 100-base single-stranded ghost probe, and 7200-base double stranded reporter, the relevant kinetic constants are

$$k_1 = 2.64 \times 10^5 \text{ L/mol/s}$$

$$k_2 = 6.55 \times 10^5 \text{ L/mol/s}$$

$$k_3 = 3.99 \times 10^5 \text{ L/mol/s}$$

$$k_4 = 1.91 \times 10^5 \text{ L/mol/s}$$

[0439] Numerically solving the system of differential equations with these kinetic constants (assuming at least a 10-fold excess of probes over target) yields the prediction that 5 pM reporter and 5 pM ghost probe will drive hybridization to 10% of completion in an overnight reaction (16-18 hours). At concentrations lower than 5 pM, the amount of completely hybridized molecules is likely impractical to measure. Thus, in a preferred embodiment, the lower concentration of a nanoreporter component (ghost probe and/or reporter probe) is 5pM.

12.3 ENTANGLEMENT OF REPORTERS

[0440] As probe concentrations increase, theory predicts that hybridization kinetics speed up without bound - the only limit being the solubility of the probes. However, the reporter probe can be very large compared to the target-specific sequence in the nanoreporter systems of the invention. Without being bound by any theory, the inventors believe that by its attachment to the reporter probe the kinetics of the target-specific sequence are altered from classical solution hybridization kinetics. Because the reporter probe is a large, polymeric molecule, it can have long-lived interactions (entanglements) with other nanoreporters when they come into contact. At low concentration the probability of two polymers becoming entangled is small, but as the concentration and/or size of a polymer in solution increases, these interactions become more and more common. In the extreme case of very long molecules at very high concentration the polymers form a permanent network, or gel, in solution. For solution hybridization to occur, a probe (e.g., a nanoreporter probe)/target pair must diffuse through solution until they contact one another and a hybridization nucleus forms. Classically, hybridization reactions are not diffusion limited because the translational diffusion of the molecules is faster than the nucleation of the hybridization (i.e., the probe and target diffuse together and interact many times before a nucleation occurs). In dilute solution its large size will slow the translational diffusion of the reporter probe, but may not significantly affect the kinetics. At some intermediate concentration, the reporter probes take up almost all of the space in the solution, effectively forming a permanently entangled gel, and can no longer diffuse in solution. However, the ghost probe and the targets are smaller molecules that are believed to still diffuse through the entangled reporter probes, allowing hybridization to take place (although possibly at a slower rate). The inventors also believe that at some higher concentration the reporter probe in solution will also hinder the movement of the ghost probe and the targets to the point that the reaction becomes diffusion limited. This concentration (which is not quantitatively known and depends upon the reporter probe structure, the ghost probe structure, and the target size) is the upper limit of the useful concentration range in the nanoreporter system, and can be empirically determined by one of skill in the art guided by the principles described herein.

12.4 LENGTH DEPENDENCE OF KINETICS

[0441] Since the limiting upper concentration for hybridization depends upon both the reporter structure and ghost probe structure (of which there are many possible variations), a theoretical framework to predict the permutations of useful concentration ranges is useful in the practice of the invention. Classical theory predicts that hybridization kinetics depend only on the size of the smaller probe. Theory would therefore predict that the size of the reporter will not play a role in the hybridization kinetics as long as both the target molecule and the ghost probe are significantly smaller. Theory then predicts that the rate of hybridization (for a constant target length) depends on $1/L^{1/2}$, where L is the length of the ghost probe, due to steric inhibition of hybridization. Consequently, the kinetics of hybridization will be faster with smaller ghost probes. As the ghost probe length increases, the hybridization rate should decrease as $1/L^{1/2}$. If a constant ghost probe length is assumed, then the range of reporter lengths and concentrations that will result in a measurable amount of hybridization events can be defined. Once a reporter size has been defined, then the approximate range of ghost probe sizes can be determined. This is an iterative process, but may give good starting points from which to gather data to generate detailed empirical guidelines, given that the theories that the inventors' rationale is based upon were generated from hybridization data in systems that do not employ a reporter probe.

12.5 ENTANGLEMENT THRESHOLD

[0442] A reporter probe is essentially a polymer in free solution, which behaves as a random coil. The volume occupied by a single reporter, V_p , can be calculated from polymer physics theories according to the Freely-Jointed Chain model (FJC, for a flexible polymer, such as single-stranded DNA or RNA) or the Worm-Like Chain model (WLC, for a stiff polymer such as double-stranded DNA or a reporter). For either model

$$V_p = \frac{4}{3} \pi R_g^3$$

where R_g is the radius of gyration. For the FJC

$$R_g = b \left(\frac{N}{6} \right)^{0.6}$$

Where b is the segment length and N is the number of segments in the chain. For the WLC

$$R_g = \sqrt{\frac{1}{6} N b^2 - \frac{b^2}{4} + \frac{b^2}{4N} \left(1 + \frac{1}{2N} (e^{-2N} - 1) \right)}$$

The entanglement threshold concentration is defined as the concentration where the entire volume of the solution is occupied by the reporters.

$$C^* = \frac{3}{4\pi R_g^3 N_A}$$

where N_A is Avogadro's number. Above this concentration it is assumed that the translational diffusion of the reporters is severely restricted. The entanglement threshold concentration varies with the reporter structure. As the reporter length increases, the entanglement threshold decreases (as $1/L^{1.5}$). From the equations above, the theoretical entanglement threshold for reporter probes with different spot sizes and different lengths can be calculated. The result of such calculations is shown in Figure 17, which shows that for a 7200 bp RNA/DNA hybrid reporter probe with 8 label attachment regions of about 900 bp each, the entanglement threshold is about 70 nM.

[0443] If both the target and the ghost probe are much smaller than the reporters, then they will most likely be free to diffuse through the solution even at these high concentrations of reporters. Initial data indicates that hybridization kinetics do not slow appreciably up to a concentration of 80 nM with a 7200-bp reporter probe, a 100-base target, and a 100-base ghost probe.

12.6 EFFECT OF ENTANGLEMENT THRESHOLD ON MULTIPLEXING

[0444] Assuming that the maximum concentration for reporters in a hybridization reaction is C^* , then the concentration of each reporter (specific to a particular target) is equal to C^*/M , where M is the multiplex of the reaction (number of different targets being addressed simultaneously). Conversely, the possible multiplex level for a particular reporter structure can be calculated from the lower limit of probe concentration (C_p from kinetics ~10 nM) and the entanglement threshold

$$M = \frac{C^*}{C_p}$$

[0445] If the number of nanoreporter codes available does not depend on reporter probe size, then the multiplexing of the nanoreporter depends primarily on the reporter probe size and concentration (since it is much larger than the ghost probe). Because the ghost probe makes an insignificant contribution to entanglement during hybridization, it is the inventors' belief that the concentration of the ghost probe can be increased far above the concentration of the reporter probe. In Table 4 below, the maximum total ghost probe concentration ($[G]$) is set to 1000 nM for all reporter concentrations. This difference in concentration of ghost probe and reporter probe is an adjustable parameter. Preliminary experiments show that in a multiplex hybridization reaction with a 7200 bp reporter and 100b ghost, 40pM of each reporter probe and 200 pM of each ghost probe results in near complete hybridization in an overnight reaction.

12.7 OPTIMAL SIZE AND CONCENTRATION RANGES

[0446] Below in Table 4 is a summary of the optimal useful size and concentration ranges of the ghost probe and reporter probe at different multiplexing as approximated by the above theories. It is the inventors' belief that ghost probes up to about 200 bases will be practical for most applications.

Table 4: Optimal size and concentration ranges of reporter probe, ghost probe and target, as well as multiplicity of probes, in the nanoreporter systems of the invention.

Reporter Length (bp)	Ghost Length (b)	Minimum [R] (pM)	Minimum [G] (pM)	Maximum [R] (nM)	Maximum [G] (nM)	Max Multiplex
2000	100	5	5	603	1000	114417
2000	50	4	4	603	1000	161811
2000	200	7	7	603	1000	80905
3000	100	6	6	292	1000	45182
3000	50	5	5	292	1000	63897
3000	200	9	9	292	1000	31948
4000	100	7	7	178	1000	23912
4000	50	5	5	178	1000	33817
4000	200	11	11	178	1000	16908
5000	100	8	8	123	1000	14746
5000	50	6	6	123	1000	20854
5000	200	12	12	123	1000	10427
6000	100	9	9	91	1000	9988
6000	50	6	6	91	1000	14125
6000	200	13	13	91	1000	7062
7200	100	10	10	68	1000	6792
7200	50	7	10	68	1000	6792
7200	200	14	10	68	1000	6792
8000	100	11	11	57	1000	5444
8000	50	7	7	57	1000	7699
8000	200	15	15	57	1000	3850
10000	100	12	12	40	1000	3419
10000	50	8	8	40	1000	4835
10000	200	17	17	40	1000	2417

13. EXAMPLE 8: EXEMPLARY EMBODIMENTS FOR DUAL NANOREPORTER ASSEMBLY

[0447] This section describes an embodiment for assembly of a dual nanoreporter in which one probe is a ghost probe and the other probe is a reporter probe comprising color RNA segments assembled on an M 13 backbone. The ghost probe is attached to a biotinylated F-hook and the reporter probe is attached to a biotinylated G-hook. The dual nanoreporter is hybridized to a biomolecular sample to detect and quantify a target molecule. The steps below do not have to be performed in the order presented. Moreover, each particular step represents a specific embodiment that may be combined with embodiments other than those presented below.

13.1 PREPARATION OF THE M13 SCAFFOLD

[0448] Single-stranded circular M13mp18 DNA (USB Corporation) is annealed to a 5-fold molar excess of an oligonucleotide complementary to the Bam HI recognition site (Bam Cutter oligo) and cut with Bam HI restriction enzyme to yield a linear single-stranded DNA backbone. An oligonucleotide complementary to the Bam Cutter oligonucleotide (anti-Bam oligonucleotide) is subsequently added in 50-fold excess to sequester free Bam Cutter oligonucleotide and thus prevent recircularization of the M13 during later steps.

[0449] The linear M13 molecule serves as a scaffold onto which RNA patches, or RNA segments, with incorporated fluorophores can be annealed.

13.2 ATTACHMENT OF A TARGET-SPECIFIC SEQUENCE TO THE SCAFFOLD

[0450] An oligonucleotide comprising a sequence (of, *e.g.*, 30-70 nucleotides) complementary to the target nucleic acid of interest, plus 9 bp of additional sequence used for ligation to the M13 scaffold, is generated and ligated to the 3' end of the linearized M13 scaffold.

13.3 ATTACHMENT OF G-TAGS TO THE SCAFFOLD

[0451] A G-tag (*e.g.*, an oligonucleotide having the sequence 5'-AACATCACACAGACC AACATCACACAGACC AACATCACACAGACC AACATCACACAGACC AGCCCTTTG-3', which includes 4 copies of the complement of the G-hook 5'-GGTCTGTGTGATGTT-3', followed by 9 bases of ligator sequence, and which is complementary to the G-hook) is attached to the 5' end of the linearized single-stranded M13 backbone to allow for (1) purification of the reporter following ligation and/or annealing of segments; and (2) immobilization of the reporter once it is "stretched" on a solid surface. The sequence of the ligator for attaching G-tag to the 5' end of single-stranded M13 which has been linearized at the BamHI site can be 5'-CTCTAGAGGATCCAAAGGGCT-3'. The ligation reaction can be performed according to the following protocol to produce approximately 80 pmol of G-tag/M13 ligation product:

Materials:**[0452]**

[100uM] anti-G4 tag oligo
 [100uM] anti-G4 tag ligator oligo
 [80 nM] Linear single-stranded M13
 10X T4 DNA Ligase Buffer (Fermentas)
 T4 DNA Ligase (Fermentas)
 20X SSC (Ambion)
 DEPC H2O (Ambion)

Method:**[0453]**

1. Pre-anneal the G-tag and ligator:

25uM 2:1 G/Glig in 1X SSC
 20 μ l [100uM] G-tag Ligator
 40 μ l [100uM] G-tag
 4 μ l 20X SSC
 16 μ l DEPC H2O

* Anneal on the MJ Thermocycler
 95°C, 3min; 72°C, 30sec, -1°C/cycle, x 68 cycles; hold at 4°C

2. Ligate the G-tag to the linear M13:

64nM M13-G4 in 1X Lig Buffer
 1000 μ l [80 nM] Linear M13
 80 μ l [25uM] 2:1 G/Glig in 1X SSC

EP 1 963 531 B1

124 μ l 10X T4 DNA Ligase Buffer
40 μ l T4 DNA Ligase

* Ligate in an aluminum heat block covered with foil at 37°C for 2hr then at 65°C for 15 minutes to inactivate the enzyme.

13.4 PREPARATION OF RNA SEGMENTS

[0454] Ten sets of oligonucleotide primer pairs are designed to create 10 different regions along the M13 scaffold. Each pair contains one primer which has a T7 RNA polymerase promoter at the 5' end. Regions 2-7 are designed to be 900 bases (approximately 300 nm) long, as this is the approximate size of a diffraction-limited spot (the smallest spot that can be achieved with standard optics). Regions 1 and 8 have both long and short versions: the long versions cover the whole 900-base region, while the short versions cover only a portion of the 900-base region to allow a target-specific sequence to be ligated. Thus a target-specific sequence can be attached to either end. The ends can also be used for attachment of anchors or tags.

[0455] PCR is performed using Taq polymerase and 0.5 ng of double-stranded M13mp18 (USB Corporation) as a template. Reactions are cleaned up using a Qiaquick purification kit from Qiagen. Each PCR reaction yields a double-stranded fragment corresponding to one specific segment as illustrated below. These fragments are used as templates for the in vitro transcription of the RNA segments.

[0456] Using the PCR products described above as double-stranded templates, RNA segments are generated using an in vitro transcription kit from Ambion (Megascript T7 kit). The products of the transcription reactions are purified (including treatment with DNase I to remove template) using a RNeasy Kit from Qiagen.

13.5 LABELING OF THE RNA SEGMENTS

[0457] Using the PCR products described above as double-stranded templates, RNA segments for later dye-coupling are generated using an in vitro transcription kit from Ambion (MessageAmp ARENA kit). Aminoallyl-modified UTP nucleotides are incorporated into the RNA segments during transcription. The products of the transcription reactions are purified (including treatment with DNase I to remove template) using a RNeasy Kit from Qiagen.

[0458] 20-100 μ g of aminoallyl-modified RNA segment is coupled with NHS-ester dyes using Ambion Aminoallyl Labeling Kit. Dyes used include Alexa 488, Alexa 594 and Alexa 647 (Invitrogen/Molecular Probes) as well as Cy3 (Amersham).

[0459] Each segment is made separately in 4 colors so that each position on the scaffold can be filled with a segment in any of the four colors; thus different colors can be added at different positions to create many unique color combinations.

[0460] In this particular embodiment, adjacent segments are of different colors or there may be dark segments interspersed so that each segment is detected as an individual 'spot'. Dark segments may be used as part of the nanoreporter code.

13.6 ANNEALING OF THE RNA SEGMENTS TO THE SCAFFOLD

[0461] Segments for each position are annealed in a 2:1 ratio of segment to M13 scaffold in 1X SSPE buffer at 70°C for 2 hours. An assembled nanoreporter with labeled RNA segments is depicted in Figure 3A-3B. Figure 3A depicts a nanoreporter in which only alternate "spots" (1, 3, 5 and 7) are labeled, and Figure 3B depicts a nanoreporter in which every spot is labeled.

13.7 PREPARATION OF THE GHOST PROBE

[0462] One or more oligonucleotides comprising sequences (of, e.g., 30-70 nucleotides) complementary to different regions of the target nucleic acid(s) of interest than those to which the target-specific sequences of the reporter probe are complementary, are generated. Optionally, F-tags for F-hook attachment are ligated to the 5' end of the ghost probe using a ligator oligonucleotide that is complementary to a short sequence on the 3' end of the F-hook as well as a short sequence on the 5' end of the ghost probe. The sequences that are complementary to the ligator oligonucleotide are not part of the F-hook sequence or the probe sequence, but are additional nucleotides added to those oligos in order to facilitate ligation.

13.8 ATTACHMENT OF F-TAGS TO THE GHOST PROBE

[0463] An F-tag (e.g., an oligonucleotide having the sequence 5'-GATGGAGAC GTCTATCATCACAGC GTCTATCATCACAGC-biotin -3', which includes 2 copies of the complement of the F-hook 5'-GCTGTGATGATAGAC -3', followed

EP 1 963 531 B1

by 9 bases of ligator sequence and is complementary to the F-hook) is attached to the 3' end of the ghost probe to allow for (1) purification of the ghost-probe-target-reporter hybridization complex; and (2) attachment of the hybridization complex on the slide via the biotin moiety. The sequence of the ligator for attaching F-tag to the 3' end of the ghost probe can be 5'-GTCTCCATCTTCCGACAG-3'.

Materials:

[0464]

100uM F-biotin tag
100uM F ghost probe ligator
Fermentas 10X T4 DNA Ligase Buffer
1 uM ghost probes
Fermentas T4 DNA Ligase

Method:

[0465]

1. Pre-anneal the hook and ligator:

5uM F-biotin tag /ligator mix
5 μ l [100uM] F-biotin tag
5 μ l [100uM] F-ghost probe ligator
10 μ l 10X T4 DNA Ligase Buffer
80 μ l DEPC H₂O

Anneal on the MJ Thermocycler (95°C, 3min; 72°C, 30sec, -1°C/cycle x 68 cycles; hold at 4°C).

2. Set up the following ghost probe ligation:

300 nM anti-F2-biotin -GP
6.0 μ l [1uM] Ghost Probe
4.8 μ l [5uM] anti-F2-biotin tag /ligator mix
1.52 μ l 10X T4 DNA Ligase Buffer
3.68 μ l DEPC H₂O
4.0 μ l T4 DNA Ligase

Ligate on the MJ Thermocycler (37°C, 18hr; 65°C, 15 minutes; hold at 4°C)

3. QC the ligation on a 15% Novex TBE-Urea gel:

Prepare the following loading solutions:

Ligation	Neg Control-Ghost Probe
3.33 μ l [300 nM] ligation	1 μ l [1uM] ghost probe
1.67 μ l DEPC H ₂ O	0.33 μ l 10X T4 DNA Ligase Buffer
5 μ l 2X Loading Buffer	3.67 μ l DEPC H ₂ O
	5 μ l 2X Loading Buffer

Neg Control- F-biotin tag /ligator mix
2 μ l [0.5uM] F-biotin tag /ligator mix
0.33 μ l 10X T4 DNA Ligase Buffer
2.67 μ l DEPC H₂O
5 μ l 2X Loading Buffer

(continued)

50 bp Oligo Ladder
 4 μ l Ladder
 6 μ l 2X loading buffer

Run of a 15% Novex TBE-Urea gel at 180V for 50 minutes.
 Stain with SYBR Gold for 30 minutes.

13.8.1 ALTERNATIVE EMBODIMENTS

[0466] Rather than covalently coupling biotin to the single-stranded F-tag, the biotinylation of the ghost probe can also be accomplished by annealing a biotinylated oligonucleotide (DNA or RNA) with a sequence complementary to the common portion of the ghost probe. Such a sequence could be the F sequence itself, or another sequence which is added to the ghost probe in addition to the F sequence. If such an additional sequence is added, it could be from 10-100 bases long, from 1-10 copies, with the preferred configuration being a single copy from 50-100 bases long.

13.9 BIOTINYLATION OF TARGET MRNA

[0467] There are a number of commercially available kits available for the direct labeling of an mRNA sample including *Label IT*® μ ArrayTM Biotin Mirus #MIR 8010 and Biotin-Chem-Link (Roche (1 812 149)). Following manufacturer's procedures biotin labeled mRNA is added to the hybridization reaction as described in Section 3d (below) with the following modifications: Since most protocols suggest the use of poly A+ mRNA, the amount of RNA used could be reduced below the 100 ng total RNA in a typical hybridization to 10 ng and possibly 1ng. No ghost probe should be added to this reaction. F bead post-hybridization purification is no longer required. G-bead post-hybridization purification should be used to remove unhybridized biotinylated mRNA that might compete for binding to the slide. Depending on the amount of RNA used, this may or may not be required. Alternatively, total RNA could be biotinylated without the need for purification of the poly A+ fraction. In this case, the original amount of total RNA should be used (100 ng). The use of total RNA might require modifications of the manufacturer's protocol to increase labeling efficiency.

[0468] An alternative approach would be to enzymatically generate biotinylated 1st strand cDNA or biotinylated amplified RNA (aRNA) using commercially available kits and use these in place of total or mRNA. This approach would require a redesign of the reporter probes to be in the sense orientation. Both ghost probe and F-bead post-hybridization reactions would be omitted while G-bead purification would remain for removal of non-hybridized RNA.

13.10 HYBRIDIZATION OF DUAL NANOREPORTER TO TARGET

[0469] Many hybridization conditions are sufficient for achieving gene expression data. To shorten hybridization times while maintaining reasonable hybridization efficiency, several parameters can be altered: i) increasing ghost probe and reporter concentrations, ii) fragmenting of total RNA to average size range of 200-500 bp while lowering the pH of hybridization to 6.5, iii) using more total RNA in same hybridization volume, iv) lowering hybridization volume to approximately 10 μ l. Blocking reagents such as Denhardt's and ssDNA can be removed without deleterious effects on hybridization efficiency or cross hybridization to mRNAs from different species.

[0470] The following protocol has been performed successfully with multiplexing from 1 to >500 nanoreporters with ghost probes (an example demonstrating a nanoreporter assay utilizing 25 nanoreporters is described in Example 6 above, and another example demonstrating a nanoreporter assay utilizing 509 nanoreporters is described in Example 9 below). The final concentration of all nanoreporters varies depending on 1) the concentration of each reporter and 2) the number of genes being multiplexed.

[0471] Typical total nanoreporter concentrations range from 40pM (1 gene @ 40pM) to 20nM (500 genes @40pM). Ghost probe concentrations also vary from 200pM (1 gene @ 200pM) to 100nM (500 genes @ 200pM). The example that follows describes a single multiplexed hybridization containing approximately 500 endogenous genes with positive and negative controls. Add, 11.1 μ l of 2.7X hybridization mix [13.5X SSPE pH 7.5 (USB #75890), 0.27 μ g/ μ l sheared salmon sperm DNA (Sigma #D-7656), 0.27% tween 20 (Sigma #P-1379), and 13.5X Denhardt's reagent (Sigma D-2532)], 5 μ l of gene Nanoreporter mix (0.24nM each or 123nM total, includes 509 endogenous genes and 8 hybridization controls), 4.6 μ l 513 gene ghost probe mix (1.3nM each or 667 nM total, includes 509 endogenous genes and 8 hybridization controls), 1 μ l of purification control reporter mix (0.5pM), 1 μ l of total cellular RNA (100ng/ μ l), 1 μ l of 30X spike target mix (1.5nM-3fM) and 6.3 μ l of DEPC treated water (Ambion #9922) to a 0.2ml thin wall tube (final volume 30 μ l).

[0472] Final concentration of hybridization reagents should be 5X SSPE, 0.1% tween 20, 100ng/ μ l sheared salmon

sperm DNA, 5X Denhardt's reagent, 40 pM each Nanoreporter (~20nM total), 200pM each ghost probe (~100nM total) and 33ng/ μ l of total cellular RNA. Control spike targets typically vary in range from 50 M down to 0.1 fM in a single reaction. All reagents are most preferably free of all nuclease activity. For optimal results, all reagents should be free of nuclease activity.

[0473] Mix reagents well and incubate in temperature block with heated lid for 20 hours. After hybridization purify the nanoreporters with affinity reagents for both the ghost probe and the reporter probe.

13.10.1 ALTERNATIVE EMBODIMENT: HYBRIDIZATION PROTOCOL FOR WITHOUT SSDNA AND DENHARDT'S REAGENT

[0474] This protocol has been performed successfully with multiplexing from 1-500 nanoreporters and ghost probes. Removal of ssDNA and Denhardt's reagent from hybridizations performed with human reagents (Nanoreporters and ghost probes) had no effect on cross hybridization with mouse total RNA when compared to a hybridization containing ssDNA and Denhardt's. In addition, removal of ssDNA and Denhardt's does not result in an increased background signal (based on negative hybridization controls). Finally, there is no significant loss (or gain) of signal for endogenous genes hybridized in the presence or absence of ssDNA and Denhardt's (509 genes, R^2 value=0.998).

13.10.2 ALTERNATIVE EMBODIMENT: HYBRIDIZATION CONDITIONS FOR FRAGMENTED CELLULAR MRNA

[0475] Fragmentation of cellular RNA has been achieved by both thermal and cation catalyzed protocols. These protocols were designed to obtain fragment lengths between 100 and 700 bp (on average). Thermal fragmentation: Dilute total RNA sample to 200 ng/ μ l in RNase free water. Heat sample to 95°C in temperature block with heated lid. Stop fragmentation by placing sample on ice. Use immediately or store at -80°C until use. Fragmentation via cation catalyzed reaction modified from manufacturer's protocol. (Ambion). Bring volume of RNA sample up to 9 μ l with RNase free water. Final concentration of total RNA should be between 0.2 and 2 μ g/ml. Add 1 μ l of 10X fragmentation buffer (Ambion 10X fragmentation buffer). Incubate at 70°C for 5 minutes in temperature block. Longer times will result in smaller fragment size on average. Stop reaction by addition of 1 μ l 200 mM EDTA. Use immediately or store at -80°C until use.

[0476] Fragmented RNA samples are hybridized as described herein except for the following modifications: i) pH of SSPE is reduced to 6.5 and ii) the time of reaction is reduced to 6 hours (for hybridization reactions in which reporter probe and ghost probe concentrations are 200 pM).

13.11 PURIFICATION OF NANOREPORTER-TARGET COMPLEXES

[0477] Post-hybridization purification is preferred when the total reporter probe concentration is above 1 nM. Purification significantly decreases non-specific binding and increases specific binding efficiency to the slide at higher reporter and ghost probe concentrations. In the example provided above, a single F-bead purification is described (purifies hybridized complexes from the ghost-probe end). As described in Example 9 below, optimal results at high ghost probe concentrations (>5nM total) are obtained via a subsequent G-bead purification which purifies the hybridization complexes from the 5' end of the reporter effectively removing excess non-hybridized ghost probes. The preferred order of purification is F-bead, then G-bead but the order can be reversed and the protocols optimized accordingly. The exact sequences used in these affinity purifications can likely be changed and optimized in alternative embodiments of the technology. These affinity purification steps and reagents are currently nucleic acid based but could theoretically be any sort of binding pairs that exhibit specific binding to one another and can be released by chemical treatment or alteration of binding conditions such that the interaction is disrupted and released. For example, an antibody/antigen pair, a protein/metal interaction, or ligand/receptor interaction, *etc.*

[0478] One example of purification is provided below.

[0479] After hybridization is complete, the salt of a hybridization sample (30 μ l, starting at 5x SSPE = 825mM Na⁺) is adjusted to a final concentration of approximately 1x SSPE. The diluted sample is added to 30 μ l F-hook MyOne Dynabeads (F-MODB) and bound for 15 minutes at room temperature while rotating. The beads are sequestered with a magnet and the supernatant removed. The beads are washed twice with 150 μ l 0.1xSSPE + 0.1% Tween at room temperature for 15 minutes w/ rotation and discarded. The purified reporters are eluted in 30 μ l 0.1x SSPE at 45 °C for 15 minutes w/ rotation. At this point the hybridized reporters are purified from the contaminating un-hybridized reporters. The elution still contains contaminating un-hybridized ghost probes which will compete with the reporters for biotin-binding sites on the streptavidin coated slide. The 30 μ l is added to 130 μ l of 1xSSPE + 0.1% Tween to increase salt concentration. The sample (150 μ l) is then loaded onto 30 μ l of G-MODB and bound for 15 min at room temperature. The supernatant is discarded and the beads washed with 150 μ l 0.1xSSPE + 0.1% Tween at room temperature for 15 minutes w/ rotation. The wash is discarded and the fully purified reporters eluted with 25 μ l 0.1x SSPE at 45°C for 15 minutes w/ rotation. At this point only targets molecules that are hybridized to both a ghost probe (containing the anti-F

sequence) and a reporter (containing the anti-G sequence) will remain in solution.

13.12 IMMOBILIZATION AND STRETCHING AND IMAGING OF NANOREPORTER-TARGET COMPLEXES

[0480] Attachment to the slide and immobilization of the stretched complex may be achieved via a biotin-streptavidin interaction. In alternative embodiments, immobilization and stretching are achieved with other interaction pairs provided one of the two could be immobilized on the slide and the other attached to either the ghost probe or the reporter. Stretching does not have to be achieved via electrophoresis but can be done mechanically. The addition of bis-tris propane to the sample before binding is not required. The technology is not limited to the use of particular label monomers exemplified herein as long as the different label monomers can be separated by image processing.

[0481] One example of an immobilization and stretching protocol is provided below.

[0482] After purification, the hybridization products are loaded directly into an open well of a microfluidic device. The liquid is pulled into a microfluidic channel by capillary action where the hybridized molecules bind to the streptavidin-coated slide through the biotinylated ghost probe. The microfluidic device then intermittently tilts along the axis perpendicular to the length of the channels in alternating directions in order to force the reaction mixture to repeatedly pass through the channel and increase the binding efficiency.

[0483] After binding the hybridization reaction, the channel is washed with 1x TAE for 5 minutes by tilting the device at an angle. Fresh TAE is then added to each well to a level sufficient to contact platinum electrodes which are inserted in the wells (30 microliters in our current geometry). An electrical potential of 200V is then applied between the two wells connected by the microfluidic channel, stretching the reporters. After one minute of pre-electrophoresis to remove any remaining contaminating un-bound reporter molecules in the channel, a solution of 0.5 μ M G-hooks in 1x TAE is added to the cathodic well (60 microliters of this solution). The electrical potential draws the G-hooks through the channel toward the anodic well. As they pass through the channel, the hooks hybridize with the free G-tag sequences on the free-end of the reporters which are bound to the surface and stretched. The streptavidin on the surface then binds the biotin on the G-hook and immobilizes the free end. When the potential is removed, the reporters remain stretched for imaging.

14. EXAMPLE 9: HYBRIDIZATION OF 509 CELLULAR GENES TO 100NG TOTAL RNA FROM A549 CELLS USING NANOSTRING REPORTER SYSTEM.

14.1 HYBRIDIZATION REACTION

[0484] Detection of 509 endogenous cellular genes was carried out in single multiplexed hybridization reaction. Eight non-human control sequences were spiked into each reaction that corresponded to approximately 0.1, 0.5, 1, 5, 10, 50, and 100 copies per cell as well as two reporters with no target (negative controls). There were also 4 reporters added that served as positive (3) and negative (1) controls for the post-hybridization purification process. A set of negative control hybridization was also performed containing the entire Nanostring reporter library but lacking cellular RNA.

[0485] Each sample was hybridized in triplicate. Final concentrations of the hybridization reagents were as follows: 20.8 nM total Nanoreporters (521 individual Nanoreporters at 40pM each), 103 nM total ghost probe (517 individual ghost probes @ 200pM each), 5X SSPE (pH 7.5), 5X Denhardt's reagent, 100ng/ul sheared salmon sperm DNA, 0.1 % tween 20, 50fM S11 spike target DNA, 10fM S10 spike target DNA, 5fM S9 spike target DNA, 1fM S8 spike target DNA, 0.5 fM S7 spike target DNA, 0.1fM S6 spike target DNA. S3 and S4 were added as negative controls. RNA was obtained from A549 lung epithelial cells under two different conditions. The final concentration of total RNA per hybridization was 33ng/ul. No total RNA was added to the negative control hybridizations. The final volume of the reaction was 30ul. Reagents were mixed and incubated at 65°C in thermocycler block with heated lid for 20 hours.

Master mix	(1 Reaction)	(9.3 reactions)
2.7 X hybridization mix*	11.1 μ l	103.2 μ l
513 endogenous gene reporters (0.24nM each)	5 μ l	46.5 μ l
513 endogenous gene ghostprobes (1.3nM each)	4.6 μ l	42.9 μ l
Purification Control reporters (0.6nM each)	1 μ l	9.3 μ l
30X control target mix	1 μ l	9.3 μ l
Total	22.7 μ l	211.2 μ l
*Hybridization mix (13.5X SSPE, 13.5X Denhardt's reagent, 270 ng salmon sperm DNA, 0.27% tween 20)		

EP 1 963 531 B1

Reactions	1	2	3	4	5	6	7	8	9
Master mix	22.7	22.7	22.7	22.7	22.7	22.7	22.7	22.7	22.7
48.5ng/ μ l RNA #1	2.1	2.1	2.1	0	0	0	0	0	0
48.4ng/ μ l RNA #2	0	0	0	2.1	2.1	2.1	0	0	0
H ₂ O	5.2	5.2	5.2	5.2	5.2	5.2	7.3	7.3	7.3
Total Rxn volume	30 μ l	30 μ l	30 μ l	30 μ l	30 μ l	30 μ l	30 μ l	30 μ l	30 μ l
Incubate reactions in thermocycler with heated lid overnight (20hours).									

14.2 POST-HYBRIDIZATION PURIFICATION

[0486] Hybridization reactions were purified to remove unhybridized reporters using an oligonucleotide complimentary to ghost probe attached to magnetic beads (F-bead). Hybridization reactions were diluted 5 fold in 0.1 % Tween-20/TE to bring the final salt concentration to 1X SSPE. The diluted hybridization solution was then added to 100ul of F-beads (in 0.1% Tween-20) and allowed to bind to the beads at room temperature for 30min with continuous rotation. The beads were then washed three times in 150ul of 0.1X SSPE/0.1% Tween-20 and eluted in 100ul of 0.1X SSPE/0.1% Tween-20 for 15min at 45°C.

[0487] After F-bead elution, samples were purified from the opposite end of the hybridized complex using G-beads. Elutions were brought to a final concentration of 1X SSPE by the addition of 50ul of 3X SSPE/0.1% Tween-20 and bound to 30ul of G-beads (in 0.1% Tween-20) for 15min at room temperature with rotation. Beads were then washed as above and eluted in 30ul of 0.1X SSPE/Tween-20 and prepared for binding as described below.

14.3 BINDING, STRETCHING, AND IMMOBILIZATION

[0488] The samples were prepared for binding by addition of 1 ul of 1/5000 dilution of 0.1 uM Tetraspec™ fluorescent microspheres (product # T7279, Molecular Probes). Samples were loaded into a Nanostring fluidic device and attached to Accer18 Optichem® slide coated with streptavidin (product #TB0200) by tilting the device 45deg for 15mm and repeated a total of 4 times. After loading, slide surface was washed once with 90ul of 1X TAE. After wash buffer is removed the sample is prepared for electrostretching by addition of 40ul of TAE to each well. Attached complexes were stretched by applying 200V across the fluidic channel. After 1 minute the samples were immobilized in the stretched position by adding 60ul of 500nM of G-hook oligo solution to the well containing the negatively charged electrode while continuing to apply voltage for 5 minutes. After immobilization the TAE solution is removed and replaced with anti-photobleaching reagent for imaging.

14.4 IMAGING

[0489] Slides were imaged on Nikon Eclipse TE2000E equipped with a metal halide light source (X-cite 120, Exfo Corporation) and a 60X oil immersion lens (1.4 NA Plan Apo VC, Nikon). For each field of view, 4 images at different excitation wavelengths (480, 545, 580 and 622) were acquired with an Orca Ag CCD camera (Hamamatsu) under control of either Metamorph (Universal Imaging Corporation) or custom software. Images were processed with custom image processing software.

14.5 DATA ANALYSIS

[0490] Raw data was extracted from processed images using custom software. Data was normalized to the average counts for control spikes in each sample. To determine if a gene was "detected" by the system, the counts obtained for each gene from hybridizations containing RNA were compared to average counts of the two negative controls using a Student's test. The number of genes detected was 441 (87%) and 445 (88%) in sample #1 and #2, respectively.

[0491] A scatter plot (Figure shows normalized and average log₂ signal values from each positive sample (n=3) for all 509 genes. The genes that were significantly different in the two samples were identified by a T-test of signal values in sample #2 against sample #1. In the graph below, the solid lines indicate the 2 fold upregulated threshold (black line) and 2-fold downregulated threshold (gray line) relative to sample #1. Genes with significant fold changes (p-value < 0.05) are shown in solid black diamonds. Genes whose fold change p-values were above this threshold are shown in open black squares.

15. EXAMPLE 10: DETECTION OF SMALL SPOTS

[0492] As mentioned above, the label attachment regions of a nanoreporter scaffold region have a length anywhere from 10 nm to 10,000 nm, but preferably corresponds closely to the smallest spot that can be detected with standard optics, which is about 300 nm. Spots of different color (spectrally distinguishable) are spatially resolvable at closer spacing than spots of the same color. It is possible to fit one, two, three or four spots of different colors between two spots of the same color, and yet spectrally and spatially resolve all the spots. It is also possible to significantly reduce the distance between two spots of the same color.

[0493] The limits of spatial resolution, i.e., differentiating closely spaced spots of the same color, are often thought of as hard limits, i.e., the Rayleigh Criteria (Inoue, S., Spring, Video Microscopy (Plenum Press, 1997), p 30.) There are many techniques to drive beyond these limits that involve different imaging and/or image processing techniques. On the imaging side, structured illumination is one method to resolve spots of the same color that are spaced closer together. 50 nm has been demonstrated but, in theory, resolution with structured illumination is unlimited (Gustafsson, 2005, Proc. Nat'l. Acad. Sci. U.S.A. 102:13081-13086). On the image processing side, mixture modeling is an effective technique to push beyond commonly accepted limits (Thomann et al, 2002, J. Microsc. 211:230-248). The combination of these techniques allows for drastically smaller nanoreporters with smaller spots, corresponding to label attachment regions of less than 50 nm.

[0494] These smaller spot spacing could allow for drastically shorter and more stable reporters, a larger number of codes, as well as a higher degree of multiplexing before the entanglement threshold is passed (for an explanation of entanglement thresholds, see Example 9 (described in Section 14) above. The tradeoff of making the spots much smaller and the reporters much shorter would be decreased signal and slower scan times. However, other technical advances, such as brighter light sources, and more efficient CCDs may offset the increased scan times making these approaches reasonable.

Claims

1. A probe pair comprising a first probe and a second probe,

(a) said first probe being a complex comprising:

- (i) a first molecule comprising a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal and a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal;
- (ii) a first target-specific sequence attached to the first molecule, and
- (iii) an affinity tag attached to said first molecule,

(b) said second probe being a second molecule, comprising

- (i) a second target-specific sequence;
- (ii) optionally, a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal; and
- (iii) an affinity tag attached to said second molecule;

wherein the first target-specific sequence and the second target-specific sequence bind to different regions of the same target molecule,

wherein the target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA,

wherein when said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitute at least part of a code that identifies the target molecule,

wherein the first, second, and third label attachment regions are predetermined nucleotide sequences,

wherein the first, second, and third signals are spatially distinguishable, and

wherein a first DNA or RNA molecule is hybridized to the first label attachment region, to which said first DNA or RNA molecule is bound said one or more label monomers that emit light constituting said first signal; and wherein a second DNA or RNA molecule is hybridized to the second label attachment region, to which said second DNA or RNA molecule is bound said one or more label monomers that emit light constituting said second signal.

2. A probe pair comprising a first probe and a second probe,

(a) said first probe being a complex comprising:

- 5 (i) a first molecule comprising a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal and a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal;
- 10 (ii) a first target-specific sequence attached to the first molecule, and
- (iii) an affinity tag attached to said first molecule,

(b) said second probe being a second molecule, comprising

- 15 (i) a second target-specific sequence;
- (ii) optionally, a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal; and
- (iii) an affinity tag attached to said second molecule;

20 wherein the first target-specific sequence and the second target-specific sequence bind to different regions of the same target molecule,
 wherein the first signal, the second signal and the third signal are spatially distinguishable; wherein the target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA,
 25 wherein when said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitute at least part of a code that identifies the target molecule,
 wherein the first, second, and third label attachment regions are predetermined nucleotide sequences, and
 wherein a first DNA or RNA molecule is hybridized to the first label attachment region, to which said first DNA or RNA molecule is bound said one or more label monomers that emit light constituting said first signal; and wherein
 30 a second DNA or RNA molecule is hybridized to the second label attachment region, to which said second DNA or RNA molecule is bound said one or more label monomers that emit light constituting said second signal.

3. The probe pair of claims 1 or 2, wherein a plurality of first DNA or RNA molecules are hybridized to the first label attachment region, to which DNA or RNA molecules are bound said one or more label monomers that emit light constituting said first signal; and wherein a plurality of second DNA or RNA molecules are hybridized to the second
 35 label attachment region, to which second DNA or RNA molecules are bound said one or more label monomers that emit light constituting said second signal.

4. The probe pair of any one of claims 1-3, wherein the second probe comprises a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal.
 40

5. The probe pair of claim 4, wherein said third label attachment region is a predetermined nucleotide sequence, and wherein a third RNA molecule is hybridized to the third label attachment region, to which third RNA molecule are attached said one or more label monomers that emit light constituting said third signal.

45 6. The probe pair of claim 5, wherein when the probe pair is bound to its target molecule, the code comprises the identity of the first signal, second signal and third signal and their locations relative to each other.

7. The probe pair of claims 1 or 2, wherein the code comprises the identity of the first and second signals, their locations relative to each other, and the size of the spot resulting from at least one of said signals.
 50

8. The probe pair of any one of claims 1-3, wherein the first and second molecules are nucleic acid molecules.

9. The probe pair of claim 8, wherein the label attachment regions and target-specific sequences are predetermined nucleotide sequences.
 55

10. The probe pair of claim 9 in which said one or more label monomers are covalently bound to nucleic acids hybridized to their respective attachment regions.

11. The probe pair of claim 10, wherein said nucleic acids are hybridized to the respective attachment regions by way of one or more non-covalently bound bridging nucleic acids.
12. The probe pair of claims 1 or 2 in which the first and second target-specific sequences are unlabeled with any one or more of said label monomers.
13. The probe pair off any one of claims 1-3, wherein the second molecule further comprises a fourth label attachment region, which is non-overlapping with the third label attachment region, to which is attached one or more label monomers that emit light constituting a fourth signal.
14. The probe pair of claim 13, wherein when the probe pair is bound to its target molecule, the code comprises the identity of the first signal, second signal, third signal and fourth signal and their locations relative to each other.
15. The probe pair of any one of claims 1-3, wherein a first RNA molecule is hybridized to the first label attachment region, to which first RNA molecule are bound said one or more label monomers that emit light constituting said first signal; and wherein a second RNA molecule is hybridized to the second label attachment region, to which second RNA molecule are bound said one or more label monomers that emit light constituting said second signal.
16. The probe pair of any one of claims 1-3, wherein the affinity tags are covalently attached to the first molecule or to the second molecule.
17. The probe pair of claims 1 or 2, wherein the label monomers attached to the first label attachment region emit light at the same wavelength, which light constitutes said first signal, wherein the label monomers attached to the second label attachment region emit light at the same wavelength, which light constitutes the second signal and wherein the label monomers attached to the third label attachment region, if present, emit light at the same wavelength, which light constitutes the third signal.
18. The probe pair of claims 1 or 2, wherein at least one of the first signal and the second signal comprises light at a plurality of different wavelengths.
19. The probe pair of claims 1 or 2, wherein the first, second and third signals are spectrally distinguishable.
20. The probe pair of any one of claims 1-3, wherein the first and second RNA molecules are each 100 to 3,000 nucleotides in length or wherein the first and second RNA molecules are each 500 to 1,500 nucleotides in length.
21. The probe pair of any one of claims 1-3, wherein the second probe is a nucleic acid complex comprising:
 - (a) the second molecule, wherein the second molecule comprises a third label attachment region to which is hybridized a third RNA molecule, to which third RNA molecule are bound one or more label monomers that emit light constituting a third signal; and
 - (b) the second target-specific sequence attached to the second molecule,wherein the code comprises the identity of the first, second and third signals and their locations relative to each other.
22. The probe pair of claim 21, wherein the label monomers attached to the first label attachment regions emit light at the same wavelength, which light constitutes said first signal; wherein the label monomers attached to the second label attachment region emit light at the same wavelength, which light constitutes the second signal; and wherein the label monomers attached to the third label attachment region emit light at the same wavelength, which light constitutes the third signal.
23. The probe pair of claim 21, wherein at least one of the first signal, the second signal and the third signal comprises light at a plurality of different wavelengths.
24. The probe pair of claim 21, in which the first target-specific sequence and the second target-specific sequence are unlabeled with any one or more of said label monomers.
25. The probe pair of claim 21, wherein first, second and third signals are spectrally distinguishable.

26. The probe pair of claim 21, wherein first and third signals emit at the same wavelength or wavelengths.
27. The probe pair of claim 21, wherein the affinity tag is covalently attached to the first molecule.
- 5 28. The probe pair of claim 21, wherein the affinity tag is covalently attached to the second molecule.
29. The probe pair of claim 21, wherein the first, second and third RNA molecules are each 100 to 3,000 nucleotides or wherein the first, second and third RNA molecules are each 500 to 1,500 nucleotides,
- 10 30. The probe pair of claims 21-30, wherein the RNA molecules comprise a single region comprising a regularly-repeated base.
31. The probe pair of claim 30, wherein the regularly repeated base is aminoallyl-modified uracil.
- 15 32. A method of detecting a target molecule in a biomolecular sample comprising:
- (i) contacting said sample with a probe pair according to any one of claims 1-31 under conditions that allow binding of the first target-specific sequence and the second target-specific sequence to the target molecule and
- (ii) detecting the code that identifies the target molecule.
- 20 33. The method of claim 32, further comprising quantitating the amount of said target molecule in said biomolecular sample.
34. The method of claims 32 or 33, wherein the concentration of each of said first and second target-specific sequences is in excess of the concentration of said target molecule.
- 25 35. The method of claims 32-34, further comprising the step of stretching the bound probe pair and target molecule prior to detecting the code that identifies the target molecule.
- 30 36. The method of claim 35, wherein the stretching is performed using a technique selected from the group consisting of flow-stretching, a receding meniscus technique, electrostretching and constriction in the flow of a liquid containing the nanoreporter in conjunction with an oscillating electric field.
- 35 37. A method of detecting a plurality of target molecules in a biomolecular sample comprising:
- (i) contacting said samples with a population of probe pairs according to any one of claims 1-31, under conditions that allow binding of the first target-specific sequence and the second target-specific sequence of each probe pair to their respective target molecule, wherein each probe pair in said population when bound to its respective target molecule is associated with a distinguishable code, and
- (ii) detecting the codes that identify the plurality of target molecules.
- 40 38. The method of claim 37, further comprising quantitating the amount of each of said plurality of target molecules in said biomolecular sample.
- 45 39. The method of claims 37-38, wherein the concentration of each of said first and second target-specific sequences of each probe pair of said population is in excess of the concentration of its respective target molecule.
40. The method of claims 37-39, further comprising the step of stretching the bound probe pairs and their respective target molecules prior to detecting the codes that identifies the plurality of target molecules.
- 50 41. The method of claim 40, wherein the stretching is performed using a technique selected from the group consisting of flow-stretching, a receding meniscus technique, electrostretching and constriction in the flow of a liquid containing the nanoreporter in conjunction with an oscillating electric field.

55

Patentansprüche

1. Sondenpaars, umfassend eine erste Sonde und eine zweite Sonde,

(a) worin die erste Sonde einen Komplex darstellt, umfassend:

- (i) ein erstes Molekül, umfassend eine erste Region zur Markierungsbefestigung, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines ersten Signals, und eine zweite Region zur Markierungsbefestigung, welche nicht mit der ersten Region zur Markierungsbefestigung überlappt, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines zweiten Signals;
- (ii) eine erste, zielspezifische Sequenz, welche an dem ersten Molekül befestigt ist, und
- (iii) einen Affinitäts-Tag, welcher an dem ersten Molekül befestigt ist,

(b) worin die zweite Sonde ein zweites Molekül ist, umfassend

- (i) eine zweite, zielspezifische Sequenz;
- (ii) optional eine dritte Region zur Markierungsbefestigung, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines dritten Signals; und
- (iii) einen Affinitäts-Tag, welcher an dem zweiten Molekül befestigt ist;

worin die erste zielspezifische Sequenz und die zweite zielspezifische Sequenz an unterschiedliche Regionen des gleichen Zielmoleküls binden,

worin das Zielmolekül ein natürlich vorkommendes Molekül oder eine cDNA eines natürlich vorkommenden Moleküls oder das Komplement der cDNA ist,

worin, wenn das Sondenpaar an sein Zielmolekül gebunden ist, die Identität des ersten und zweiten Signals und ihre Positionen relativ zueinander zumindest einen Teil eines Codes bilden, welcher das Zielmolekül identifiziert,

worin die erste, zweite und dritte Region zur Markierungsbefestigung vorher festgelegte Nukleotidsequenzen sind,

worin das erste, zweite und dritte Signal räumlich unterscheidbar sind, und

worin ein erstes DNA- oder RNA-Molekül mit der ersten Region zur Markierungsbefestigung hybridisiert ist, worin an das erste DNA- oder RNA-Molekül ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des ersten Signals;

und worin ein zweites DNA- oder RNA-Molekül mit der zweiten Region zur Markierungsbefestigung hybridisiert ist,

worin an das zweite DNA- oder RNA-Molekül ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des zweiten Signals.

2. Sondenpaar, umfassend eine erste Sonde und eine zweite Sonde,

(a) worin die erste Sonde einen Komplex darstellt, umfassend:

- (i) ein erstes Molekül, umfassend eine erste Region zur Markierungsbefestigung, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines ersten Signals, und eine zweite Region zur Markierungsbefestigung, welche nicht mit der ersten Region zur Markierungsbefestigung überlappt, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines zweiten Signals;
- (ii) eine erste, zielspezifische Sequenz, welche an dem ersten Molekül befestigt ist, und
- (iii) einen Affinitäts-Tag, welcher an dem ersten Molekül befestigt ist,

(b) worin die zweite Sonde ein zweites Molekül ist, umfassend

- (i) eine zweite, zielspezifische Sequenz;
- (ii) optional eine dritte Region zur Markierungsbefestigung, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines dritten Signals; und
- (iii) einen Affinitäts-Tag, welcher an dem zweiten Molekül befestigt ist;

worin die erste zielspezifische Sequenz und die zweite zielspezifische Sequenz an unterschiedliche Regionen des gleichen Zielmoleküls binden,

worin das erste Signal, das zweite Signal und das dritte Signal räumlich unterscheidbar sind;

worin das Zielmolekül ein natürlich vorkommendes Molekül oder eine DNA eines natürlich vorkommenden Moleküls oder das Komplement der cDNA ist,

worin, wenn das Sondenpaar an sein Zielmolekül gebunden ist, die Identität des ersten und zweiten Signals und ihre Positionen relativ zueinander zumindest einen Teil eines Codes bilden, welcher das Zielmolekül identifiziert,

worin die erste, zweite und dritte Region zur Markierungsbefestigung vorher festgelegte Nukleotidsequenzen sind, und

worin ein erstes DNA- oder RNA-Molekül mit der ersten Region zur Markierungsbefestigung hybridisiert ist, worin an das erste DNA- oder RNA-Molekül ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des ersten Signals;

und worin ein zweites DNA- oder RNA-Molekül mit der zweiten Region zur Markierungsbefestigung hybridisiert ist, worin an das zweite DNA- oder RNA-Molekül ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des zweiten Signals.

3. Sondenpaar nach Anspruch 1 oder 2, worin eine Vielzahl von ersten DNA- oder RNA-Molekülen mit der ersten Region zur Markierungsbefestigung hybridisiert sind, worin an die DNA- oder RNA-Moleküle ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des ersten Signals; und worin eine Vielzahl von zweiten DNA- oder RNA-Molekülen mit der zweiten Region zur Markierungsbefestigung hybridisiert sind, worin an die zweiten DNA- oder RNA-Moleküle ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des zweiten Signals.
4. Sondenpaar nach einem der Ansprüche 1-3, worin die zweite Sonde eine dritte Region zur Markierungsbefestigung umfasst, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines dritten Signals.
5. Sondenpaar nach Anspruch 4, worin die dritte Region zur Markierungsbefestigung eine vorher festgelegte Nukleotidsequenz ist, und worin ein drittes RNA-Molekül mit der dritten Region zur Markierungsbefestigung hybridisiert ist, worin an dem dritten RNA-Molekül ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung des dritten Signals.
6. Sondenpaar nach Anspruch 5, worin, wenn das Sondenpaar an sein Zielmolekül gebunden ist, der Code die Identität des ersten Signals, des zweiten Signals und des dritten Signals und ihre Positionen relativ zueinander umfasst.
7. Sondenpaar nach Anspruch 1 oder 2, worin der Code die Identität des ersten und zweiten Signals, ihre Positionen relativ zueinander und die Größe des aus mindestens einem der Signale resultierenden Spots umfasst.
8. Sondenpaar nach einem der Ansprüche 1-3, worin das erste und zweite Molekül Nukleinsäuremoleküle sind.
9. Sondenpaars nach Anspruch 8, worin die Regionen zur Markierungsbefestigung und die zielspezifischen Sequenzen vorher festgelegte Nukleotidsequenzen sind.
10. Sondenpaar nach Anspruch 9, worin ein oder mehrere Markierungsmonomere kovalent an Nukleinsäuren gebunden sind, welche mit deren betreffenden Regionen zur Befestigung hybridisiert sind.
11. Sondenpaar nach Anspruch 10, worin die Nukleinsäuren mit den betreffenden Regionen zur Befestigung mittels einer oder mehreren nicht kovalent gebundenen brückenbildenden Nukleinsäuren hybridisiert sind.
12. Sondenpaar nach Anspruch 1 oder 2, worin die erste und zweite zielspezifische Sequenz nicht markiert sind mit einem oder mehreren Markierungsmonomeren.
13. Sondenpaar nach einem der Ansprüche 1-3, worin das zweite Molekül weiter eine vierte Region zur Markierungsbefestigung umfasst, welche nicht mit der dritten Region zur Markierungsbefestigung überlappt, an welcher ein oder mehrere Markierungsmonomere befestigt sind, welche Licht emittieren unter Bildung eines vierten Signals.
14. Sondenpaar nach Anspruch 13, worin, wenn das Sondenpaar an sein Zielmolekül gebunden ist, der Code die Identität des ersten Signals, des zweiten Signals, des dritten Signals und des vierten Signals und ihre Positionen relativ zueinander umfasst.
15. Sondenpaar nach einem der Ansprüche 1-3, worin ein erstes RNA-Molekül mit der ersten Region zur Markierungsbefestigung hybridisiert ist, worin an das erste RNA-Molekül ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des ersten Signals, und worin ein zweites RNA-Molekül mit der zweiten Region zur Markierungsbefestigung hybridisiert ist, worin an das zweite RNA-Molekül ein oder mehrere Markierungsmonomere gebunden sind, welche Licht emittieren unter Bildung des zweiten Signals.

16. Sondenpaar nach einem der Ansprüche 1-3, worin die Affinitäts-Tags kovalent an dem ersten Molekül oder an dem zweiten Molekül befestigt sind.
- 5 17. Sondenpaar nach Anspruch 1 oder 2, worin die Markierungsmonomere, welche an der ersten Region zur Markierungsbe-
festigung befestigt sind, Licht mit der gleichen Wellenlänge emittieren, worin das Licht das erste Signal
bildet, worin die Markierungsmonomere, welche an der zweiten Region zur Markierungsbe-
festigung befestigt sind, Licht mit der gleichen Wellenlänge emittieren, worin das Licht das zweite Signal bildet, und worin die Markierungs-
monomere, welche an der dritten Region zur Markierungsbe-
festigung befestigt sind, falls vorhanden, Licht mit der
10 gleichen Wellenlänge emittieren, worin das Licht das dritte Signal bildet.
18. Sondenpaar nach Anspruch 1 oder 2, worin zumindest eines des ersten und des zweiten Signals Licht mit mehreren
unterschiedlichen Wellenlängen umfasst.
19. Sondenpaar nach Anspruch 1 oder 2, worin das erste, zweite und dritte Signal spektral unterscheidbar sind.
- 15 20. Sondenpaar nach einem der Ansprüche 1-3, worin das erste und zweite RNA-Molekül jeweils 100 bis 3000 Nukleotide
lang sind, oder worin das erste und zweite RNA-Molekül jeweils 500 bis 1500 Nukleotide lang sind.
21. Sondenpaar nach einem der Ansprüche 1-3, worin die zweite Sonde ein Nukleinsäurekomplex ist, umfassend:
- 20 (a) das zweite Molekül, worin das zweite Molekül eine dritte Region zur Markierungsbe-
festigung umfasst, mit
welcher ein drittes RNA-Molekül hybridisiert ist, worin an das dritte RNA-Molekül ein oder mehrere Markie-
rungsmonomere gebunden sind, welche Licht emittieren unter Bildung eines dritten Signals, und
25 (b) die zweite zielspezifische Sequenz, befestigt an dem zweiten Molekül, worin der Code die Identität des
ersten, zweiten und dritten Signals und ihre Positionen relativ zueinander umfasst.
22. Sondenpaar nach Anspruch 21, worin die Markierungsmonomere, welche an den ersten Regionen zur Markierungs-
be-
festigung befestigt sind, Licht mit der gleichen Wellenlänge emittieren, worin das Licht das erste Signal bildet,
worin die Markierungsmonomere, welche an der zweiten Region zur Markierungsbe-
festigung befestigt sind, Licht mit der gleichen Wellenlänge emittieren, worin das Licht das zweite Signal bildet, und worin die Markierungsmono-
mere, welche an der dritten Region zur Markierungsbe-
festigung befestigt sind, Licht mit der gleichen Wellenlänge
30 emittieren, worin das Licht das dritte Signal bildet.
23. Sondenpaar nach Anspruch 21, worin zumindest eines des ersten Signals, des zweiten Signals und des dritten
35 Signals Licht mit mehreren unterschiedlichen Wellenlängen umfasst.
24. Sondenpaar nach Anspruch 21, worin die erste zielspezifische Sequenz und die zweite zielspezifische Sequenz
nicht markiert sind mit einem oder mehreren Markierungsmonomeren.
- 40 25. Sondenpaar nach Anspruch 21, worin das erste, zweite und dritte Signal spektral unterscheidbar sind.
26. Sondenpaar nach Anspruch 21, worin das erste und dritte Signal mit der gleichen Wellenlänge oder den gleichen
Wellenlängen emittieren.
- 45 27. Sondenpaar nach Anspruch 21, worin der Affinitäts-Tag kovalent an dem ersten Molekül befestigt ist.
28. Sondenpaar nach Anspruch 21, worin der Affinitäts-Tag kovalent an dem zweiten Molekül befestigt ist.
29. Sondenpaar nach Anspruch 21, worin das erste, zweite und dritte RNA-Molekül jeweils 100 bis 3000 Nukleotide
50 aufweisen oder worin das erste, zweite und dritte RNA-Molekül jeweils 500 bis 1500 Nukleotide aufweisen.
30. Sondenpaar nach Anspruch 21-30, worin die RNA-Moleküle eine einzelne Region umfassen, welche eine sich
regelmäßig wiederholende Base umfasst.
- 55 31. Sondenpaar nach Anspruch 30, worin die sich regelmäßig wiederholende Base Aminoallyl-modifiziertes Uracil ist.
32. Verfahren zum Nachweisen eines Zielmoleküls in einer biomolekularen Probe, umfassend:

(i) Inkontaktbringen der Probe mit einem Sondenpaar nach einem der Ansprüche 1 - 31 unter Bedingungen, welche eine Bindung der ersten zielspezifischen Sequenz und der zweiten zielspezifischen Sequenz an das Zielmolekül ermöglicht, und

(ii) Nachweisen des Codes, welcher das Zielmolekül identifiziert.

5

33. Verfahren nach Anspruch 32, weiter umfassend Quantifizieren der Menge des Zielmoleküls in der biomolekularen Probe.

10

34. Verfahren nach Anspruch 32 oder 33, worin die Konzentration sowohl der ersten als auch der zweiten zielspezifischen Sequenz größer ist als die Konzentration des Zielmoleküls.

35. Verfahren nach Anspruch 32-34, weiter umfassend den Schritt Strecken des gebundenen Sondenpaars und des Zielmoleküls vor dem Nachweisen des Codes, welcher das Zielmolekül identifiziert,

15

36. Verfahren nach Anspruch 35, worin das Strecken durchgeführt wird unter Verwendung eines Verfahrens, ausgewählt aus der Gruppe bestehend aus Strömungsstrecken ("Flow-Stretching"), einem Verfahren mittels zurückgehendem Meniskus, Elektrostrecken und Einengung in der Strömung einer Flüssigkeit, enthaltend den Nanoreporter, in Verbindung mit einem oszillierenden elektrischen Feld.

20

37. Verfahren zum Nachweisen mehrerer Zielmoleküle in einer biomolekularen Probe, umfassend:

(i) Inkontaktbringen der Probe mit einer Population von Sondenpaaren nach einem der Ansprüche 1-31, unter Bedingungen, welche eine Bindung der ersten zielspezifischen Sequenz und der zweiten zielspezifischen Sequenz jedes Sondenpaars an ihr entsprechendes Zielmolekül ermöglicht, worin jedes Sondenpaar in der Population bei Bindung mit seinem jeweiligen Zielmolekül mit einem unterscheidbaren Code verbunden wird, und

25

(ii) Nachweisen der Codes, welche die mehreren Zielmoleküle identifizieren.

38. Verfahren nach Anspruch 37, weiter umfassend Quantifizieren der Menge jedes der mehreren Zielmoleküle in der biomolekularen Probe.

30

39. Verfahren nach Anspruch 37-38, worin die Konzentration sowohl der ersten als auch der zweiten zielspezifischen Sequenzen jedes Sondenpaars der Population größer ist als die Konzentration ihres entsprechenden Zielmoleküls.

35

40. Verfahren nach Anspruch 37-39, weiter umfassend den Schritt Strecken der gebundenen Sondenpaare und ihrer entsprechenden Zielmoleküle vor dem Nachweisen der Codes, welche die mehreren Zielmoleküle identifizieren.

40

41. Verfahren nach Anspruch 40, worin das Strecken durchgeführt wird unter Verwendung eines Verfahrens, ausgewählt aus der Gruppe bestehend aus Strömungsstrecken ("Flow-Stretching"), einem Verfahren mittels zurückgehendem Meniskus, Elektrostrecken und Einengung in der Strömung einer Flüssigkeit, enthaltend den Nanoreporter, in Verbindung mit einem oszillierenden elektrischen Feld,

Revendications

45

1. Paire de sondes comprenant une première sonde et une deuxième sonde,

(a) ladite première sonde étant un complexe comprenant :

50

(i) une première molécule comprenant une première région d'attachement de marqueurs à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un premier signal et une deuxième région d'attachement de marqueurs, qui n'est pas en chevauchement avec la première région d'attachement de marqueurs, à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un deuxième signal ;

55

(ii) une première séquence à spécificité de cible attachée à la première molécule, et
(iii) une étiquette d'affinité attachée à ladite première molécule,

(b) ladite deuxième sonde étant une deuxième molécules comprenant :

- (i) une deuxième séquence à spécificité de cible;
- (ii) éventuellement une troisième région d'attachement de marqueurs à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un troisième signal ; et
- (iii) une étiquette d'affinité attachée à ladite deuxième molécule ;

5

dans laquelle la première séquence à spécificité de cible et la deuxième séquence à spécificité de cible se lient à des régions différentes de la même molécule cible, dans laquelle la molécule cible est une molécule naturelle ou un ADNc d'une molécule naturelle ou le complément dudit ADNc,

10

dans laquelle, quand ladite paire de sondes est liée à sa molécule cible, l'identité des premier et deuxième signaux et leurs positions l'un par rapport à l'autre constituent au moins une partie d'un code qui identifie la molécule cible, dans laquelle les première, deuxième et troisième régions d'attachement de marqueurs sont des séquences de nucléotides prédéterminées,

15

dans laquelle les premier, deuxième et troisième signaux sont spatialement distinguables, et dans laquelle une première molécule d'ADN ou d'ARN est hybridée à la première région d'attachement de marqueurs, lesdits un ou plusieurs marqueurs monomères qui émettent une lumière constituant ledit premier signal étant liés à ladite première molécule d'ADN ou d'ARN ; et dans laquelle une deuxième molécule d'ADN ou d'ARN est hybridée à la deuxième région d'attachement de marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit deuxième signal étant liés à ladite deuxième molécule d'ADN ou d'ARN.

20

2. Paire de sondes comprenant une première sonde et une deuxième sonde,

(a) ladite première sonde étant un complexe comprenant :

25

(i) une première molécule comprenant une première région d'attachement de marqueurs à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un premier signal et une deuxième région d'attachement de marqueurs, qui n'est pas en chevauchement avec la première région d'attachement de marqueurs, à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un deuxième signal ;

30

(ii) une première séquence à spécificité de cible attachée à la première molécule, et
(iii) une étiquette d'affinité attachée à ladite première molécule,

(b) ladite deuxième sonde étant une deuxième molécule comprenant :

35

(i) une deuxième séquence à spécificité de cible ;
(ii) éventuellement une troisième région d'attachement de marqueurs à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un troisième signal ; et
(iii) une étiquette d'affinité attachée à ladite deuxième molécule ;

40

dans laquelle la première séquence à spécificité de cible et la deuxième séquence à spécificité de cible se lient à des régions différentes de la même molécule cible, dans laquelle le premier signal, le deuxième signal et le troisième signal sont spatialement distinguables ; dans laquelle la molécule cible est une molécule naturelle ou un ADNc d'une molécule naturelle ou le complément dudit ADNc,

45

dans laquelle, quand ladite paire de sondes est liée à sa molécule cible, l'identité des premier et deuxième signaux et leurs positions l'un par rapport à l'autre constituent au moins une partie d'un code qui identifie la molécule cible, dans laquelle les première, deuxième et troisième régions d'attachement de marqueurs sont des séquences de nucléotides prédéterminées, et

50

dans laquelle une première molécule d'ADN ou d'ARN est hybridée à la première région d'attachement de marqueurs, lesdits un ou plusieurs marqueurs monomères qui émettent une lumière constituant ledit premier signal étant liés à ladite première molécule d'ADN ou d'ARN ; et dans laquelle une deuxième molécule d'ADN ou d'ARN est hybridée à la deuxième région d'attachement de marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit deuxième signal étant liés à ladite deuxième molécule d'ADN ou d'ARN.

55

3. Paire de sondes selon la revendication 1 ou 2, dans laquelle une pluralité de premières molécules d'ADN ou d'ARN sont hybridées à la première région d'attachement de marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit premier signal étant liés aux molécules d'ADN ou d'ARN ; et dans laquelle une pluralité de deuxièmes molécules d'ADN ou d'ARN sont hybridées à la deuxième région d'attachement de

marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit deuxième signal étant liés aux deuxième molécules d'ADN ou d'ARN.

- 5 4. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle la deuxième sonde comprend une troisième région d'attachement de marqueurs à laquelle sont attachées un ou plusieurs monomères marqueurs qui émettent une lumière constituant un troisième signal.
- 10 5. Paire de sondes selon la revendication 4, dans laquelle ladite troisième région d'attachement de marqueurs est une séquence de nucléotides prédéterminée, et dans laquelle une troisième molécule d'ARN est hybridée à la troisième région d'attachement de marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit troisième signal étant attachés à la troisième molécule d'ARN.
- 15 6. Paire de sondes selon la revendication 5, dans laquelle, quand la paire de sondes est liée à sa molécule cible, le code comprend l'identité du premier signal, du deuxième signal et du troisième signal, et leurs positions les uns par rapport aux autres.
- 20 7. Paire de sondes selon la revendication 1 ou 2, dans laquelle le code comprend l'identité des premier et deuxième signaux, leurs positions l'un par rapport à l'autre, et la taille du point résultant d'au moins l'un desdits signaux.
- 25 8. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle les première et deuxième molécules sont des molécules d'acide nucléique.
- 30 9. Paire de sondes selon la revendication 8, dans laquelle les régions d'attachement de marqueurs et les séquences à spécificité de cible sont des séquences de nucléotides prédéterminées.
- 35 10. Paire de sondes selon la revendication 9, dans laquelle lesdits un ou plusieurs monomères marqueurs sont liés de manière covalente à des acides nucléiques hybridés à leurs régions d'attachement respectives.
- 40 11. Paire de sondes selon la revendication 10, dans laquelle lesdits acides nucléiques sont hybridés aux régions d'attachement respectives au moyen d'un ou plusieurs acides nucléiques de pontage liés de manière non covalente.
- 45 12. Paire de sondes selon la revendication 1 ou 2, dans laquelle les première et deuxième séquences à spécificité de cible ne sont pas marquées par un ou plusieurs quelconques desdits monomères marqueurs.
- 50 13. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle la deuxième molécule comprend en outre une quatrième région d'attachement de marqueurs, qui n'est pas en chevauchement avec la troisième région d'attachement de marqueurs, à laquelle sont attachés un ou plusieurs monomères marqueurs qui émettent une lumière constituant un quatrième signal.
- 55 14. Paire de sondes selon la revendication 13, dans laquelle, quand la paire de sondes est liée à sa molécule cible, le code comprend l'identité du premier signal, du deuxième signal, du troisième signal et du quatrième signal, et leurs positions les uns par rapport aux autres.
15. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle une première molécule d'ARN est hybridée à la première région d'attachement de marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit premier signal étant liés à la première molécule d'ARN ; et dans laquelle une deuxième molécule d'ARN est hybridée à la deuxième région d'attachement de marqueurs, lesdits un ou plusieurs monomères marqueurs qui émettent une lumière constituant ledit deuxième signal étant liés à la deuxième molécule d'ARN.
16. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle les étiquettes d'affinité sont attachées de manière covalente à la première molécule ou à la deuxième molécule.
17. Paire de sondes selon la revendication 1 ou 2, dans laquelle les monomères marqueurs attachés à la première région d'attachement de marqueurs émettent une lumière à la même longueur d'onde, laquelle lumière constitue ledit premier signal, dans laquelle les monomères marqueurs attachés à la deuxième région d'attachement de marqueurs émettent une lumière à la même longueur d'onde, laquelle lumière constitue le deuxième signal, et dans laquelle les monomères marqueurs attachés à la troisième région d'attachement de marqueurs, s'ils sont présents,

émettent une lumière à la même longueur d'onde, laquelle lumière constitue le troisième signal.

18. Paire de sondes selon la revendication 1 ou 2, dans laquelle au moins l'un parmi le premier signal et le deuxième signal comprend une lumière à une pluralité de longueurs d'onde différentes.
19. Paire de sondes selon la revendication 1 ou 2, dans laquelle les premier, deuxième et troisième signaux sont spectralement distinguables,
20. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle les première et deuxième molécules d'ARN ont chacune une longueur de 100 à 3000 nucléotides, ou dans laquelle les première et deuxième molécules d'ARN ont chacune une longueur de 500 à 1500 nucléotides.
21. Paire de sondes selon l'une quelconque des revendications 1 à 3, dans laquelle à deuxième sonde est un complexe d'acide nucléique comprenant :
 - (a) la deuxième molécule, laquelle deuxième molécule comprend une troisième région d'attachement de marqueurs à laquelle est hybridée une troisième molécule d'ARN, un ou plusieurs monomères marqueurs qui émettent une lumière constituant un troisième signal étant liés à la troisième molécule d'ARN ; et
 - (b) la deuxième séquence à spécificité de cible attachée à la deuxième molécule,
 dans laquelle le code comprend l'identité des premier, deuxième et troisième signaux et leurs positions les uns par rapport aux autres.
22. Paire de sondes selon la revendication 21, dans laquelle les monomères marqueurs attachés aux premières régions d'attachement de marqueurs émettent une lumière à la même longueur d'onde, laquelle lumière constitue ledit premier signal ; dans laquelle les monomères marqueurs attachés à la deuxième région d'attachement de marqueurs émettent une lumière à la même longueur d'onde, laquelle lumière constitue le deuxième signal ; et dans laquelle les monomères marqueurs attachés à la troisième région d'attachement de marqueurs émettent une lumière à la même longueur d'onde, laquelle lumière constitue le troisième signal.
23. Paire de sondes selon la revendication 21, dans laquelle au moins l'un parmi le premier signal, le deuxième signal et le troisième signal comprend une lumière à une pluralité de longueurs d'ondes différentes.
24. Paire de sondes selon la revendication 21, dans laquelle la première séquence à spécificité de cible et la deuxième séquence à spécificité de cible ne sont pas marquées par un ou plusieurs quelconques desdits monomères marqueurs.
25. Paire de sondes selon la revendication 21, dans laquelle les premier, deuxième et troisième signaux sont spectralement distinguables.
26. Paire de sondes selon la revendication 21, dans laquelle les premier et troisième signaux émettent à la ou aux mêmes longueurs d'onde.
27. Paire de sondes selon la revendication 21, dans laquelle l'étiquette d'affinité est attachée de manière covalente à la première molécule.
28. Paire de sondes selon la revendication 21, dans laquelle l'étiquette d'affinité est attachée de manière covalente à la deuxième molécule.
29. Paire de sondes selon la revendication 21, dans laquelle les première, deuxième et troisième molécules d'ARN ont chacune 100 à 3000 nucléotides ou dans laquelle les première, deuxième et troisième molécules d'ARN ont chacune 500 à 1500 nucléotides.
30. Paire de sondes selon les revendications 21 à 30, dans laquelle les molécules d'ARN comprennent une seule région comprenant une base régulièrement répétée.
31. Paire de sondes selon la revendication 30, dans laquelle la base régulièrement répétée est l'uracile à modification aminoallyle.

32. Procédé pour détecter une molécule cible dans un échantillon biomoléculaire, comprenant :

- 5 (i) la mise en contact dudit échantillon avec une paire de sondes selon l'une quelconque des revendications 1 à 31 dans des conditions qui permettent la liaison, à la molécule cible, de la première séquence à spécificité de cible et de la deuxième séquence à spécificité de cible, et
(ii) la détection du code qui identifie la molécule cible.

10 33. Procédé selon la revendication 32, comprenant en outre la quantification de la quantité de ladite molécule cible dans ledit échantillon biomoléculaire,

34. Procédé selon la revendication 32 ou 33, dans lequel la concentration de chacune desdites première et deuxième séquences à spécificité de cible est supérieure à la concentration de ladite molécule cible.

15 35. Procédé selon les revendications 32 à 34, comprenant en outre l'étape consistant à étirer la paire de sondes liée et la molécule cible avant détection du code qui identifie la molécule cible.

20 36. Procédé selon la revendication 35, dans lequel l'étirement est effectué au moyen d'une technique choisie dans l'ensemble constitué par un étirement par écoulement, une technique de ménisque de recul, un électro-étirement et une constriction dans l'écoulement d'un liquide contenant le nanorapporteur conjointement avec un champ électrique oscillant.

37. Procédé pour détecter une pluralité de molécules cibles dans un échantillon biomoléculaire, comprenant :

- 25 (i) la mise en contact dudit échantillon avec une population de paires de sondes selon l'une quelconque des revendications 1 à 31, dans des conditions qui permettent la liaison de la première séquence à spécificité de cible et de la deuxième séquence à spécificité de cible de chaque paire de sondes à leur molécule cible respective, chaque paire de sondes dans ladite population, lorsqu'elle est liée à sa molécule cible respective, étant associée à un code distinguable, et
(ii) la détection du code qui identifie la pluralité de molécules cibles.

30 38. Procédé selon la revendication 37, comprenant en outre la quantification de la quantité de chacune parmi ladite pluralité de molécules cibles dans ledit échantillon biomoléculaire.

35 39. Procédé selon les revendications 37 et 38, dans laquelle la concentration de chacune parmi lesdites première et deuxième séquences à spécificité de cible de chaque paire de sondes de ladite population est supérieure à la concentration de sa molécule cible respective.

40 40. Procédé selon les revendications 37 à 39, comprenant en outre l'étape consistant à étirer les paires de sondes liées et leurs molécules cibles respectives avant détection des codes qui identifient la pluralité de molécules cibles.

45 41. Procédé selon la revendication 40, dans lequel l'étirement est effectué au moyen d'une technique choisie dans l'ensemble constitué par un étirement par écoulement, une technique de ménisque de recul, un électro-étirement et une constriction dans l'écoulement d'un liquide contenant le nanorapporteur conjointement avec un champ électrique oscillant.

50

55

FIGURE 1A-1C

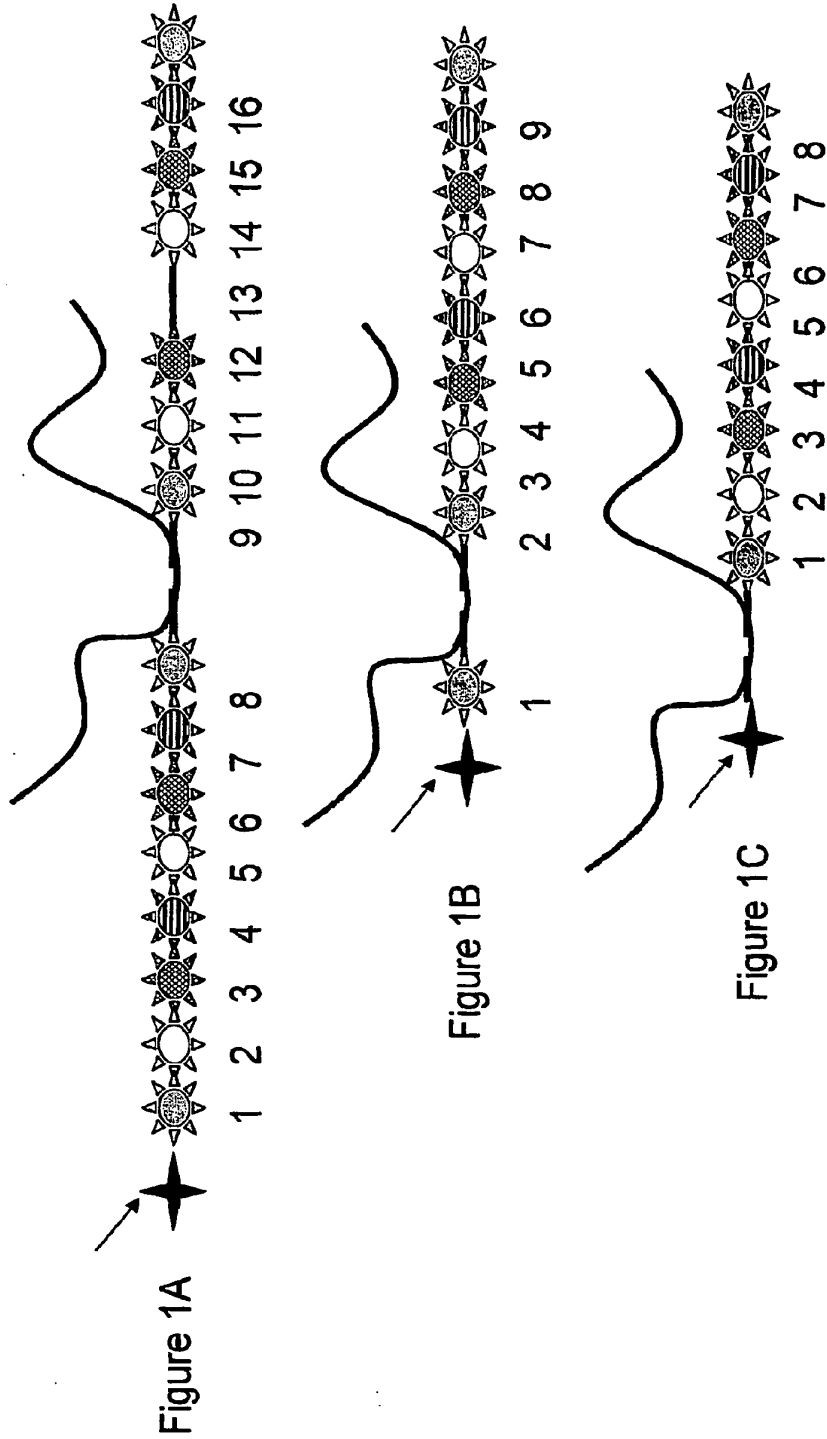


FIGURE 1D-1F

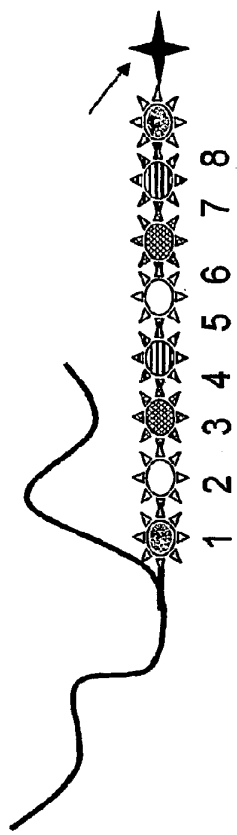


Figure 1D

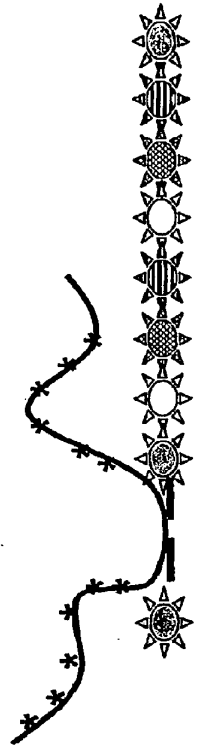


Figure 1E

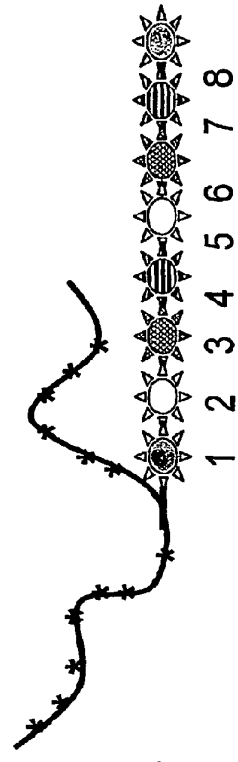


Figure 1F

FIGURE 2

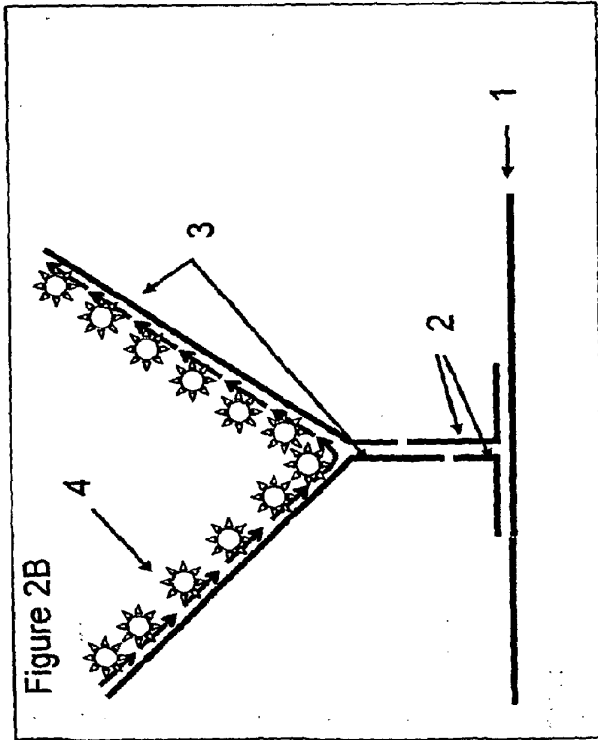


Figure 2C

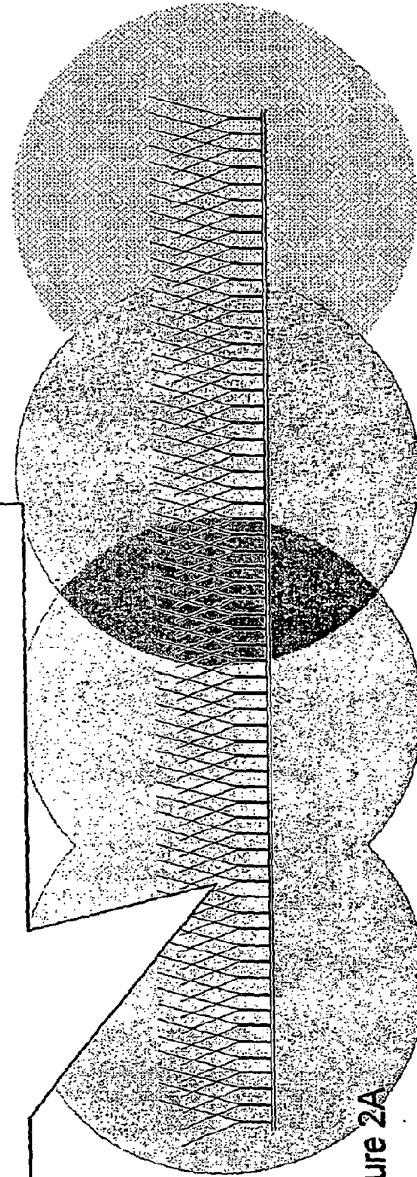
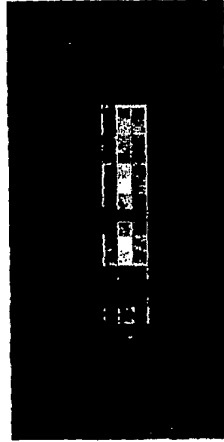


Figure 2A

FIGURE 3

Figure 3A

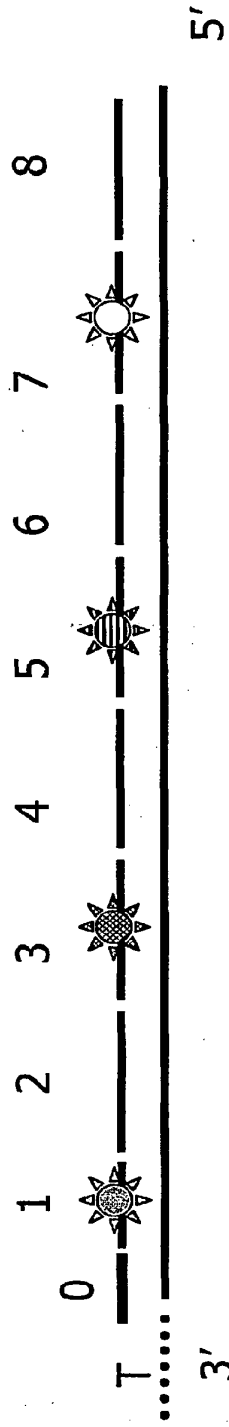


Figure 3B

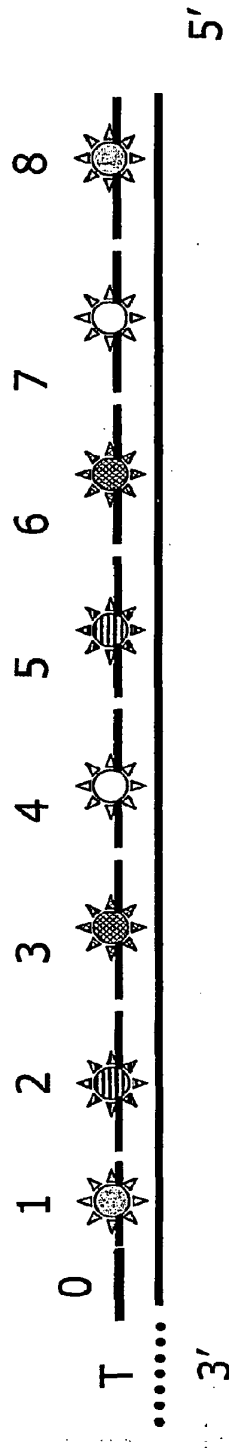


FIGURE 4

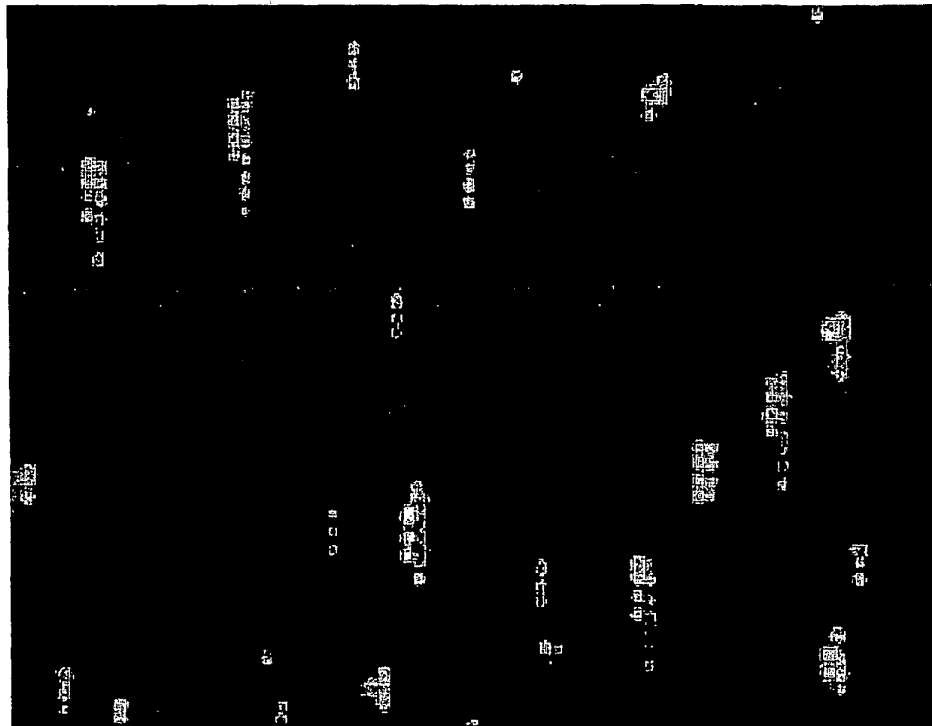


FIGURE 5

Figure 5A

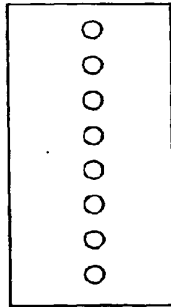


Figure 5B

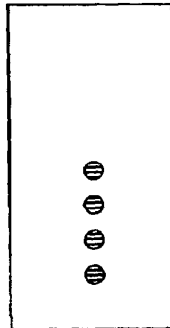


Figure 5C

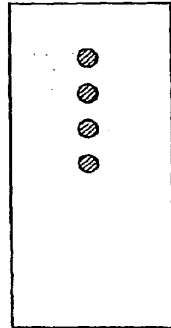


Figure 5D

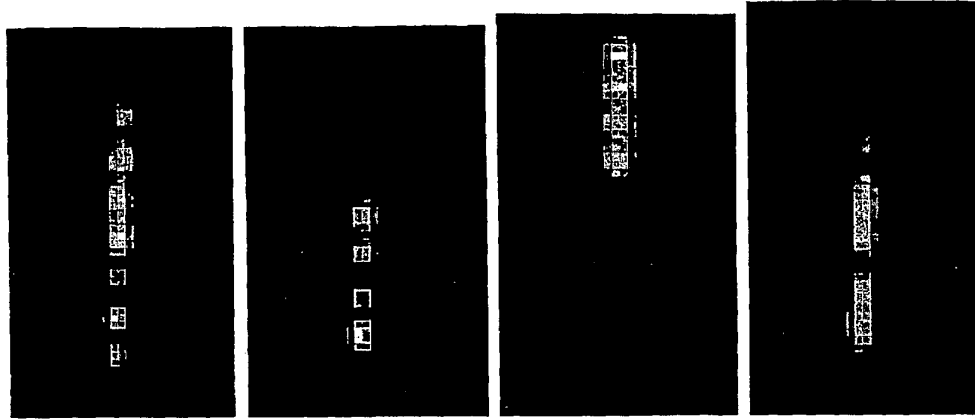
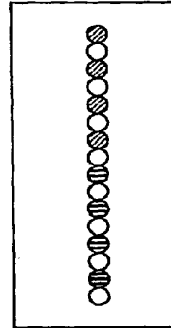


FIGURE 6

Figure 6A

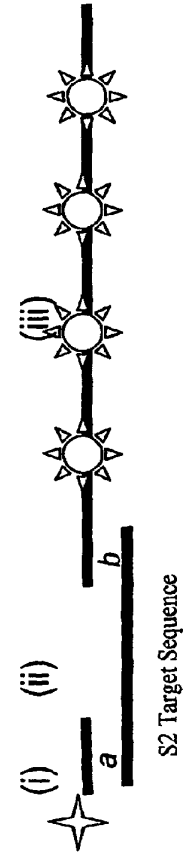


Figure 6E

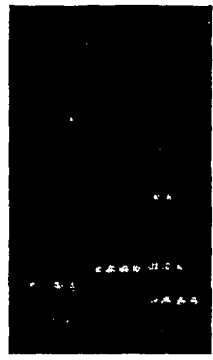


Figure 6D

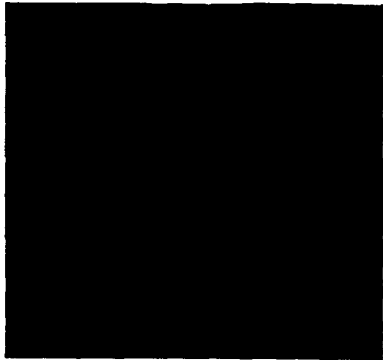


Figure 6C

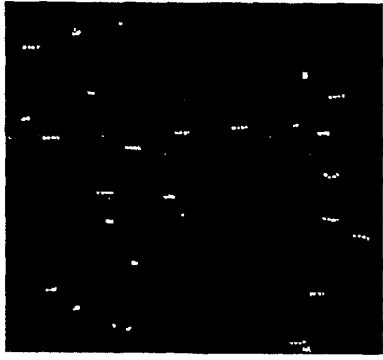


Figure 6B

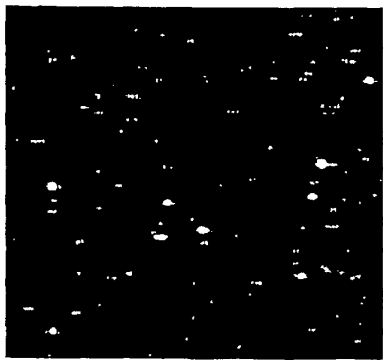


FIGURE 7A-D

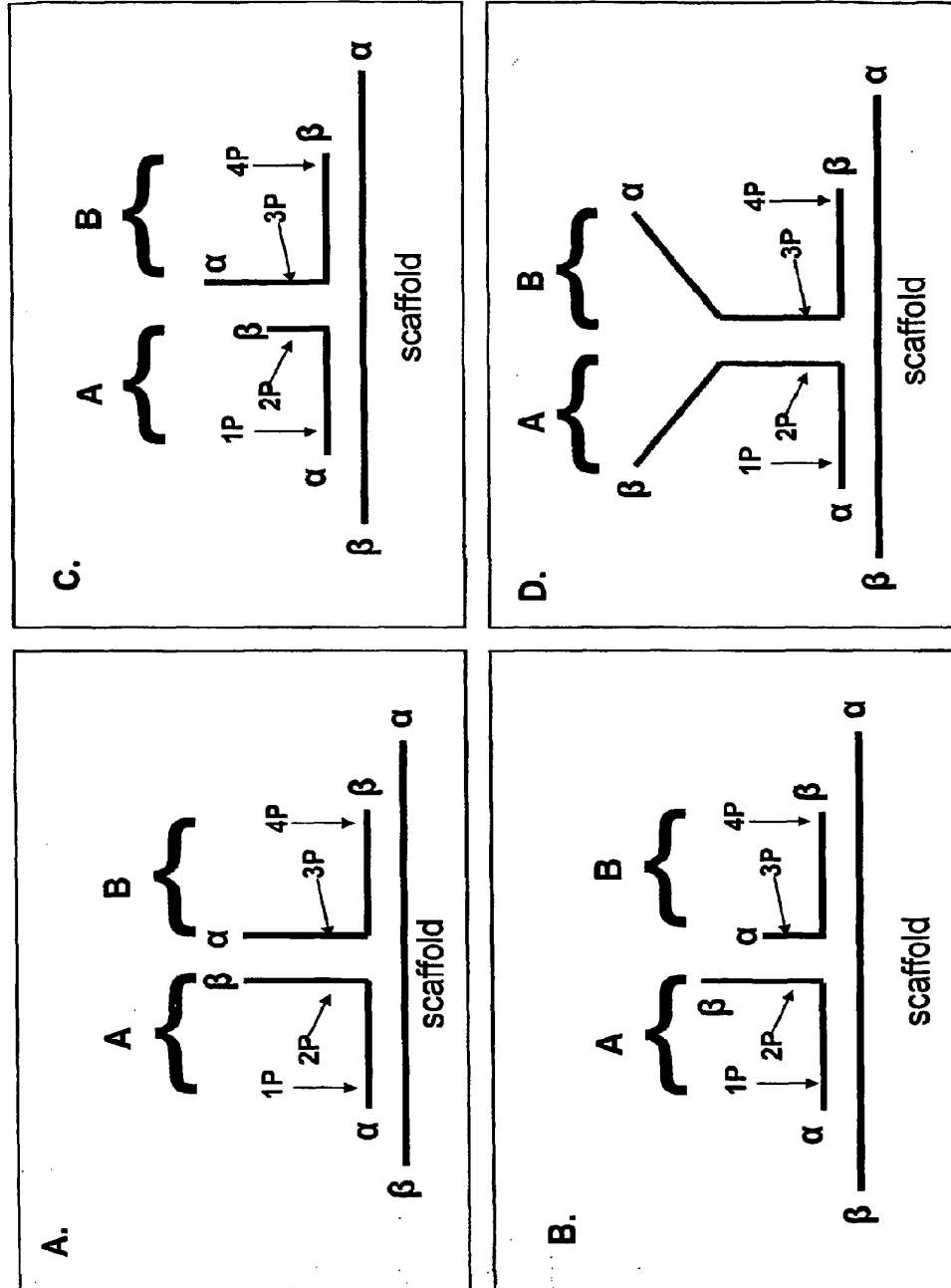


FIGURE 7E-G

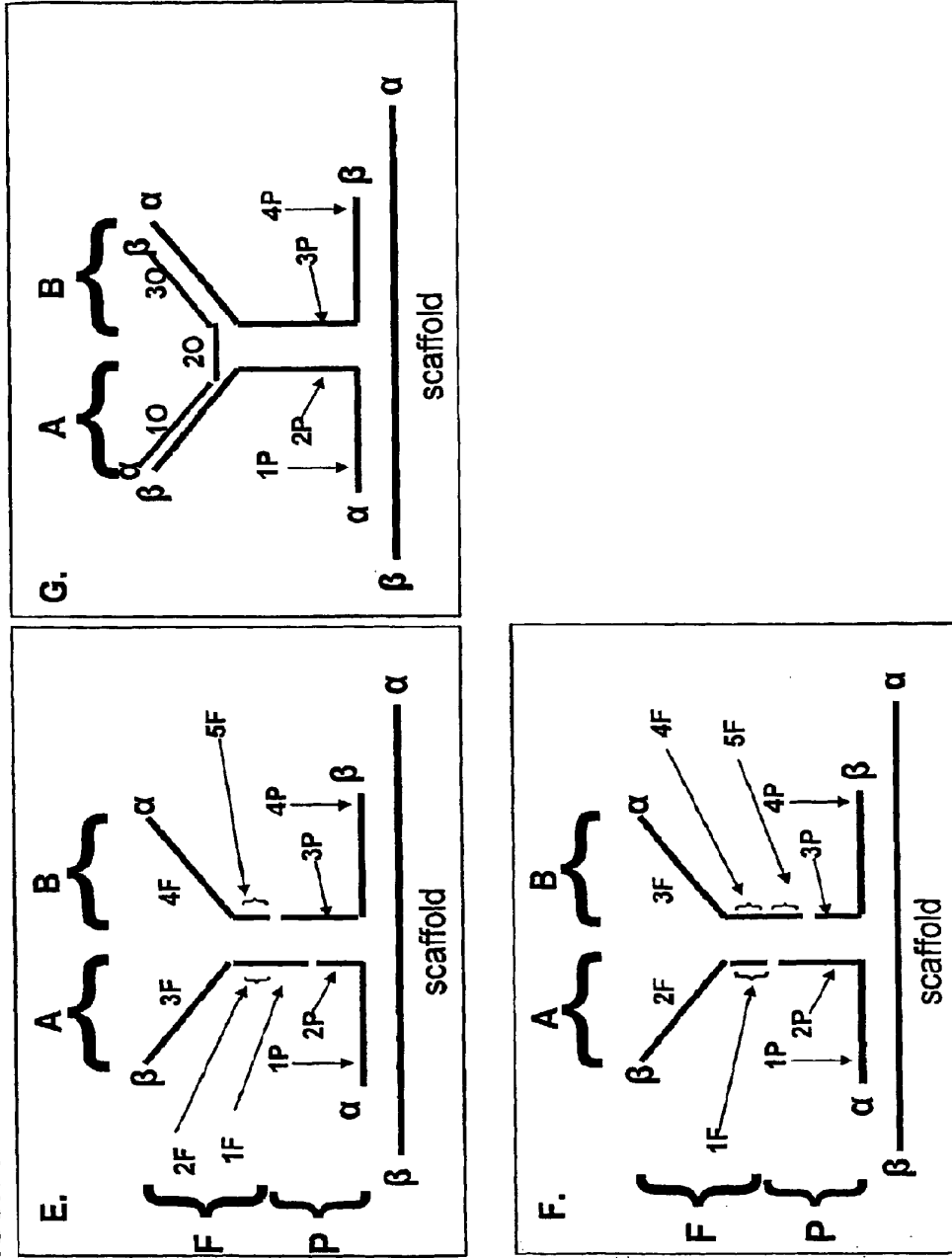


FIGURE 8

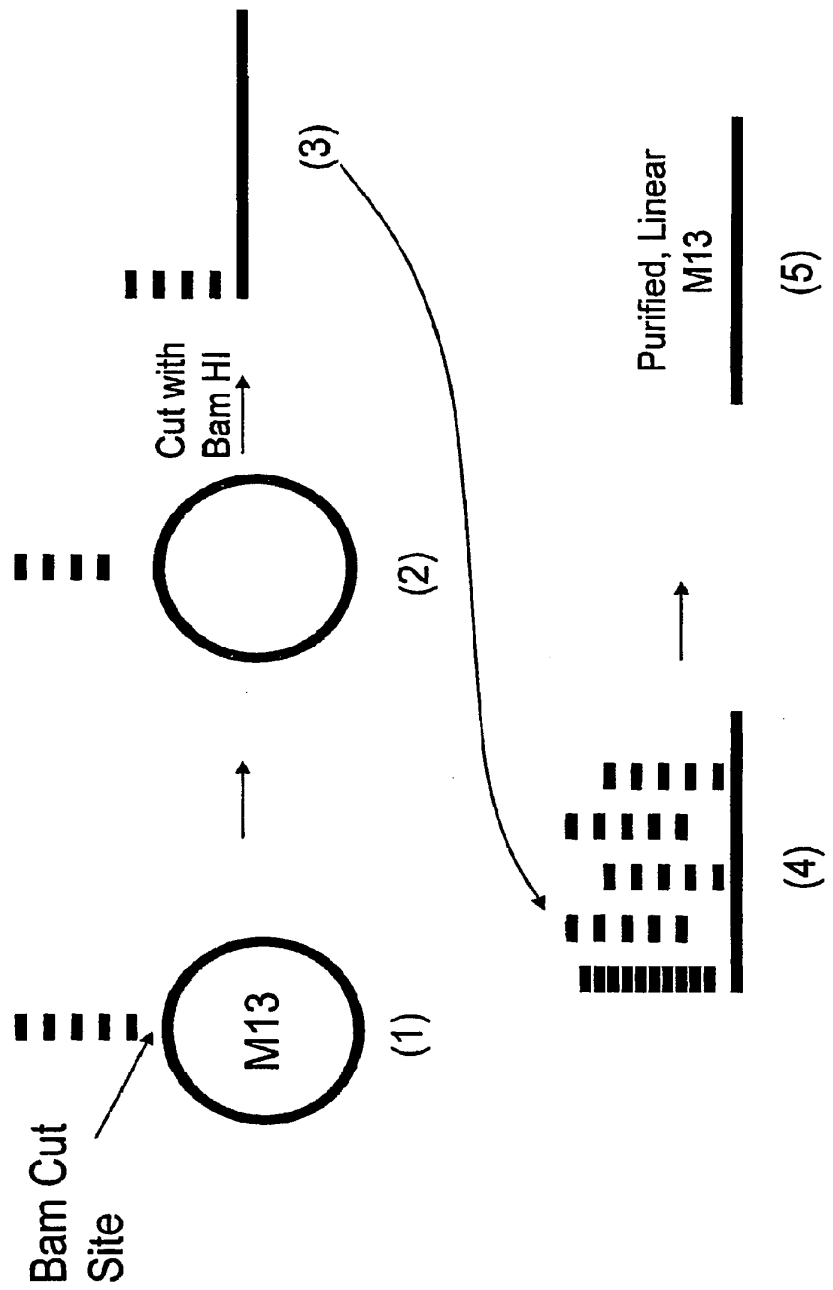


FIGURE 9

Figure 9A

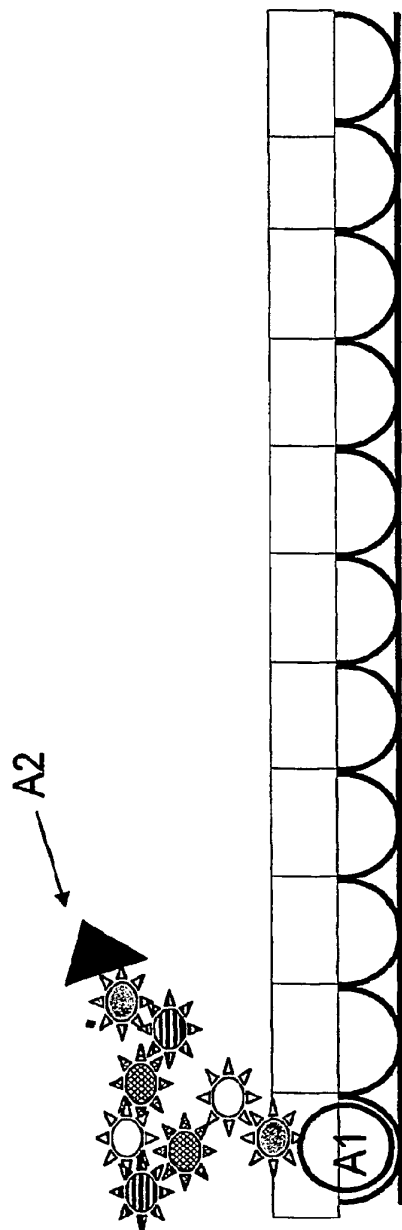


Figure 9B

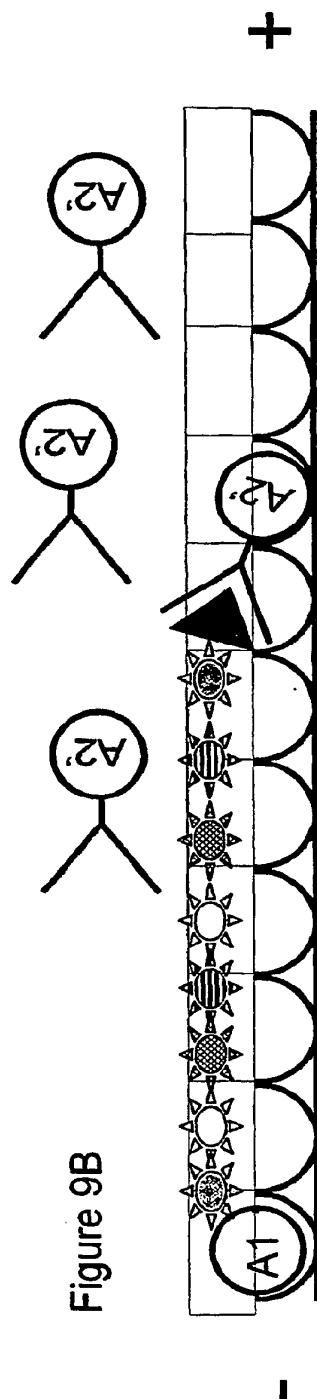


FIGURE 10

Figure 10A



Figure 10B



Figure 10C

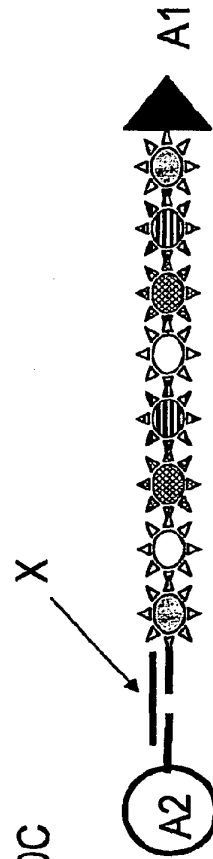


FIGURE 11

Figure 11A

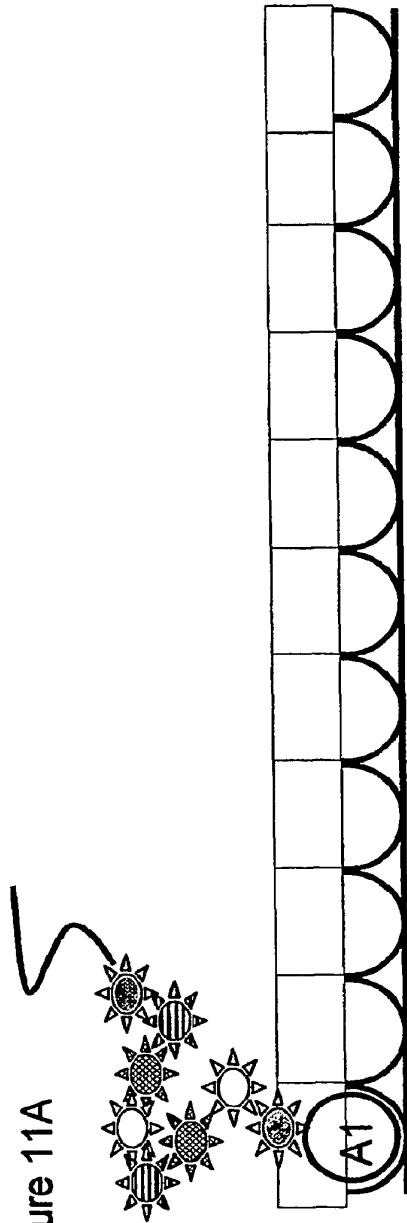


Figure 11B

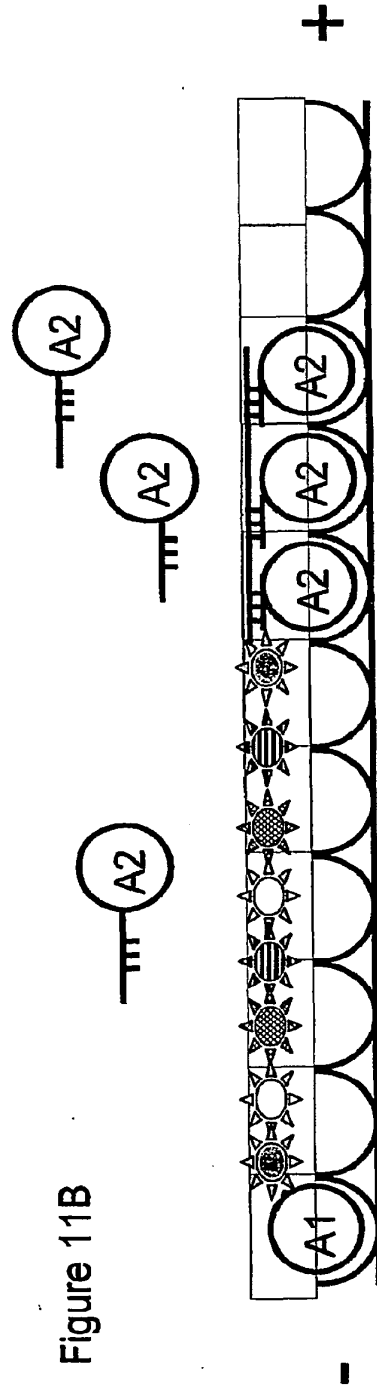


FIGURE 12A-12B

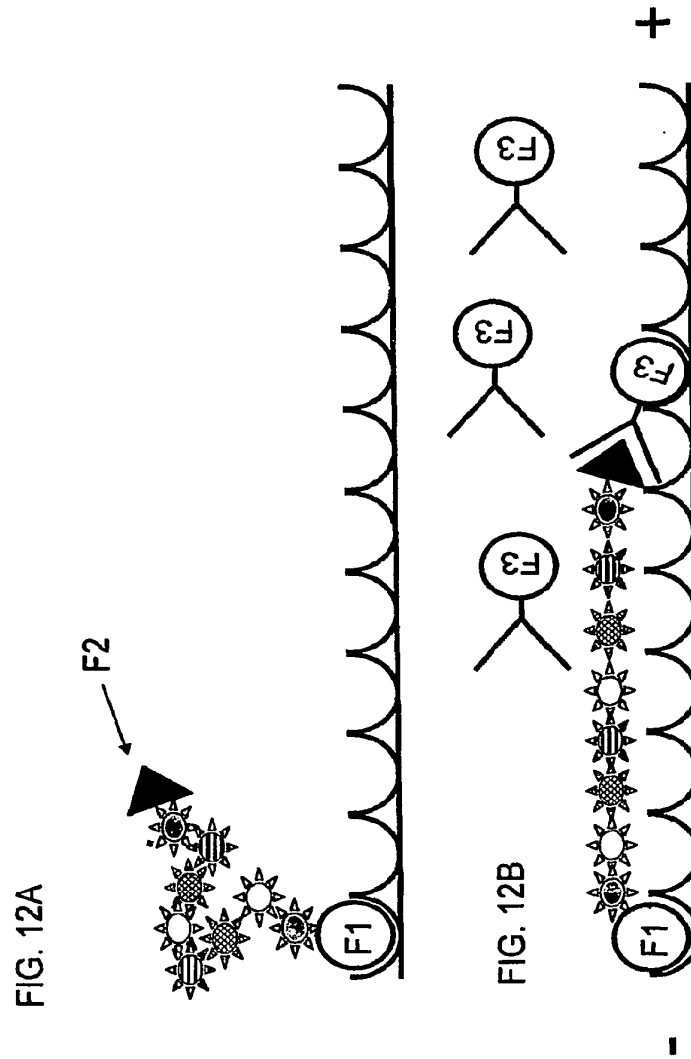


FIGURE 13A-13C

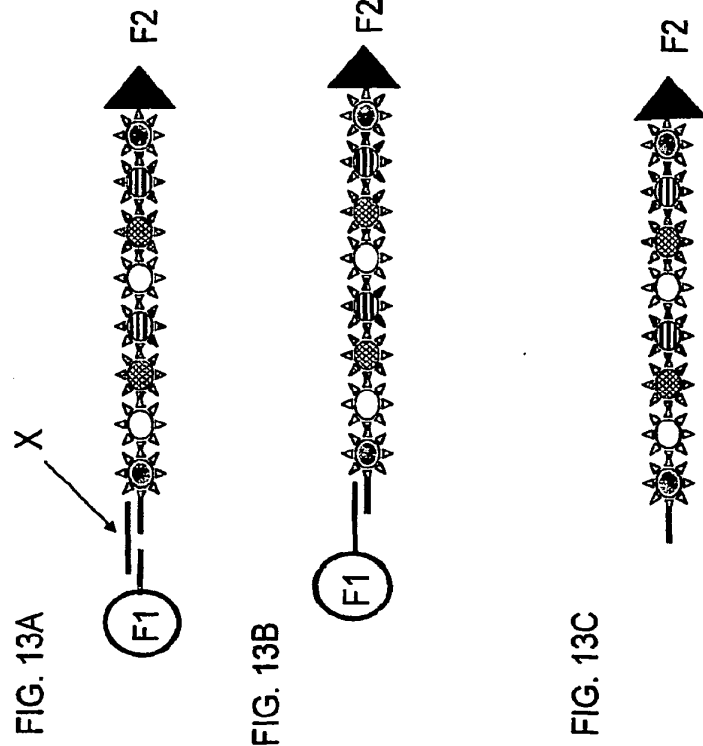


FIGURE 14A-14B

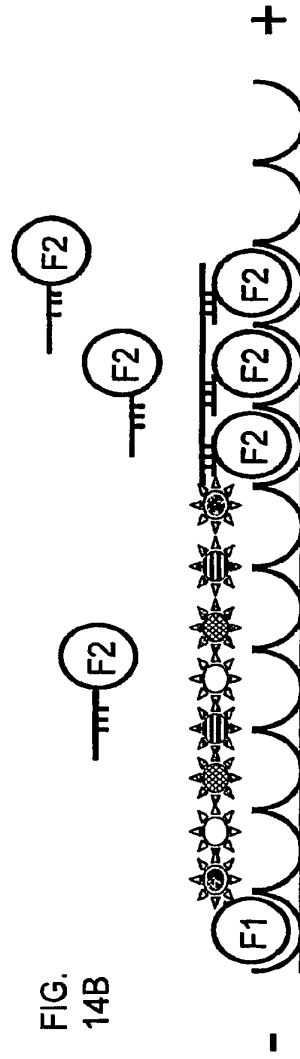
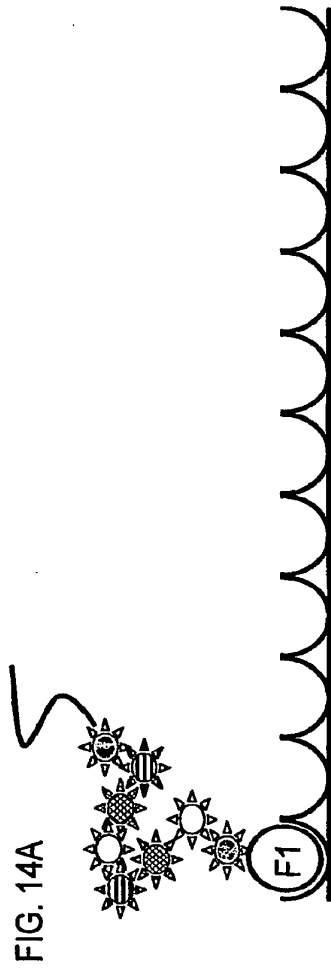


FIGURE 14C-14D

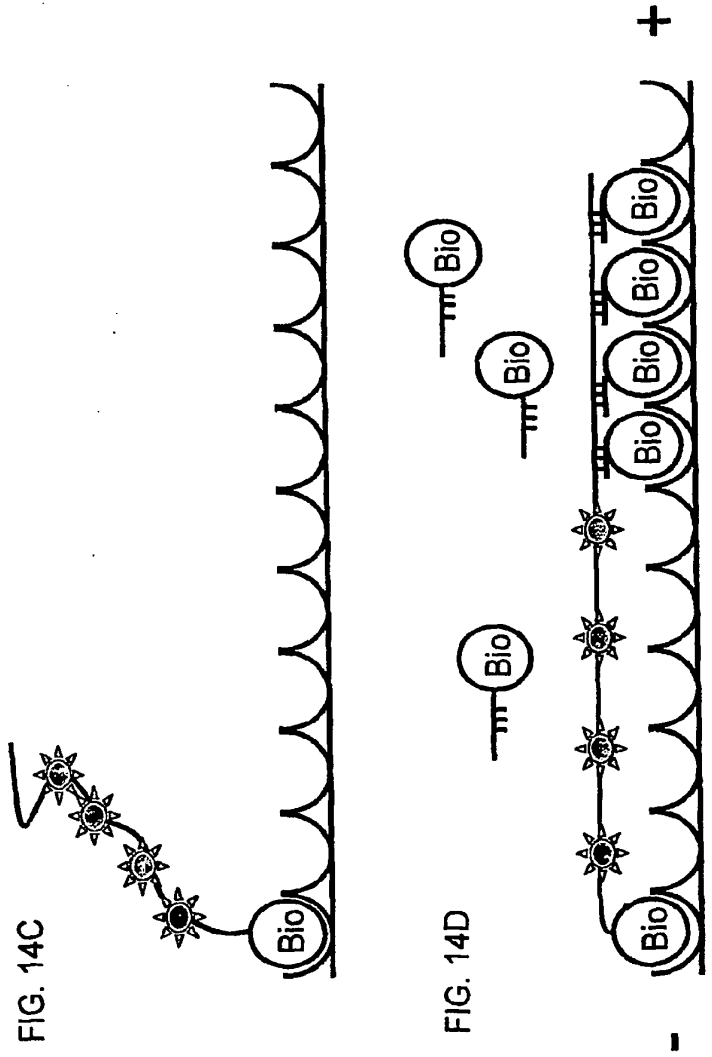


FIGURE 15A

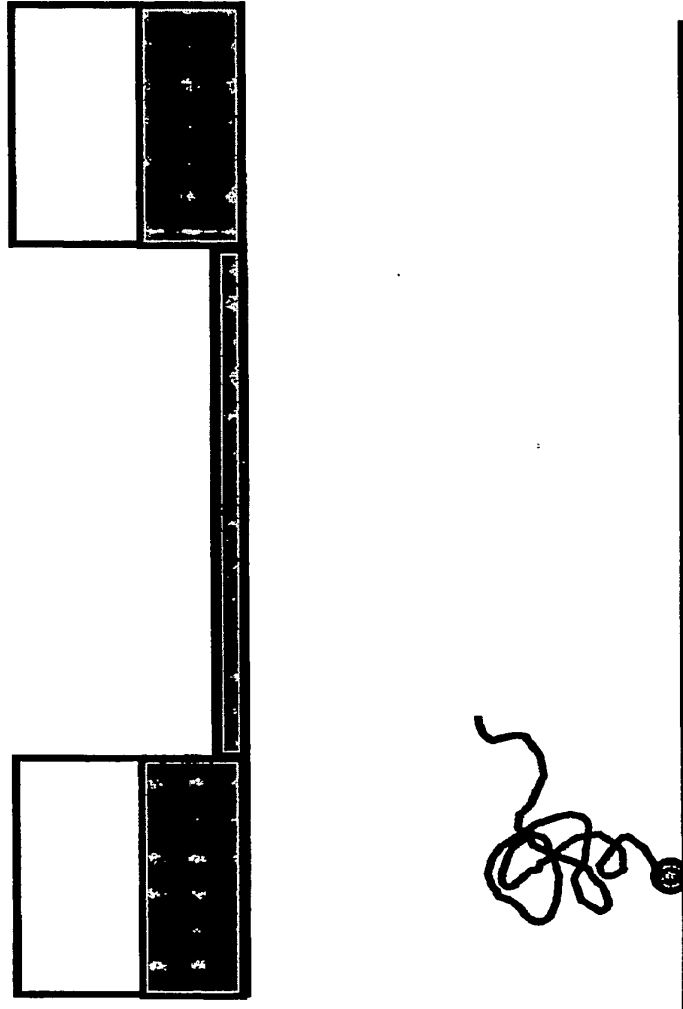


FIGURE 15B

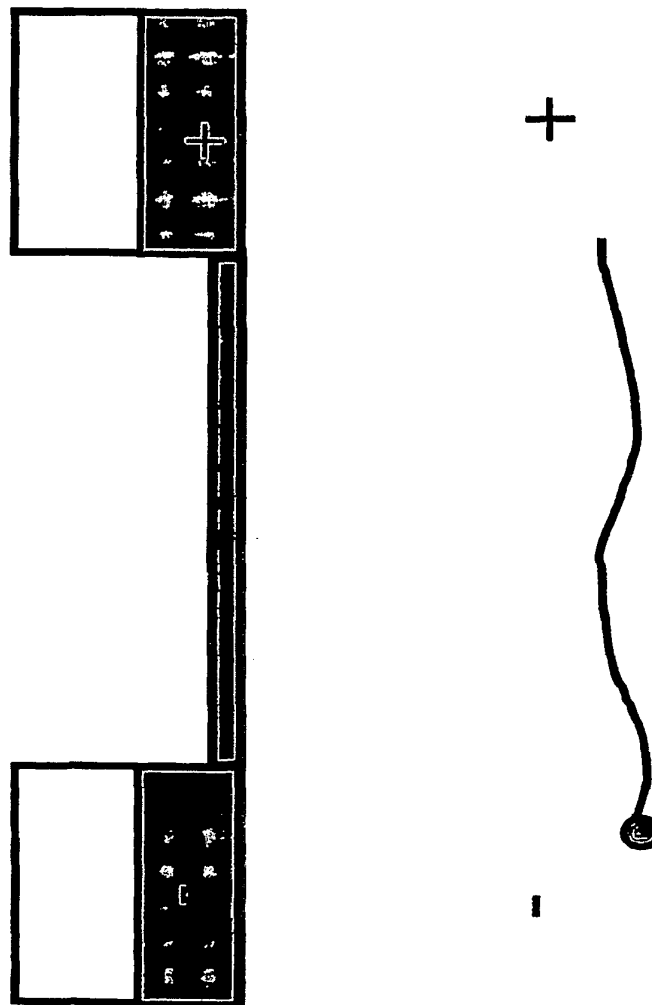
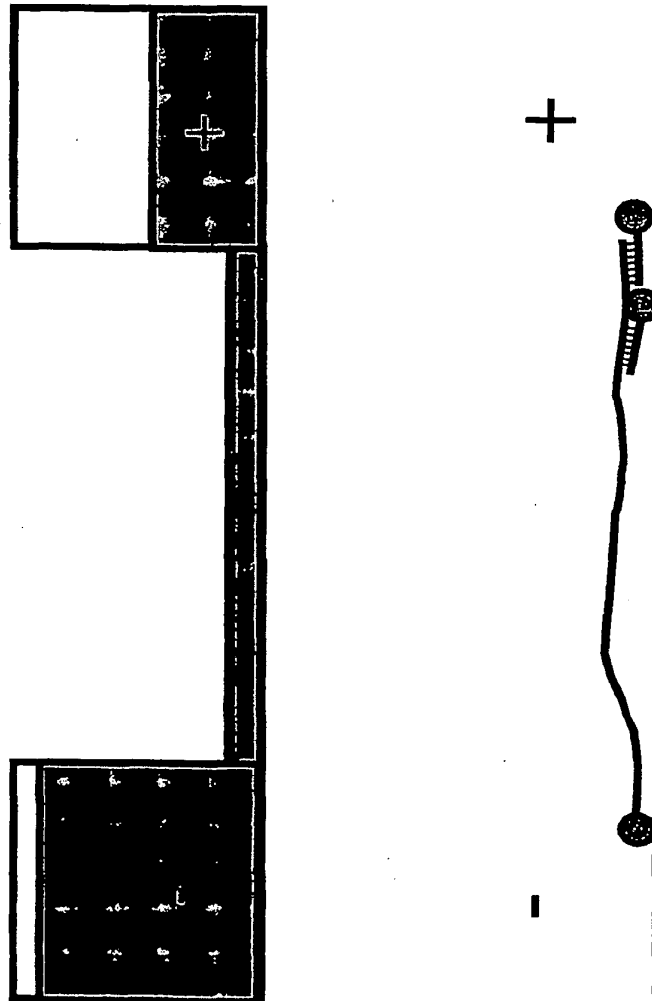


FIGURE 15C



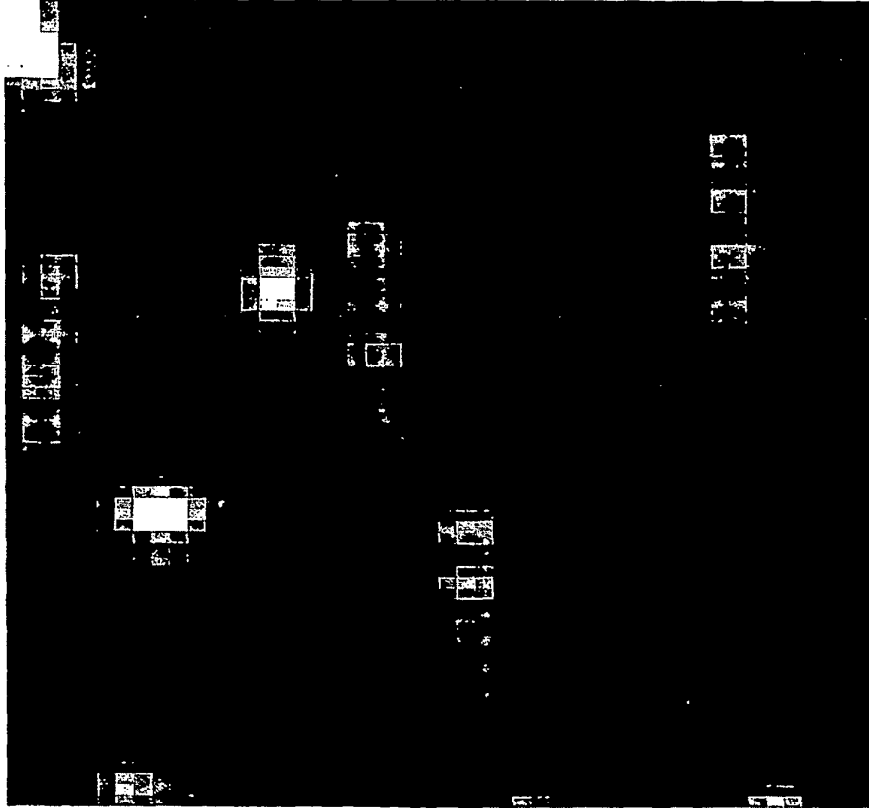


FIGURE 16

FIGURE 17

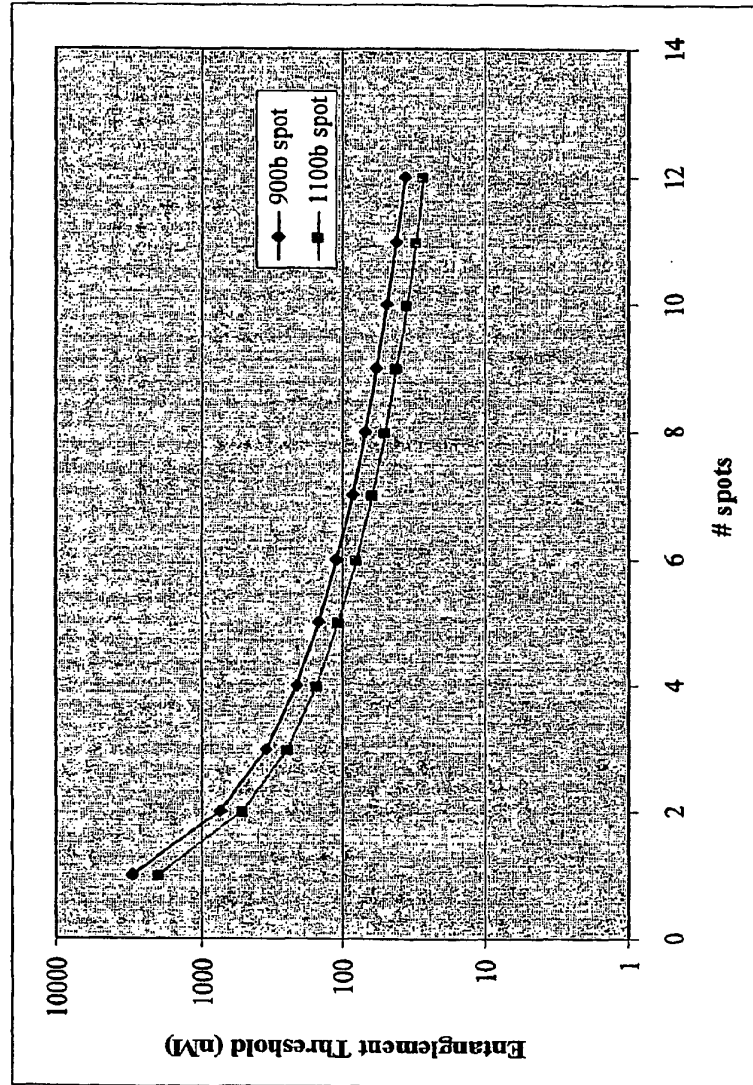
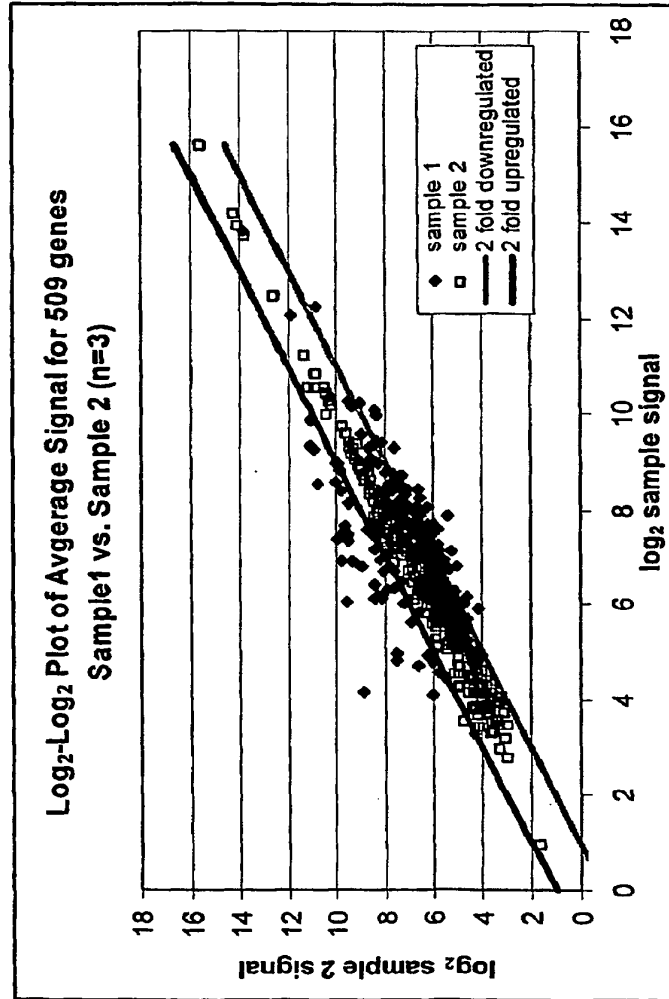


FIGURE 18



REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 2005071401 A [0006]
- US 6268147 B [0007]
- WO 03003810 A [0008]
- US 20100261026 A1 [0078]
- US 5293050 A [0175]
- US 5354707 A, Chapple Sokol, [0175]
- US 5496934 A [0196]
- US 5202247 A [0196]
- US 5137819 A [0196]
- US 6001983 A [0253]
- US 6548255 B [0272] [0274]
- US 6344319 B [0272] [0274]
- US 6303296 B [0272] [0274]
- US 6265153 B [0272] [0274]
- US 6225055 B [0272] [0274]
- US 6054327 B [0272] [0274]
- US 5840862 B [0272] [0274]

Non-patent literature cited in the description

- Schena et al. *Science*, 1995, vol. 270, 467-470 [0009]
- Lockhart et al. *Nature Biotechnology*, 1996, vol. 14, 1675-1680 [0009]
- Blanchard et al. *Nature Biotechnology*, 1996, vol. 14, 1649 [0009]
- Zuker. *Nucleic Acids Res.*, 2003, vol. 31 (13), 3406-15 [0135]
- Mathews et al. *J. Mol. Biol.*, 1999, vol. 288, 911-940 [0135]
- Hyrup et al. *Bioorganic & Medicinal Chemistry*, vol. 4 (1), 5-23 [0140]
- Hyrup et al. *Bioorganic & Medicinal Chemistry*, 1996, vol. 4 (1), 5-23 [0140]
- Perry-O'Keefe et al. *Proc. Natl. Acad. Sci. USA*, 1996, vol. 93, 14670-675 [0140]
- Metal-chelate affinity chromatography. Petty et al. *Current Protocols in Molecular Biology*. Greene Publish. Assoc. & Wiley Interscience, 1996, vol. 2 [0196]
- Smith. *Methods Mol. Cell Bio.*, 1993, vol. 4, 220-229 [0196]
- Guan et al. *Gene*, 1987, vol. 67, 21-30 [0196]
- Tomme et al. *Protein Eng.*, 1994, vol. 7, 117-123 [0196]
- Henegariu et al. *Biotechniques*, 2001, vol. 31, 246-50 [0224]
- Yokota et al. *Nuc. Acids Res.*, 1997, vol. 25, 1064-1070 [0224]
- Matsuura et al. *Nuc. Acids Res.*, 2001, vol. 29, E79 [0224]
- Uhlman ; Peyman. *Chemical Review*, 1990, vol. 90 (4), 544 584 [0253]
- Goodchild. *Bioconjugate Chem.*, 1990, vol. 1 (3), 165 186 [0253]
- Egholm et al. *J. Am. Chem. Soc.*, 1992, vol. 114, 1895 [0253]
- Gryaznov et al. *J. Am. Chem. Soc.*, vol. 116, 3143 3144 [0253]
- Fasman. *CRC Practical Handbook of Biochemistry and Molecular Biology*. 1985, 385-392 [0253]
- Beilstein's *Handbuch der Organischen Chemie*. Springer Verlag [0253]
- Matsuura et al. *JBiomol Struct Dyn.*, 2002, vol. 20 (3), 429-36 [0271]
- Ferree ; Blanch. *Biophys J.*, 2003, vol. 85 (4), 2539-46 [0271]
- Stigter ; Bustamante. *Biophys J.*, 1998, vol. 75 (3), 1197-210 [0271]
- Matsuura et al. *Nucleic Acids Res.*, 2001, vol. 29 (16 [0271]
- Ferree ; Blanch. *Biophys J.*, 2004, vol. 87 (1), 468-75 [0271]
- Bensimon et al. *Science*, 1994, vol. 265, 2096-2098 [0272]
- Henegariu et al. *BioTechniques*, 2001, vol. 31, 246-250 [0272]
- Kraus et al. *Human Genetics*, 1997, vol. 99, 374-380 [0272] [0273]
- Michalet et al. *Science*, 1997, vol. 277, 1518-1523 [0272] [0273]
- Yokota et al. *Nucleic Acids Res.*, 1997, vol. 25 (5), 1064-70 [0272] [0273]
- Otobe et al. *Nucleic Acids Research*, 2001, vol. 29, 109 [0272]
- Zimmerman ; Cox. *Nucleic Acids Res.*, 1994, vol. 22 (3), 492-7 [0272]
- Ashkin. *Optics Letters*, 1986, vol. 11, 288-290 [0275]
- Ashkin et al. *Science*, 1987, vol. 235, 1517-1520 [0275]
- Ashkin et al. *Nature*, vol. 330, 769-771 [0275]

EP 1 963 531 B1

- **Perkins et al.** *Science*, 1994, vol. 264, 822-826 [0275]
- **Simmons et al.** *Biophysical Journal*, 1996, vol. 70 (1813-1), 22 [0275]
- **Block et al.** *Nature*, 1990, vol. 348, 348-352 [0275]
- **Grier.** *Nature*, 2003, vol. 424, 810-816 [0275]
- **Asbury et al.** *Electrophoresis*, 2002, vol. 23 (16), 2658-66 [0305]
- **Kabata et al.** *Science*, 1993, vol. 262 (5139), 1561-3 [0305]
- **Asbury ; van den Engh.** *Biophys J.*, 1998, vol. 74, 1024-30 [0305]
- Surface Treatment. Wiley Encyclopedia of Packaging Technology. John Wiley & Sons, 1997, 867 874 [0311]
- **Atkinson ; Smith.** Oligonucleotide Synthesis: A Practical Approach. IRL Press, 1984, vol. Solid Ph, 45 49 [0311]
- **Pease.** *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, 5022 5026 [0311]
- **Lloyd Williams et al.** Chemical Approaches to the Synthesis of Peptides and Proteins. CRC Press, 1997 [0311]
- **Wetmur.** *J. Annu. Rev. Biophys. Bioeng.*, 1976, vol. 5, 337-361 [0436]
- **Britten et al.** *Methods in Enzymology*, 1974, vol. 29E, 363-406 [0437]
- **Inoue, S.** Video Microcopy. Plenum Press, 1997, 30 [0493]
- **Gustafsson.** *Proc. Nat'l. Acad. Sci. U.S.A.*, 2005, vol. 102, 13081-13086 [0493]
- **Thomann.** *J. Microsc.*, 2002, vol. 21 (1), 230-248 [0493]

(19)



(11)

EP 2 224 017 A1

(12)

EUROPEAN PATENT APPLICATION
published in accordance with Art. 153(4) EPC

(43) Date of publication:

01.09.2010 Bulletin 2010/35

(51) Int Cl.:

C12Q 1/68 (2006.01)**C12N 15/09 (2006.01)**(21) Application number: **08865229.2**

(86) International application number:

PCT/JP2008/073537(22) Date of filing: **25.12.2008**

(87) International publication number:

WO 2009/081967 (02.07.2009 Gazette 2009/27)

(84) Designated Contracting States:

**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT
RO SE SI SK TR**

Designated Extension States:

AL BA MK RS• **MAJIMA Satoshi****Kyoto-shi, Kyoto 601-8045 (JP)**• **HOSOMI Toshiya****Kyoto-shi, Kyoto 601-8045 (JP)**(30) Priority: **26.12.2007 JP 2007334360**(71) Applicant: **ARKRAY, Inc.****Kyoto-shi, Kyoto 601-8045 (JP)**(74) Representative: **Jones, Elizabeth Louise et al****Dehns****St Bride's House****10 Salisbury Square****London****EC4Y 8JD (GB)**

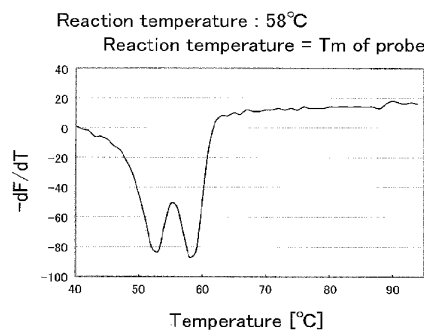
(72) Inventors:

• **HIRAI Mitsuharu****Kyoto-shi, Kyoto 601-8045 (JP)**

(54) **METHOD FOR AMPLIFYING TARGET NUCLEIC ACID SEQUENCE AND PROBE USED FOR THE SAME**

(57) The present invention provides a method for amplifying a target sequence while suppressing inhibition of an amplification reaction in nucleic acid amplification in the presence of the probe. At the time of amplifying the target sequence, as the probe caused to coexist in amplification of the target sequence, a probe having a base sequence in which a melting temperature of the double-stranded nucleic acid is equal to or lower than a reaction

temperature of the elongation reaction is used. In the presence of such a probe, for example, annealing of a primer and an elongation reaction from the primer are hardly inhibited by the presence of the probe so that amplification of the target sequence can be conducted sufficiently. Therefore, when a polymorphism of a target site in the target sequence is analyzed by a T_m analysis or the like, high reliability can be achieved.

**FIG. 2****EP 2 224 017 A1**

Description

Technical Field

5 **[0001]** The present invention relates to a method for amplifying a target nucleic acid sequence in the presence of a probe and a probe used therefor. In addition, the present invention relates to a method for suppressing the inhibition of amplification of a target nucleic acid sequence in the presence of a probe and a method for analyzing a target nucleic acid sequence.

10 Background Art

[0002] In recent years, as a method for detecting a mutation or a polymorphism in a gene, a method for analyzing a melting temperature (T_m) of a double-stranded nucleic acid composed of a target nucleic acid sequence and a detection probe has been put to practical use. The foregoing method is, for example, referred to as a melting profile analysis or a T_m analysis because the method is conducted by analyzing a melting profile of the double strand. A T_m generally is defined as follows. The absorbance at 260 nm increases as a solution containing a double-stranded nucleic acid such as a double-stranded DNA or the like is heated. This increase is caused by the fact that the hydrogen bond between both the strands in a double-stranded nucleic acid is unbound by heating, and the double-stranded nucleic acid is dissociated into single-stranded nucleic acids (melting of a nucleic acid). When every double-stranded nucleic acid is dissociated into single-stranded nucleic acids, the solution exhibits an absorbance about 1.5 times as large as the absorbance at the time when the heating was initiated. (the absorbance of the solution containing only the double-stranded nucleic acid), whereby it can be determined that the melting is completed. Based on this phenomenon, a melting temperature, (T_m) ($^{\circ}\text{C}$), generally is defined as a temperature at the time when the amount of increase in absorbance reaches 50% of the total amount of increase in absorbance.

25 **[0003]** A single nucleotide polymorphism occurs by substitution of a single base (point mutation) in a site to be detected (hereinafter simply referred to as "target site"). Therefore, in order to analyze the type of the polymorphism in the target site, it is necessary to distinguish the difference in a single base. In the T_m analysis described above, this difference in a single base is analyzed by the difference in T_m value. It is described specifically below. A polymorphism in a target site is referred to as a mutant type or a wild type for the sake of convenience.

30 **[0004]** The T_m value becomes higher as the homology of a double-stranded nucleic acid becomes higher and becomes lower as the homology of a double-stranded nucleic acid becomes lower. Therefore, for example, in the case where a detection probe that is perfectly complementary to a target nucleic acid sequence having a mutant-type target site (hereinafter referred to as "mutant-type probe") is designed, the $T_{m_{\text{mut}}}$ value of a double-stranded nucleic acid composed of the mutant-type probe and the target sequence with a mutant-type target site is higher than the $T_{m_{\text{wild}}}$ value of a double-stranded nucleic acid composed of the mutant-type probe and a target sequence with a wild-type target site. According to the T_m analysis utilizing this nature, a polymorphism can be analyzed in the following manner. First, a $T_{m_{\text{mut}}}$ value and a $T_{m_{\text{wild}}}$ value in the case where the mutant-type probe is used are previously determined as evaluation criteria. Subsequently, a double-stranded nucleic acid composed of a target nucleic acid sequence to be analyzed having an unknown base in a target site and the mutant-type probe is formed, a heat treatment is conducted to the formed double-stranded nucleic acid, and signal values such as absorbance and the like showing disassociation of the double-stranded nucleic acid with increase in temperature are measured. Further, a melting profile showing variations in signal value with changes in temperature is prepared, and whether a peak is present at the previously determined $T_{m_{\text{mut}}}$ value or at the previously determined $T_{m_{\text{wild}}}$ value is checked. When the peak is present at around the $T_{m_{\text{mut}}}$ value, the target nucleic acid sequence is a 100% match to the mutant-type probe, whereby the target site in the target nucleic acid sequence can be determined as having a mutant-type polymorphism. On the other hand, when the peak is present at around the $T_{m_{\text{wild}}}$ value, the target nucleic acid sequence is a mismatch to the mutant-type probe in a single base, whereby the target site in the target nucleic acid sequence can be determined as having a wild-type polymorphism. Further, whether the polymorphism is homozygous or heterozygous also can be determined. That is, in an analysis of a polymorphism in a pair of alleles, when peaks are present at both around the $T_{m_{\text{mut}}}$ value and around the $T_{m_{\text{wild}}}$ value, the polymorphism can be determined as being heterozygous. On the other hand, when a peak is present at only around the $T_{m_{\text{mut}}}$ value, the polymorphism can be determined as being a mutant-type homozygous, and when a peak is present at only around the $T_{m_{\text{wild}}}$ value, the polymorphism can be determined as being a wild-type homozygous. Further, also in the case where a detection probe that is perfectly complementary to a target nucleic acid sequence having a wild-type target site (hereinafter referred to as "wild-type probe") is used, by setting, as evaluation criteria, the $T_{m_{\text{wild}}}$ value of a double-stranded nucleic acid composed of the wild-type probe and a target sequence with a wild-type target site and the $T_{m_{\text{mut}}}$ value of a double-stranded nucleic acid composed of the wild-type probe and the target sequence with a mutant-type target site, the determination can be made in the same manner. It is to be noted that, in the case of the wild-type probe, the $T_{m_{\text{wild}}}$ value of a double-stranded nucleic acid composed of the wild-type probe and the target

sequence with a wild-type target site is higher than the $T_{m_{mut}}$ value of a double-stranded nucleic acid composed of the wild-type probe and the target sequence with a mutant-type target site.

[0005] In a method for analyzing a polymorphism utilizing such a T_m analysis, in order to further simplify the analysis, an amplification reaction of the target nucleic acid sequence is conducted in the presence of the detection probe, and signal values are measured using the resultant reaction solution after the amplification reaction as it is. According to such a method, a simplified analysis can be achieved because, after the amplification reaction, a step of further adding the detection probe into the resultant reaction solution can be omitted. Further, when the detection probe is previously added in the reaction solution, airtight state of the reaction solution can be maintained, for example, whereby prevention of contamination can be achieved.

Patent Citation 1: JP 2005-58107 A

Disclosure of Invention

[0006] However, comparing a method in which an amplification reaction is conducted in the absence of a probe and a T_m analysis is conducted by adding the probe to the resultant reaction solution and a method in which an amplification reaction is conducted in the presence of a probe and a T_m analysis is conducted using the resultant reaction solution, it was found that, for example, even if the same template nucleic acid and probe are used, the following problem may arise in the latter method. Specifically, the problem is that, even in the case where a target nucleic acid sequence with a known polymorphism is used, i.e., it is known whether the peak is found at either the $T_{m_{wild}}$ value or the $T_{m_{mut}}$ value, it is difficult to distinguish the peak because it is too small, or the peak is not found. Especially, in the case of a sample known to be heterozygous, it is theoretically considered that peaks comparable to each other are seen at both the $T_{m_{wild}}$ value and the $T_{m_{mut}}$ value. However, in fact, there is a case where a peak of a target sequence perfectly complementary to the probe (for example, in the case where the mutant-type probe is used, it is a peak at the $T_{m_{mut}}$ value) is hardly distinguishable.

[0007] Hence, the present invention is intended to provide an amplification method that can amplify a target nucleic acid sequence even in the presence of a probe, a method for suppressing amplification inhibition caused by the probe, a method for analyzing a target nucleic acid sequence, and the probe used therefor.

[0008] In order to achieve the aforementioned object, the amplification method of the present invention is a method for amplifying a target nucleic acid sequence in a template nucleic acid in the presence of a probe, including the step of: amplifying the target nucleic acid sequence by an elongation reaction from a primer in the presence of a probe that can hybridize to the target nucleic acid sequence. In this method, used as the probe is a probe that forms, with a strand complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction. In the present invention, hereinafter, a target nucleic acid sequence also is referred to as a target sequence.

[0009] The method for suppressing amplification inhibition of the present invention is a method for suppressing, in amplification of a target nucleic acid sequence in a template nucleic acid in the presence of a probe, inhibition of the amplification of the target nucleic acid sequence, wherein a method for amplifying the target nucleic acid sequence is the amplification method of the present invention.

[0010] The analysis method of the present invention is a method for analyzing a target nucleic acid sequence using a probe that can hybridize to the target nucleic acid sequence, including the steps of: (A) amplifying the target nucleic acid sequence in a template nucleic acid in the presence of the probe by the amplification method according to the present invention; and (B) after the step (A), measuring a signal value showing a molten state of a double-stranded nucleic acid composed of a resultant amplification product and the probe while changing a temperature of a reaction solution in the step (A).

[0011] The probe of the present invention is a probe used for the amplification method of the present invention. The probe is a probe that forms, with a strand complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction.

[0012] The inventors of the present invention conducted earnest studies to elucidate the cause of the above-described problem in the case where an amplification reaction is conducted in the presence of the probe and a T_m analysis is conducted using the resultant reaction solution. As a result of the studies, they found a possibility that a probe previously added to the reaction solution before the amplification reaction hybridizes to a template nucleic acid at the time of the amplification reaction, thereby inhibiting, for example, annealing of a primer or elongation from the annealed primer. It is considered that, if annealing of a primer to the template nucleic acid or elongation from the primer is inhibited, even when the target sequence is present, the amplification thereof does not occur sufficiently, thereby causing a problem of a peak being difficult to detect. Based on this finding, the inventors found that the problem can be avoided by designing a probe so that the T_m value thereof and the elongation reaction temperature satisfy the relationship described above. This is for the following reasons.

[0013] In a polymorphism analysis using a probe, in order to enhance the accuracy of the analysis, a probe is usually designed so as to hybridize specifically to a target sequence. In order to allow a probe to hybridize specifically to a target sequence, a generally employed approach is to make a probe relatively long. However, as the probe becomes relatively long, the T_m value thereof (i.e., for example, the T_m value of a double-stranded nucleic acid composed of the probe and a strand perfectly complementary to the probe) becomes relatively high, so that, once a double strand is formed, the probe is difficult to dissociate from a template unless a higher temperature is applied thereto. The inventors found that the conventional problem of difficulty in detecting a peak is caused by the fact that a probe that is designed so as to specifically hybridize to the target nucleic acid for a T_m analysis of a polymorphism after nucleic acid amplification hybridizes to the template nucleic acid at the time of nucleic acid amplification, thereby inhibiting annealing of a primer and an elongation reaction from the annealed primer. Thus, in order to make a probe less prone to hybridize to a template nucleic acid at least in an elongation reaction in nucleic acid amplification, the base sequence of the probe is set so that the T_m value thereof, i.e., the T_m value of a double-stranded nucleic acid composed of the probe and a strand complementary to the probe (for example, a strand perfectly complementary to the probe), is equal to or lower than the elongation reaction temperature, thereby achieving the present invention. According to the present invention, even if amplification of a target sequence is conducted in the presence of a probe, the probe is less prone to hybridize to a template nucleic acid at a reaction temperature of the elongation reaction. Therefore, nucleic acid amplification of a target sequence hardly is inhibited by the probe, specifically, for example, annealing of a primer to a template nucleic acid and an elongation reaction from the primer are hardly inhibited. As a result, it becomes possible to amplify the target sequence sufficiently. In particular, comparing the T_m value of a double-stranded nucleic acid composed of the probe and a strand perfectly complementary to the probe and the T_m value of a double-stranded nucleic acid composed of the probe and a strand complementary to the probe except for a single base, the former T_m value is higher than the latter T_m value. Thus, conventionally, it is considered that a target sequence perfectly complementary to a probe is more susceptible to influence by the probe than a target sequence complementary to the probe except for a single base. However, according to the probe of the present invention, such a problem is avoided and hence, for example, even in the case of a heterozygous template nucleic acid as described above, it becomes possible to achieve the same amplification efficiency and also, it becomes possible to detect peaks in a T_m analysis. Therefore, according to the present invention, for example, in a T_m analysis, the presence or absence of a peak can be distinguished with higher reliability than ever before. Consequently, it can be said that the present invention is very useful technology especially in the field of gene analysis.

[0014] The problem caused in the T_m analysis in the case where a nucleic acid amplification is conducted in the presence of a probe and the cause of the problem are newly found by the inventors. Further, although there has been a probe designed so that the T_m value is relatively high in order to allow the probe to specifically hybridize to a target sequence, an attempt to design a probe considering the temperature relation between the T_m value and the reaction temperature of the elongation reaction in order to avoid the influence of the probe on nucleic acid amplification is made first in the present invention.

Brief Description of Drawings

[0015]

[FIG. 1] FIG. 1 is a graph showing a result of a T_m analysis in a comparative example.

[FIG. 2] FIG. 2 is a graph showing a result of a T_m analysis in an example of the present invention.

[FIG. 3] FIG. 3 is a graph showing another result of a T_m analysis in an example of the present invention.

Best Mode for Carrying out the Invention

<Amplification method>

[0016] The amplification method of the present invention is, as described above, a method for amplifying a target nucleic acid sequence in a template nucleic acid in the presence of a probe, including the step of: amplifying the target nucleic acid sequence by an elongation reaction from a primer in the presence of the probe described above. The probe used in this method is the probe of the present invention, i.e., a probe that forms, with a strand complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction. The probe used in the present invention is, hereinafter, referred to as a probe of the present invention. In an amplification method of the present invention, the probe preferably is a probe that forms, with a strand perfectly complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction.

[0017] By conducting a nucleic acid amplification reaction in the presence of the probe of the present invention, as described above, for example, amplification of a target sequence can be conducted while suppressing annealing of a

primer to a template nucleic acid and inhibition of elongation from the annealed primer. It is to be noted that the amplification method of the present invention is characterized by using a probe that satisfies the condition described above, and types, conditions, and the like of the nucleic acid amplification method are not limited at all.

[0018] In the present invention, in order to identify a probe to be used, the probe is defined as having a property such that, when a double-stranded nucleic acid composed of the probe and a strand perfectly complementary to the probe is formed, "a melting temperature of the double-stranded nucleic acid composed of the probe and the strand perfectly complementary to the probe is equal to or lower than a reaction temperature of the elongation reaction". That is, the "perfectly complementary strand" merely is a condition to identify the property of a probe to be used, and it is not required that the probe forms a double-stranded nucleic acid with the perfectly complementary strand at the time of use. As described above, for example, it is not necessary that the present invention includes a step of forming a double-stranded nucleic acid from the probe and a strand perfectly complementary to the probe. In addition, for example, in a method for analyzing a target nucleic acid sequence to be described later and the like, when a double strand composed of an amplification product after an amplification reaction and the probe is formed, an object to which the probe hybridizes is not limited to the strand perfectly complementary to the probe, and the probe may hybridize to a complementary strand with a single-base mismatch or multiple-base mismatches.

[0019] In the present invention, hereinafter, the state of being perfectly complementary to the probe is referred to as "perfect match", a complementary strand perfectly matching to the probe is referred to as "perfectly matched complementary strand", a double-stranded nucleic acid composed of the probe and the perfectly matched complementary strand is referred to as "double-stranded nucleic acid of perfect match", and a T_m value of the double-stranded nucleic acid of perfect match is referred to as " $T_{m_{Per}}$ value". The state of being perfectly complementary to the probe except for a single base or multiple bases is referred to as "mismatch", a complementary strand mismatching the probe is referred to as "mismatched complementary strand", a double-stranded nucleic acid composed of the probe and the mismatched complementary strand is referred to as "double-stranded nucleic acid of mismatch", and a T_m value of the double-stranded nucleic acid of mismatch is referred to as " $T_{m_{Mis}}$ value". In addition, in the present invention, a sequence amplified by a nucleic acid amplification method may be a target sequence and a sequence complementary thereto, or may be a sequence composed of a region including the target sequence and a sequence complementary thereto.

[0020] The probe of the present invention may be a probe designed such that, for example, the relation between the $T_{m_{Per}}$ value of the double-stranded nucleic acid of perfect match and the elongation reaction temperature satisfies the condition " $T_{m_{Per}}$ value $\leq$ elongation reaction temperature ($^{\circ}\text{C}$)", and a base sequence of the probe or the number of bases in the same is not limited at all. Nucleic acid amplification generally includes three steps, namely, melting of a double-stranded nucleic acid (disassociation to single-stranded nucleic acids), annealing of primers to the single-stranded nucleic acids, and an elongation reaction of complementary strands from the annealed primers. The annealing is a precondition of the elongation reaction, and in the nucleic acid amplification, generally, the temperature of annealing and the reaction temperature of an elongation reaction are set to be the same. In this case, "elongation reaction temperature" in the present invention can be replaced with the annealing temperature, for example.

[0021] For example, when the reaction temperature of an elongation reaction has been set to a particular value, the base sequence of the probe of the present invention can be designed so that the T_m value is equal to or lower than the elongation reaction temperature. The method for designing a probe taking the T_m value into account is not particularly limited, and a conventionally known method can be adopted. The sequence of the probe or the T_m value can be determined by the conventionally known MELT-CALC software (<http://www.meltcalc.com/>), the conventionally known Nearest Neighbor Method, or the like. With respect to the probe designed using the foregoing software or the like, the $T_{m_{Per}}$ value is preferably checked, for example, under the actual condition of conducting nucleic acid amplification. Further, the probe of the present invention may be adapted to satisfy the condition of the present invention by setting the elongation reaction temperature depending on the $T_{m_{Per}}$ value of the probe, instead of determining the base sequence depending on the elongation reaction temperature as described above. That is, with respect to the probe of the present invention, it is only necessary that the $T_{m_{Per}}$ value of the probe and the elongation reaction temperature eventually satisfy the formula " $T_{m_{Per}}$ value $\leq$ elongation reaction temperature", and the means for allowing the formula to be satisfied may be either adjusting the base sequence of the probe or adjusting the elongation reaction temperature.

[0022] The $T_{m_{Per}}$ value of a double-stranded nucleic acid composed of the probe and the perfectly matched complementary strand may satisfy the condition " $T_{m_{Per}}$ value $\leq$ elongation reaction temperature". The difference between the $T_{m_{Per}}$ value and the elongation reaction temperature is, for example, 0°C or higher (about 0°C or higher). The upper limit of the difference is not particularly limited, and is, for example, preferably from 0°C to 30°C , more preferably from 0°C to 20°C , still more preferably from 0°C to 10°C , and particularly preferably from 0°C to 2°C .

[0023] Since the homology of the mismatched complementary strand to the probe is lower than that of the perfectly matched complementary strand, the $T_{m_{Mis}}$ value is lower than the $T_{m_{Per}}$ value ($T_{m_{Mis}}$ value $< T_{m_{Per}}$ value $\leq$ elongation reaction temperature ($^{\circ}\text{C}$)). With respect to the probe of the present invention, the difference between the $T_{m_{Per}}$ value of a double-stranded nucleic acid composed of the probe and the perfectly matched complementary strand and the $T_{m_{Mis}}$ value of a double-stranded nucleic acid composed of the probe and the mismatched complementary strand is,

for example, preferably 1°C or higher (about 1°C or higher), more preferably 3°C or higher (about 3°C or higher), and particularly preferably 5°C or higher (about 5°C or higher). By providing enough difference between the $T_{m_{Per}}$ value and the $T_{m_{Mis}}$ value, for example, when a melting profile to be described later is prepared, a peak of the $T_{m_{Per}}$ value and a peak of the $T_{m_{Mis}}$ value can be determined easily, whereby the accuracy of analysis is enhanced further.

[0024] As long as the $T_{m_{Per}}$ value of the probe and the elongation reaction temperature satisfy the relation described above, the length of the probe is not particularly limited. It is preferable that the sequence of the probe is relatively long because the probe can hybridize to a target sequence more specifically. The length is, for example, from 5 to 50 bases, and preferably from 10 to 30 bases.

[0025] The probe of the present invention may be, for example, a non-labeled probe, or a labeled probe labeled with a labeling substance, and specifically, the labeled probe is preferred. According to such a labeled probe, for example, in the method for analyzing a target sequence to be described later, signals of the labeling substance can be measured as signal values showing molten states of a double-stranded nucleic acid composed of the amplification product and the probe.

[0026] Examples of the labeling substance include fluorescent substances such as a fluorescence dye, a fluorophore, and the like. As a specific example of the labeled probe, for example, a probe labeled with the fluorescent substance, showing fluorescence independently and exhibiting reduced (for example, quenched) fluorescence by hybrid formation is preferred. A probe utilizing such a quenching phenomenon is generally called a fluorescence quenching probe. Among them, preferred is the probe labeled with the fluorescent substance at the 3' terminal of a polynucleotide or the 5' terminal of the same, and the terminal base that is labeled is preferred to be C. In this case, preferably, the base sequence of the labeled probe is designed so that, in a base sequence to which the labeled probe hybridizes, a base to be paired with the terminal base C of the labeled probe or a base that is 1 base to 3 bases away therefrom is G. The foregoing probe is generally called a guanine quenching probe, and it is known as the so-called QProbe (registered trademark). When such a guanine quenching probe hybridizes to a base sequence, the phenomenon that light emission of the fluorescent substance becomes weak (the fluorescence intensity is reduced) as the terminal C labeled with a fluorescent substance approaches the G in the base sequence is occurs. By using such a probe, hybridization and disassociation can be recognized easily by variations in signal.

[0027] The fluorescent substance is not particularly limited, and examples thereof include fluorescent dyes such as fluorescein, phosphor, rhodamine, polymethine dye derivatives, and the like. Examples of a commercially available fluorescent substance include fluorescent dyes such as BODIPY FL (trademark, produced by Molecular Probes, Inc.), FluorePrime (product name, produced by Amersham Pharmacia Biotech, Inc.), FluoreDite (product name, produced by Millipore Corporation), FAM (produced by ABI Inc.), Cy 3 and Cy 5 (produced by Amersham pharmacia), TAMRA (produced by Molecular Probes, Inc.), and the like.

[0028] When two or more types of the probes of the present invention are used in one reaction system, for example, the probes are preferably labeled with different fluorescent substances from each other (fluorescent substances detected at different wavelengths from each other). The combination of the fluorescent substances is not particularly limited as long as, for example, the fluorescent substances can be detected under different conditions from each other. Examples thereof include the combination of Pacific Blue (the detection wavelength is from 450 to 480 nm), TAMRA (the detection wavelength is from 585 to 700 nm), and BODIPY FL (the detection wavelength is from 515 to 555 nm), and the like.

[0029] Further, in a labeled probe labeled with a fluorescent dye at the 5' terminal, for example, a phosphate group further may be added to the 3' terminal to prevent elongation of the probe itself in an amplification reaction.

[0030] In the present invention, the nucleic acid amplification method is not particularly limited, and examples thereof include a polymerase chain reaction (PCR) method, a nucleic acid sequence based amplification (NASBA) method, a transcription-mediated amplification (TMA) method, a strand displacement amplification (SDA) method, and the like. Among them, the PCR method is preferred. Hereinafter, the present invention will be explained with the PCR method as an example. However, the present invention is not limited to this example.

[0031] A sample to which the present invention is applied is not particularly limited as long as it contains a nucleic acid serving as a template. Specific examples thereof include: whole blood, somatic cells such as oral cells (for example, oral mucosa), nails, hair, and the like; germ cells; sputa; amniotic fluid; paraffin-embedded tissues, urine, gastric juice (for example, gastric lavage fluid), and the like, suspensions thereof, and the like.

[0032] The proportion of the sample added in a reaction solution for nucleic acid amplification is not particularly limited. As a specific example, when the sample is a biological sample (for example, whole blood sample), the lower limit of the proportion thereof in the reaction solution is preferably 0.01% by volume or more, more preferably 0.05% by volume or more, and still more preferably 0.1% by volume or more. Further, the upper limit of the proportion thereof also is not particularly limited, and is, for example, preferably 2% by volume or less, more preferably 1% by volume or less, and still more preferably 0.5% by volume or less.

[0033] Further, as will be described later, when optical detection is conducted with respect to a reaction solution of nucleic acid amplification, especially when an optical detection using a labeled probe is conducted, preferably, the proportion of a biological sample such as a whole blood sample or the like added in the reaction solution is, for example,

set to from 0.1% to 0.5% by volume. In a PCR, usually a heat treatment is conducted to denature a nucleic acid (disassociation to single-stranded nucleic acids). However, there is a risk that, by this heat treatment, sugars, proteins, and the like contained in the sample are denatured, resulting in the generation of insoluble deposits, turbidity, or the like. Therefore, when the presence or absence of an amplification product or a polymorphism of the same is checked by an optical method, the generation of insoluble deposits or turbidity as described above may influence the accuracy of measurement. However, when the proportion of a whole blood sample added in a reaction solution is set in the range described above, for example, the influence of generated insoluble deposits or the like caused by the denaturation can be prevented sufficiently, although the mechanism thereof is unknown, whereby the accuracy of the measurement by an optical method can be enhanced. Further, since the inhibition of a PCR by foreign substances in the whole blood sample also is suppressed sufficiently, amplification efficiency can be enhanced further.

[0034] Further, the proportion of a whole blood sample in the reaction solution also can be expressed as percentage by weight of hemoglobin (hereinafter referred to as "Hb"), instead of percentage by volume as described above (for example, from 0.1% to 0.5% by volume). In this case, the proportion of the whole blood sample in the reaction solution is converted to the amount of Hb, which is, for example, preferably in a range of from 0.565 to 113 g/L, more preferably in a range of from 2.825 to 56.5 g/L, and still more preferably in a range of from 5.65 to 28.25 g/L. Note here that the proportion of a whole blood sample added in the reaction solution may satisfy both the above-described percentage by volume and the percentage by weight of Hb or either one of them, for example.

[0035] Whole blood may be, for example, any of hemolyzed whole blood, non-hemolyzed whole blood, anticoagulant whole blood, whole blood including coagulation fraction, and the like.

[0036] In the present invention, a template nucleic acid contained in a sample is, for example, DNA. The DNA may be, for example, DNA inherently contained in a sample such as a biological sample or the like or amplification product DNA obtained through amplification by nucleic acid amplification. In the latter case, the amplification product DNA may be cDNA produced by a reverse transcription reaction from RNA (total RNA, mRNA, and the like) that is contained inherently in the sample.

[0037] In the method for producing an amplification product of the present invention, for example, prior to the initiation of nucleic acid amplification, albumin preferably is further added to the reaction solution. By such addition of albumin, for example, the influence of generated insoluble deposits or turbidity described above can be reduced further, and amplification efficiency also can be enhanced further.

[0038] The proportion of albumin added in the reaction solution is, for example, from 0.01% to 2% by weight, preferably from 0.1% to 1% by weight, and more preferably from 0.2% to 0.8% by weight. The albumin is not particularly limited, and examples thereof include bovine serum albumin (BSA), human serum albumin, rat serum albumin, horse serum albumin, and the like. These albumins may be used alone or in a combination of two or more of them.

[0039] Next, the method for producing an amplification product of the present invention will be explained with reference to an example where, with respect to a whole blood sample, amplification of a target nucleic acid is conducted using DNA as a template nucleic acid through a PCR using a primer for amplifying the target nucleic acid in the presence of the probe of the present invention. It is to be noted that the present invention is characterized by using the probe of the present invention, and other configurations and conditions are not limited at all.

[0040] First, a PCR reaction solution is prepared. Although the proportion of a probe added in the reaction solution is not particularly limited, it is preferable that the probe is added to the reaction solution so that its concentration is in a range of 10 to 400 nmol/L, more preferably in a range of 20 to 200 nmol/L. When a fluorescent substance is used as a label of a probe, a non-labeled probe having the same sequence as that of the labeled probe may be used in combination in order to adjust the fluorescence intensity to be detected, for example. In this non-labeled probe, phosphoric acid may be added to the 3' terminal. In this case, it is preferable that the molar ratio of the labeled probe to the non-labeled probe is, for example, from 1 : 10 to 10 : 1.

[0041] Although the proportion of a primer added in the reaction solution is not particularly limited, it is preferable to add the primer so that its concentration is from 0.1 to 2 $\mu\text{mol/L}$, more preferably from 0.25 to 1.5 $\mu\text{mol/L}$, and particularly preferably from 0.5 to 1 $\mu\text{mol/L}$. Further, as the primer, a paired primer set composed of a forward primer and a reverse primer can be used, and in this case, both the primers preferably are added so that their concentrations are in the range described above. The ratio between the forward primer (F) and the reverse primer (R) added (molar ratio F : R) is not particularly limited, and is, for example, preferably from 1 : 0.25 to 1 : 4, more preferably from 1 : 0.5 to 1 : 2. One type or two or more types of primer or primer set may be used, for example. In the present invention, the primer can be used as appropriate depending on the number of target sequences that are desired to be amplified in one reaction solution, for example.

[0042] Although the proportion of a whole blood sample in the reaction solution is not particularly limited, it is preferred to be in the range described above. The whole blood sample may be added to the reaction solution as it is or may be added to the reaction solution after being diluted with a solvent such as water, a buffer solution, or the like. When the whole blood sample is previously diluted, the dilution factor is not particularly limited, and for example, it can be set so that the final proportion of the added whole blood in a reaction solution is in the range described above. The dilution

factor is, for example, from 100-fold to 2000-fold, and preferably from 200-fold to 10000-fold.

[0043] Other composition components in the reaction solution are not particularly limited and may be conventionally known components. The proportions of the components are also not particularly limited. Examples of the composition components include DNA polymerase, nucleotide (nucleoside triphosphate (dNTP)), and a solvent. Further, as described above, it is preferable that the reaction solution further contains albumin. It is to be noted that, in the reaction solution, the order of adding each composition component is not limited at all.

[0044] The DNA polymerase is not particularly limited, and for example, conventionally known thermotolerant bacterium-derived polymerase can be used. As specific examples, *Thermus aquaticus*-derived DNA polymerase (USP. Nos. 4,889,818 and 5,079,352) (product name "Taq polymerase"), *Thermus thermophilus*-derived DNA polymerase (WO/1991/009950) (rTth DNA polymerase), *Pyrococcus furiosus*-derived DNA polymerase (WO/1992/009689) (Pfu DNA polymerase: produced by Stratagene), *Thermococcus litoralis*-derived DNA polymerase (EP-A 455 430) (trademark "Vent": produced by New England Biolabs Inc.), *Thermococcus kodakaraensis*-derived DNA polymerase (product name "KOD DNA polymerase"), and the like are commercially available. Among them, *Thermus aquaticus*-derived thermotolerant DNA polymerase is preferred.

[0045] The proportion of DNA polymerase added in the reaction solution is not particularly limited, and is, for example, from 1 to 100 U/mL, preferably from 5 to 50 U/mL, and more preferably from 20 to 30 U/mL. With regard to the unit of activity (U) of DNA polymerase, 1 U generally is defined as an activity for incorporating 10 nmol of an entire nucleotide into acid-insoluble precipitate per 30 minutes at 74°C in a reaction solution for activity measurement using activated salmon sperm DNA as a template primer. The composition of the reaction solution for activity measurement is, for example, 25 mmol/L TAPS buffer (pH 9.3, 25°C), 50 mmol/L KCL, 2 mmol/L MgCl₂, 1 mmol/L mercaptoethanol, 200 μmol/L dATP, 200 μmol/L dGTP, 200 μmol/L dTTP, 100 μmol/L [α -³²P] dCTP, and 0.25 mg/mL activated salmon sperm DNA.

[0046] The nucleotide triphosphate may be generally dNTP (dATP, dCTP, dGTP, dTTP, and/or dUTP). The proportion of dNTP added in the reaction solution is not particularly limited, and is, for example, from 0.01 to 1 mmol/L, preferably from 0.05 to 0.5 mmol/L, and more preferably from 0.1 to 0.3 mmol/L.

[0047] Examples of the solvent include buffer solutions such as Tris-HCl, Tricine, MES, MOPS, HEPES, CAPS, and the like, and commercially available buffer solutions for PCR or buffer solutions of commercially available PCR kits can be used.

[0048] Further, the PCR solution further may contain glycerol, heparin, betaine, KCl, MgCl₂, MgSO₄, DMSO, and the like, and the proportions of these components added may be set, for example, within a range in which the PCR is not inhibited.

[0049] The total volume of the reaction solution is not particularly limited and can be, for example, determined as appropriate depending on a device to be used (thermal cycler), a container such as a tube or a chip etc. However, the total volume is generally from 1 to 500 μL and preferably from 10 to 100 μL.

[0050] Next, a PCR is conducted. Generally, a PCR has three steps: (1) disassociation of a double-stranded nucleic acid to single-stranded nucleic acids, (2) annealing of primers to the single-stranded nucleic acids, and (3) elongation of the annealed primers (polymerase reaction), as one cycle. The temperature condition in each step is not particularly limited as long as, as described above, the elongation reaction temperature in the step (3) is higher than the T_m value of the probe. As the conditions for conducting the respective steps, the following conditions can be exemplified: for the dissociation step (1), for example, at 90°C to 99°C for 1 to 120 seconds, preferably at 92°C to 95°C for 1 to 60 seconds; for the annealing step (2), for example, at 40°C to 70°C for 1 to 300 seconds, preferably at 50°C to 70°C for 5 to 60 seconds; and for the elongation step (3), for example, at 50°C to 80°C for 5 to 60 seconds, preferably at 50°C to 75°C for 5 to 60 seconds. However the conditions for conducting the respective steps are not limited thereto.

[0051] The number of cycles in the PCR is not particularly limited, and preferably is, for example, 30 cycles or more. The upper limit of the number of cycles is not particularly limited, and is, for example, a total of 100 cycles or less, preferably a total of 70 cycles or less, and more preferably a total of 50 cycles or less.

[0052] In the above-described manner, a target sequence can be amplified in the presence of the probe of the present invention. According to this method, as described above, since annealing of a primer to a template nucleic acid and elongation from the annealed primer is hardly inhibited by the probe, amplification with higher efficiency than ever before is achieved even in either case where the target sequence contains a perfectly matched complementary strand to the probe or where the target sequence contains a mismatched complementary strand to the probe. The amplification method of the present invention also can be referred to as the method for producing an amplification product of a target sequence.

[0053] The amplification method of the present invention further may include the step of: detecting an amplification product of a target nucleic acid obtained by an amplification reaction described above. This allows the presence or absence of the amplification product or a polymorphism in the target sequence to be detected. One aspect of this embodiment will be explained below as a method for analyzing a target sequence of the present invention.

<Method for analyzing target sequence>

[0054] The method for analyzing a target sequence of the present invention is, as described above, a method for analyzing a target nucleic acid sequence using a probe that can hybridize to the target nucleic acid sequence, including the steps of: (A) amplifying the target nucleic acid sequence in a template nucleic acid in the presence of the probe by the amplification method of the present invention; and (B) after the step (A), measuring a signal value showing a molten state of a double-stranded nucleic acid composed of a resultant amplification product and the probe while changing a temperature of a reaction solution in the step (A).

[0055] According to the method for analyzing a target sequence of the present invention, for example, the presence or absence of amplification of the target sequence or a polymorphism in the target sequence can be distinguished. It is to be noted that the analysis method of the present invention is characterized by amplifying a target sequence by the amplification method of the present invention, i.e., amplifying a target sequence in the presence of the probe of the present invention, and for example, other configurations or conditions are not limited at all.

[0056] Preferably, the method for analyzing a target sequence of the present invention further includes the following step (C), for example. By checking variations in signal value with changes in temperature of the reaction solution as in the step (C), the presence or absence of amplification of the target sequence or a polymorphism can be distinguished as described above. (C) analyzing the target sequence from a variation in the signal value with a change in temperature

[0057] In the present invention, the step (C) is preferably a melting profile analysis using a melting profile showing variations in signal value with changes in temperature. A melting profile is generally a graph showing the relationship between each temperature and a signal value showing the molten state of a sample or a differential value of the signal value (hereinafter referred to as "signal differential value"). By using such a melting profile, for example, the presence or absence of a peak at the $T_{m_{per}}$ value and the $T_{m_{mis}}$ can be determined easily by visual observation, for example.

[0058] The melting profile is preferably a graph showing the relationship between the temperatures and the signal differential values because the amount of change in signal is easy to determine. The signal differential value may be represented by, for example, "dF/dT" or "-dF/dT". The dF shows the amount of change in signal value, and the dT shows the amount of change in time. When the generation of signals is suppressed by melting of a sample, a valley-shaped peak is shown in the melting profile representing the signal differential values by "dF/dT", and a mountain-shaped peak is shown in the melting profile representing the signal differential values by "-dF/dT". On the other hand, when signals are generated by melting of a sample, a mountain-shaped peak is shown in the melting profile representing the signal derivative values by "dF/dT", and a valley-shaped peak is shown in the melting profile representing the signal derivative values by "-dF/dT". The signals may be generated by non-melting of the sample and the generation of the signals may be suppressed by melting of the sample, or on the other hand, the generation of the signals may be suppressed by non-melting of the sample and the signals may be generated by melting of the sample, for example. Regardless of whether signals are generated by either melting or non-melting of a sample and whether signal differential values are represented by either of the formulae described above, the magnitude of the peak can be evaluated by the magnitude of the absolute value of the signal differential value.

[0059] When a polymorphism of the target sequence is analyzed by the analysis method of the present invention, the target sequence is, for example, a sequence having a target site in which a polymorphism occurs, and the probe of the present invention is a sequence complementary to the sequence. In the step (C), the polymorphism in the target site of the target sequence can be analyzed from variations in the signal value. As described above, in the case where the probe is a mutant-type probe, when a peak is detected at around the $T_{m_{per}}$ value (i.e., $T_{m_{mut}}$ value), the polymorphism can be determined as a mutant type, and when a peak is detected at around the $T_{m_{mis}}$ value (i.e., $T_{m_{wild}}$ value), the polymorphism can be determined as a wild type. Further, when a peak is detected at either one of the T_m values, the polymorphism can be determined as a homozygous polymorphism, and when a peak is detected at both of the T_m values, the polymorphism is determined as a heterozygous polymorphism. It is to be noted that, in the present invention, polymorphisms in a target site are referred to as a mutant type and a wild type. However, this is merely for the sake of convenience and by no means limits the present invention.

[0060] In the step (B), the measurement of signal values showing the molten states of a double-stranded nucleic acid composed of the amplification product and the probe may be measurement of the absorbance at 260nm, or measurement of signals of a labeling substance. Preferably, the former is adopted, for example, when the probe of the present invention is a non-labeled probe. Preferably, the latter is adopted, for example, when the probe of the present invention is a labeled probe labeled with a labeling substance. The labeled probe may be, for example, a labeled probe that shows signals independently and shows no signals when it forms a hybrid, or a labeled probe that shows no signals independently and shows signals when it forms a hybrid. The former probe does not show signals when it forms a hybrid (double-stranded nucleic acid) with a target sequence and shows signals when it is released by heating. On the other hand, the latter probe shows signals when it forms a hybrid (double-stranded nucleic acid) with a target sequence, and the signals are reduced (quenched) when the probe is released by heating. Therefore, by detecting signals by this labeling substance under the condition specific to the signals (absorption wavelength and the like), the determination of the progress of

melting of the hybrid and the T_m values, and the like can be achieved as in the case of the measurement of absorbance at 260 nm.

[0061] In the present invention, as described above, it is possible to conduct an analysis of two or more target sequences or a polymorphism analysis with respect to two or more target sites in one target sequence in the same reaction solution. In such cases, to a reaction solution of nucleic acid amplification of the step (A), two or more types of the probes of the present invention may be added previously depending on the target sequences or target sites to be analyzed. When two or more types of the probe are used as above, it is preferable that the probes are respectively labeled with different labeling substances that can be detected under different conditions. As above, by using the different labeling substances, analysis of the respective target sequences or the respective target sites can be achieved even in the same reaction solution by changing a detection condition.

[0062] Next, the analysis method of the present invention will be explained with reference to an example where a labeled probe is used as the probe of the present invention. However, the present invention is not limited to this example. It is to be noted that the analysis method of the present invention can be conducted in the same manner as the amplification method of the present invention unless otherwise stated.

[0063] First, in the same manner as described above, a PCR solution containing a labeled primer and a primer for amplifying a target sequence is prepared, and a PCR is conducted to amplify the target sequence. The reaction solution contains the probe of the present invention, a primer, DNA polymerase, dNTP, a sample containing a template nucleic acid, and the like, for example. Besides these, the reaction solution may contain a variety of additives that can be used for nucleic acid amplification.

[0064] Next, disassociation of the obtained amplification product (for example, disassociation from a double-stranded DNA and the like into single-stranded nucleic acids) and hybridization of the disassociated single-stranded nucleic acid and the labeled probe are conducted using the PCR solution as it is. For example, this can be achieved by changing the temperature of the reaction solution.

[0065] The heating temperature in the disassociation step is not particularly limited as long as the amplification product can be disassociated at the temperature, and is, for example, from 85°C to 95 °C. The heating time is also not particularly limited, and is, generally, from 1 second to 10 minutes, preferably from 1 second to 5 minutes.

[0066] The hybridization of the disassociated single-stranded nucleic acid and the labeled probe can be conducted by, for example, after the disassociation step, lowering the heating temperature in the disassociation step. Although the temperature condition is not particularly limited, when checking whether or not the target sequence is a perfect match to the probe, the temperature is preferably lower than the $T_{m_{Per}}$ value of the labeled probe. When checking whether or not the target sequence is a mismatch to the probe, and further, when checking whether the target sequence is a perfect match or a mismatch, the temperature is preferably lower than the $T_{m_{Mis}}$ value. The hybridization temperature is not particularly limited, and is generally from 40°C to 50 °C, for example.

[0067] Then, the temperature in the reaction solution is changed, and signal values showing the molten states of the formed hybrid composed of the amplification product and the labeled probe are measured. Specifically, for example, the reaction solution (the formed hybrid composed of the single-stranded DNA and the labeled probe) is heated, and variations in signal value with increase in temperature are measured. As described above, when the probe (guanine quenching probe) in which the terminal base C is labeled is used, fluorescence is reduced (or quenched) in the state where the probe hybridizes with a single-stranded DNA, and fluorescence is generated in the state where the probe is disassociated. Therefore, for example, the formed hybrid with reduced (quenched) fluorescence may be heated gradually, and the increase in fluorescence intensity with increase in temperature may be measured.

[0068] A temperature range at the time of measuring variations in fluorescence intensity is not particularly limited. However, for example, the initiation temperature is from room temperature to 85°C, preferably from 25°C to 70°C, and the end temperature is, for example, from 40°C to 105°C. The temperature rising rate is not particularly limited, and is, for example, from 0.1 to 20 °C/sec., preferably from 0.3 to 5 °C/sec.

[0069] Next, the target sequence is analyzed based on the variations in the signal. To this end, with respect to the probe of the present invention, a $T_{m_{Per}}$ value and a $T_{m_{Mis}}$ value are previously determined. Then, based on the fluorescence intensities obtained, the amount of change in fluorescence intensity per unit time at each temperature is calculated, and it is checked whether or not the maximum absolute value is shown either at the $T_{m_{Per}}$ value or the $T_{m_{Mis}}$. Specifically, for example, a melting profile plotting the temperatures and "-d amount of change in fluorescence intensity/dT" is prepared, and it is checked whether or not a valley-shaped peak showing the minimum value is present either at the $T_{m_{Per}}$ value or the $T_{m_{Mis}}$. For example, when a melting profile plotting the temperatures and "-d amount of change in fluorescence intensity/dT" is prepared, it is checked whether or not a mountain-shaped peak showing the maximum value is present either at the $T_{m_{Per}}$ value or the $T_{m_{Mis}}$. As a result of this, if a peak is found at around the $T_{m_{Per}}$ value, the target sequence can be determined as being a perfect match to the probe, and when a peak is found at around the $T_{m_{Mis}}$ value, the target sequence can be determined as being a mismatch to the probe. Further, when the probe is a mutant-type detection probe, it can be determined that a polymorphism is a mutant type in the case of perfect match and that a polymorphism is a wild type in the case of mismatch. On the other hand, when the probe is a wild-type detection

probe, it can be determined that a polymorphism is a wild type in the case of perfect match and that a polymorphism is a mutant type in the case of mismatch.

[0070] Further, when the probe that generates fluorescence in the state of hybridizing with a single-stranded DNA and exhibits reduced (or quenched) fluorescence in the state of being dissociated as opposed to a guanine quenching probe is used, the formed hybrid generating fluorescence may be heated gradually, and the decrease in fluorescence intensity with increase in temperature may be measured.

[0071] Further, in the present invention, instead of raising the temperature of a reaction solution containing the probe and measuring variations in signal with the increase in temperature as described above, for example, variations in signal at the time of forming a hybrid may be measured. That is, at the time of forming a hybrid by lowering the temperature of the reaction solution containing the probe, variations in signal with decrease in temperature may be measured.

[0072] As a specific example, when a labeled probe (for example, guanine quenching probe) that shows signals independently and shows no signals when it forms a hybrid is used, fluorescence is generated in the state where a single-stranded DNA and the probe are dissociated, but the fluorescence is reduced (or quenched) when a hybrid is formed by lowering the temperature. Therefore, for example, the temperature of the reaction solution may be lowered gradually, and the decrease in fluorescence intensity with decrease in temperature may be measured. On the other hand, when a labeled probe that shows no signals independently and shows signals when it forms a hybrid is used, fluorescence is not generated in the state where a single-stranded DNA and the probe are dissociated, but the fluorescence is generated when a hybrid is formed by lowering the temperature. Therefore, for example, the temperature of the reaction solution may be lowered gradually, and the increase in fluorescence intensity with decrease in temperature may be measured.

<Method for suppressing amplification inhibition>

[0073] The method for suppressing amplification inhibition of the present invention is, as described above, a method for suppressing, in amplification of a target sequence in the presence of a probe, inhibition of the amplification of the target sequence, wherein a method for amplifying the target sequence is the amplification method of the present invention. The present invention is characterized by conducting, in amplification of a target sequence in the presence of the probe of the present invention, the amplification method of the present invention, i.e., amplifying the target sequence in the presence of the probe of the present invention, and other configurations and conditions are not limited at all. It is to be noted that the suppression method of the present invention can be conducted in the same manner as the amplification method of the present invention.

<Probe>

[0074] The probe of the present invention is, as described above, a probe used for the amplification method of the present invention or the method for suppressing amplification inhibition of the present invention. The probe is a probe that forms, with a strand complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction. Preferably, the probe forms, with a strand perfectly complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction. A specific explanation therefor is as described in connection with the amplification method of the present invention. The probe of the present invention can be used for the suppression method, the amplification method, and the analysis method of the present invention.

<Amplification reagent>

[0075] An amplification reagent of the present invention is a reagent used for the amplification method, the suppression method, or the analysis method of the present invention and contains the probe of the present invention. It is only necessary that the amplification reagent of the present invention contains the probe of the present invention, and other components are not limited at all. The amplification reagent of the present invention may further contain a primer for amplifying a target sequence, polymerase such as DNA polymerase or the like, dNTP, and the like, for example.

[0076] The form of the amplification reagent of the present invention is not particularly limited, and the reagent may be, for example, a liquid reagent or a dry reagent to be suspended in a solvent before use. Further, the contents of other components such as the probe of the present invention, the primer, and the like are also not particularly limited. The amplification reagent of the present invention may be an amplification kit, and for example, respective reagents may be contained in the same container or different containers. Furthermore, the amplification kit is preferably provided with instructions.

<Method for designing probe>

[0077] A method for designing a probe of the present invention is a method for designing a probe to be used in a method for amplifying a target nucleic acid sequence in a template nucleic acid in the presence of a probe. The probe is designed so that a melting temperature of a double-stranded nucleic acid composed of the probe and a strand complementary to the probe is equal to or lower than a reaction temperature of the elongation reaction. In the present invention, preferably, the probe is designed so that a melting temperature of a double-stranded nucleic acid composed of the probe and a strand perfectly complementary to the probe is equal to or lower than a reaction temperature of the elongation reaction. Specifically, the design method of the present invention can be conducted in the same manner as the method described in connection with the amplification method of the present invention. The present invention also can be referred to as a probe producing method and the method includes the step of designing a probe in a manner described above.

[0078] Next, the present invention will be explained with reference to examples. However, the present invention is not limited to these examples.

Example 1

[0079] From blood of a subject showing a heterozygous CYP2C9*3 polymorphism, a genome was purified using a kit (product name "QIAamp DNA Mini kit", produced by QIAGEN Co., Ltd.). The purified genome was diluted to 1/10 (volume) with TE (10 mmol/L Tris-HCl and 1 mmol/L EDTA). 1 μ L of this diluted purified genome was added to 49 μ L of a PCR reagent of the following composition, and a PCR was conducted using a thermal cycler. The PCR condition was as follows: a treatment at 95°C for 60 seconds, followed by 50 cycles, with a dissociation treatment of a double strand for 1 second and an elongation reaction treatment for 15 seconds as 1 cycle. The temperature of the dissociation treatment was set at 95°C, and the temperature of the elongation reaction treatment (including an annealing treatment) was set at the predetermined temperature (52°C, 58°C, or 60°C). After the PCR, the resultant reaction solution further was treated at 95°C for 1 second and at 40°C for 60 seconds, followed by heating from 40°C to 95°C at a temperature rate of 10°C/3 sec., and the change in fluorescence intensity with time was measured. The measurement wavelength was set at 515 nm to 555 nm (detection of fluorescence dye BODIPY FL).

[0080]

[Table 1]

(PCR reaction solution: unit μ L)

Distilled water	27.6
10x Gene Taq buffer (Mg Free) ⁽¹⁾	5
10% NaN ₃	0.2
5 U/ μ L Gene Taq FP ⁽¹⁾	0.25
80% Glycerol	9.37
100 mmol/L MgCl ₂	0.75
2.5 mmol/L dNTP	4
100 μ mol/L CYP2C9 F-primer	0.5
100 μ mol/L CYP2C9 R-primer	0.2
100 μ mol/L Probe for CYP2C9*3	0.04
20% BSA	1
Total 49 μ L	
(1) Product name "Gene Taq FP", produced by NIPPON GENE CO., LTD.	

[0081]

(Primer set)

CYP2C9*3 F-primer (SEQ ID NO: 1) 5'-gagcccctgcattgcaa-3'

CYP2C9*3 R-primer (SEQ ID NO: 2) 5'-gatactatgaattgggacttcgaa-3'

(Mutant-type detection probe)

EP 2 224 017 A1

CYP2C9*3 Probe (SEQ ID NO: 3) 5'-gggagaagGtcaAGgTatc-(BODIPY FL)-3'

[0082] The $T_{m_{per}}$ value of a hybrid (perfectly matched double strand) composed of a mutant-type detection probe for CYP2C9*3 and a strand perfectly complementary to the probe is 58°C. The $T_{m_{mis}}$ value of a hybrid (mismatched double strand) composed of the probe for the CYP2C9*3 and a strand perfectly complementary to the probe except for a single base in a target site is 53°C.

[0083] The results are shown in FIGs. 1 to 3. FIGs. 1 to 3 are graphs of melting profiles showing changes in fluorescence intensity with increase in temperature. Differential values on the vertical axis show "-d amount of increase in fluorescence intensity/dT" (-dF/dT), and the horizontal axis shows temperatures. FIG. 1 is a graph showing a melting profile in the case where an elongation reaction temperature of the PCR was set at 52°C, FIG. 2 is a graph showing a melting profile in the case where the elongation reaction temperature was set at 58°C, and FIG. 3 is a graph showing a melting profile in the case where the elongation reaction temperature was set at 60°C.

[0084] When the elongation reaction temperature is 52°C, the relationship between the $T_{m_{per}}$ value of the perfectly matched double strand and the elongation reaction temperature is " $T_{m_{per}}$ value > elongation reaction temperature" and does not satisfy the condition of the present invention. Therefore, as shown in FIG. 1, although the peak of the mismatched double strand at 53°C could be identified well, the peak of a perfectly matched double strand at 58°C was very small, and hardly distinguishable. On the other hand, when the elongation reaction temperature is 58°C, the relationship between the $T_{m_{per}}$ value of the perfectly matched double strand and the elongation reaction temperature is " $T_{m_{per}}$ value = elongation reaction temperature", and when the elongation reaction temperature is 60°C, the relationship between the T_m value of the perfectly matched double strand and the elongation reaction temperature is " $T_{m_{per}}$ value < elongation reaction temperature". In both the cases, the relations satisfy the condition of the present invention. Accordingly, as shown in FIGs. 2 and 3, as well as the peak of perfect match at 58°C, the peak of mismatch at 53°C could be detected.

Industrial Applicability

[0085] As described above, according to the present invention, even when a target sequence is amplified in the presence of a probe, the probe is less prone to hybridize to a template nucleic acid at a reaction temperature of the elongation reaction, whereby annealing of a primer to a target nucleic acid or an elongation reaction from the primer is hardly inhibited by the probe. As a result, it becomes possible to amplify the target sequence sufficiently. Therefore, according to the present invention, for example, in a T_m analysis, the presence or absence of a peak can be distinguished with higher reliability than ever before. Consequently, it can be said that the present invention is very useful technology especially in the field of gene analysis.

[Sequence listing]

[0086] TF08044-01. ST25. txt

SEQUENCE LISTING

<110> ARKRAY, Inc.
 5 <120> Methods for amplification of target nucleic acid seaquence and probe
 using thereof
 <130> P27605-WO-EP
 <140> 08865229.2
 <141> 2008-12-25
 10 <150> JP2007/334360
 <151> 2007-12-26
 <160> 3
 15 <170> PatentIn version 3.1
 <210> 1
 <211> 16
 <212> DNA
 <213> Artificial
 20 <220>
 <223> forward primer
 <400> 1
 gagccctgc atgcaa 16
 25 <210> 2
 <211> 26
 <212> DNA
 <213> Artificial
 30 <220>
 <223> reverse primer
 <400> 2
 gatactatga atttggggac ttcgaa 26
 35 <210> 3
 <211> 19
 <212> DNA
 <213> Artificial
 40 <220>
 <223> probe
 <400> 3
 gggagaaggt caaggtatc 19
 45
 50
 55

EP 2 224 017 A1

SEQUENCE LISTING

<110> ARKRAY, Inc.

5 <120> Methods for amplification of target nucleic acid sequence and probe
using thereof

<130> TF08044-01

10 <150> JP2007/334360
<151> 2007-12-26

<160> 3

15 <170> PatentIn version 3.1

<210> 1
<211> 16
<212> DNA
<213> Artificial

20 <220>
<223> forward primer

<400> 1
gagcccctgc atgcaa 16

25 <210> 2
<211> 26
<212> DNA
<213> Artificial

30 <220>
<223> reverse primer

<400> 2
gatactatga atttggggac ttcgaa 26

35 <210> 3
<211> 19
<212> DNA
<213> Artificial

40 <220>
<223> probe

<400> 3
gggagaaggt caaggatc 19

45

50

Claims

1. A method for amplifying a target nucleic acid sequence in a template nucleic acid in the presence of a probe,
comprising the step of:
- 55 amplifying the target nucleic acid sequence by an elongation reaction from a primer in the presence of a probe
that can hybridize to the target nucleic acid sequence,
wherein used as the probe is a probe that forms, with a strand complementary to the probe, a double-stranded

nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction.

2. The amplification method according to claim 1, wherein used as the probe is a probe that forms, with a strand perfectly complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction.
3. The amplification method according to claim 2, wherein a difference between the melting temperature of the double-stranded nucleic acid composed of the probe and the strand perfectly complementary to the probe and the reaction temperature of the elongation reaction is about 0°C or higher.
4. The amplification method according to claim 2, wherein the probe is a probe that forms, with a strand perfectly complementary to the probe except for a single base, a double-stranded nucleic acid whose melting temperature is different from the melting temperature of the double-stranded nucleic acid composed of the probe and the strand perfectly complementary to the probe by about 1°C or higher.
5. The amplification method according to claim 1, wherein the probe is a labeled probe labeled with a labeling substance.
6. The amplification method according to claim 4, wherein the labeling substance is a fluorescent substance.
7. A method for suppressing, in amplification of a target nucleic acid sequence in a template nucleic acid in the presence of a probe, inhibition of the amplification of the target nucleic acid sequence, wherein a method for amplifying the target nucleic acid sequence is the amplification method according to claim 1.
8. A method for analyzing a target nucleic acid sequence using a probe that can hybridize to the target nucleic acid sequence, comprising the steps of:
 - (A) amplifying the target nucleic acid sequence in a template nucleic acid in the presence of the probe by the amplification method according to claim 1; and
 - (B) after the step (A), measuring a signal value showing a molten state of a double-stranded nucleic acid composed of a resultant amplification product and the probe while changing a temperature of a reaction solution in the step (A).
9. The analysis method according to claim 8, further comprising the step of:
 - (C) analyzing the target sequence from a variation in the signal value with a change in temperature.
10. The analysis method according to claim 9, wherein the step (C) is a melting profile analysis using a melting profile showing the variation in the signal value with the change in temperature.
11. The analysis method according to claim 9, wherein the target nucleic acid sequence comprises a target site to be detected where a polymorphism occurs, and in the step (C), the polymorphism in the target site of the target nucleic acid sequence is analyzed from the variation in the signal value.
12. The analysis method according to claim 9, wherein in the step (C), the presence or absence of amplification of the target nucleic acid sequence is analyzed from the variation in the signal value.
13. The analysis method according to claim 8, wherein the probe is a labeled probe labeled with a labeling substance.
14. The analysis method according to claim 13, wherein the labeling substance is a fluorescent substance, and the signal value in the step (B) is a fluorescence intensity.
15. A probe to be used for the amplification method according to claim 1, wherein the probe is a probe that forms, with a strand complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than a reaction temperature of the elongation reaction.
16. The probe according to claim 15, wherein the probe is a probe that forms, with a strand perfectly complementary to the probe, a double-stranded nucleic acid whose melting temperature is equal to or lower than the reaction temperature of the elongation reaction.

EP 2 224 017 A1

17. The probe according to claim 16, wherein a difference between the melting temperature of the double-stranded nucleic acid composed of the probe and the strand perfectly complementary to the probe and the reaction temperature of the elongation reaction is about 0°C or higher.

5 18. The probe according to claim 16, wherein a difference between the melting temperature of the double-stranded nucleic acid composed of the probe and the strand perfectly complementary to the probe and a melting temperature of a double-stranded nucleic acid composed of the probe and a strand perfectly complementary to the probe except for a single base is about 1°C or higher.

10 19. The probe according to claim 15, wherein the probe is a labeled probe labeled with a labeling substance.

20. The probe according to claim 19, wherein the labeling substance is a fluorescent substance.

15 21. A method for designing a probe to be used for a method for amplifying a target nucleic acid sequence in a template nucleic acid in the presence of a probe, wherein the probe is designed so that a melting temperature of a double-stranded nucleic acid composed of the probe and a strand complementary to the probe is equal to or lower than a reaction temperature of an elongation reaction.

20 22. The probe design method according to claim 21, wherein the probe is designed so that a melting temperature of a double-stranded nucleic acid composed of the probe and a strand perfectly complementary to the probe is equal to or lower than the reaction temperature of the elongation reaction.

25

30

35

40

45

50

55

Reaction temperature : 52°C

T_m of probe > Reaction temperature

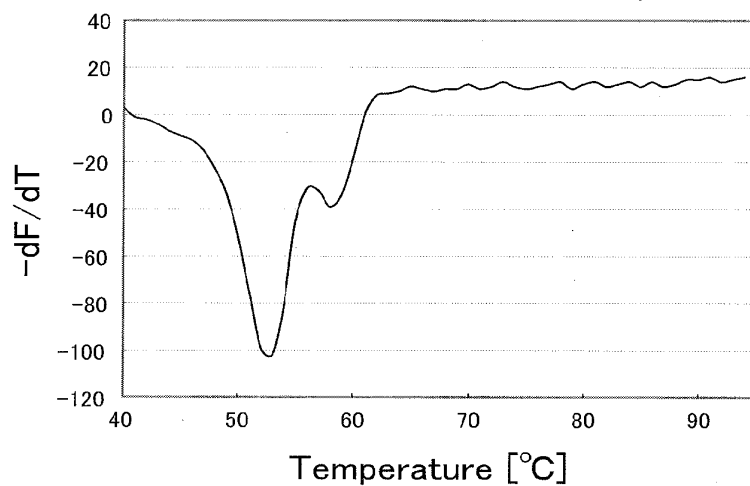


FIG. 1

Reaction temperature : 58°C

Reaction temperature = T_m of probe

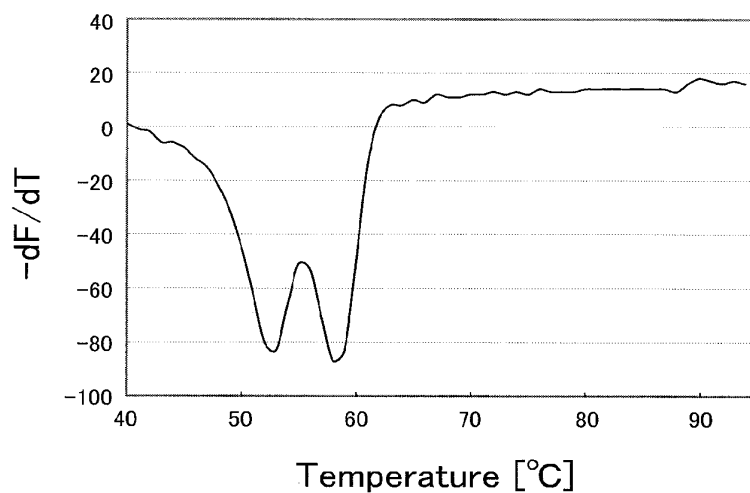


FIG. 2

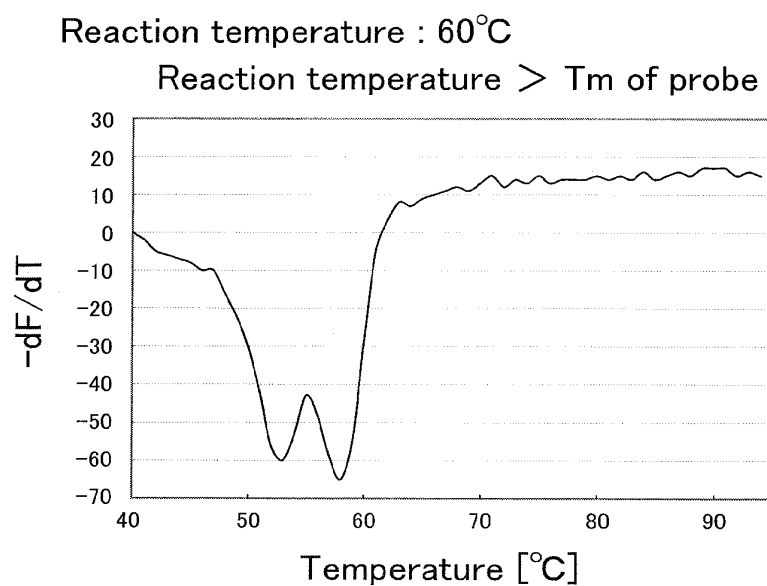


FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/073537

A. CLASSIFICATION OF SUBJECT MATTER C12Q1/68(2006.01) i, C12N15/09(2006.01) i		
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) C12Q1/68, C12N15/00-15/90		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) JSTPlus/JMEDPlus/JST7580 (JDreamII), G-Search		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2005-204515 A (Arkray, Inc.), 04 August, 2005 (04.08.05), (Family: none)	1-22
P, X	WO 2008/066163 A1 (Arkray, Inc.), 05 June, 2008 (05.06.08), (Family: none)	1-22
A	Hyojun Gijutsushu (Kakusan no Zofuku Oyobi Kenshutsu), Heisei 11 Nendo, Tokkyocho Hakko, pages 9 to 10 [Class of Technique] 1-1 Probe Primer to shite Tekishita Hairetsu no Sentaku Hoho [Title of Technique] Tm	1-22
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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 14 January, 2009 (14.01.09)		Date of mailing of the international search report 27 January, 2009 (27.01.09)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

Form PCT/ISA/210 (second sheet) (April 2007)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- JP 2005058107 A [0005]
- US 4889818 A [0044]
- US 5079352 A [0044]
- WO 1991009950 A [0044]
- WO 1992009689 A [0044]
- EP 455430 A [0044]

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 February 2001 (01.02.2001)

PCT
(10) International Publication Number
WO 01/07640 A2

- (51) International Patent Classification⁷: C12Q 1/00 Service, Priory House, Gooch Street North, Birmingham B5 6QQ (GB).
- (21) International Application Number: PCT/GB00/02795
- (22) International Filing Date: 24 July 2000 (24.07.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
9917307.2 23 July 1999 (23.07.1999) GB
0009187.6 14 April 2000 (14.04.2000) GB
- (71) Applicant (*for all designated States except US*): **THE SECRETARY OF STATE FOR THE HOME DEPARTMENT** [GB/GB]; c/o The Forensic Science Service, Priory House, Gooch Street North, Birmingham B5 6QQ (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **GILL, Peter** [GB/GB]; The Forensic Science Service, Priory House, Gooch Street North, Birmingham B5 6QQ (GB). **HUS-SAIN, Javaid** [GB/GB]; The Forensic Science Service, Priory House, Gooch Street North, Birmingham B5 6QQ (GB). **LONG, Adam** [GB/GB]; The Forensic Science
- (74) Agent: **PAWLYN, Anthony, Neil**; Urquhart-Dykes & Lord, Tower House, Merrion Way, Leeds LS2 8PA (GB).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— Without international search report and to be republished upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/07640 A2

(54) Title: IMPROVEMENTS IN AND RELATING TO ANALYSIS OF DNA

(57) Abstract: The invention provides improved techniques for investigating DNA samples, which offers improved sensitivity and specificity. The invention provides a method of investigating single nucleotide polymorphisms in a sample of DNA, the method comprising contacting the DNA containing sample with at least one first set of primers, amplifying the DNA using those primers to give an amplified product, contacting at least a portion of the amplified product with at least one second set of primers, amplifying the DNA using those second set of primers to give a further amplified product and examining one or more characteristics of the further amplified product, one or more of the primers of the first set of primers including a locus specific portion and a further portion, the locus specific portion of one of those one or more of the primers annealing to one side of the SNP under investigation.

IMPROVEMENTS IN AND RELATING TO ANALYSIS OF DNA

This invention concerns improvements in and relating to analysis of DNA, particularly, but not exclusively to techniques using single nucleotide polymorphisms for investigative purposes.

In forensic investigations, and analysis for other purposes, it is known to make use of bi-allelic markers or single nucleotide polymorphisms (SNPs). SNPs represent single base locations where variations between the sequence for one being and another can occur. A SNP may for instance be the presence of G or C, or of A or T, in the sequence of an individual, with some of the individuals having one of the options and other individuals having the other option. By considering a large number of such SNPs at different loci, a set of SNP results for an individual can be obtained which is useful for investigative purposes. The results may be compared with the results from another sample, with the statistical occurrence of that set of results within the population as a whole or used in other ways.

As each SNPs can only vary in being one of two options, a substantial number of different locations, generally several hundred loci, need to be investigated to achieve a set of results which is statistically significant in comparisons or other uses. Analysing such a large number of loci to determine the identity of SNP's on them is highly time consuming if the loci are considered individually, and introduces significant compatibility and reliability problems if multiplexes are used to analyse a number of those loci simultaneously.

The present invention has amongst other aims a technique for improving SNP based investigations, particularly where the original sample levels are very small. Improvements in the specificity of the results and in terms of the ease with which a very large number of such SNPs can be investigated simultaneously are also considered..

According to a first aspect of the invention we provide a method of investigating single nucleotide polymorphisms in a sample of DNA, the method comprising contacting the DNA containing sample with at least one first set of primers, amplifying the DNA using those primers to give an amplified product, contacting at least a portion of the

-2-

amplified product with at least one second set of primers, amplifying the DNA using those second set of primers to give a further amplified product and examining one or more characteristics of the further amplified product.

According to a second aspect of the invention we provide a plurality of primers for investigating single nucleotide polymorphisms in a sample of DNA, the plurality of primers comprising two or more primers of a first set of primers and / or two or more primers of a second set of primers.

The first and / or second aspects of the invention may include any of the features, options or possibilities set out elsewhere in this application.

In one embodiment of the invention one or more, preferably all, of the first sets of primers may include two forward primers and a reverse primer. One or more, preferably all, of the first sets of primers may consist of two forward and a reverse primer. The forward primers and reverse primer preferably include sequences which anneal to the 3' and 5' sides respectively of the SNP at the locus incorporating the SNP under investigation.

In an alternative embodiment of the invention, one or more, preferably all of the first sets of primers may include a forward primer and a reverse primer. One or more, preferably all of the first sets of primers may consist of one forward primer and one reverse primer. The forward primer and reverse primer preferably include sequences which pair / anneal to the 3' and 5' sides respectively of the SNP at the locus incorporating the SNP under investigation.

The first set of primers may include one or more primers including a locus specific portion and a further portion. Preferably the forward primers are so provided. Preferably the further portion is attached to the 5' end of the locus specific portion, particularly in the case of forward primers. The 3' end of the forward primer is preferably provided with a SNP identifying portion. The further portion is preferably attached to the locus specific portion by a SNP related portion.

In one embodiment of the invention the locus specific portion preferably includes a sequence which matches the sequence of the locus sequence in the vicinity of the SNP under investigation. The match may occur at between 2 to 10 bases to the respective sides of the SNP under investigation. More preferably the sequence matches the locus sequence for the locus sequence adjacent to the SNP under investigation, ideally up to and including

-3-

the nucleotide before the SNP on the 3' side of the SNP. Preferably the forward primers of a first set of primers are provided with identical sequences for the locus specific portion.

In one embodiment of the invention the SNP identifying portion is preferably a single nucleotide. The SNP identifying portion may be a C for investigating an SNP where the SNP may be a G nucleotide. The SNP identifying portion may be a G nucleotide for investigating an SNP where the SNP may be a C nucleotide. The SNP identifying portion may be a T nucleotide for investigating an SNP where the SNP may be an A nucleotide. The SNP identifying portion may be an A nucleotide for investigating an SNP where the SNP may be a T nucleotide. Preferably the SNP identifying portion for one forward primer of a set is one of C or G or A or T, with the SNP identifying portion of the other forward primer of the set being one of C or G or A or T, but different from the SNP identifying portion of the first forward primer of the set. Preferably the SNP identifying portions are provided to target the two possible variations of the SNP in question, for instance C and T for the primers to investigate G or A for the SNP, C or G for the primers to investigate G or C for the SNP and so on.

Preferably the SNP identifying portion forms the 3' end of the forward primers of the first set.

An exonuclease digestion prevention unit may be provided towards the 3' end of the forward primers. The exonuclease digestion prevention unit may be phosphorothioate. The exonuclease digestion prevention unit may be provided at the junction of the locus specific portion and SNP identifying portion.

The further portion preferably includes a sequence which does not match the locus sequence on the locus's 3' side of the locus sequence matching the locus specific portion of the primer. More preferably the sequence does not match the sequence of the locus in the vicinity of the SNP under investigation. Ideally the sequence does not anneal to, and particularly does not match, the sequence of any published part, ideally any part, of the entire DNA sequence of the entity from which the DNA containing the SNP under investigation was obtained, for instance Homo Sapiens. The inability of the sequence of the further portion to amplify human DNA is a particularly preferred feature. Preferably the forward primers of a first set of primers are provided with identical sequences for the further portion.

-4-

Preferably the further portion forms the 5' end of the forward primers of the first set.

The further portion of two or more of the forward primers of the first set may have an equivalent sequence. All the forward primers of the first set may be provided with further portions of equivalent sequence.

In a preferred embodiment of the invention, the further portion of at least one of the forward primers of the first set is different from the further portion of at least one of the other forward primers of the first set, at least in part. Preferably the further portion of each forward primer of the first set is different from the further portion of each of the other forward primers of the first set, at least in part. It is preferred that the forward primers are different from one another with respect to at least 25% of the nucleotides forming the further portion of the forward primers. Differences in sequence, ranging between 25% and 100% of the nucleotides forming the further portion of the forward primers may be employed. The differences in sequence may form one or more distinguishing portions. One or more distinguishing portions may be provided as or within the further portion of the forward primers. A distinguishing portion may be provided at the 5' end of the further portion of the forward primer. The distinguishing portion may be provided at the 3' end of the further portion of the forward primer. Preferably the distinguishing portion is provided at an intermediate location within the sequence of the further portion. Preferably a 5' end portion, distinguishing portion and 3' end portion defines the further portion of the forward primers.

The further portion of one or more of the primers in the first set may be provided with one or more portions which correspond with one or more portions in the further portion of one or more of the other primers in the first set. The nucleotides of the further portion of one or more of the forward primers may be equivalent to the nucleotides of one of the other forward primers, outside the distinguishing portion of the further portion. In particular, the 5' end portion and / or 3' portion of the further portion of one or more of the forward primers may be equivalent to the corresponding further portion of one or more of the other forward primers. Preferably all of the forward primers are provided with equivalent 5' end and/or 3' end portions to one another. The equivalent portions may form between 1 and 25% of the sequence of the further portion of the primers. Preferably the

-5-

equivalent portions form between 10 and 25% of the sequence of the further portions. The reverse primer or primers of the first set may be provided with equivalent portions too.

The SNP related portion is preferably a single nucleotide. The SNP related portion is preferably identical to the SNP identifying portion of that primer. Preferably the two forward primers are provided with SNP related portions which are identical with their respective SNP identifying portions. The SNP related portion may be a C for investigating an SNP where the SNP may be a G nucleotide. The SNP related portion may be a G nucleotide for investigating an SNP where the SNP may be a C nucleotide. The SNP related portion may be a T nucleotide for investigating an SNP where the SNP may be an A nucleotide. The SNP related portion may be an A nucleotide for investigating an SNP where the SNP may be a T nucleotide. Preferably the SNP related portion for one forward primer of a set may be one of C or G or A or T, with the SNP related portion of another primer of the set being one of C or G or A or T, but different to the SNP related portion of the first primer of the set. Preferably the SNP related portions for the primers of a set are provided to match the SNP identifying portion of their respective primers.

Preferably during amplification the SNP related portion results in the amplified copies of the locus incorporating the SNP having an SNP repeat introduced into them. Ideally, the repeat has a base identity identical to that of the SNP.

Preferably the locus specific portion and SNP identifying portion of one of the forward primers anneals to the 3' side of the locus having the SNP under investigation. Preferably the locus specific portion and SNP identifying portion of another, ideally the other, of the forward primers does not anneal to the 3' side of the SNP under investigation. Preferably the annealing primer anneals due to a match between the SNP identifying portion and the SNP site, (for instance C matching to G). Preferably the non-annealing primer does not anneal due to a mis-match between the SNP identifying portion and the SNP site, (for instance, T mis-matching with T).

The SNP under investigation may be a location with variation between individuals of any two bases selected from C or G or A or T nucleotides. For instance, the SNP under investigation may be a location with variation between individuals of either a T or A nucleotide, T or C nucleotide, T or G nucleotide, A or C nucleotide, A or G nucleotide or C or G nucleotide. One possible variation may be investigated at one or

-6-

more sites, with one or more other potential variations being investigated at one or more other sites.

Two or more SNP's may be investigated using a simultaneous first amplification and/or simultaneous second amplification and/or simultaneous examination of the one or more characteristic of the further amplified product. Preferably at least the first amplification and second amplification are conducted simultaneously for a plurality of SNP investigations. The number of SNP's investigated simultaneously in one or more stages of the process may be greater than 20, preferably greater than 25, more preferably greater than 50 and ideally greater than 100.

The sample may be a sample of DNA extracted from a collected source.

The sample may be contacted with the first primer set by mixing the sample and primers together.

The sample may be a mixture. One or more contributions to the sample may be analysed as the sample itself using the present invention. The mixed sample may include male and female DNA. One of the sexes of DNA, particularly the male, may be present in low concentrations relative to the other sex. For instance, the minor sex DNA contribution may form less than 1% of the sample, potentially less than 0.1% and even less than 0.05%. The sample may contain samples from two or more sources. The method may investigate the minor sample in a mixture from two or more sources. The minor sample may form less than 1% of the mixed sample, potentially less than 0.1% of the mixed sample and even less than 0.05% of the mixed sample.

The investigation may indicate the amount of DNA in a mixed sample from one or more of the sources. The indication may be based on a comparison of the experimentally determined results, for instance the level of a distinctive unit present, compared with a set of calibration results based on investigation of known amounts of DNA in a sample.

The first amplification is preferably performed by PCR. The amplification preferably involves between 18 to 60 cycles, more preferably 20 to 40 cycles.

The amplification cycles, particularly where the first and second amplification processes are used, may have the following characteristics. Preferably the amplification cycles include a first cycle set in which the annealing temperature of the cycle is similar or above the melting temperature of the first set of primers, particularly of the locus specific

-7-

portion of the first set of primers and / or similar or above the second set of primer. The amplification cycles may include a second set of cycles, with preferably, the annealing temperature in the second set of cycles being similar or below the melting temperature of the first set of primers and / or above the melting temperature of the second set of primers. The melting temperature of the first set of primers may rise after one or two cycles. The amplification cycles may include a third set of cycles, with, preferably, the annealing temperature in the third set of cycles being below the melting temperature of the first set of primers and / or similar or above the melting temperature of the second set of primers.

It is preferred that the first set of cycles provide between 2 and 10 cycles. It is preferred that the second set of cycles provide between 3 and 15 cycles. It is preferred that the third set of cycles provide between 15 and 35 cycles. Preferably the total of cycles provided in the first, second and third sets does not exceed 40 cycles.

It is preferred that the denaturation temperature for the first and / or second and / or third set of cycles be 92 to 96°C, ideally 94°C.

It is preferred that the annealing temperature for the first and / or second and / or third set of cycles be between 60 and 62°C, ideally 61°C. It is preferred that the annealing temperature for the second set of cycles be between 70 and 78°C, ideally between 72 and 75°C.

It is preferred that the extension temperature for the first and / or second and / or third set of cycles be between 70 and 75°C, ideally 72°C.

Amplification preferably results in extension of the annealed forward primer from its 3' end towards the 5' end of the target sequence. Amplification preferably results in extension of the reverse primer from its 3' end towards the 5' end of its target sequence. Preferably further cycles of amplification result in extension of the forward primer sequence towards the 5' end of its target, including the reverse primer sequence. Preferably further cycles of amplification result in extension of the reverse primer sequence towards the 5' end of its target, including one or more or all of the forward primer sequence and particularly the SNP identifying portion, locus specific portion, SNP related portion and further portion.

A portion of the amplified product may be removed and contacted separately with the second set of primers. Contact with the second set of primers may occur in a separate vessel to the contact with the first set of primers. This is particularly preferred where

-8-

universal primers incorporating molecular beacons are used. Preferably a two tube and / or branched PCR process is used where universal primers incorporating molecular beacons are employed.

The first and second amplifications may occur in the same vessel. The first and second amplifications may occur substantially simultaneously. Preferably the method includes adding one or more of the first set of primers and one or more of the second set of primers to the sample to be amplified prior to conducting amplification cycles.

The one or more first sets of primers may be provided at a concentration of between 20 and 80nM, more preferably between 40 and 60nM and ideally at 50nM +/- 5%. Preferably the primers which do not compete and/or for which site overlap does not occur are provided at these levels. Where primer competition could occur and/or where primer site overlap occurs preferably the primer's relative concentrations are balanced. The reverse primer concentration for such a simultaneous process may be between 75nM and 125nM, for instance 100nM +/- 10%.

The second set of primers may be provided at a concentration of between 20 and 80nM, more preferably between 40 and 60nM and ideally at 50nM +/- 5%. The amount of the second set of primers added may be defined by $C_n \times L$, where C_n is the concentration of the primers and L is the number of loci under consideration +/- 2 and ideally is the number of loci under consideration, particularly where L is less than 100 or even less than 50. Preferably the maximum second set of primers concentration is 1000nM.

Particularly where the first and second sets of primers are present together, it is preferred to provide the second set of primers and first set of primers at a concentration ratio of at least 5:1. A ratio of at least 10:1, more preferably at least 20:1 and ideally at least 30:1, second set concentration: first set concentration may be provided. The first set may be provided at a concentration of between 5 and 400nM, more preferably between 10 and 200nM. The second set may be provided at a concentration of between 300nM and 5000nM, more preferably between 400 and 4000nM.

Particularly where the first and second sets of primers are present together, it is preferred to use an annealing temperature at which at least 80% of the second set of primers remain single stranded, more preferably a temperature at which at least 95% of the second set of primers remain single stranded and ideally a temperature at which at least 99% of the second set of primers remain single stranded, for some of the cycles of the

amplification process. A lower annealing temperature may be used for other cycles of the amplification process. Preferably the higher temperature annealing is used at least in cycles 3 to 30, more preferably in cycles 3 to 40. A lower annealing temperature may be used in the first two cycles. A lower annealing temperature is preferably used in at least the last two cycles. The lower annealing temperature is preferably a temperature at which at least 80%, more preferably at least 90% and ideally at least 99% of the second set of primers anneal.

The amplified product may be contacted with the second primer set by mixing the sample and primers together.

The second set of primers may include one, two, three or four forward primers. A reverse primer may be present, but the second set of primers may lack a reverse primer.

The invention may only provide one second set of primers provided.

In one embodiment of the invention preferably the one second set of primers consisting of two forward primers and a reverse primer. One or more, preferably all, of the second sets of primers may include two forward primers and a reverse primer. One of the forward primers of the second set preferably includes a sequence which anneals to the SNP incorporating strand on the 3' side of the SNP. The reverse primer of the second set preferably includes a sequence which anneals to the 3' side of the base pairing to the SNP. More preferably one of the forward primers includes a sequence which anneals to the 3' side of the SNP repeat. Preferably the other forward primer or primers does not anneal.

In the one embodiment of the invention the second set of primers may include one or more primers including a second further portion. Preferably the forward primers are so provided. Preferably the second further portion is provided with a second SNP identifying portion and / or more preferably an SNP repeat identifying portion. The second SNP or SNP repeat identifying portion may be attached to the 3' end of the second further portion, particularly in the case of forward primers. The 5' end of the forward primer is preferably provided with a distinctive unit.

In the one embodiment of the invention the second further portion preferably includes a sequence which pairs to the sequence of the amplified product in the vicinity of the SNP identifying portion and/or, more preferably, SNP repeat related portion thereof. More preferably the further portion sequence adjacent to the SNP related portion, ideally up to and including the nucleotide before the SNP related portion matches the sequence of

-10-

the amplified product adjacent to the SNP repeat, ideally up to and including the nucleotide before the SNP repeat. Preferably the forward primers of a second set of primers are provided with identical sequences for the second further portions.

In an alternative embodiment of the invention it is preferred that the one second set of primers consists of one forward primer and one reverse primer. One or more, preferably all, of the second set of primers may consist of one forward primer and one reverse primer. Preferably the forward primer of the second set includes a sequence which anneals to the SNP incorporating strand on the 3' side of the SNP. Preferably the reverse primer of the second set includes a sequence which anneals to the 3' side of the base pairing to the SNP. Most preferably the forward primer includes a sequence which anneals to the sequence which pairs to the sequence produced by the copying of the further portion of the forward primer and / or which corresponds to the sequence of the further portion of the forward primer of the first set.

In the alternative embodiment of the invention, the second set of primers may include a primer including a second further portion and more preferably consisting of a second further portion. Preferably the forward primer is so provided. Preferably the second further portion is provided with a sequence which pairs to the sequence of the amplified product in the vicinity of the sequence which pairs to the further portion of the forward primer of the first set. More preferably, the second further portion includes a sequence which matches the sequence of the first further portion and / or pairs to the sequence of the amplified product matching the first further portion.

Preferably the sequence of the second further portion does not anneal to, and particularly does not match, the sequence of any published part, ideally any part, of the entire DNA sequence of the entity from which the DNA containing the SNP under investigation was obtained, for instance Homo Sapiens. The inability of the sequence of the second further portion to amplify human DNA is a particularly preferred feature. Preferably the forward primers of a second set of primers are provided with identical sequences for the second further portion.

In the one embodiment of the invention the second SNP related portion is preferably a single nucleotide or two nucleotides.

In the one embodiment of the invention preferably the second SNP related portion of one primer of the second set is, or includes, a nucleotide which is identical to the SNP

-11-

identifying portion and/or SNP related portion of a primer of the first set. Preferably another, ideally the other, primer of the second set has a second SNP related portion which is, or includes, a nucleotide which is identical to the SNP identifying portion and/or SNP related portion of another, ideally the other, primer of the first set.

In the one embodiment of the invention where a single nucleotide forms the second SNP related portion, the second SNP related portion may be a C nucleotide when amplifying a target in which the SNP or SNP repeat is a G nucleotide. The second SNP related portion may be a G nucleotide when amplifying a target in which the SNP or SNP repeat is a C nucleotide. The second SNP related portion may be a T nucleotide when amplifying a target in which the SNP or SNP repeat is an A nucleotide. The second SNP related portion may be an A nucleotide when amplifying a target in which the SNP or SNP repeat is a T nucleotide. The second SNP related portion for one forward primer of a second set may be one of C or G or T or A with the second SNP related portion of another primer of the second set being one of C or G or A or T, but different to the second SNP related portion of the first primer of that set where the SNP or SNP repeat under investigation could be any two of C or G or T or A nucleotides.

In the one embodiment of the invention the second SNP related portion may be formed of two nucleotides. Preferably the end nucleotide of the two matches with the nucleotide of the SNP or SNP repeat of interest. Preferably the nucleotide adjacent to the end nucleotide of the second SNP related portion is a mismatch with the base adjacent to the SNP or SNP repeat in the target sequence.

In the one embodiment of the invention preferably the second SNP related portion forms the 3' end of the forward primers of the second set.

An exonuclease digestion prevention unit may be provided towards the 3' end of the forward primer or primers of the first and / or second set. The exonuclease digestion prevention unit may be phosphorothioate. The exonuclease digestion prevention unit may be provided at the end of the second further portion and / or the junction of the second further portion and second SNP related portion.

Preferably the second further portion and / or second SNP related portion of the forward primer and / or of one of the forward primers anneals to the 3' side of the SNP or SNP repeat. Preferably the second further portion and / or second SNP related portion of another, ideally the other, of the forward primer and / or of the forward primers does not

-12-

anneal to the 3' side of the SNP and / or SNP repeat. In one embodiment of the invention preferably the annealing primer anneals due to a match between the second SNP related portion and the SNP repeat and / or adjacent sequences. Preferably the non-annealing primer does not anneal due to a mis-match between the second SNP related portion and the SNP repeat. In an alternative embodiment of the invention preferably the annealing primer anneals due to a match between the second further portion and a sequence which paired to the first further portion.

The second amplification is preferably performed by PCR. The amplification preferably involves between 18 and 30 cycles, more preferably 20 to 25 cycles.

One or more of the primers of the first and/or second set may be provided with one or more portions which are complimentary to one or more portions on one or more of the other primers in that set. The complimentary portion or portions are preferably provided in the further portion of the primers of the first set. The complimentary portion or portions are preferably provided in the second further portion of the primers of the second set. Preferably a complimentary portion is provided on each of the primers of a set. Preferably at least two complimentary portions are provided on each of the primers of a set. Preferably a complimentary portion is provided at the 3' end of a primer, ideally all the primers. Preferably a complimentary portion is provided at the 5' end of a primer, ideally all of the primers. Preferably the 3' end complimentary portion of one primer is complimentary to the 5' end complimentary portion of another primer, ideally all the other primers of the set and/or both sets. Preferably the 5' end complimentary portion of one primer is complimentary to the 3' end complimentary portion of another primer, ideally all the other primers of the set and/or both sets. A locus specific portion may be provided on the further portion including the complimentary portion or portions, particularly on the 3' end. The further portion and/or second further portion may include a sequence matching the sequence of the locus under consideration, particularly provided between two complimentary portions. The complimentary portions may be at least 3 nucleotides long, more preferably between 3 and 20 nucleotides long. The complimentary portions are preferably both of the same length. The complimentary portions may form between 5 and 40% of the further portion and/or second further portion. One, two, three or four primers of a set may be provided in this way. Preferably the reverse primer or primers are similarly provided.

-13-

The further amplified product, or a portion thereof, may be removed from the vessel in which the amplification is performed to examine the one or more characteristics. Alternatively or additionally, the one or more characteristics may be examined with the further amplified product in the vessel in which amplification is performed.

The one or more characteristic of the further amplified product may be examined by means of the presence and/or absence of a distinctive unit in the further amplified product. The distinctive unit may be incorporated in the further amplified product or be associated there with. The distinctive unit may be introduced during the amplification process and / or in a subsequent step. The subsequent step may comprise hybridisation, for instance, of a component to the SNP base. The component may be a dideoxynucleotide, particularly a dideoxynucleotide incorporating a distinctive unit such as a dye.

The distinctive unit may be a dye label or colour producing molecule.

The distinctive unit may be a sequence of DNA, for instance a molecular beacon. The sequence of DNA, for instance a molecular beacon, may comprise a sequence of DNA incorporating a dye molecule. The sequence of DNA may be a single strand. The sequence of DNA may be looped by joining one part of the sequence to another. Preferably the dye molecule is in the loop, still more preferably in one part of the sequence which is joined to another. Preferably the dye molecule is in proximity with a quencher molecule. Preferably the quencher molecule prevents the dye molecule characteristic, for instance fluorescence, being visible. Preferably the dye molecule becomes visible, for instance fluorescent, upon activation. Preferably activation is caused by primer extension into the sequence of the molecular beacon. Activation preferably occurs through the opening of the loop. The molecular beacon sequence may be F-ACGCGCTCTCTTCTTCTTTGCGCG-Q where F is a distinctive unit such as a dye, and Q is a quenching unit or vice versa. Preferably the parts of the molecular beacon sequence which join to one another are the stems ACGCGC from the 5' end and GCGCG from the 3' end. Preferably the universal primer incorporating molecular beacon does not contain phosphorothioate bonding. Preferably none of the second set of primers contain phosphorothioate bonding. Ideally none of the first or second primers contain phosphorothioate bonding. Where molecular beacons are used, the amplification product may be examined for one or more characteristics in the amplification reaction vessel. For

-14-

instance, the Roche Light Cycler™ or other such instruments may be used for this purpose.

The distinctive unit may be visible under daylight or conventional lighting and/or may be fluorescent.

The distinctive unit may be an emitter of radiation, such as a characteristic isotope.

The distinctive unit is preferably provided at the 5' end of one or the primers, more preferably on a forward primer and ideally with a different distinctive unit for the other forward primer of the second set.

Preferably the distinctive unit is indicative of the nucleotide present at the SNP. Preferably a different distinctive unit is present if one nucleotide is present at the SNP and than if the other nucleotide is present at the SNP. Different distinctive units may be provided for indicating the SNP at one locus when compared with the distinctive units for indicating the SNP present at a different locus.

The examination may involve separating the further amplified product relating to one SNP from the further amplified product from one or more other SNP's. Preferably the further amplified products for each SNP are separated from one another. Electrophoresis may be used to separate one or more of the further amplified products from one another. The further amplified products may be separated from one another based on size of the further amplified products, for instance due to the different length of the further amplified products.

The examination may involve analysing the response of the further amplified product, for instance in the vessel in which amplification was performed, to radiation of various wavelengths, for instance fluorescent light.

The examination may involve the use of micro-fabricated arrays.

The further amplified product may be contacted with one or more components retained on a solid support. One or more of the components may be an oligonucleotide, preferably with its 5' end tethered to the support. Preferably the oligonucleotide has a sequence which pairs / anneals with the sequence of at least one, ideally only one, of the further amplified products.

In an embodiment the oligonucleotide may have a sequence which pairs / anneals with the sequence of at least one, ideally only one, of the further amplified products up to

-15-

the base before the base which is the SNP site. Only a portion of the further amplified product may pair / anneal to the oligonucleotide. Preferably a particular further amplified product type pairs / anneals to a particular oligonucleotide.

In an another embodiment, the oligonucleotide may have a sequence which pairs / anneals with the sequence of at least one, ideally only one, of the further amplified products along the sequence corresponding to the locus specific portion and the further portion. Preferably the further portion of the further amplified product includes a distinctive unit. The distinctive unit is preferably a dye. Preferably a different dye is present on each different further amplified product.

A plurality of such components, such as a plurality of oligonucleotides may be provided. A plurality of different oligonucleotides may be provided with each having a sequence which pairs / anneals to a further amplified product, ideally only one such product. It is particularly preferred that each oligonucleotide type pairs / anneals to a different further amplified product type from the others. The plurality of different types of oligonucleotides may be provided at a plurality of different, ideally discrete locations on the support.

The solid support may be glass, silicon, plastics, magnetic beads or other materials.

In an embodiment one form of the invention, the oligonucleotide and paired / annealed further amplified product may be contacted with one or more further components. Preferably one or more of the further components includes a dideoxynucleotide. Preferably one or more of the further components includes a distinctive unit, such as a dye. Preferably different further component types include different distinctive units. Two or more components comprising two or more different dideoxynucleotides with a different distinctive unit attached to each may be provided. The dideoxynucleotides may be A, T, C or G. Three or four dideoxynucleotides may be provided, preferably each with a different distinctive unit.

One or more, preferably only one of the further components may selectively attach to the SNP base and / or 3' end of the oligonucleotide. Preferably the selectivity of the attachment is based on the pairing of part of the further component's identity with the SNP base identity, such as the pairing of the dideoxynucleotide identity with the SNP base identity. Preferably the pairing incorporates the distinctive unit in the structure.

-16-

Preferably the pairing incorporates the distinctive unit in the structure. Preferably non-pairing further components and their distinctive units are not incorporated in the structure.

The identity of the distinctive unit attached to the component in the structure is preferably investigated. Preferably the identity of the further component and / or the identity of the SNP is derived from the identity of the distinctive unit.

In another form of the invention, the oligonucleotide and paired / annealed further amplified product may be contacted with one or more additional components. The one or more additional components may be one or more further oligonucleotides. Preferably one or more of the additional components includes an end base, preferably at its 5' end. Preferably one or more of the additional components includes a distinctive unit, such as a dye. Preferably different additional component types include different distinctive units. The additional components may comprise two or more different further oligonucleotides with a different distinctive unit and / or end base attached to each. The end base of the further oligonucleotides may be C, G, A or T. Three of four further oligonucleotides may be provided, preferably each having a different distinct unit and / or end base.

One or more, preferably only one, of the further oligonucleotides may selectively attach to the SNP base and / or 3' end of the tethered oligonucleotide. Preferably the selectivity is based on the pairing of the further oligonucleotide's end base identity with the SNP base identity. Ligase may be provided in contact with the tethered oligonucleotide and / or further oligonucleotide and / or further amplified product. Preferably ligation occurs where the SNP base and end base pair, thereby incorporating the distinctive unit in the structure. Preferably non-pairing further components and the distinctive units are not incorporated in the structure.

The identity of the distinctive unit attached to the component in the structure is preferably investigated. Preferably the identity of the additional component and / or the identity of the end base of the additional component and / or the identity of the SNP is derived from the identity of the distinctive unit.

In yet another embodiment of the invention the further amplified product may incorporate an attachment unit. Preferably the attachment unit facilitates attachment of the further amplified product to a solid support. The solid support may be glass, silicone, plastics, magnetic beads or other materials. Preferably attachment is affected by means of a covalent bond. The attachment unit may be an amino group, preferably an amino group

-17-

provided at the 5' end of the further amplified product. It is preferred that the solid support is an epoxy-silane treated support in such cases. The attachment unit may be a phosphorothiate unit, ideally provided at the 5' end of the further amplified product. In such a case, attachment to a bromo-acetomide treated solid support is preferred.

The further amplified product, attached to a solid support, is preferably contacted with one or more probes preferably having a different sequence from one another, at least in part. Preferably each probe has a common sequence portion to each other probe. It is particularly preferred that this common sequence portion correspond in sequence to the locus specific portion of the further amplified product. Preferably the probes incorporate at least one different sequence portion compared with one another. Preferably the different portions, for at least one of the probes, corresponds to the universal primer portion sequence of the further amplified product. It is preferred that contact of the probes with the further amplified product results in hybridisation of one of the probes to the further amplified product, ideally with no hybridisation of the other probe or probes. Preferably each probe has a distinctive unit attached, such as a dye unit. Preferably different distinctive units are used for each different probe.

The sample may be compared with another sample. The comparison may be based on comparing one or more of the one or more characteristic of the further amplified products for each sample. The samples may be compared to confirm a match in the characteristic between the samples. The samples may be compared to eliminate a match in the characteristic between the samples. The occurrence of the one or more further characteristic for one or more SNP's may be compared with information on the frequency of occurrence of the one or more further characteristic for the one or more SNP's in a population. The population may be a representative sample of the population of a country, an ethnic group or database.

According to another embodiment of the invention we provide a method of investigating a sample of DNA, the method comprising contacting the DNA containing sample with two or more primers, amplifying the DNA using those primers to give an amplified product and examining one or more characteristics of the amplified product, at least a first set of amplification conditions and a second set of amplification conditions

-18-

being employed, amplification by one or more of the primers being at least impaired during the use of one or the sets of amplification conditions.

Preferably amplification by one or more of the primers is inhibited during the use of one at least one of the sets of amplification conditions. Preferably annealing of the one or more primers is inhibited during the use of at least one of the sets of amplification conditions. Preferably one or more of the primers remains single stranded during the annealing step of one or more of the sets of amplification conditions.

The first set of amplification conditions may inhibit amplification by one or more of the primers. The second set of amplification conditions may inhibit amplification by one or more of the primers. One or more further sets of amplification conditions may be provided. One or more of the one or more further sets of amplification conditions may impair and / or inhibit amplification by one or more of the primers.

In one embodiment, the first set of amplification conditions may impair or inhibit amplification by one or more of the primers. Preferably the second set of amplification conditions provides for amplification by the previously impaired or inhibited one or more primers, particularly by all primers.

In another embodiment, the first set of amplification conditions may not inhibit or impair amplification by the primers, though amplification by one or more of the primers may be impaired or inhibited by other factors, the second set of conditions may inhibit or impair amplification by one or more of the primers and a third set of conditions may be provided in which the one or more primers impaired or inhibited by the second set of conditions are no longer inhibited or impaired.

In yet another embodiment, one or more primers may be inhibited during the first set of amplification conditions, not inhibited during a second set of amplification conditions, and not inhibited during a third set of amplification conditions. Particularly in such a case one or more other primers may remain inhibited during the first and second set of amplification conditions, but not during the third set of amplification conditions.

A set of amplification conditions may include a denaturation stage, annealing stage and extension stage. Preferably a set of amplification conditions is applied for a number of cycles. The extension stage may be provided during the cycles of one or more sets of conditions for between 30 seconds and 3 minutes, preferably 2 minutes +/- 10%. The denaturation stage may be provided during the cycles of one or more sets of

-19-

conditions for between 15 and 60 seconds, preferably 30 seconds +/- 10%. The annealing stage may be provided during the cycles of one or more sets of conditions for between 15 and 60 seconds, preferably for 30 seconds +/-10%. A temperature of between 70 and 80°C, more preferably between 74 and 78°C and ideally 76°C is preferably provided for the extension stage. A denaturation temperature of between 85 and 98°C, more preferably between 92 and 96°C and ideally 94°C is preferably provided.

An annealing temperature of less than 72°C is preferably provided in a set of amplification conditions allowing annealing of two or more primers and / or not inhibiting any primer annealing. Preferably a temperature of 72°C or greater is employed in one or more sets of amplification conditions intended to impair or inhibit amplification by one or more of the primers.

The amplification process may be performed for between 1 and 20 cycles, particularly the first one to three cycles, where one or more of the primers is impaired or inhibited from annealing. Preferably between 10 and 80 cycles are employed where the primers are not impaired or inhibited. The overall process may include between 40 and 80 cycles, more preferably between 45 and 70 cycles, and ideally between 50 and 70 cycles.

The method may use one or more of the primers, preferably for all of the primers which are not inhibited or impaired during all the sets of conditions, at a concentration of between 5 and 500nM, more preferably at between 10 and 200nM.

Various embodiments of the invention will now be described, by way of example only, and with reference to the accompanying drawings in which:-

Figure 1a to 1e illustrates the various parts of the first stage of a process according to the present invention;

Figure 2a illustrates one forward primer suitable for use in the present invention;

Figure 2b illustrates a second forward primer suitable for use in the present invention and intended for use with the primer of Figure 2a;

Figure 3a to 3e illustrates the various parts of the second stage of a process according to the present invention;

Figure 4a illustrates a "universal" forward primer suitable for use in the second stage of the present invention;

-20-

Figure 4b illustrates a second "universal" forward primer for use in the second stage of the present invention and intended for use with the primer of Figure 4a;

Figure 5 illustrates schematically the products of the two stage process when applied in a multiplex system;

Figure 6 illustrates amplification in the first stage using a primer according to the present invention;

Figure 7a schematically illustrates a structure for providing an examinable characteristic for the further products of the present invention;

Figure 7b illustrates the sequence of the structure of Figure 7a;

Figure 7c illustrates an "universal" forward primer incorporating a molecular beacon having one universal primer sequence;

Figure 7d illustrates a further "universal" forward primer incorporating a molecular beacon having a different primer sequence;

Figures 8a to 8f illustrate an alternative amplification process according to the present invention;

Figures 9a to 9d illustrate a micro-fabricated array investigation technique for amplifying products according to the present invention, based on genetic bit analysis;

Figures 10a to 10e illustrate a micro-fabricated array investigation for the amplified products of the present invention based on a ligation technique;

Figure 11 illustrates two forward primers and a reverse primer suitable for use in one embodiment of the first stage amplification process of the present invention;

Figures 12a to 12e illustrate various features of a hybridisation based investigation of the amplification result where the amplified strand is tethered to a glass slide;

Figures 13a and 13b illustrate a further way of investigating the results by hybridising the amplified products with oligonucleotides tethered to glass slides;

Figure 14 illustrates results for three samples and a control, at four loci, using the present invention;

-21-

Figure 15 illustrates the sensitivity and specificity of the technique of the present invention;

Figure 16a primer 416a tested against Gc 1S1S; 1F-1F; 2-2 and negative control, illustrating that only 1F-1F gives a signal and no background was observed with the other samples;

Figure 16b illustrates a simplex reaction to test specificity of the forward primer 420G, which detects both Gc1 polymorphisms, the primer being tested against a series of individuals, Gc1S-1S; 1F-1F; 2-2 and a negative control, with only the first two samples giving a signal with the remainder being clear;

Figure 16c illustrates a simplex reaction to test specificity of the forward primer 420T which detects Gc2, with samples taken from 1S-1S, 1F-1F, 2-2 and negative control, again with no background being detected;

Figure 16d demonstrates the low levels of sample which can be detected, the test being conducted on a series of samples taken from individual 1S-1S and prepared at 1NG, 200PG, 400PG and 8PG respectively, (Figures top to bottom in order), with a cocktail of GC420; GC416 primers and the universal primers used in the PCR reaction, the results indicating, as both bits are green, that priming by 416C and 420G thoroughly occurred, this being consistent with the known genotype of the individual, with consequently no background from 416A or 420T being observed;

Figure 17 is a graph of fluorescence verses cycle number for samples amplified with Gc1s primer and varying concentrations of second round universal primer;

Figure 18 is a minigel result for PCR products according to an amplification performed according to the invention;

Figure 19 is a graph of fluorescence verses cycle number for reduced concentrations of Gc1s + ve DNA compared to a control, SDW;

Figure 20a is a graph of fluorescence against amplification cycle number for Gc1 primer with universal G beacon for DNA samples 1, 2 and control sample 3;

Figure 20b is a graph of fluorescence against amplification cycle number for Gc2 primer with universal C beacon for the samples of Figure 17a;

-22-

Figure 20c is a graph of fluorescence against amplification cycle number for Gc1s primer with universal G beacon for the samples of Figure 17a;

Figure 20d is a graph of fluorescence against amplification cycle number for Gc1f primer with universal C beacon for the samples of Figure 17a;

Figure 21 is an illustration of the primer binding sites of Gc1f primer in respect of codon positions 416 and 420 on samples that are Gc1s, Gc1f and Gc2 phenotypes;

Figure 22 is a graph of fluorescence against amplification cycle number for male DNA sample with amelo Y primer and universal G beacon;

Figure 23a is a graph of fluorescence against amplification cycle number for the major component of a 2 DNA sample mixture coding for the Gc2 polymorphism and varying major to minor component ratios;

Figure 23b is a graph of fluorescence against cycle number showing the detection of the minor component of a 2 sample DNA mixture coding for the Gc1s polymorphism with varying major to minor concentration ratios;

Figure 24 is an illustration of experimental results with amplification at different annealing temperatures illustrating extensive primer dimer formation at 70°C and 72°C, diminished by the primer dimer formation at 74°C and substantially no primer dimer formation at 76°C;

Figure 25a illustrates primers susceptible to primer dimer formation;

Figure 25b illustrates two primers for which a primer dimer formation cannot occur;

Figure 25c illustrates primer sequences according to the concept of Figure 25b;

Figure 26 is a graph of fluorescence verses cycle numbers for different samples containing different amounts of DNA, a DNA negative sample and two sterile distilled water samples;

Figure 27 is a graph of fluorescence against cycle number for samples amplified using a particular primer set, again illustrating samples containing various levels of DNA, a DNA negative sample and two sterile distilled water samples; and

-23-

Figure 28 illustrates a graph of fluorescence verses cycle number for samples amplified using a particular primer set, various samples containing varying levels of DNA being provided, together with a DNA negative sample and two samples of sterile distilled water.

The nucleotide sequence of humans and other biological entities is in a large part consistent between individuals. Locations are known, however, at which variation occurs. One such form of variation is known as single nucleotide polymorphisms or bi-allelic markers, where the identity of a single nucleotide at a specific location is one of four possibilities from any of the four bases available, A, T, G or C. In many cases the variation is only bi-allelic and hence only one or two possibilities applies. Thus, some individuals may have a sequence incorporating a C base at a particular position, whereas other individuals will have a G base at that position; the surrounding sequences for both individuals being identical.

Medical diagnostics, forensic investigations and other DNA tracing applications make use of such single nucleotide polymorphisms (SNPs) for identification purposes. As the variation between individuals can only be between one of two options, a very substantial number of such locations, loci, must be considered for a statistically significant result, for instance the statistical significance of a match between a collected sample and an individual's makeup to be obtained.

Investigating such a large number of loci, frequently several hundred, on an individual basis is extremely time consuming. To reduce the time taken, it might be desirable to construct multiplexes which allow a substantial number of loci to be investigated simultaneously based on PCR or other amplifying techniques. The design of reliable constructs for a large number of loci, however, is extremely difficult due to problems in interactions between the primers needed for the different loci, different conditions for suitably efficiency amplification of the different primers and a variety of other issues.

The technique of the present invention is designed to simplify SNP based and other investigations, and particularly to facilitate the rapid development of multiplexes suitable for investigating a large number of such loci simultaneously, due to the flexibility offered by the present technique.

-24-

The technique is based around two amplification stages, generally achieved through PCR, with both of the stages offering specificity in terms of the SNPs identified and amplified. The two amplification stages can be conducted separately or simultaneously and the amplification products can be analysed in a variety of ways.

Figure 1 illustrates, according to one embodiment of the invention, a series of stages involved in the first amplification process based around a target template 1 with a potential C or G single nucleotide polymorphism 3 in one strand 5 of that target template 1. As illustrated in step A, the target template strand 5 of the particular individual under consideration has a C nucleotide at the SNP site 3.

The first step in this amplification stage involves contacting the template target 1 with two different forward primers 7 and 9, and a reverse primer 11. The forward primers 7 and 9 are locus specific primers, described in more detail below.

Forward locus specific primer 7 is terminated by a G nucleotide thus rendering it a match with the C nucleotide at the SNP site 3 and resulting in annealing of that primer 7 with the strand 5. The reverse primer 11 is non-specific and anneals to the other strand 13 of the template 1 at the appropriate location.

In step B, the specific forward primer 7 and the reverse primer 11 extend to produce the strands 14 and 16 through primer extension.

Denaturation of the strands results in the separation of the strands 5, 13 from their respective copied strands 14 and 16. The copied strand 14 only is shown in step C and the illustration of the subsequent steps.

Subsequent primer annealing, step D, is then performed again using the two forward primers 7, 9 and reverse primer 11. As we are considering strand 14 it is the reverse primer 11 which attaches to the strand 14 due to its sequence. The specific forward primer 7 would attach to strand 16, once again annealing in alignment with the site of the SNP 3 in that strand's sequence, not shown.

In subsequent primer extension, stage E, the reverse primer 11 extends the sequence of new strand 18 with the appropriate sequence given the sequence of strand 14, including the extension to produce tail portion 19 which arose as the strand 14 included the tail portion 21 of the forward specific primer 7. Due to the G base in the sequence of strand 14, the new strand 18 includes an opposing C base so as to match the identity of the SNP at site 3 in original strand 5. Due to the G base in the sequence of strand 14, due to

-25-

the SNP related base 10, the new strand 18 includes an opposing C base 20 so as to match the identity of the SNP related site 10 in the originally copied strand 14.

Repetition of steps A through E over 20 to 25 cycles produces many millions of copies of sequences incorporating the same SNP identity, SNP repeat and surrounding sequence as the target template 1.

Figure 2a and b illustrate two locus specific forward primers, suitable for use in the stage detailed above, for use in investigating an SNP which could be either G or C. Each of the locus specific forward primers 30 consists of a locus specific portion 32 which has a sequence corresponding to the sequence of the loci under consideration up to the SNP site. The 3' end 34 of the locus specific forward primers ends in a G nucleotide 34a for one of the primers, Figure 2a, and in a C nucleotide 34b for the other primer, Figure 2b. Due to this different nucleotide used at the position corresponding to the SNP, then depending upon the identity of the SNP actually encountered, one of the locus specific forward primers will anneal thereto, but the other will not. Thus it is the forward primer of Figure 2a which anneals to the target in the example of Figure 1. This selectivity in annealing gives consequential specificity in the subsequent amplification cycles of the first stage.

In addition to the locus specific portion 32 the locus specific forward primer 30 includes a "universal" primer portion 36. The "universal" primer portion 36 consists of a nucleotide sequence which is identical for each of the two loci specific forward primers, save for a single nucleotide location 38 at the junction between the universal primer portion 36 and loci specific portion 32 of the primer 30. The nucleotide at the location 38 is identical to the 3' end nucleotide 34 of the locus specific portion 32 of the respective primer 30. Thus, the "universal" primer of Figure 2a incorporates G in its sequence at location 38 to reflect the G nucleotide present at the 3' end 34. The "universal" primer portion of Figure 2b, on the other hand, includes a C at location 38 to reflect the fact that a C nucleotide forms the 3' end 34 of this primer 30.

Whilst it is the locus specific portion 32 of the forward primers 30 which determines whether a primer anneals or not to the target, in the second and subsequent copying stages of the amplification process of stage 1, primer extension causes copying of the "universal" primer portion 36 of the primer sequence also and hence copying of the SNP equivalent nucleotide identity at location 38 too.

-26-

As previously stated the amplification process of the first stage results in a large number of copy sequences, including the SNP identity reflecting nucleotide and the matching nucleotide at location 38.

In the second stage of amplification, illustrated in Figure 3, a further specific amplification process is performed. It is much preferred that the second stage of amplification be conducted in the same vessel as the first, substantially simultaneous with the first amplification process. Such a possibility is described in more detail below.

For this stage, an aliquot of the amplification products from the first stage, described above, are taken and contacted with a pair of "universal" forward primers and a "universal" reverse primer. These "universal" primers are described in more detail below.

In step A, the strands 40 and 42 (copy strands which are equivalent to strands 14, 18 produced in the first stage as illustrated above) produced by the first stage 1 are denaturated and contacted with the two "universal" forward primers 50, 52 and reverse "universal" primer 54.

The two "universal" forward primers differ in terms of the 3' terminal end nucleotide 55 and in terms of a dye unit D or other form of label provided on the 5' end 56. The 3' end nucleotide 55 for the forward "universal" primers in this example is either C, "universal" primer 50, or G, "universal" primer 52.

As the strands 40 and 42 represent the outcome of copies of copies of the originals being taken, unlike strands 14, 18, they both have tail portions 44, 46 respectively which arise from the copying of the "universal" primer portions of the locus specific forward primer and reverse primer in the first stage.

The "universal" primers 50, 52 each have a sequence corresponding to the "universal" primer portion 34 of the first stage locus specific primers 30 up to location 38 of the locus specific forward primers 30. At location 55 the forward primers 50, 52 of the second stage have a base corresponding in identity to the identity of the nucleotide pairing to the SNP repeat in the stage 1 process, in one case, and in the other case corresponding to the identity of the other option for the SNP repeat. The nucleotide identity for the "universal" primers 50, 52 at location 55, corresponding to location 38, is thus different for the two primers 50, 52, with one providing one of the options and the other providing the other.

-27-

In the illustrated example, primer 50 carries a C and primer 52 carries a G nucleotide at position 55.

The sequence of the primers 50, 52 and particularly the identity at position 55 determines whether or not that primer 50, 52 anneals to the tail portion 44 of the strand 42 or not. In the illustrated case, strand 42 carries the SNP nucleotide C at site 63 as this was a copy of the identity of the SNP at site 3 in the original target strand 5. The C identity is also repeated in the tail portion 44 at site 65 as this was copied due to the copying of the tail of the original primer 7 by the reverse primer 11 in the first stage. As a consequence the sequence of the tail portion 44 of strand 42 provides an annealing site for "universal" primer 52, but not primer 50. The reverse primer 54 anneals to the tail portion 46 of strand 40 due to the sequence matching.

Primer extension, step B, results in the production of strand 60 by matching strand 40, including SNP site copy C, and in the production of strand 62, including the match for the SNP, G, by matching strand 42 by the "universal" reverse primer 54 and specific "universal" forward primer 52 respectively. The SNP repeats are also copied.

Thermal denaturation is then used to separate the strands, step C, and from here on strands 60 and 62 only are considered although similar processes apply to the other strands too.

In annealing step D, the specific "universal" forward primer 52 anneals to the tail 64 of strand 60 due to the presence of a C nucleotide at the relevant position 65 in strand 60 and the consequential pairing to the "universal" forward primer 52. The reverse primer 54 anneals to the tail portion 66 of the strand 62.

In the further extension step E, the forward primer 52, which brings with it the label D1, extends the sequence of new strand 68, including tail portion 70. The reverse primer 54 extends the sequence of new strand 72, (thereby reproducing the SNP identity at site 74), including tail portion 76, (thereby reproducing the nucleotide corresponding to the SNP repeat 75 in that part too). Strand 62 already incorporates the label D1 from its start as the primer 52 in step A

Once again, repeating stages A to E gives substantial amplification of the sequences and produces a great number of sequences label with a dye D1, the dye being selectively taken up as only one primer anneals and thus takes the dye into the sequence with it.

-28-

As described above, the second stage of the process uses a pair of "universal" primers on their own, illustrated in Figures 4a and 4b. These consist of a portion 80 having a sequence identical with the "universal" primer portion 32 of the locus specific primers 30 up to the single nucleotide variation at the end of the "universal" primer portion 32. The ends 82 of the universal primers of Figures 4a and 4b are different from one another and have an identity consistent with one of the two SNP possibilities, as is the case for the primers of Figures 2a and 2b. Thus, one "universal" primer 52, Figure 4a, is provided with G at its terminal 3' end 82 and the other "universal" primer 50, Figure 4b, is provided with C at its terminal 3' end 82.

During stage 2 of the process, these "universal" primers will selectively anneal to the amplification products of the first stage depending upon whether the tail portions extended and amplified during that stage incorporates the G or C variation.

Of course, equivalent primer types could be used with T or A variations in the above mentioned processes to investigate an SNP having potential T or A variation.

It is desirable for both the "universal" locus specific primers and the "universal" primers themselves to be provided with a phosphorothioate residue at the 3' end in order to protect the final base against exonuclease digestion by the Taq polymerase. This maximises the potential of the system to discriminate polymorphisms.

The different "universal" forward primers are provided with different labels/markers, in this case a JOE dye label and an FAM dye label respectively. The dye labels are provided at the 5' end of the forward primer in the second stage of the process. Of course, other different dyes and other forms of marking, such as radio nuclides could be used.

The "universal" primers were carefully designed to give desirable characteristics in terms of their melting temperatures, particularly a melting temperature of around 60°C. The sequences were also checked to ensure minimal hairpin formation and checked for minimal primer dimer formation. The sequences were also checked against human DNA sequence records and / or samples to ensure that human DNA is not amplified and to avoid any correspondence to any published sequence and particularly any part of the human DNA sequence. Primer dimer formation was also taken into account so as to keep such formation minimal.

-29-

An example of a suitable "universal" forward primer is provided (written 5' to 3') by the sequence :-

CGA CGT GGT GGA TGTG CTAR,

where R equals G or C or A or T depending upon the SNP to be detected for; and a suitable "universal" reverse primer is provided (written 5' to 3') by the sequence :-

TGA CCT GG CTG ACT CGA CTG.

The melting conditions for these primers, under 50mM primer and 50mM salt is 59°C with a GC content of 60%.

The specificity of the forward primers can be increased still further by making the ending for the primers of the first stage as follows, for a G detecting primer -TC and/or for a C detecting primer -AG and/or for a T detecting primer -CT and/or for a A detecting primer -GA.

By virtue of the different dyes or other indicators provided on the SNP specific "universal" forward primers an indication as to which of the possible SNP variations occurs at the SNP under consideration is given by the amplified product.

Although the technique has an advantageously low background noise at each locus, it is possible to incorporate one or more typed samples as controls.

In Figure 5, the amplification products arising from a significant number of such a two stage processes are indicated. In this example, the sample subjected to the multiplex has produced amplification products with indications at:-

- loci 1 with SNP G, indicated by dye X rather than dye Y;
- loci 2 with SNP T, indicated by dye W rather than dye V;
- loci 3 with SNP C, indicated by dye Y rather than dye X;
- loci 4 with SNP G, indicated by dye Z rather than dye U;
- loci 5 with SNP G, indicated by dye X rather than dye Y;
- loci 6 with SNP A, indicated by dye V rather than dye W.

As the lengths of the sequences forming the amplification products at different loci are designed to be of different length, it is possible to separate those amplification

-30-

products based on their size, for instance, using electrophoretic techniques and thus produce a series of lines on a gel whose colour is indicative of the SNP variation at the particular loci. Where the length of the sequence for one loci may be close to another, different dyes for each of the possibilities can be used. This in the example of Figure 5 loci 1 and 5 may have the same SNP variation but are sufficiently separable for the same dyes to be used. Loci 4 on the other hand is potentially close to loci 5 results and so different dyes for the results of the G, C variation for loci 4 and 5 are used.

Figure 6 provides a detailed illustration of the locus specific primer annealing and subsequent extension where the SNP in the genomic DNA target template has a C SNP and as a consequence binds to the G incorporating locus specific primer.

As a further modification over the basic incorporation of a dye on the primer, it is possible to incorporate a so-called "molecular beacon". This consists of a single stranded DNA molecule that possesses a stem and loop structure. When the hairpin loop structure, illustrated in Figure 7a, is formed, the dye molecule is brought into close proximity to the quenching molecule, and as a consequence does not fluoresce when investigated. 5-carboxyfluorescein (FAM) offers a suitable fluorescent dye for use in such a situation with 4-4' dimethylaminophenylazo benzoic acid (DABCYL) offering a suitable quencher moiety. Methyl red could also be used as an alternative quencher, particularly branched methyl red.

A molecular beacon of this type, when the reverse primer extends into the beacon stem, becomes destabilised and gives rise to unfolding of the molecule. This moves the fluorescent dye label sufficiently away from the quenching molecule that quenching of the dye molecule no longer occurs. Subsequent investigation of the system would give fluorescence, therefore, the fluorescence being indicative of the particular dye and hence the particular SNP considered.

The particular sequence and components of the structure of Figure 7a are illustrated in Figure 7b. Further illustrations of universal primers in conjunction with molecular beacons are illustrated in Figure 7c, for a universal molecular G beacon with an AG sequence at its 3' end, and in Figure 7d for a universal molecular C beacon with a TC at its 3' end. The universal primer reverse sequence used with the universal molecular beacons of Figures 7c and 7d is:-

TGC CGT GGC TGA CCT GAG AC

-31-

It is preferred that none of the universal molecular beacon incorporating primers contain phosphorothioate bonding.

A variety of ways of implementing the technique are possible, thus whilst PCR followed by direct inspection of the results by an instrument and / or inspection following electrophoretic investigations of the amplification products in gels could be undertaken, the technique is also suited to use with solid or other supports. Solid support media includes silicon, glass, plastics and magnetic beads. In particular, the technique is suited for implementation of the micro fabricated array system.

Micro fabricated arrays for implementing the present invention use oligonucleotides which are designed to terminate at the base which is penultimate to the SNP polymorphism under consideration. The oligonucleotides are attached to a solid support such as glass following a method, such as Z; Guilfoyle RA; Theil AJ; Wang R and Smith LM (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Research (1994), vol 22: 5456-5465.

In such a method microscope slides are immersed in 1% 3-aminopropyltrimethoxysilane solution in 95% acetone/water for 2min. The slides are washed 10 times with acetone, 5 min per wash, dried for 45 min at 110C, treated for 2h with a solution of 0.2% 1,4- phenylene diisothiocyanate (PDC) solution in 10% pyridinedimethyl formamide and washed with methanol and acetone.

These activated glass slides can be stored indefinitely in a vacuum desiccator.

The oligonucleotides are laid down in a grid pattern on the glass slide (using a robot it may be possible to lay down several hundred different oligonucleotides). The oligonucleotide contains a 5' end amino group introduced using the reagent – trifluoroacetyl-6-aminoethyl-2-cyanoethyl N', N'-diisopropylphosphoamidite. When the oligonucleotide is applied directly to the glass slide, the 5' end covalently binds to the glass.

The microfabricated arrays can be used in a variety of ways, some of which are described in more detail below, and frequently in conjunction with amplified products produced according to the following technique. In the alternative technique, as shown in Figure 8a, a target sample 800 including SNP site 802, which has an identity of a C base, is under consideration in conjunction with corresponding strand 804. To achieve

-32-

amplification a locus specific forward primer 806 and locus specific reverse primer 808 are introduced. Each of these primers is of the general type described above in comprising a locus specific primer portion 810 and 812 respectively and universal portion 814 and 816 respectively. The locus specific portions 810 and 812 are provided with sequences which match to sequences of the strands 803 and 804 respectively, but in contrast to the method described above, away from the site 802 of the SNP. Consistent with the format of the forward and reverse primers described above (see the description relating to Figures 2a and 2b, for instance), the universal primer portions have a sequence which does not hybridise to the locus in question and preferably does not hybridise to any naturally occurring DNA sequence.

Once hybridised, as shown in Figure 8b, extension causes the forward primer 806 to generate a copy strand 818 and causes the reverse primer 808 to form a copy strand 820. The strand 820 includes a base having an identity with the SNP and a strand 818 has a base pairing to the SNP identities.

Following denaturation of these strands, as shown in Figure 8c, strand 818 anneals with a reverse primer 808 and strand 820 anneals with a forward primer 806.

As illustrated in Figure 8d primer extension results in further copy strands 822, which includes a copy of the SNP identity and in terms of strand 824 which includes a copy of the base identity pairing to the SNP. These amplification products also include tail portions 826, 828 corresponding to the sequences introduced by the universal portions of the forward and reverse primers during earlier stages.

In the second amplification process of Figures 8e and 8f, which amplification can occur substantially simultaneously with the first stage and / or separately further amplification occurs. In this case, however, the amplification products for the first stage, strands 824 and 822 from the illustration of Figure 8d are contacted with forward and reverse primer. In this case, the forward primer 830 has a sequence corresponding to the sequence of the universal portion 814 of the forward primer 806 of the first stage. Similarly, the reverse primer 832 has a sequence identical with the universal portion 816 of the reverse primer 808 of the first stage. As a consequence of these sequences, the forward primer anneals to the tail portion 826 of strand 822 and the reverse primer 832 anneals to the tail portion 828 of strand 824.

-33-

Primer extension, Figure 8f, results in further copies of the strands 824, 822 being generated in terms of strands 834 and 836 respectively.

In the second stage of the process, as is true for the amplification process described previously, the fact that the amplification products from a large variety of loci will include the tail portions of the first stage primers means that a single forward and reverse primer can be used to further amplify all of the loci amplification products present, rather than having to use forward and reverse primers which are specific to the loci to achieve amplification. This renders the technique particularly suitable for amplifying a large number of target sequences from a large number of different loci simultaneously, without concerns as to coordinating melting temperatures and other properties.

The amplification products of the process described above can be analysed in a variety of ways.

By way of example, and with reference to Figures 9a to 9d, it is possible to analyse these amplification products using genetic bit analysis. In this example, a solid support 900, such as glass, is provided for the array. Tethered to the array by its 5' end is a tailored oligonucleotide 902 which is designed to have a pairing sequence with the amplification products incorporating the SNP of interest. The tethered oligonucleotide 902 has a 3' end 904 which is one base short of the base which pairs to the SNP base.

In use, the amplification products are brought into contact with the oligonucleotides 902 of the array (substantial number of such oligonucleotides are provided with the illustrations only representing one such oligonucleotide for clarity). As a result of this contact, the amplification product, for instance represented by strand 834 from Figure 8f, hybridises to the tethered oligonucleotide 902 by virtue of the pairing sequence. As the oligonucleotide 902 is one base short of the SNP location 906, however, no base pairing to the SNP is provided during this hybridisation process.

In a subsequent step, 9c, dideoxynucleotides labelled with dyes are introduced. In the Figure 9c example a dideoxynucleotide 908 is provided labelled with a dye A and a dideoxynucleotide 910 is provided with a dye B. Dideoxynucleotide 908 is a C base, whereas dideoxynucleotide 910 is a G base. Addition of these dideoxynucleotides in conjunction with TAQ results in the TAQ polymerase adding the appropriate base on to the tethered nucleotide 902 as shown in Figure 9d. In this case, as the base pairing to the

-34-

SNP is a G base, it is dideoxynucleotide 910 which is added and as a consequence, with it, dye B.

In a subsequent stage, not shown, the un-fixed dideoxynucleotides and their dyes are washed from the micro-fabricated array and inspection of the array reveals that it is dye B and as a consequence dideoxynucleotide 910 which has been added to the tethered nucleotide 902 and that as a consequence the identity of SNP 906 is a C base.

Investigations for other bases can be achieved simultaneously using different dyes for each of the four possibilities and with different tethered oligonucleotides being provided at different locations in micro-fabricated arrays. Thus a substantial number of loci can simultaneously be investigated using loci specific tethered nucleotides in conjunction with a different dye attached to each of the four possible dideoxynucleotides.

The method of Genetic Bit Analysis (GBA) is used to detect the polymorphism, as described in Nikiforov TT; Rendle RB; Goellet P; Rogers YH, Kotewicz MI, Anderson S, Trainor GL, Knapp MR (1994) Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res* 22:4167-75.

In the hybridization solution, there may be 4 dideoxy bases (A,G,C,T) which are dye-labelled with different dyes e.g. TAMARA, JOE, FAM, HEX. Taq or Klenow polymerase may be included in the reaction mix. This extends the oligonucleotide by just one base, with the complementary dideoxy - which subsequently fluoresces a colour which is dependent upon the complementary base.

As an alternative to genetic bit analysis a ligation assay can be used using such arrays. To investigate the identity of the SNP in the type of amplification product produced by the process described above in relation to Figures 8a to 8f. In this method, tethered oligonucleotides 1001 are provided on the support 1003 and micro-fabricated array is contacted with the amplification products.

As the tethered oligonucleotides 1001 are provided with a sequence which pairs to the sequence of the SNP incorporating strands, for instance strand 834 from Figure 8f, that strand hybridises to the tethered oligonucleotide 1001. The sequence of the tethered oligonucleotide 1001 stops one base short of the SNP site 1005. As a consequence, as shown in Figure 10b, no base pairing to the SNP site 1005, in this case base C, is provided.

In the subsequent step illustrated in Figure 10c to allele specific primers are brought into contact with the tethered oligonucleotide 1001 and the hybridised strand 834.

-35-

These primers are provided with 3' end bases representing the possible base identities which would match with the SNP base. Thus, oligonucleotide 1007 is provided with a 3' end G base and oligonucleotide 1009 is provided with a 3' end C base. The oligonucleotides 1007 and 1009 also incorporate different dye molecules from one another. By providing these oligonucleotides 1007 and 1009 together with ligase then only the oligonucleotide 1007 or 1009 which has a 3' base pairing to the SNP base identity will be introduced to the end of the tethered oligonucleotide 1001, see Figure 10d.

Following the successful ligation, as illustrated in Figure 10e, the temperature is raised to melt away the oligonucleotides carrying the unincorporated dyes and the amplification product, strand 834 itself. As a consequence, only the tethered oligonucleotide 1001 and oligonucleotide 1007 incorporating its dye remain. Investigating the colours present for the various arrays of the micro-fabricated array then reveals the identity of the dye and hence the 3' base identity for the oligonucleotide and hence the base identity and the SNP 1001 for each of the amplification products under consideration.

Again, a substantial number of loci can be considered simultaneously by the use of different tethered oligonucleotide 1001.

The basic technique of the invention could also be used in conjunction with mass spectrometry and/or micro titre plate based assays, including Taq-man assay and molecular beacons.

To improve the technique's ability to distinguish accurately SNP differences following amplification, the following modifications can be made. In the techniques described above the "universal" primer portion of the forward primers of the first stage add equivalent sequences in each case. The only difference was in the single nucleotide identity located at the junction between the universal primer portion and the loci specific portion of the primer. In this modification, however, the universal primer portion of the forward primer is provided with a distinct sequence in each case when compared with other forward primers for different SNP identities. Given that there are only a maximum of four possible SNP identities, a maximum of four different forward primers, each having a different universal primer portion are required for any one SNP. The reverse primer also includes a universal primer portion having a different universal primer portion to the portions of the forward primers. The function of the reverse primers and forward primers during amplification is equivalent to that described above.

-36-

Examples of suitable forward primers, 1000 and 1002, are illustrated in Figure 11 together with a suitable reverse primer, 1004. The forward primer 1000 is locus specific to alpha-1-antitrypsin M1/S polymorphism with the aim of detecting a T based variation whereas the forward primer 1002 is intended to detect an A based variation at the SNP. This is reflected in the identities of the SNP identifying portion 1006 and 1008 respectively. The locus specific portion, 1010, 1012 respectively, of these two forward primers, is attached to a universal primer portion 1014, 1016 respectively, with the universal primer portion incorporating a distinguishing portion 1018, 1020 respectively in each case. In this example the distinguishing portions are provided within the overall sequence of the universal primer portion for reasons described in more detail below. However, the entire universal primer portion or one end of the universal primer portion could be used to provide the distinguishing sequence.

A similar structure is provided for the first primer in terms of locus specific sequence 1022, universal primer portion 1024 and distinguishing portion 1026.

The universal primer portions are preferably different from one another by at least 6 bases. Universal primer portions having a sequence difference of between 25 and 100% when compared with one another is advantageous.

One of the principal advantages of this alteration to the universal primer portion sequence is that it introduces multiple base changes into the amplification products, and as a consequence renders the amplification products more suitable for investigation and detection using hybridisation on to solid support surfaces, such as glass. This technique is described in more detail below.

Having produced amplification products which are substantially different in terms of their bases from one another, the amplification products are particularly suited to investigation and analysis using the type of technique illustrated in Figure 12.

As illustrated in Figure 12a, the amplification product consists of the forward primer sequence 1200, including the universal primer portion 1202 and locus specific portion 1204 and the reverse primer portion 1206. By providing an amino end group 1208 on the terminal 5' end of the reverse primer the strand can be covalently attached to an epoxy-silane treated glass slide, 1210. The preparation of such epoxy-silane slides is based on the method of Beatty et al, Molecular Biology, Volume 4, 1995, 213-225.

-37-

Other attachment chemistry, for instance the use of 5' ends labelled with phosphorothiate can be used to attach such strands to bromo-acetamide slides for instance.

Generally, the amplified product is purified for instance using a centricon filtration to remove an incorporated primer and dNTP's. It is then extracted with water before spotting on to the glass slides, for instance using an Amersham generation 3 micro array spotter. The slides can be incubated in an high humidity chamber for between 30 minutes to two hours at 20°C to 40°C before washing in water at 95°C, 10mM triethyamine (pH 9.2) at room temperature, and two further washes with water at 60°C before being stored dry at room temperature.

To effect the detection step the single strands must be contacted with suitable fluorescent probe constructs. In Figure 12b a biallelic system is being introduced, and as a consequence two different fluorescent probes are introduced. Each fluorescent probe has a different dye label 1211, common locus specific portion 1212 and different universal portions 1214 and 1216 respectively. Hybridisation is carried out at low temperature, 40°C. Probe specificity is controlled by carrying out post-hybridisation washes at higher temperatures.

As illustrated in Figure 12c only one of the fluorescent probes is sufficiently complimentary to hybridise completely to the fixed strand. By conducting the post-hybridisation washes at a sufficiently high temperature (60°C to 75°C) specificity of the probes is maintained as at such temperatures the fact that the locus specific portion 1212 of the other fluorescent probe corresponds is insufficient to remain hybridised due to the difference between the universal primer portion 1202 and 1216, Figure 12d. As a consequence the second fluorescent dye label is not retained by the single strand.

On a similar basis, Figure 12e, if the universal primer portions are complimentary, but the locus specific region is different, a fluorescent probe for a different locus, hybridisation cannot occur sufficiently for the fluorescent probe to be retained on a similar basis.

Experimental conditions for such a system are described below.

As an alternative analysis technique, synthetic oligonucleotides can be spotted on to the slide and covalently bound thereto using an amino-linker at 5' end. Such a process is illustrated in Figure 13a. Once contacted with the amplification products in cases where the oligonucleotide 1300 has a sequence for its locus specific portion 1302 which matches

-38-

the locus specific portion of the amplified product 1304 and the universal primer portion 1306 matches that amplified product then hybridisation will occur. The universal reverse primer sequence 1308 does not participate in the hybridisation. Again the dye label molecules 1310 can be inspected in such cases to indicate the identity.

EXAMPLE - Multiplex amplification

As well as the above mentioned two distinct stage process, a single vessel process is also envisaged. This involves adding all first reaction primers at a concentration of 50nM with the exception of any primers for which there would be competition, for instance due to primer site overlap. This is exemplified by forward primers Gc 420G/T and 416C/A for instance of the specific primers identified. When competition potentially occurs it is necessary to balance the reaction by trial and error experimentation of the concentrations. For the above mentioned specific examples the optimum concentration for the Gc 416C/A primers were 25nM and the 420G/T primers were at 50nM. The reverse primer optimum concentration was 100nM for all cases.

To this mixture the universal primers used in the second stage are then added. As a rule of thumb, the amount of universal primers used must be increased according to the number of first reaction primer sets utilised. The concentration (C_n) of each universal primer is $C_n \times L$ where L is the number of loci (or SNPs) analysed. Thus for a 5 locus system the concentration of each of the universal primers is $50 \times 5 = 250\text{nM}$. The amount of DNA is held constant at 1ng for optimum amplification but it is possible to analyse lower quantities of DNA. At higher numbers of loci it is believed that the concentration of primer which is usefully deployed will plateau out.

The overall combination was then subjected to the amplification process as follows:-

Summary of conditions:

50ul PCR reaction total volume.

Buffer = Perkin Elmer PCR buffer

$\text{MgCl}_2 = 1.5\text{mM}$

DNTPs = 200uM each

Taq Gold (Perkin Elmer) 1.25units

-39-

Cycling conditions:

94°C for 30 seconds

61°C for 30 seconds

72°C for 90 seconds

6 cycles

94°C for 30 seconds

72 - 75°C for 30 seconds

for 5 to 10 cycles

94°C for 30 seconds

61°C for 30 seconds

72°C for 90 seconds

for 24 to 29 cycles

The resulting amplified products were then considered in the manner outlined above, electrophoresis based separation for instance, to determine which label and hence which SNP were present at each loci under consideration.

The underlying technique of this invention offers a significant number of advantages:-

- 1) the universal primers themselves do not prime human DNA, thus artifacts mediated by mis-priming of degraded DNA are minimal and the pull-up artifact is easily recognised since the electrophoretic migration rates of different SNP's is different;
- 2) the use of a two stage reaction process improves specificity;
- 3) mis-priming is virtually negligible, thus virtually eliminating background noise and rendering the technique useful for very low level mixtures;
- 4) the need for dye or other labels on the universal primers only significantly reduces production costs for the reagents;
- 5) large numbers of different loci may be amplified in the second stage of the process using only one set of primers thus rendering large number multiplexing possible;

-40-

6) the use of only two forward primers in the second stage gives a balanced reaction without competition between primers;

7) the use of a two stage process enables very small levels of feed material to the first stage, sub-nanogram levels, to be amplified to the extent where several aliquots are produced for use in potentially different second stage processes.

8) by providing a one tube reaction the conduct of the reaction is simplified and speeded up.

These and other advantages, and features of the present invention are apparent from the following practical demonstrations of the invention.

EXAMPLE - Dye based

Mitochondrial DNA

Tully et al (1996) Genomics 34, 107-113, described a minisequencing approach to analyse mitochondrial DNA SNPs. The SNPs listed in table 1 were analysed using the approach described above but with the primers amended according to the present invention to provide universal G or universal C attached to the 5' end of the primers listed.

The sizes of each DNA fragment are known and when run on a gel, bands indicative of a loci are produced and these are either JOE (green) for universal primer E or FAM (blue) for universal primer C labelled depending on the SNP identity, so allowing visualisation of the results.

TABLE 1

Forward Primers Position	Primer Sequence Used With Universal C	3' Polymorphism Used With Universal G
73	GTATTTTCGTCTGGGGGGTA	G
146	GTCTGTCTTTGATTCCTGCCC	T
152	TTTGATTCCTGCCTCATCCC	T
195	ATATTACAGGCGAACATACC	T

-41-

247	GCTTG TAGGACATAATAACAATT A	G
Reverse Primer 326	CAGAGATGTGTTTAAGTGCTGT	

Reaction conditions:

For each separate reaction:

DNTPs were at a final concentration of 35mM

Perkin Elmer (PE) buffer was at a final concentration of 0.375mM with 0.375mM MgCl₂.

0.25 AmpliTaq (PE) was added to 50ul reaction.

Primer concentrations are detailed separately with examples given.

All phenotypes were verified by independent analysis using the mini-sequencing method of Tully et al (1996).

Example 1

Multiplexed mitochondrial DNA

Reaction conditions:

DNTPs all at 10mM;

Final concentration of 35mM.. PE buffer 15mM 15mM MgCl₂ per reaction MgCl₂ =.375mM. AmpliTaq =.25ul in 50ul

In the following example, 1 uM of each of the forward primers and 2uM of the reverse primer listed in table 4 was used in the reaction mixture. A 50ul reaction containing 0.3ng of genomic DNA was amplified through 8 cycles at 94C for 30scc; 57C for 30sec and 72C for 90sec. An aliquot of 5ul of the reactant was then transferred into a second tube containing 1uM of each forward universal primer and 1uM of the reverse universal primer. This was amplified for 22cycles at 94C for 30sec, 62C for 30sec and 72C for 90sec. Samples were electrophoresed on a ABD 377 automated sequencer with Rox 500 sizing standard. The negative control was treated under the same conditions, except that no DNA was added to the reaction.

-42-

The results are illustrated for the four samples in Figure 11, with each of the three samples being represented by one trace and with the control sample (no DNA added) being the fourth trace. The peaks labelled a are size standards, those labelled b indicated green labels and those labelled c indicating blue labels. These samples were analysed against 4 mitochondrial loci, 247, 195, 152, 146. The identities revealed for the four loci are summarised in Table 4.

Table 4

Sample	Loci 247	Loci 195	Loci 152	Loci 146
One	G	T	T	T
Two	G	C	C	C
Three	G	T	C	C
Control	none	none	none	none

Example 2

Elucidation of a mixture where the minor component is <10pu DNA (genomic equivalent).

In the next example, the results for which are illustrated in Figure Y2, mixtures were prepared with the major component coding for the mt0073A polymorphism (2ng genomic DNA) and the minor component coding for the mt00326 polymorphism (0-50pg). Amplified with forward primers, either mt0073-G or mt0073-A (1uM) and the reverse primer mt00326 (1uM), the cycling conditions were the same as described previously but the second round amplification was just 3 cycles.

In the first experiments, left hand series, (a) primers used were mt0073-G (1uM) and mt00326 (1uM) whereas in experiments, right hand series, (b) primers were mt0073-A (1uM) and mt00326 (1uM). The results (a) showed that even in the presence of very great excess of mt0073-G template, there was no mt0073-A background product detected. Similarly, in experiment (b) using just primer m00O73A there was no mt0073-G detected. The high specificity of the reaction demonstrated discrimination of minor components in mixtures down to extremely low levels of 12,5pg in a total - a mixture ratio of 1:200.

-43-

In the figures, peaks a are size standards, peaks b indicate a green response and peaks c indicate a blue response.

Example 3

Genomic DNA - Group Specific Component (Gc)

The Gc single nucleotide polymorphisms have been well characterised (Braun et al, 1992) Hum Eanet, 89:401-406. In addition a large number of rare variants have been identified - the test described here only detects the common alleles - Gc 2, Gc1F and Gc1S. Reynolds and Sensabaugh (1990) in Polesby et al (eds) Advances in Forensic Haemogenetics, Vol. 3 Springer, Berlin, Heidelberg, New York pp 158-161 compared cDNA sequences of Yang et al (1985) Ploc - Nat - Accad - Sci USA 82:7994-7998 and Cooke N. E. and David E. V. (1985) Serum D - binding protein is a 3rd member of the albumin and alpha-feto protein gene family, J. Clin. Invest. Vol. 76 pg. 2420 - 2429. Although polymorphisms were observed at 4 different sites, the most informative are at codons 416 and 420, where single base changes result in an amino acid change. At triple 416, GAT codes for an aspartic acid residue in the Gc2 and Gc1F phenotypes, whereas GC1S has a glutamic acid residue determined by codon CAG. Amino acid 420 is a lysine residue in the Gc2 phenotype coded by AAG; a threonine residue in both Gc1 phenotypes is coded by ACG.

Four different forward primers were prepared to distinguish between the various polymorphisms (table A, B). These primers were attached at the 5' end to universal primers as described previously.

codon	sequence
Forward primers	
420G/T	ACCAGCTTTGCCAGTTCCR
416C/A	TTCCGTGGGTGTGGCX
Reverse primer	GGCAGAGCGACTAAAAGCAA
	A

-44-

Table A: Sequence of primers used to detect Gc1F, Gc1S and Gc2 polymorphisms. R=G or T; X=C or SA. 420T and 416A were attached to FAM labelled universal primer G; 420G and 416C were attached to JOE labelled universal primer C.

416		420	
		G	T
	A	Gc1F	Gc2
	C	Gc1S	

Table B: The GC phenotypes are dependent upon the codon mutations detected. Note that 416A and 420T do not exist together in coupling. The 420G primer detects Gc1 phenotypes; 420T detects Gc2; 416C detects Gc2 or Gc1F (dependent on codon 420 sequence); 416A detects Gc1S.

Using these details a series of examples were undertaken with two aspects being tested, specificity, and sensitivity. To carry out specificity tests, a series of singleplex reactions were carried out. In Figure 12a primer 416A was shown to be a specific test for the Gc1F polymorphism. Similarly, primer 420G was specific for Gc1 polymorphisms (Figure 12b); 420T was specific for Gc2 polymorphisms (Figure 12c). The system was demonstrated to work with all primers in a cocktail mix, Figure 12d. Furthermore sensitivity of detection was c. 8pg genomic DNA (Figure 12d). A mixture was analysed this demonstrated that mixtures are easily interpreted, the different molecular weights of FAM and JOE confer different molecular weights on the DNA fragments the same size, and this facilitates interpretation. It is easy to distinguish the artefact pull-up from a true allele.

Reaction conditions

The reagent concentrations were the same as described for mitochondrial DNA. Primer concentrations used were 125nM for the locus specific forward primers and the reverse primer. The universal forward primers were at 100nM, and the universal reverse primer at 288nM. Locus specific and universal primers were admixed in a single tube

-45-

reaction. The cycling conditions used were 94 C for 30 sec; 61 C for 30 sec; 72 C for 90 sec for 35 cycles, followed by 72 C for 10min.

All phenotypes were verified by independent analysis using conventional isoelectric focussing.

EXAMPLES - Molecular beacon based

To demonstrate the benefits of the universal primer incorporating beacons, a series of experimental designs were investigated using a Roche light cycler to analyse the fluorescence resulting. The experiments were performed using a two-tube or branched PCR approach to give first and second round amplifications.

The following protocol for amplification in both the first and second round PCRs was used for the various examples which follow.

1st Round PCR

Primer	Optimum Concentration
1 st round forward primer	200nM
1 st round reverse primer	200nM
MgCl ₂	1.5mM
BSA	0.2μl/reaction
PE Buffer	10%
d NTP's	200μM
Taq Gold	0.1μl/reaction
DNA	2ng/reaction
SDW	

Reaction Volume = 20μl

NB Final [MgCl₂] = 3mM

1st round amplification was carried out as follows:-

Denaturation	95°C for 10 mins	}	for 36 cycles in each case
Amplification	95°C for 10 sec's		
	60°C for 10 sec's		
	72°C for 10 sec's		
Cool	35°C for 1 min		

2nd Round PCR

Primer	Optimum Concentration
2 nd round forward beacon	200nM
2 nd round reverse primer	200nM
MgCl ₂	1.5mM

-46-

BSA	0.2 μ l/reaction
PE Buffer	10%
d NTP's	200 μ M
Taq Gold	0.1 μ l/reaction
1 st round PCR product	2 μ l/reaction
SDW	
Reaction Volume = 20 μ l	NB Final [MgCl ₂] = 3mM

2nd round amplification was carried out as follows:-

Denaturation	95°C for 10 mins	} for 24 cycles in each case
Amplification	95°C for 5 sec's	
	60°C for 10 sec's	
	72°C for 10 sec's	
	50°C for 10 sec's	
Cool	35°C for 1 min	

A reading of fluorescence was taken at the end of each 50°C hold.

Example A

In this example, DNA was amplified with primers directed towards the Gc1s polymorphism in the 1st round. Subsequent 2nd round PCR, amplification was directed to this 1st round product using increasing concentrations, (200nM, 400nM, 600nM, 800nM and 1000nM), of 2nd round primers. The respective first round forward universal primers and reverse primer and second round universal primer incorporating a molecular beacon and universal reverse primer are as listed below. The stem part of the molecular beacon is underlined in the universal primer sequence incorporating the molecular beacon.

The results, displayed as a graph of fluorescence against 2nd round amplification cycle number for the various concentrations of 2nd round primer and a Gc1s polymorphism 1st round universal primer are displayed in Figure 14. Very substantial discrimination between the polymorphism featuring samples and the controls can be seen.

In order to confirm the finding that the universal primer beacons amplify DNA specifically all PCR products were run on 3% Nusieve minigel. As can be seen from the typical results set out in Figure 15 for the minigel image, no product is seen for the SDW control and two bands are present for each DNA sample, 1a and 1b. The upper band sizes is at 112bp which is consistent with Gc1s amplified product. The additional band is approximately 12bp smaller and is thought to be the result of single stranded 2nd round

-47-

product where the beacon has adopted its intra molecular secondary structure. However, this additional band does not interfere with interpretation and appears not to compromise the PCR in any way.

1st round primers

Primer name	Primer sequence
Gc1s uni 9 G	CGACGTGGTGGATGTGCTAGGTTCCGTGGGTGTGGCC
Gc reverse uni 11	TGACGTGGCTGACCTGAGACGGCAGAGCGACTAAAAGCAAA

2nd round primers

Primer name	Primer sequence
Universal G beacon	F- <u>ACGCGCTCTCTTCTTTTGC</u> GCG-Q-CGACGTGGTGGATGTGCTAG
Universal 11 reverse	TGACGTGGCTGACCTGAGAC

Example B

The molecular beacon approach is capable of amplifying 2ng of DNA. Using exactly the same approach, an experiment was performed to gain a measure of the sensitivity of the method. Various dilutions of a Gc1s + ve DNA sample were made and amplified as described in the previous example. The results are shown in Figure 16 which again illustrates a graph of fluorescent verses 2nd round amplification cycle number for various concentrations of Gc1s + ve DNA when compared with a control, SDW.

Whilst there is some degree of non-specific fluorescence seen in the SDW control, this is clearly distinguishable even from the 0.02ng Gc1s DNA + ve and therefore has no effect with respect to interpretation of samples that contain even substantially sub-nanogram quantities of DNA. Very low levels of sample can thus be successfully analysed using this technique.

Example C

As demonstrated above the universal G beacon can clearly identify and amplify DNA that is primarily amplified using the Gc1s primer. To demonstrate its truly universal application, both universal G and universal C beacons are demonstrated in this example as being able to amplify other first round products derived from different 1st round

-48-

amplification primers. To this end, 1st round amplifications were performed using Gc1, Gc2, Gc1s and Gc1f primers as specified below and the 2nd round amplification was carried out using universal G or C beacon together with universal 11 reverse. The 2nd round amplification cycle numbers were reduced to 20 in order to minimise overamplification.

1st round primers

Primer name	Primer sequence
Gc1 uni 9 G	CGACGTGGTGGATGTGCTAGACCAGCTTTGCCAGTTCCG
Gc2 uni 9 C	CGACGTGGTGGATGTGCTTACCAGCTTTGCCAGTTCCT
Gc1s uni 9 G	CGACGTGGTGGATGTGCTAGGTTCCGTGGGTGTGGCC
Gc1f uni 9 C	CGACGTGGTGGATGTGCTTCGTTCCGTGGGTGTGGCA
Gc reverse uni 11	TGACGTGGCTGACCTGAGACGGCAGAGCGACTAAAAGCAAA

2nd round primers

Primer name	Primer sequence
Universal G beacon	F- <u>ACGCGCTCTTCTTCTTTGCGCG</u> -Q-CGACGTGGTGGATGTGCTAG
Universal C beacon	F- <u>ACGCGCTCTTCTTCTTTGCGCG</u> -Q-CGACGTGGTGGATGTGCTTC
Universal 11	TGACGTGGCTGACCTGAGAC

The results from this example are displayed in Figures 17a, 17b, 17c and 17d. The results from 17a and 17b show sample 1 to have a Gc1 polymorphism and sample 2 to have a Gc2 polymorphism. Example 3 is an SDW control which produces a negative result throughout. There are two known alleles associated with the Gc1 polymorphism, designated Gc1s and Gc1f and an individual can be either homozygote for Gc1s or Gc1f, or heterozygote for Gc1s1f. As shown by Figures 17c and 17d, sample 1 amplifies in both instances and therefore indicates a Gc1s1f phenotype. As sample 2 does not amplify with Gc1s primer, this confirms its assignment as Gc2 phenotype. The amplification of sample 2 with Gc1f primers is due to the specific binding site base sequence of the Gc region. An individual who is Gc1, has a C base position 2 of codon 420, whilst a Gc2 individual has an A base at the same position. Gc1 individuals are differentiated into Gc1s and Gc1f phenotypes characterised by the base core at position 3 of codon 416, being a G or T

-49-

respectively. Individuals who are Gc2 only have a T base at position 3 in codon 416 and therefore the Gc1f primer is able to bind complimentary at this position despite having a base mismatch with codon 420. Such occurrences result in non-specific amplification as illustrated in more detail in Figure 18.

Example D

Further investigation were carried out on another locus, namely Amelogenin. The 1st round primers are as follows:-

1st round primers

Primer name	Primer sequence
Amelo X	CGACGTGGTGGATGTGCTTCTGAGCCAATGGTAAACCTGCC
Amelo Y	CGACGTGGTGGATGTGCTAGTGAGCCAATGGTAAACCTGCA
Amelo reverse	TGACGTGGCTGACCTGAGACCATAGGAAGXGTACTGGTGAGAAACA

2nd round primers

Primer name	Primer sequence
--------------------	------------------------

2nd round primers

Primer name	Primer sequence
Universal G beacon	F- <u>ACGCGCTCTCTTCTTTGCGCG</u> -Q-CGACGTGGTGGATGTGCTAG
Universal C beacon	F- <u>ACGCGCTCTCTTCTTTGCGCG</u> -Q-CGACGTGGTGGATGTGCTTC
Universal 11	TGACGTGGCTGACCTGAGAC

Amplification conditions are the same as before. A male sample was amplified with the Amelo Y primer and detected using universal G beacon. The results are illustrated in Figure 19.

Once again, product was run on a 3% Nusieve minigel to confirm that the amplified product was the correct size. As seen before there were two bands present that differed by 12bp. The longer of the two bands confirmed that the amplified product was the correct size and thus proved that the additional band was likely to be the result of intramolecular modification as previously outlined. The presence of this phenomenon

-50-

with two different loci implies that it is a feature of the universal molecular beacon primer structure rather than artifact derived from mispriming.

Clearly, the universal reporter primer principle (URP), is able to amplify multiple alleles within a given locus to facilitate accurate genotyping. Furthermore, they enable multiple loci to be determined making them truly universal.

Example E

Whilst the techniques sensitivity has been demonstrated with respect to being able to amplify and detect sub-nanogram levels of DNA, this examples takes the demonstration one step further in the context of sensitivity and specificity of the technique with respect to mixtures.

In the following example, 2 DNA's were mixed together in ratios of 1:2, 1:4, 1:6, 1:8, 1:16, 1:50, 1:100 and 1:300. In each case, the total amount of DNA present was 2ng. The mixtures were prepared with the major component coding for the Gc2 polymorphism and the minor component coding for the Gc1s polymorphism. Amplified with forward primers, either Gc2 or Gc1s respectively and a reverse Gc primer, the cycling condition were as described earlier. In the 2nd round, 1st round product was amplified with either universal C beacon or universal G beacon respectively, together with universal 11. Cycling conditions were as before however, the actual number of cycles for the 2nd round was reduced to 12. the results are shown in the graphs of Figures 20a and 20b.

In Figure 20a all sample have amplified with the exception of the male -ve control and SDW. All other sample have amplified at the same time and rate. Slight differences between samples with respect to their absolute fluorescence is attributed to the relative increase in major component as the ratios shift from 1:2 to 1:300, (minor component to major component respectively).

Figure 20b shows that as the ratio of minor component decreases, the time taken for these sample to begin amplification also decreases, as does the rate of amplification. At 1:300, the amount of starting material is approximately 6pg but the increase in fluorescence attributed to DNA amplification in this mixture is clearly defined above baseline.

-51-

As is demonstrated we can successfully discriminate and amplify both the major component and minor component in mixed DNA source samples successfully over a wide range of minor to major ratios.

These results thus demonstrate the universal reporter primer principle and the following advantages.

- a) The universal molecular beacons like the universal primers do not prime human DNA.
- b) A two stage reaction or branched PCR may be the method of choice and this improves specificity.
- c) Mispriming is negligible, helped by the branched PCR approach and this virtually eliminates background noise.
- d) The ability of the universal beacons to amplify multiple loci in the second PCR means that only this one primer set is required. This has cost saving implications.
- e) The universal molecule beacons facilitate real time detection of PCR product without the need for additional post PCR elucidative techniques ie. polyacrylamide gel electrophoresis (PAGE).
- f) There is a distinct advantage with using this technique when sample quantity is limited, (a problem commonly encountered in forensic biology when sub-nanogram levels of material may be recovered). 1st round PCR produces significant product. Aliquots of this product can then be used for several subsequent second stage PCR reactions.
- g) The universal beacons are extremely sensitive with respect to mixtures and can accurately determine a minor component equal to 1:300 dilution (6pg).

EXAMPLE - Distinct universal primer portions, attachments to glass slides and hybridisation

As described above, particularly powerful tests can be obtained if the first set of primers include within their universal primer portions, distinguishing portions, with the subsequent amplification primers being attached to glass slides and then subjected to hybridisation based analysis.

More details of the experimental method for such a system follow.

-52-

PCR CONDITIONS

For multiplex reactions

Locus specific primer concentrations	10nM-40nM
Universal reporter primer concentrations	100-1000nM
Mg-ion concentration	1.5nM
Buffer	Perkin Elmer PCR buffer X1 concentration
DNA	1nG
Reaction volume	50uL
dNTP's	200uM
Taq gold	1.25U/50uL

Cycling conditions:

95°C for 10 minutes - Taq activation

94°C for 30 seconds

60°C for 30 seconds

72°C for 90 seconds

performed for 32 to 40 cycles.

Hybridisation conditions

Hybridisation was performed at a temperature of 40°C to 60°C to promote hybridisation of the probes. The hybridisation solution consists of a solution containing - 0.5M di sodium hydrogen orthophosphate, pH7.2, 1mM EDTA and 7% SDS.

Hybridisation probes were dissolved in the solution at a concentration of 0.1-5uM. A volume of 80-120ul of hybridisation solution was applied to the slide which was then covered with a cover slip, placed in a humid atmosphere at 40°C to 60°C for two hours.

After hybridisation the slides were washed in the solution of 0.1 to 4 x 55C and 0.1%SDS at 60°C t 75°C.

When dry, the slides are scanned with a molecular dynamics erase scanner in order to detect the fluorescent probe.

Multiplex analysis

Such systems are suited for use in multiplex conditions. The amplified products are spotted on to a slide as previously described, and each spot is then visualised for specific loci using successive hybridisations. Other loci present within the amplification products will not hybridise and hence will not contribute to the detected results during each of the respective hybridisations and investigations. After the determination of a particular locus, slides can be immersed in an appropriate solution to remove all of the hybridisation probes bound to the target DNA present from that hybridisation. The covalent bonding of the strands means that they are maintained on the slide.

Suitable controls can be used to demonstrate that DNA is actually present in each spot. Controls comprising a mixture of oligonucleotides that are complementary to each universal primer portion utilised (omitting the specific primer region) are envisaged. The slide would hybridise with the control probes, and the signal would be proportional to the amount of DNA actually present in the spot.

To obtain the full benefits of the present invention's technique it is desirable to deploy it in multiplex reactions so that a large number of loci can be considered simultaneously. A problem with multiplex reactions in general is that it is difficult to prevent significant primer dimer formation as different primers are used, and as the concentration of primers increases within the reaction mix. Once a primer dimer forms the PCR reaction tends to produce primer dimer preferentially and hence the reaction becomes very inefficient. Primer dimer forms arise when primers are themselves complementary to other primers in the multiplex reaction mix.

Two techniques to address this issue have been developed by the applicants and those will now be described.

In the first technique, locus specific primers are used in a first set of primers and universal primers are used in a second set. Both sets are incorporated into a single reaction, however. The second set of primers are provided at a far higher concentration, however, than the locus specific primers of the first set. Concentrations of 400nM-4000nM compared with 10-200nM are deployed in the reaction mix. The universal primers of the second set employ dye labels as previously mentioned.

The key to avoiding primer dimer formation is the temperature employed in the PCR reaction during the various cycles.

-54-

During the first two cycles of PCR amplification only the locus specific portion of the locus specific universal primers initiate priming. Once the complimentary sequence to the universal portion has formed as a result of the first two cycles in subsequent rounds of amplification the entire first stage primer is able to prime as now both the locus specific and universal primer portions of the primer are present in the PCR product.

Increasing the annealing phase temperature to 72°C to 76°C during the early amplification cycles, for instance cycles 4 to 24, means that priming due to the first stage primer occurs but that annealing of the second stage primer, the universal primer carrying the dye label, is inhibited entirely. The annealing temperature is simply too high for the second stage primer to anneal. Once sufficient amplification has occurred, the annealing temperature for subsequent stages is reduced to around 60°C and this allows the universal primers to prime. As a consequence subsequent amplification cycles result in priming of this second stage primer also and hence in the incorporation of the fluorescent dye labels into the amplification products. Usually only a couple of cycles are needed to effect this part of the process.

The overall result of the use of low concentrations and the first primer stage at high annealing temperatures, with the universal primers switched off as a result is that primer dimer formation is minimised. The impact of careful temperature control during annealing on the process is highlighted in Figure 24 where at 70 and 72°C for the annealing temperature very substantial primer dimer formation occurs. By 74°C for the annealing temperature, this is decreased very substantially, and at 76°C is almost eliminated.

The appropriate annealing temperature to obtain such a benefit for a primer can be established through a series of experiments at different annealing temperatures.

Providing the amplification process in this way, with relatively low concentrations of the first stage primers, is also beneficial in addressing the variation in efficiencies of amplification which exist between primers. Starting from low concentrations means that the most efficient primers will be rapidly depleted as the reactions progress whereas the less efficient primers will be less rapidly depleted. Over a suitable number of reaction cycles this allows the less efficient primers to catch up and provide generally equivalent levels of amplification.

As an alternative or additional technique to assist in the avoidance of primer dimer formation, it is possible to carefully design the second stage universal primers. Primer

-55-

dimer formation occurs if two primers can hybridise with one another such that the 5' ends overhang, thereby allowing extension from the 3' end, Figure 25a. Primer dimer formation cannot occur from 3' end overhangs, Figure 25b. As a consequence the present invention aims to ensure that the primers can only possibly hybridise with one another to give 3' end overhangs.

The design process for the universal primers involves the provision of self complimentary 3' end and 5' end sequences. As shown in Figure 25c, two universal primers are provided to address a biallelic situation which is being analysed. If all four possibilities for the SNP apply, then four similarly provided universal primers can be employed.

The primers consist of the two complimentary end portions and an allele specific sequence provided therebetween. If desired for use as the first stage primer, a locus specific primer portion can be attached to either 3' end of these primers. Such sequences are illustrated in Figure 11 above and this is the reason why the discriminating portion of the further portion is an intervening sequence rather than provided at the end.

Examples - Absence of Second Stage Reverse Primer

As a refinement of some of the examples discussed above, analysis using the same primers as described on page 38 above was performed, but without the universal 11 reverse primer.

The PCR conditions and cycles are as follows:-

Primer	Optimal Concentration	
	Beacon Assay	Sybr Green Assay
1 st round forward primer	40nM	40nM
1 st round reverse primer	100nM	100nM
2 nd round forward primer	200nM	-
2 nd round reverse primer	-	- (OMITTED IN THIS METHOD)
Sybr Green	-	10%
MgCl ₂	1.5mM	1.5mM
PE Buffer	10%	10%
d NTP's	200μM	200μM

-56-

Taq Gold	0.1U/reaction	0.1U/reaction
DNA	2ng/reaction	2ng/reaction
SDW		

Reaction Volume = 20 μ l

Denaturation	95°C for 10 mins	
Amplification	95°C for 30 sec's	for 50 cycles
1	60°C for 30 sec's	
	72°C for 30 sec's	
	50°C for 30 sec's	
Cool	35°C for 1 min	

In this example, the following amounts of DNA, (1ng, 0.5ng, 0.05ng and 0.01ng), were amplified with the redesigned Gc1s primers in the presence of the molecular beacon assay, a graph showing fluorescence against cycle number for the various sample sizes amplified is illustrated in Figure 26

The beacons that incorporate universal 9 have provided a much better result under the revised amplification protocol. All products are sizing at approximately 112bp and there is no detectable presence of primer dimer.

The Ct value is defined as the cycle number at which the reporter fluorescence increases above a base line or threshold level. This level is usually set at a position directly at the start of exponential amplification when PCR amplification is optimised. This assay can unambiguously differentiate <0.01ng (10pg) genomic DNA from SDW. However, note that the DNA negative control comprises 2ng of a Gc2 individual. This is in great excess compared to the quantities of DNA analysed. With this amount of DNA a small amount of mis-priming will inevitably occur. 2ng of Gc2 gives an apparent ct that is equivalent to c.50pg (which is only 1/40th the actual amount). In real casework it is very unlikely that such large quantities of DNA will be used - this means that the test is highly allele specific. Note that the sterile distilled water negatives are completely clean.

-57-

SYBR GREEN ASSAY

The same primers were used as described for the molecular beacon assay except that the molecular beacon itself is not used. All other protocols are described were described in the previous section.

The results in Figure 27 show very similar sensitivity to that previously described for the molecular beacon experiment previously described.

DEMONSTRATION OF NON-5' END OVERHANG PRIMERS FOR SYBR GREEN ASSAY

The same protocols as previously described were followed - the following primers are used:-

Primer set C

Gc1s URP 13.1 G CTAGCTGGTGGCTGTGCTAGGTTCCGTGGGTGTGGCC

Gc reverse URP 13.1 CTAGCTGGTGGCTGTGCTAGGGCAGAGCGACTAAAAGCAAA

Results are shown in Figure 28. Again results are similar to those previously described and achieved good results even for very small samples.

To summarise, we demonstrate that the Sybr Green, a dye, and molecular beacon assays that utilise the principals set out above are able to analyse concentrations of DNA <100pg. Furthermore the reaction shows considerable allele specificity. A negative control that comprises a different allele (Gc2) at a very high level of 2 ng showed a small amount of mispriming with the Gc1s primer at approximately the 10pg level.

The technique of the present invention, as exemplified through its various embodiments in particular, is suitable for analysing very low levels of DNA contained in a mixture. The technique has the ability to pick out DNA from a source which is the minor contributor to a mixture by a long way. Thus, the technique enables mixtures in which one source only contributes to the sample in the ratio 1:5000 to be successfully picked out. This is useful where samples from mixed sources but one party is likely to contribute in a small amount. The technique is also applicable in enabling very small quantities of male DNA to be picked out and analysed in the presence of substantial quantities of female

-58-

DNA. The sample including 0.02% male DNA, or even less, can successfully be analysed using this invention even when 99.98% of the sample is female DNA.

In any cases where low levels of sample are being analysed, the level of sample present can be considered by comparing the fluorescence level arising from the unknown sample against a series of calibration results, forming a curve, with those results being based on different levels of DNA as a portion of the overall DNA sample.

-59-

CLAIMS:

1. A method of investigating single nucleotide polymorphisms in a sample of DNA, the method comprising contacting the DNA containing sample with at least one first set of primers, amplifying the DNA using those primers to give an amplified product, contacting at least a portion of the amplified product with at least one second set of primers, amplifying the DNA using those second set of primers to give a further amplified product and examining one or more characteristics of the further amplified product, one or more of the primers of the first set of primers including a locus specific portion and a further portion, the locus specific portion of one of those one or more of the primers annealing to one side of the SNP under investigation.
2. A method according to claim 1 in which one or more of the primers are provided with an SNP identifying portion, the SNP identifying portion being different for each different primer, the primer with an SNP identity portion which pairs to the SNP, annealing one side of the SNP.
3. A method according to claim 1 or claim 2 in which one or more of the second set of primers includes a second further portion, the second further portion being provided with a sequence equivalent to the sequence of the further portion of one or more of the primers of the first set which are provided with a locus specific portion and a further portion.
4. A method according to any of claims 1 to 3 in which the locus specific portion of the primers of the first set includes a sequence which matches the sequence of the locus sequence in the vicinity of the SNP under investigation, the match between the locus specific portion and sequence of the locus commencing at between one and ten bases to the respective sides of the SNP under investigation.
5. A method according to any preceding claim in which the first set of primers includes a reverse primer and further includes a forward primer for each possible identity of the SNP under investigation.

-60-

6. A method according to any preceding claim in which the further portion of a primer is attached to the locus specific portion of the primer by an SNP related portion.
7. A method according to claim 6 in which the SNP identifying portion and SNP related portion of a primer have equivalent identity.
8. A method according to any preceding claim in which the locus specific portion of the primers in a set are provided with identical sequences in each primer.
9. A method according to any preceding claim in which the further portion includes a sequence which does not match the locus sequence on the locus' 3' side of the locus with sequence matching the locus specific portion of the primer.
10. A method according to claim 9 in which the sequence of the further portion does not anneal to a sequence of any published part of the entire DNA sequence of homo sapiens.
11. A method according to any preceding claim in which the second set of primers includes a reverse primer and further includes a different forward primer for each potential identity of the SNP under investigation.
12. A method according to any preceding claim in which the second further portion is attached to a second SNP identifying portion and / or an SNP repeat identifying portion.
13. A method according to any preceding claim in which the second further portion includes a sequence which pairs to the sequence of the amplified product in the vicinity of the SNP identifying portion and / or SNP repeat related portion.
14. A method according to any preceding claim in which the sequence of the second further portion does not anneal to the sequence of any published part of the DNA sequence of homo sapiens.

-61-

15. A method according to any preceding claim in which the SNP repeat identifying portion and / or second SNP identifying portion is a single nucleotide or two nucleotides, at least one of the nucleotides being identical to the SNP identifying portion and / or SNP related portion of a primer of the first set.
16. A method according to any preceding claim in which a plurality of first sets of primers are provided to amplify a plurality of SNP loci, the amplification products resulting being of different lengths.
17. A method according to any preceding claim in which one or more characteristics of the further amplified products are investigated by means of the presence and / or absence of a distinctive unit in the further amplified product.
18. A method according to claim 17 in which the distinctive unit is a dye, dye label, colour producing molecule, molecular beacon, emitter of radiation, characteristic isotope.
19. A method according to claim 17 or claim 18 in which the distinctive unit is provided at the 5' end of the forward primers, a different distinctive unit being provided for each forward primer of the second set.
20. A method according to any of claims 17 to 19 in which the distinctive unit is indicative of the nucleotide presence of the SNP.
21. A method according to any preceding claim in which the further portion of at least one of the forward primers of the first set is different from the further portion of at least one of the other forward primers of the first set, at least in part.
22. A method according to claim 21 in which the forward primers are different from one another with respect to at least 25% of the nucleotides forming the further portion of the forward primers.
23. A method according to claim 21 or claim 22 in which the distinguishing portion is provided at an intermediate location within the sequence of the further portion.

-62-

24. A method according to any preceding claim in which the further portion of one or more of the primers in the first set is provided with one or more portions which correspond with one or more portions in the further portion of one or more of the other primers of the first set.
25. A method according any of claims 21 to 24 in which the nucleotides of the further portion of the forward primers are equivalent to the nucleotides of the other forward primers, outside the distinguishing portion of the further portion.
26. A method according to any preceding claim in which the first and second set of primers are present together and in which the concentration of the second set of primers is provided in a ratio relative to the concentration of the first set of primers of at least 5:1.
27. A method according to any preceding claim in which the first and second set of primers are present together, the first set of primers is provided at a concentration of between 10 and 200nM and the second set is provided at a concentration of between 400 and 4000nM.
28. A method according to any preceding claim in which the first and second set of primers are present together and the annealing temperature for at least some of the cycles of the amplification process is such that at least 80% of the second set of primers remain single stranded.
29. A method according to claim 28 in which the annealing temperature is so provided and used at least in cycles 3 to 30.
30. A method according to any preceding claim in which an annealing temperature is used in at least the last two cycles, the annealing temperature allowing at least 80% of the second set of primers to anneal.
31. A method according to any preceding claim in which the annealing temperature is at least 72°C for cycles 3 to 30 of the amplification process.

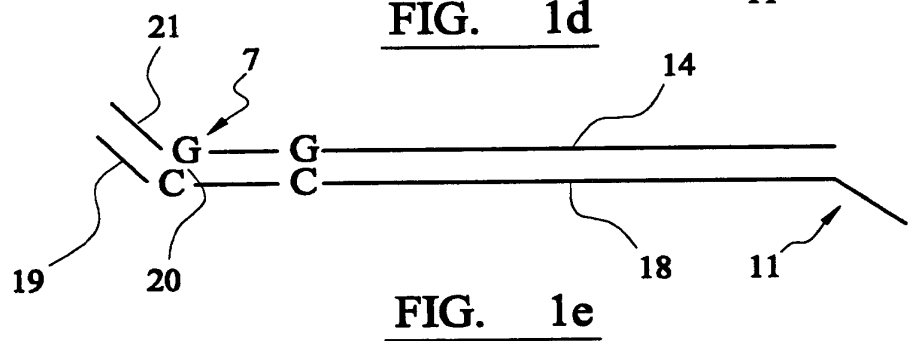
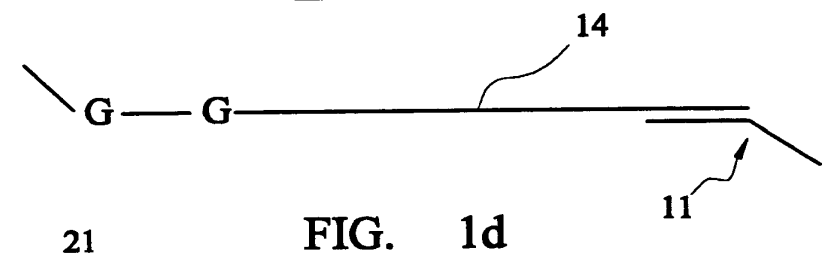
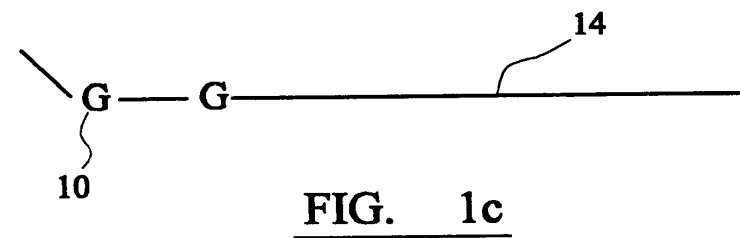
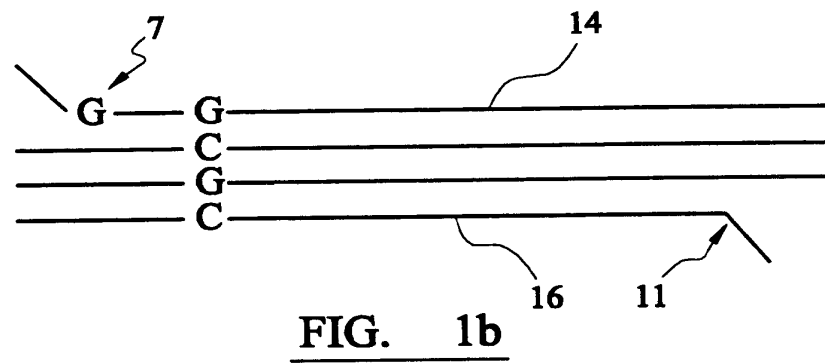
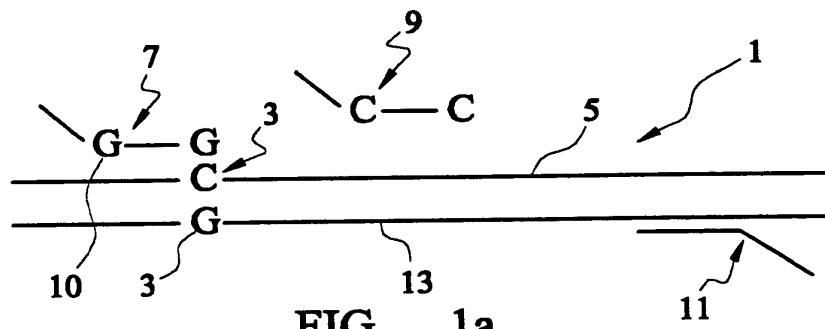
-63-

32. A method according to any preceding claim in which the annealing temperature for at least the last two cycles of the amplification process is 62°C or less.
33. A method according to any preceding claim in which the amplification products of two or more first sets of primers and one or more second sets of primers are separated from one another using electrophoresis.
34. A method according to any preceding claim in which the further amplified product is contacted with one or more components retained on a solid support, the one or more components having a sequence which anneals with at least part of the sequence of one of the further amplified products.
35. A method according to claim 34 in which the retained component anneals with the further amplified product up to the base before the base which is the SNP side.
36. A method according to claim 34 in which the retained component anneals to the further amplified product along the sequence corresponding to the locus specific portion and further portion of the further amplified product.
37. A method according to any of claims 1 to 34 in which a plurality of different retained components, preferably PCR products and / or oligonucleotides, are provided at discrete locations on a support, different retained components annealing to different further amplified products.
38. A method according to claim 37 in which the retained component and annealed further amplified product are contacted with one or more further components to introduce a distinctive unit.
39. A method according to claim 37 in which the retained component and annealed further amplified product are contacted with one or more additional components, the one or more additional component being one or more further oligonucleotides which include a distinctive unit.

-64-

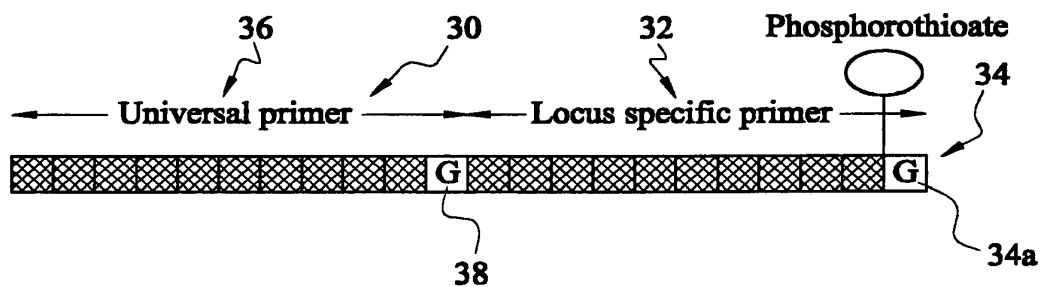
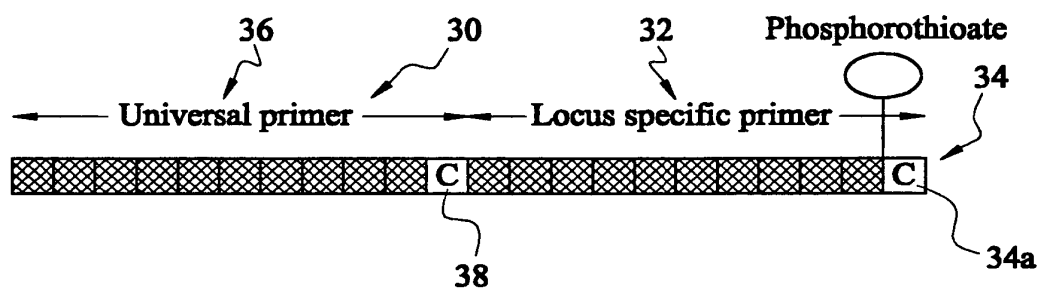
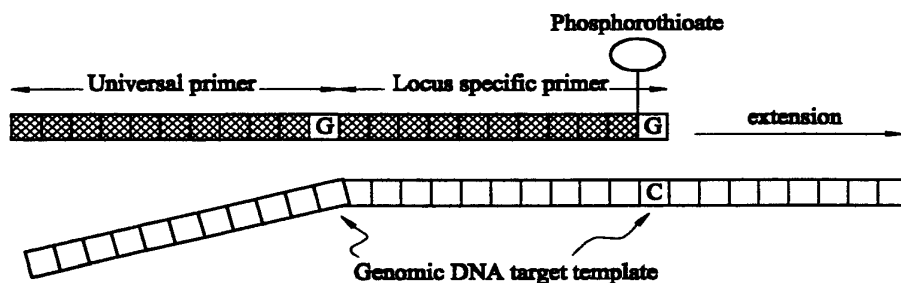
40. A method according to claim 39 in which the end base of the further oligonucleotide is one of the four possible identities for the SNP.
41. A method according to any of claims 1 to 34 in which the further amplified product includes an attachment unit and the attachment unit facilitates attachment of the further amplified product to a solid support.
42. A method according to claim 41 in which the attached further amplified product is contacted with one or more probes having different sequences from one another, at least in part.
43. A method according to claim 41 or claim 42 in which each probe has a common sequence portion to each other, the common sequence portion corresponding in sequence to the locus specific portion of the further amplified product.
44. A method according to any of claims 41 to 43 in which the probes incorporate at least one different sequence portion compared with one another, the different portion of at least one of the probes corresponding to the further primer portion sequence of the further amplified product.
45. A method according to any of claims 41 to 44 in which contact of the probes with the further amplified product results in hybridisation of one of the probes to the further amplified product, each probe having a distinctive unit relative to one another.
46. A plurality of primers for investigating single nucleotide polymorphisms in the sample of DNA, the plurality of primers comprising two or more primers of a first set of primers and / or two or more primers of a second set of primers, the primers having one or more of the features provided in claims 1 to 45.

-1/27-



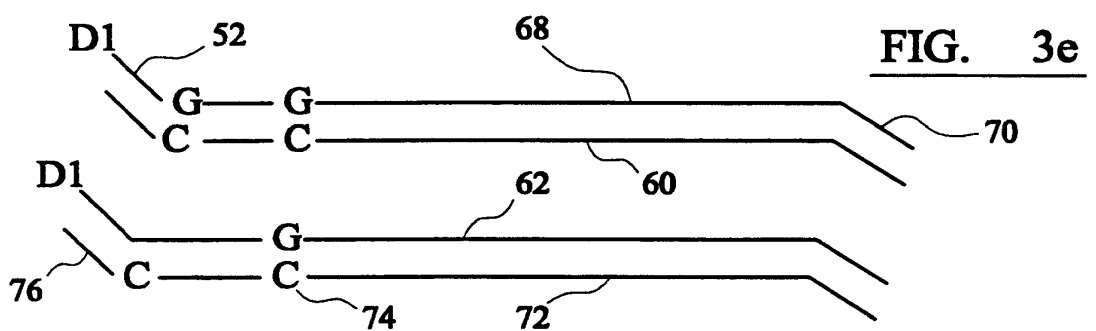
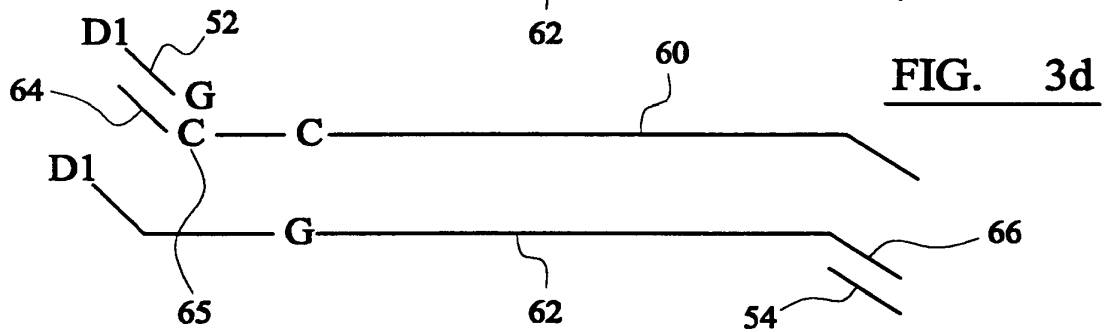
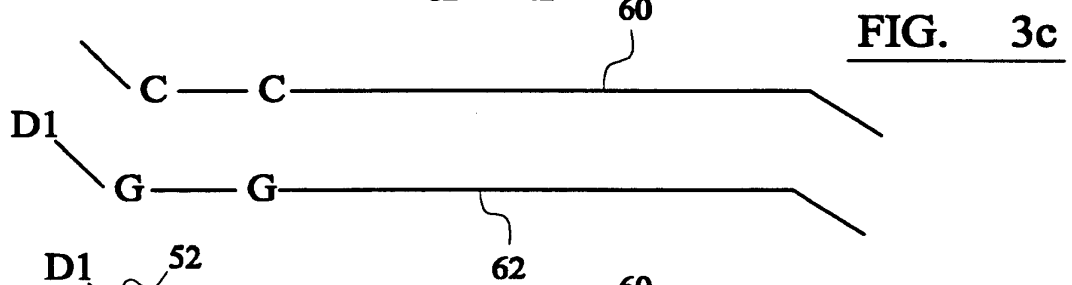
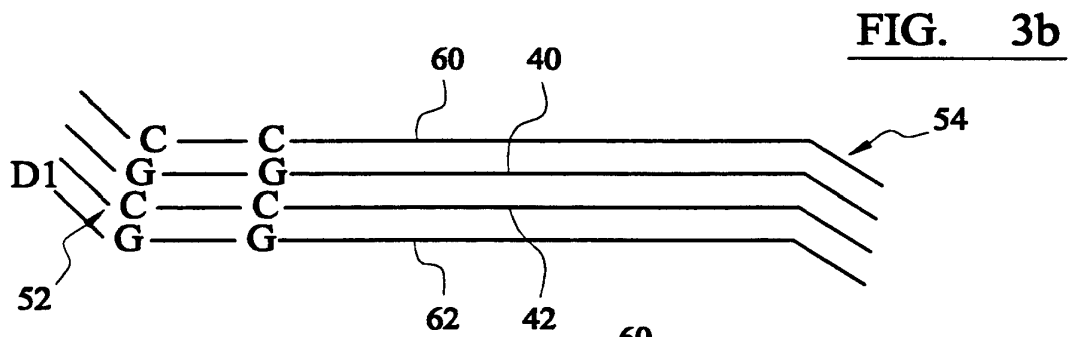
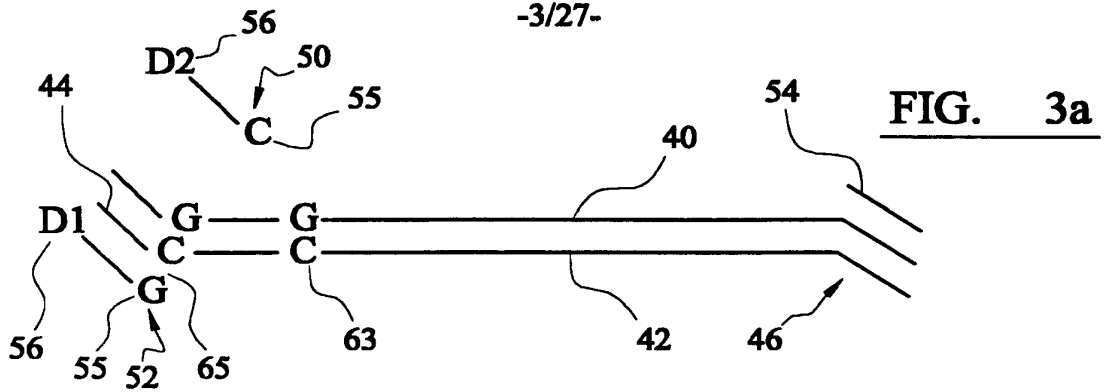
SUBSTITUTE SHEET (RULE 26)

-2/27-

FIG. 2aFIG. 2bFIG. 6

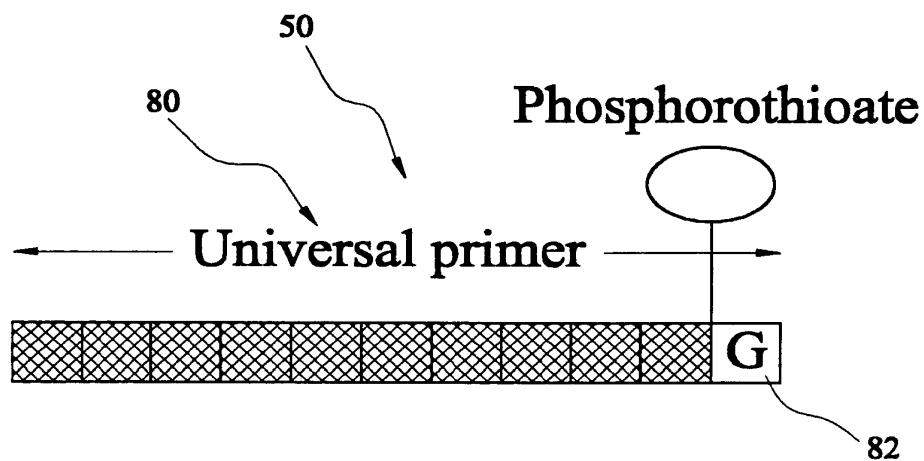
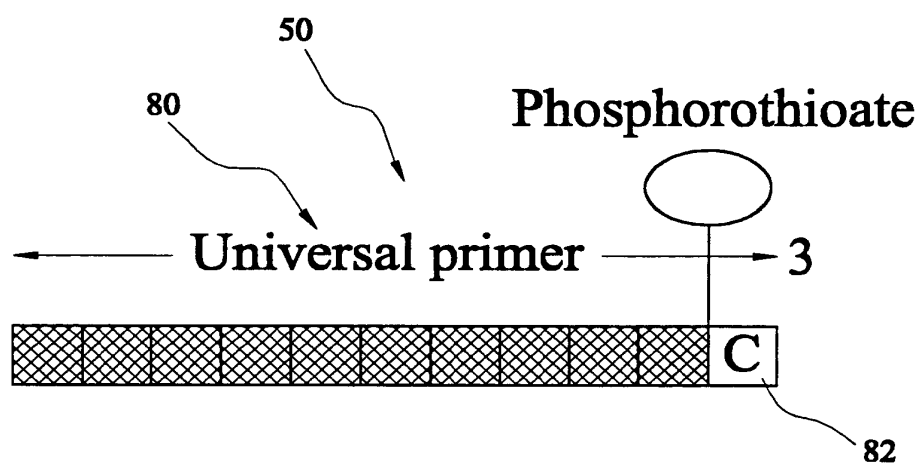
SUBSTITUTE SHEET (RULE 26)

-3/27-

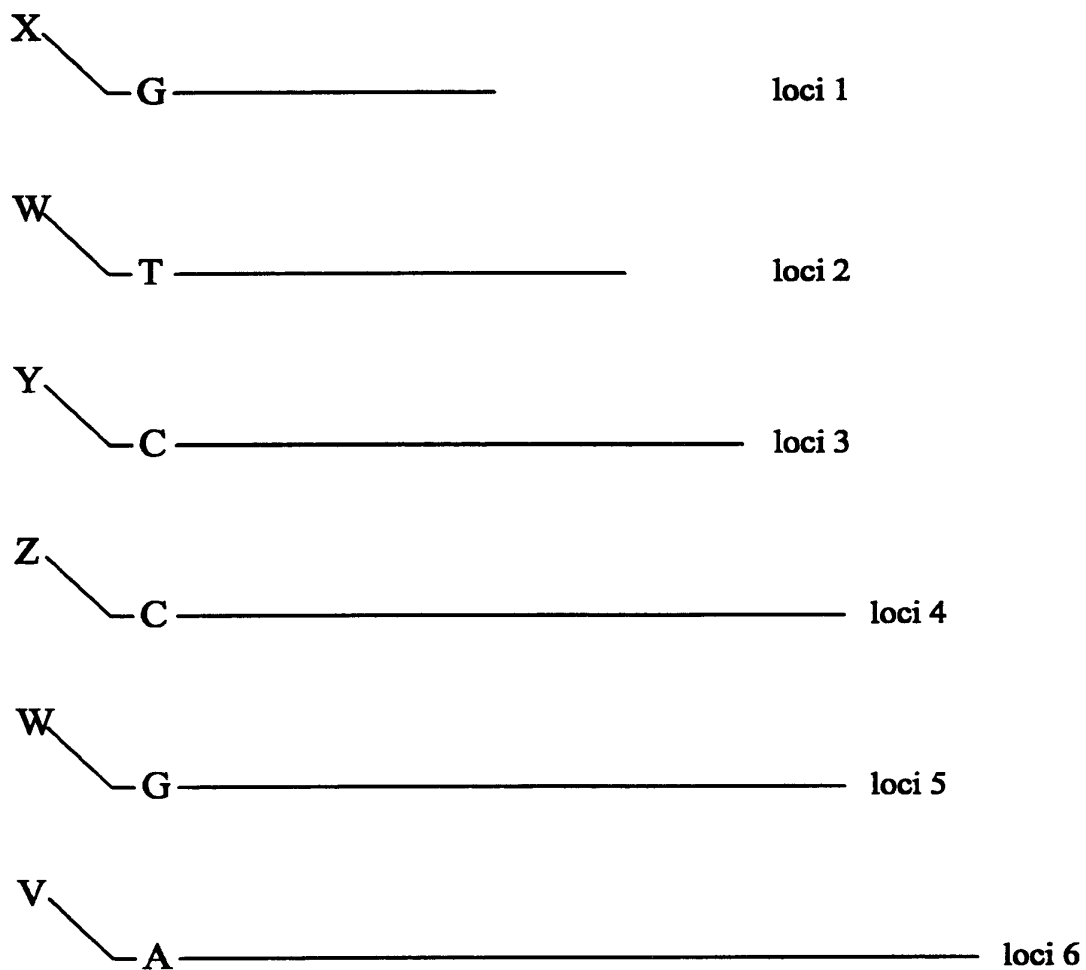


SUBSTITUTE SHEET (RULE 26)

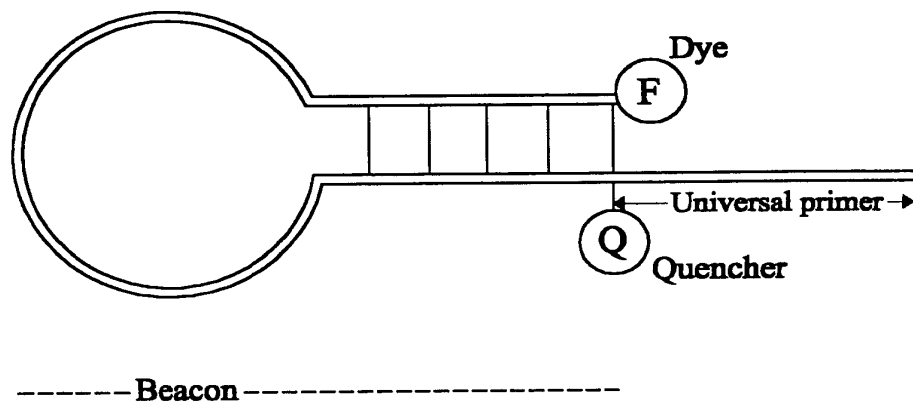
-4/27-

FIG. 4aFIG. 4b

-5/27-

FIG. 5

-6/27-



F= dye label(FAM or JOE) ; ‡ = Quencher - Methyl red; R = AG or TC

FIG. 7a

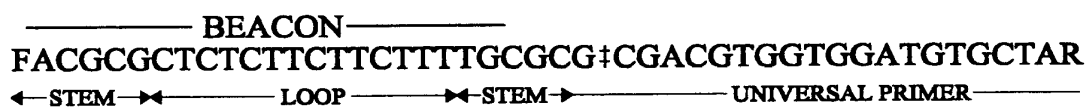


FIG. 7b

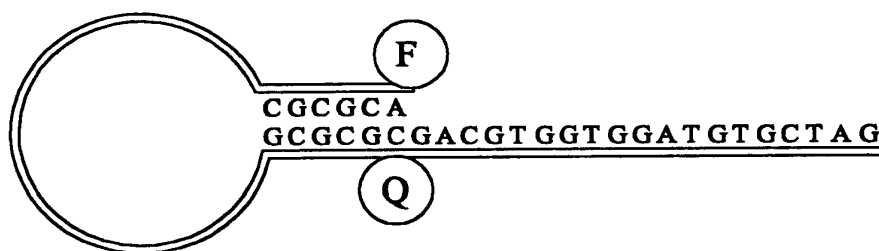


FIG. 7c

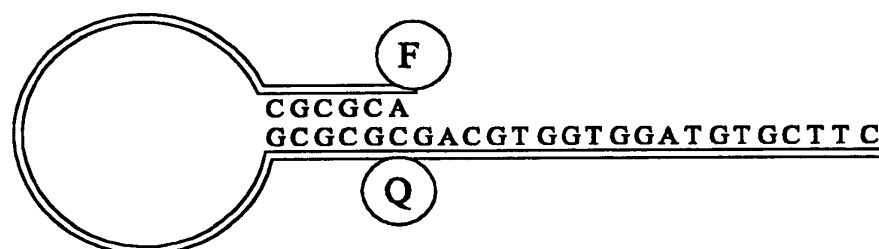
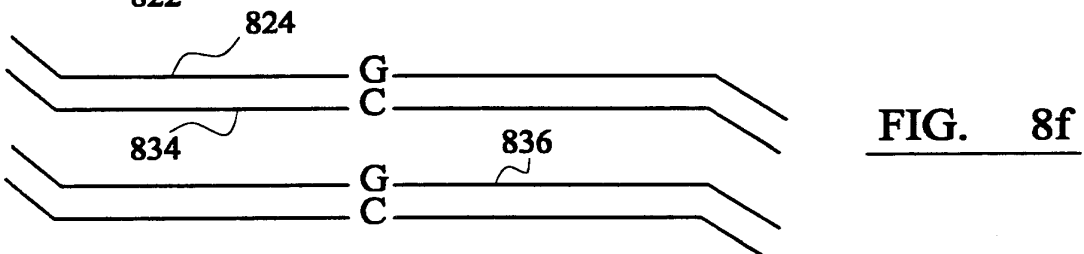
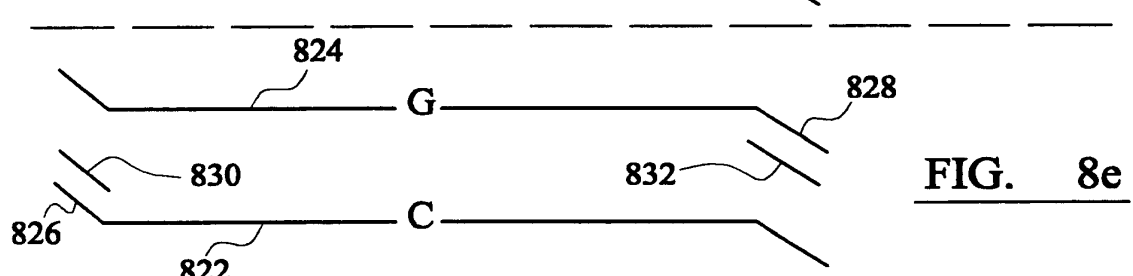
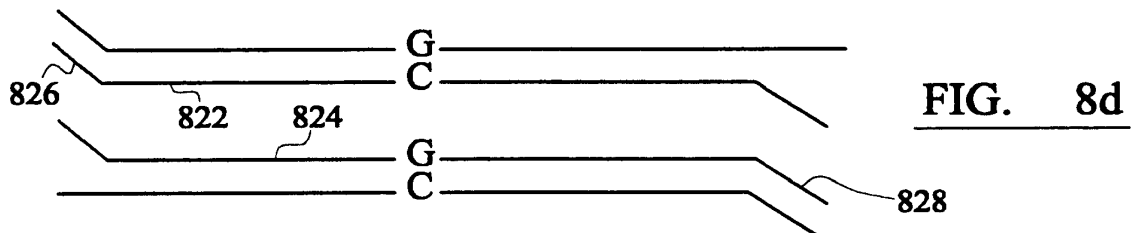
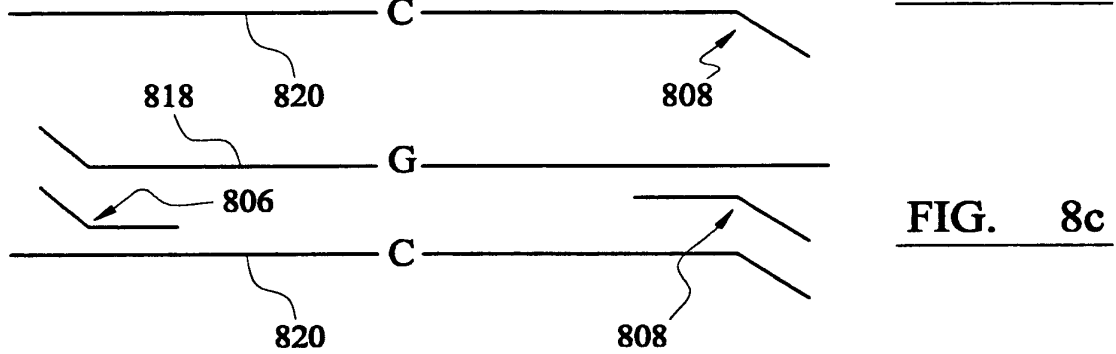
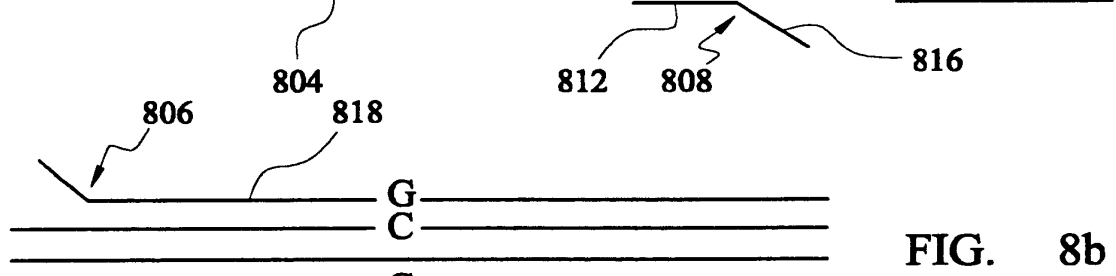
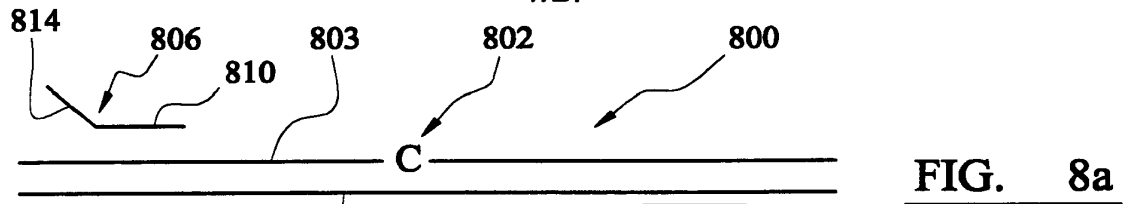


FIG. 7d

-7/27-



SUBSTITUTE SHEET (RULE 26)

-8/27-

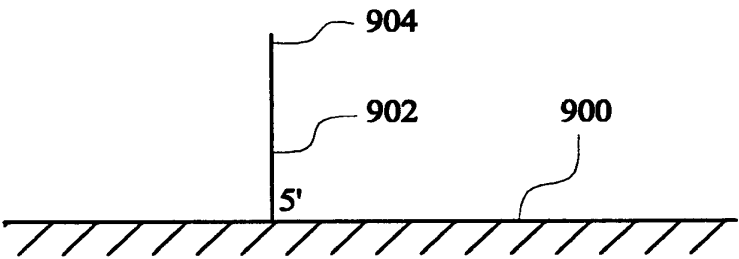


FIG. 9a

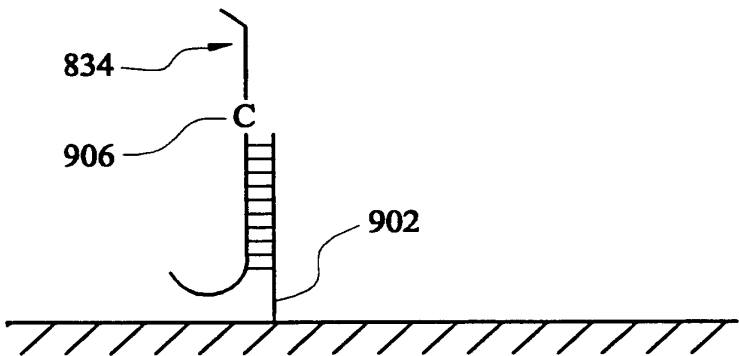


FIG. 9b

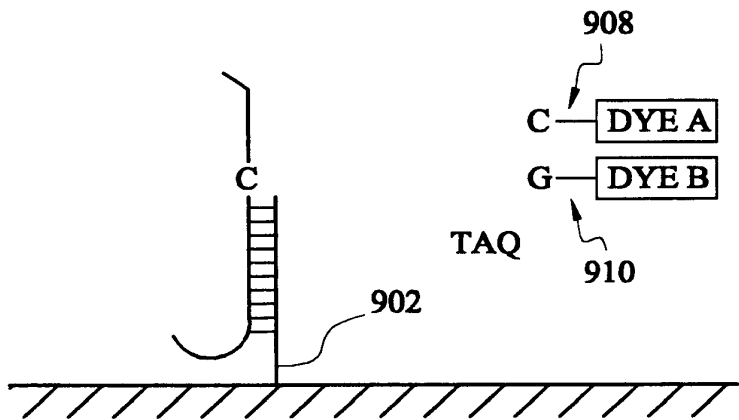


FIG. 9c

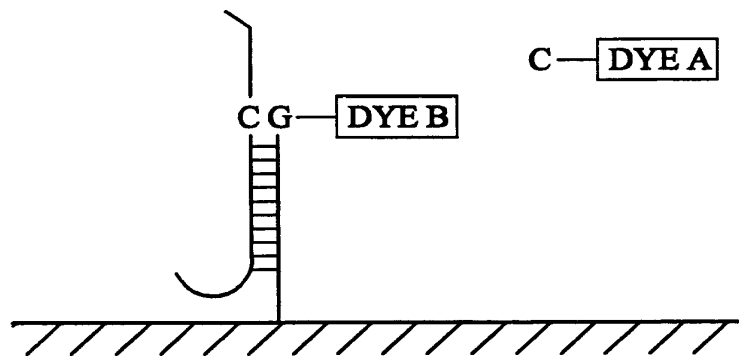


FIG. 9d

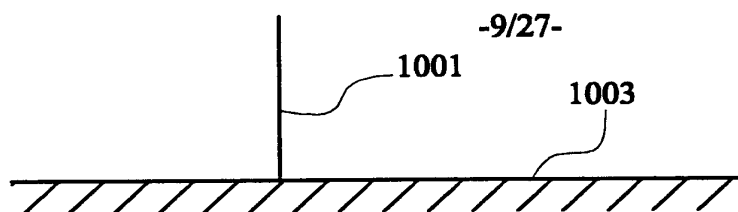


FIG. 10a

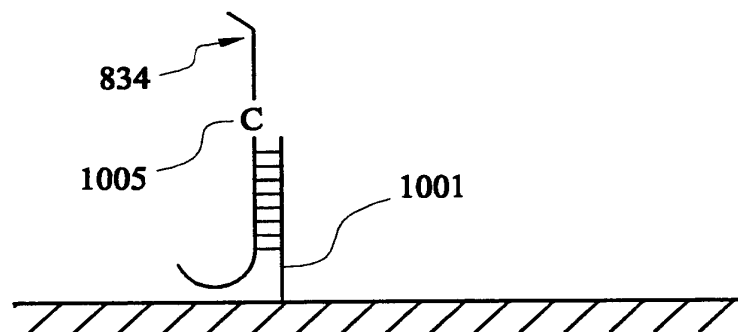


FIG. 10b

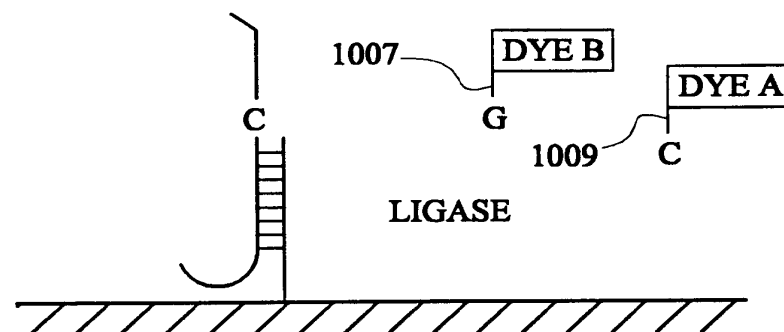


FIG. 10c

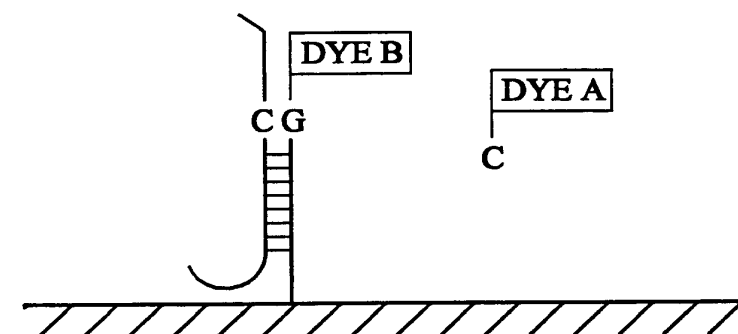


FIG. 10d

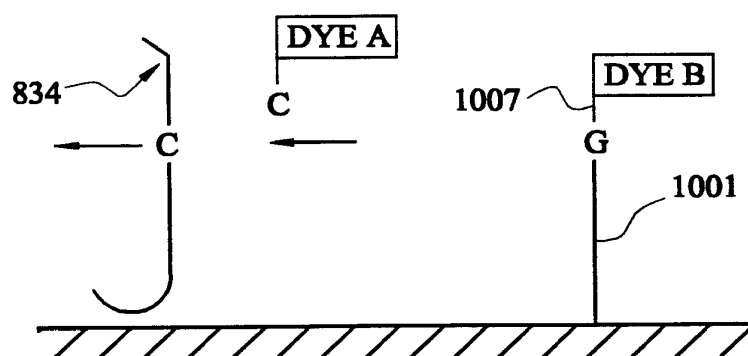
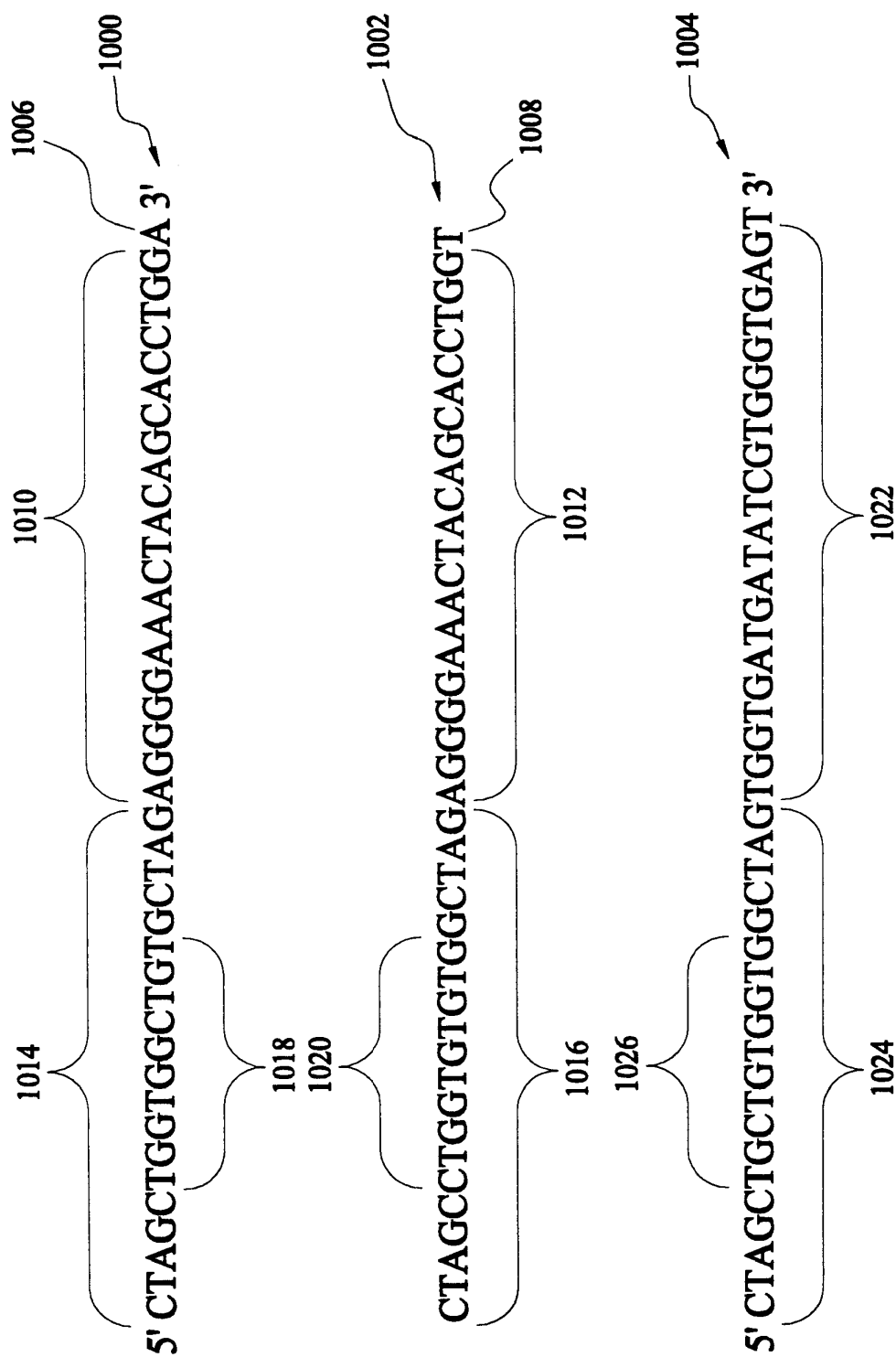


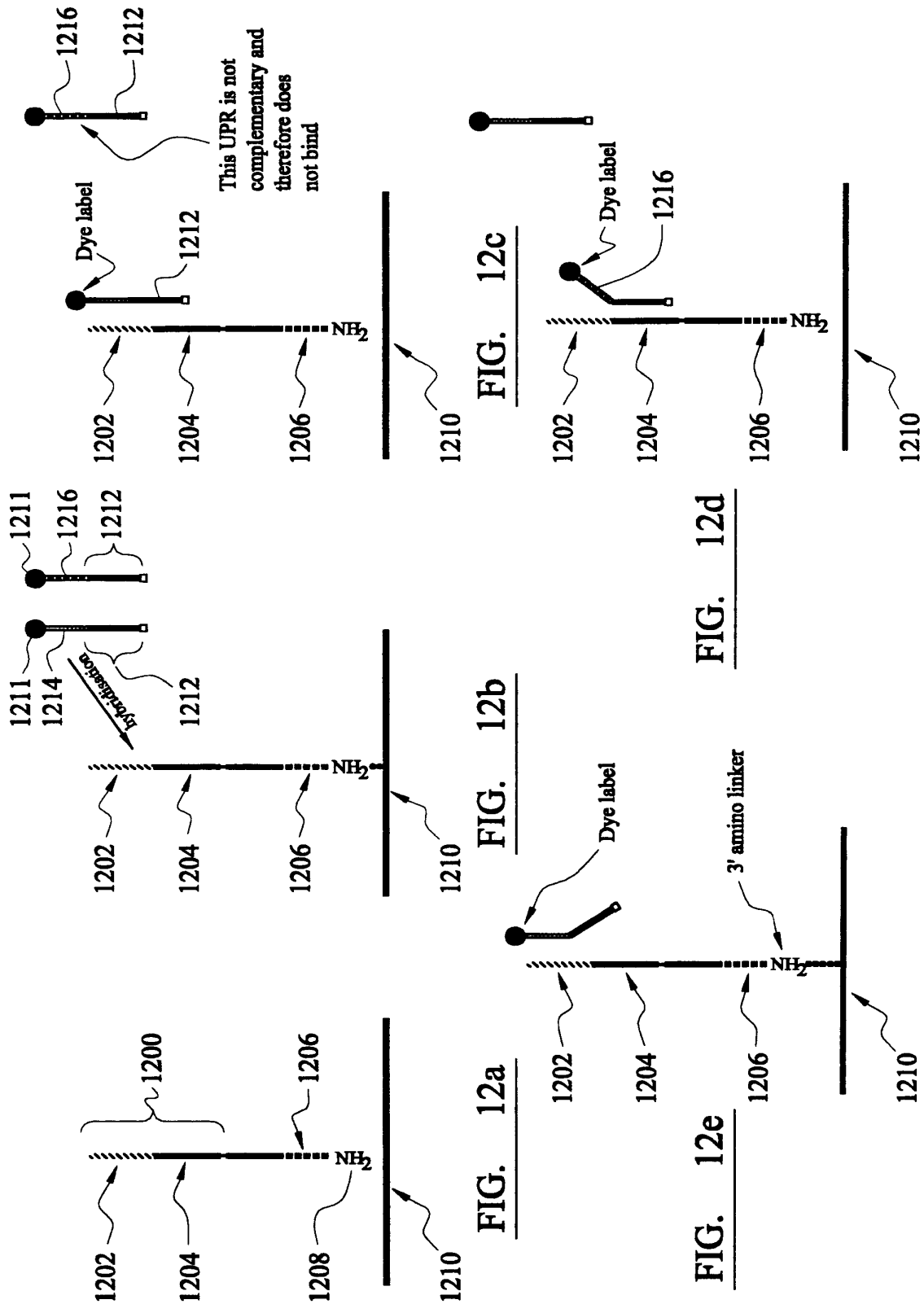
FIG. 10e

SUBSTITUTE SHEET (RULE 26)

-10/27-

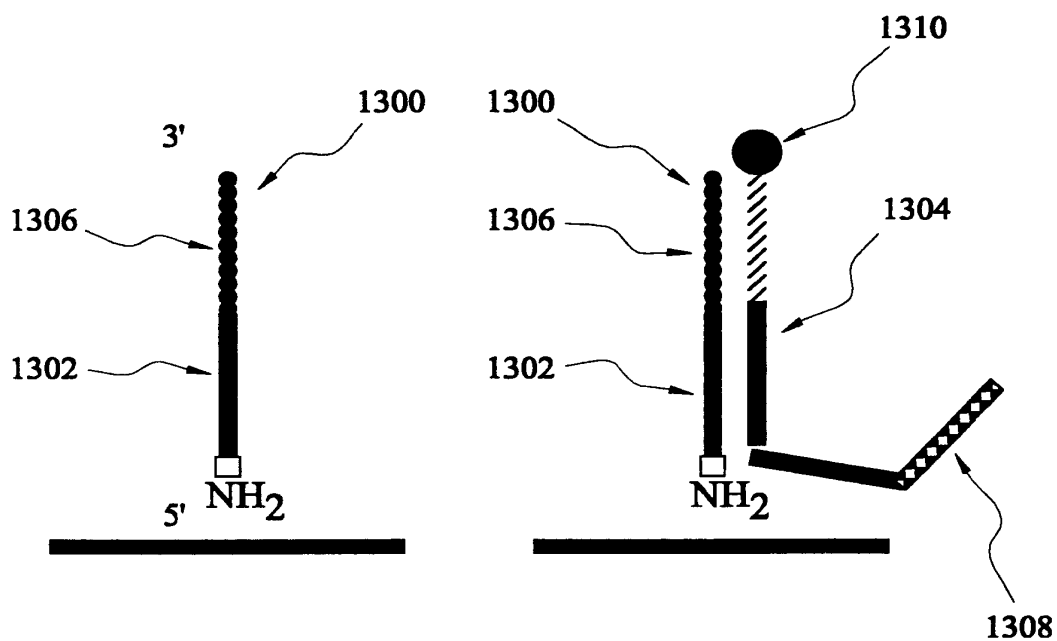
FIG. 11

-11/27-

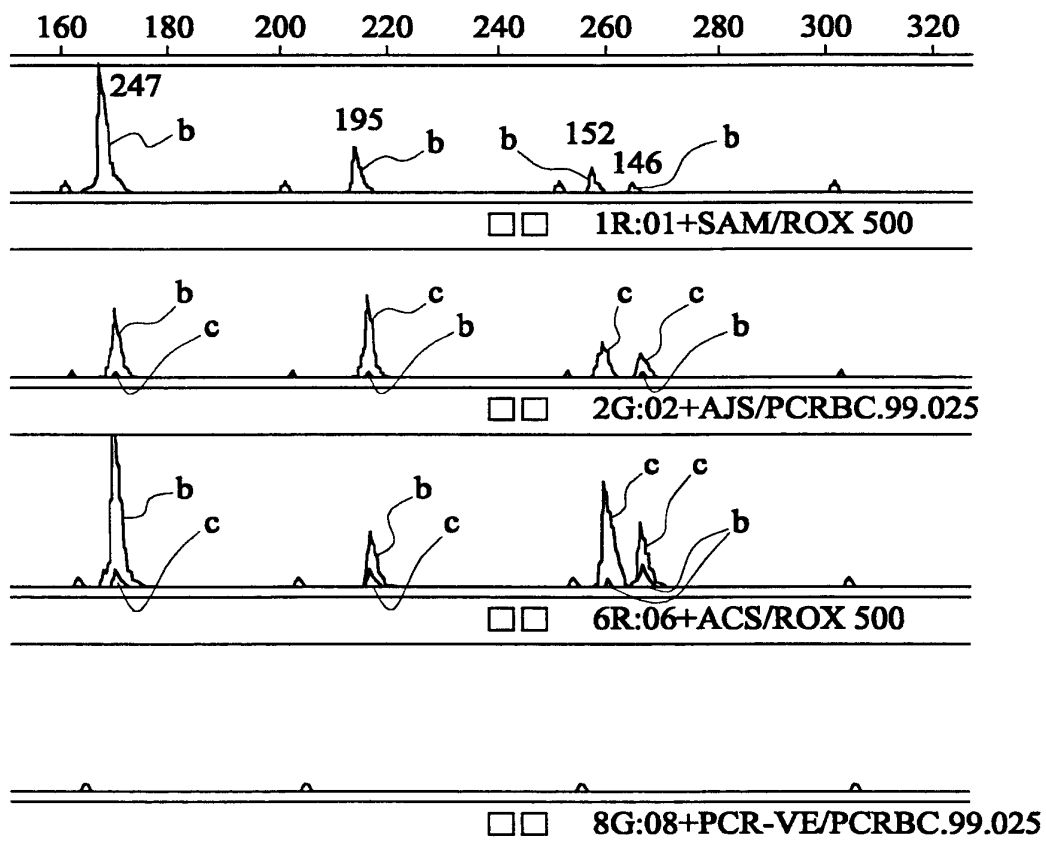


SUBSTITUTE SHEET (RULE 26)

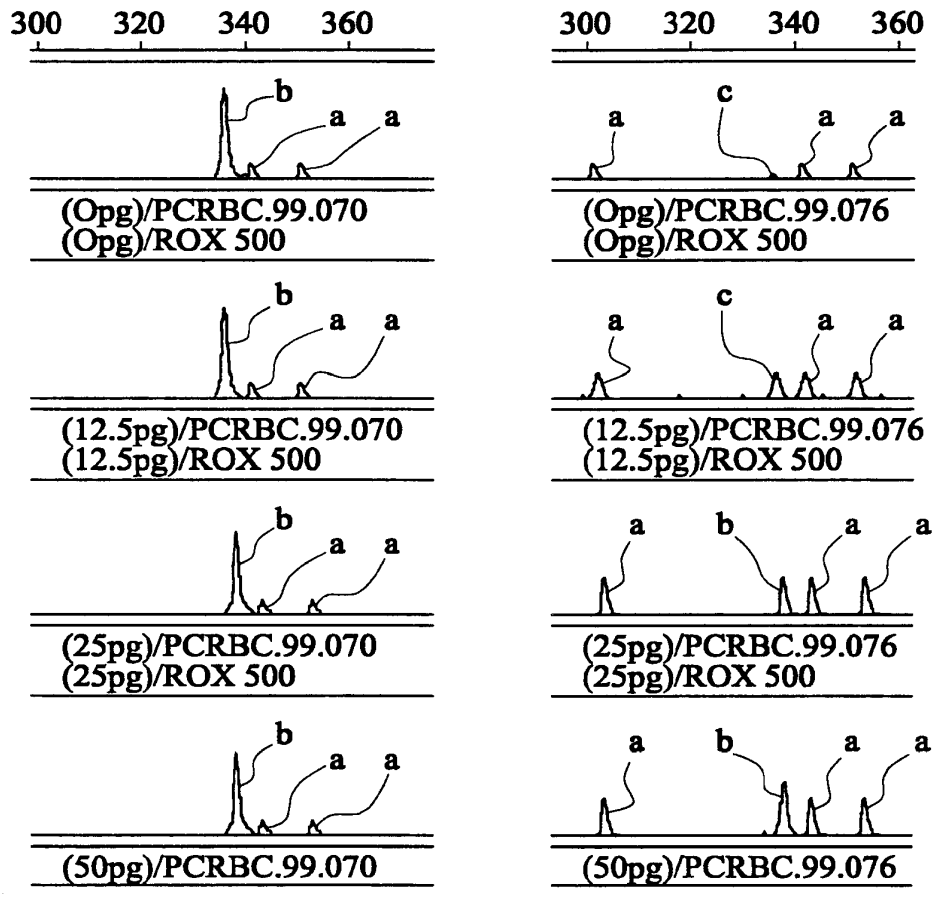
-12/27-

FIG. 13aFIG. 13b

-13/27-

FIG. 14

-14/27-

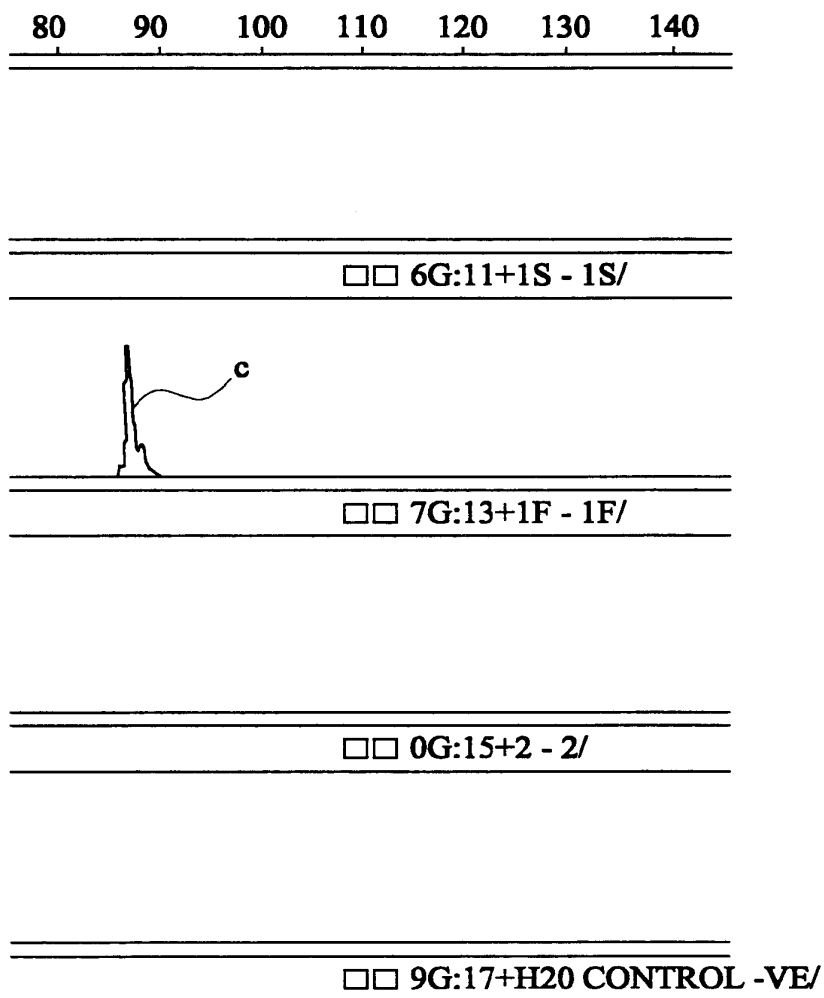


Demonstration of amplification specificity and sensitivity of detection of mitochondrial DNA (a) left hand series shows detection of the major component of the mixture coding for the mt0073G polymorphism, whereas the right hand series shows detection of the minor component of the mixture, coding for the mt0073A polymorphism.

The sensitivity of the test is c. 12.5pg genomic DNA. The red peaks 'a' are standard size; the peaks 'b' are green; and the peaks 'c' are blue.

FIG. 15

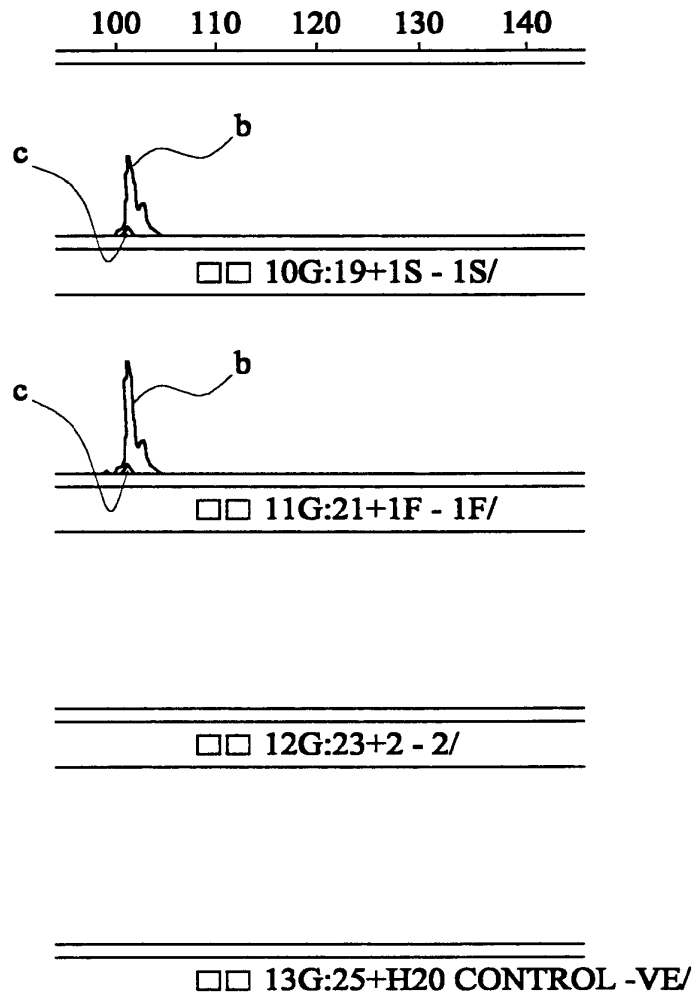
-15/27-



Primer 416A tested against Gc 1S - 1S; 1F - 1F; 2 - 2 and negative control. Only the 1F - 1F gives a signal. No background was observed with other samples.

FIG. 16a

-16/27-



Singleplex reaction to test specificity of the forward primer 420G. This detects both Gc 1 polymorphisms. The primer was tested against a series of individuals - Gc 1S - 1S; 1F - 1F; 2 - 2 and a negative control. Only the first two samples gave a signal, the remainder were clear.

FIG. 16b

-17/27-

100 110 120 130 140

□□ 14G:27+1S - 1S/

□□ 15G:29+1F - 1F/



□□ 16G:31+2 - 2/

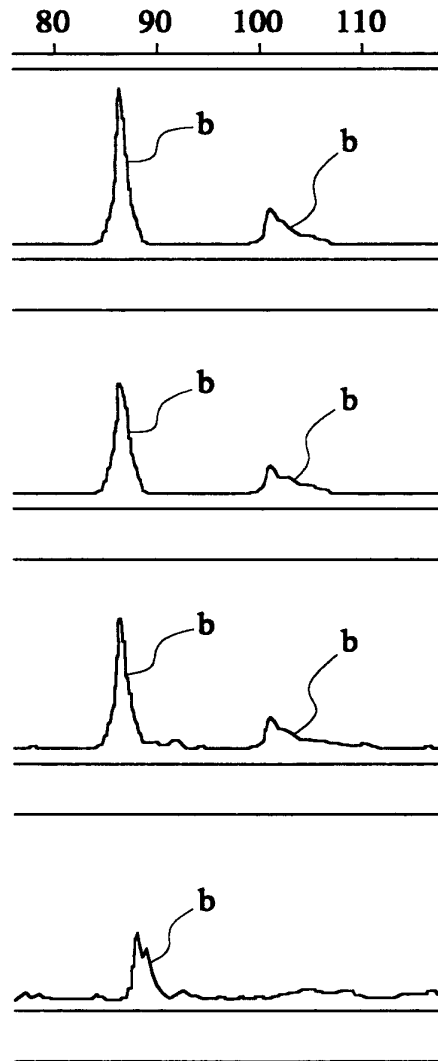
□□ 17G:83+H2O CONTROL -VE/

Singleplex reaction to test specificity of the forward primer 420T which detects Gc2. Samples are from 1S - 1S; 1F - 1F; 2 - 2 and negative control. No background was detected.

FIG. 16c

SUBSTITUTE SHEET (RULE 26)

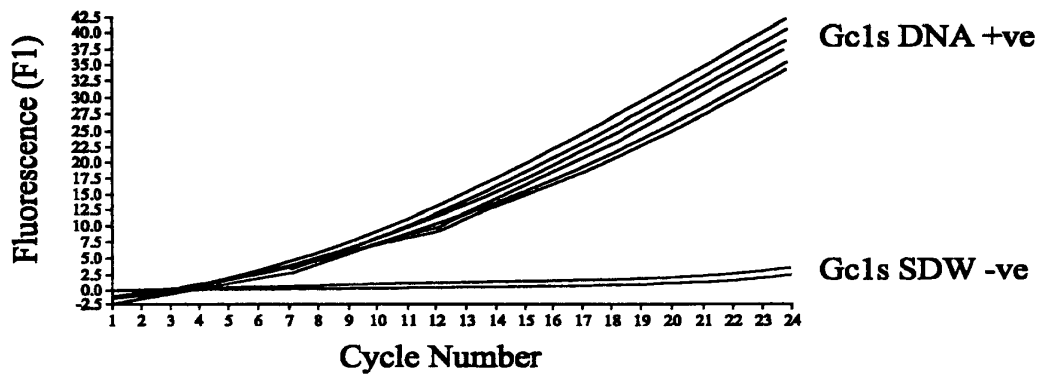
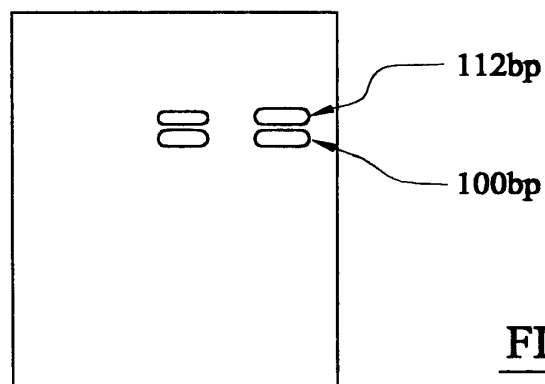
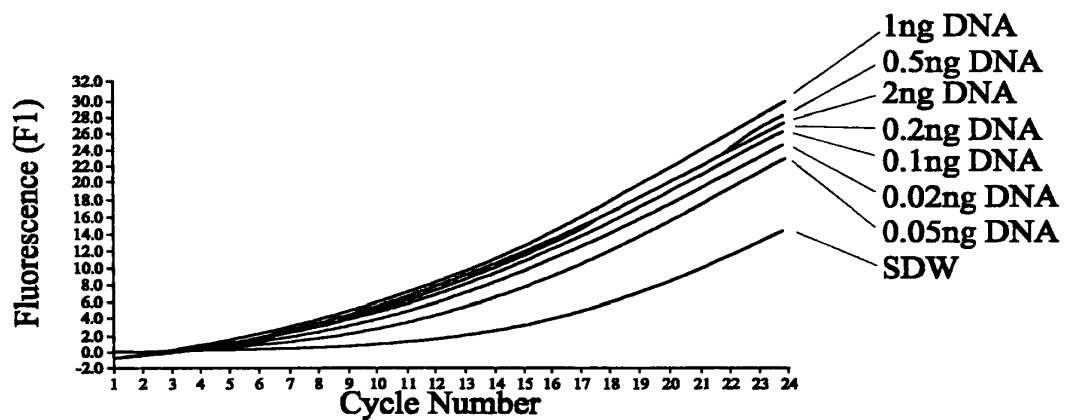
-18/27-



Demonstration of the limits of detection. A dilution series of a 1S - 1S individual was prepared (1ng, 200pg, 40pg, 8pg, respectively). A cocktail of Gc 420; Gc416 primers and the universal primers were used in the PCR reaction. Both peaks are green, indicating priming by 416C and 420G only. This is consistent with the known phenotype of the individual. No background from 416A or 420T was observed.

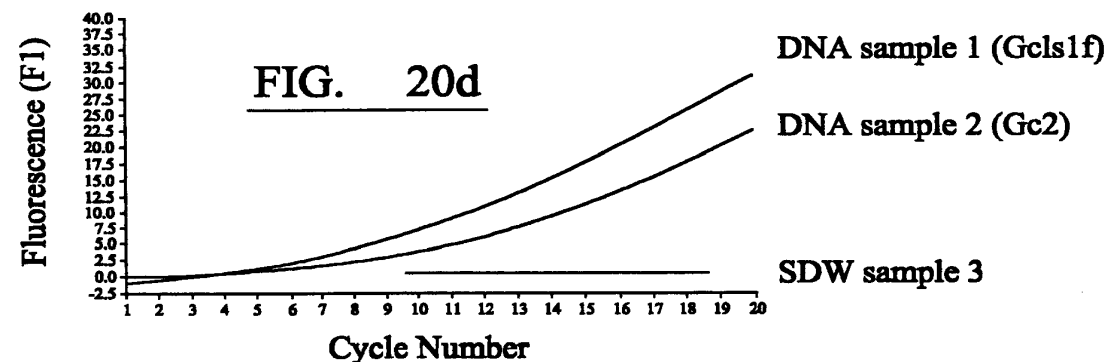
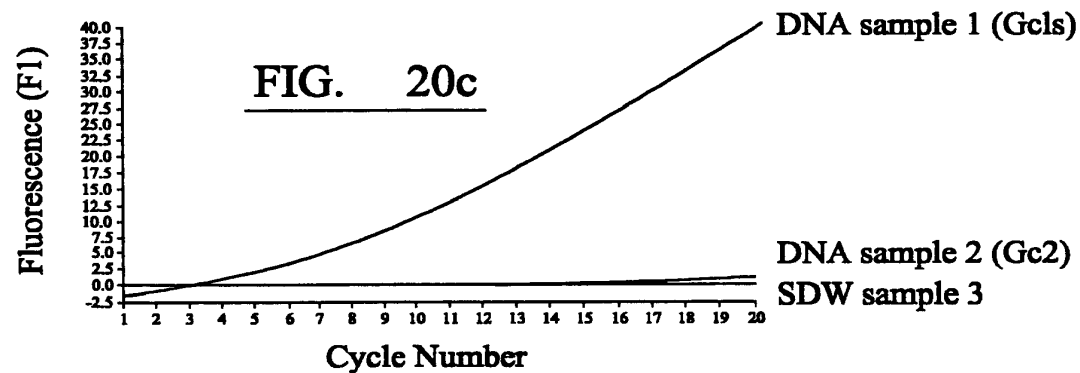
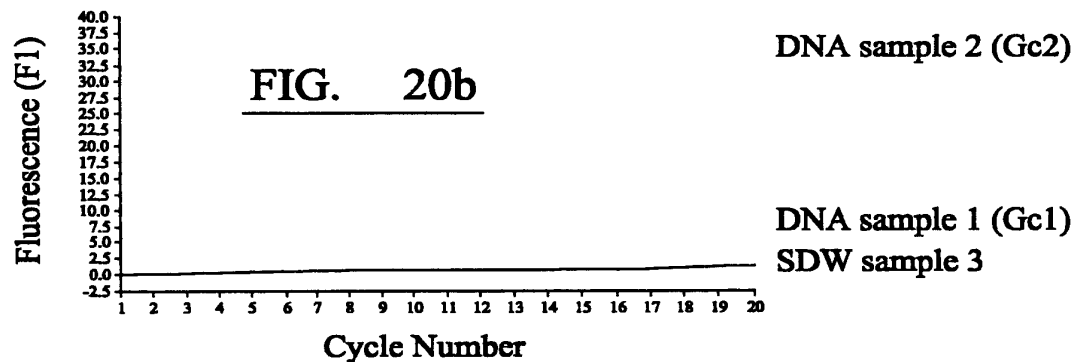
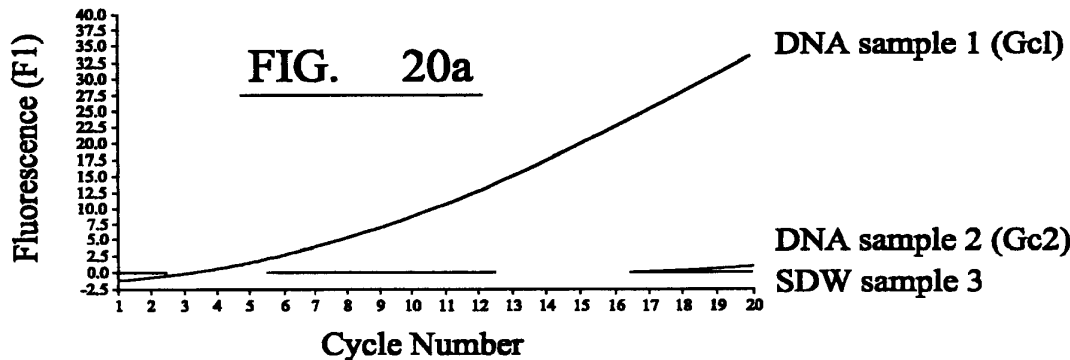
FIG. 16d

-19/27-

FIG. 17FIG. 18FIG. 19

SUBSTITUTE SHEET (RULE 26)

-20/27-



-21/27-

	416	420	
	CCTGAAGCCACACCCACGGA	ACTGGCA	Gc1s DNA
NO AMPLIFICATION	CCGGTGTGGGTGCCTTGA		Gc1f primer
	CCTGAGGCCACACCCACGGA	ACTGGCA	Gc1f DNA
SPECIFIC AMP.	←CCGGTGTGGGTGCCTTGA		Gc1f primer
	CCTGAGGCCACACCCAAGGA	ACTGGCA	Gc2 DNA
NON-SPECIFIC AMP.	←CCGGTGTGGGTCCCTTGA		Gc1f primer

FIG. 21

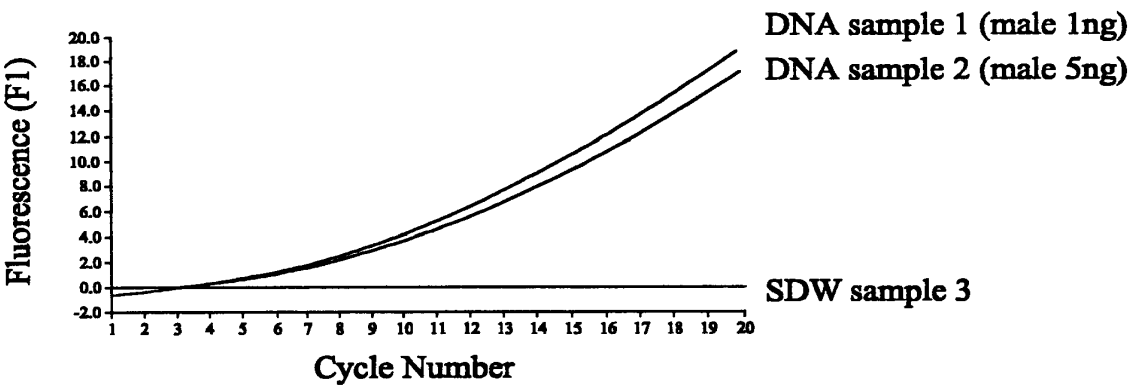


FIG. 22

-22/27-

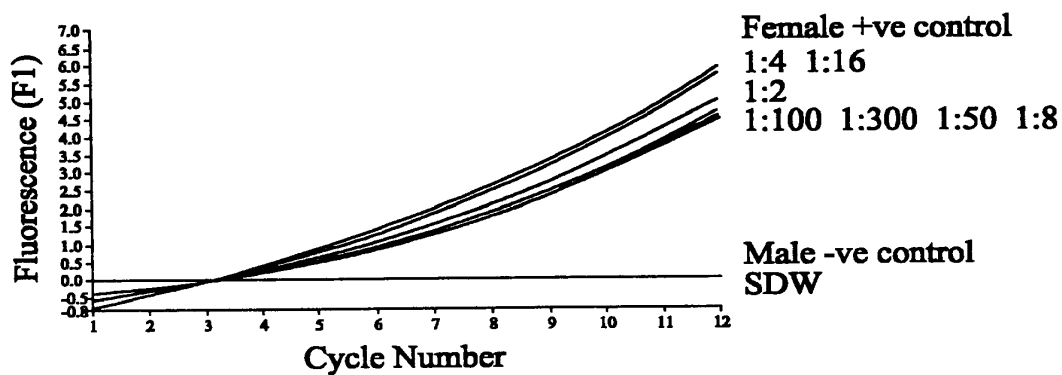


FIG. 23a

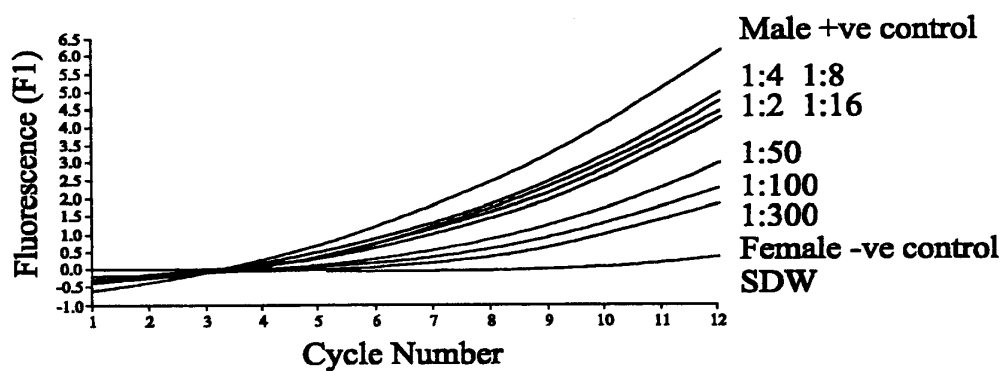


FIG. 23b

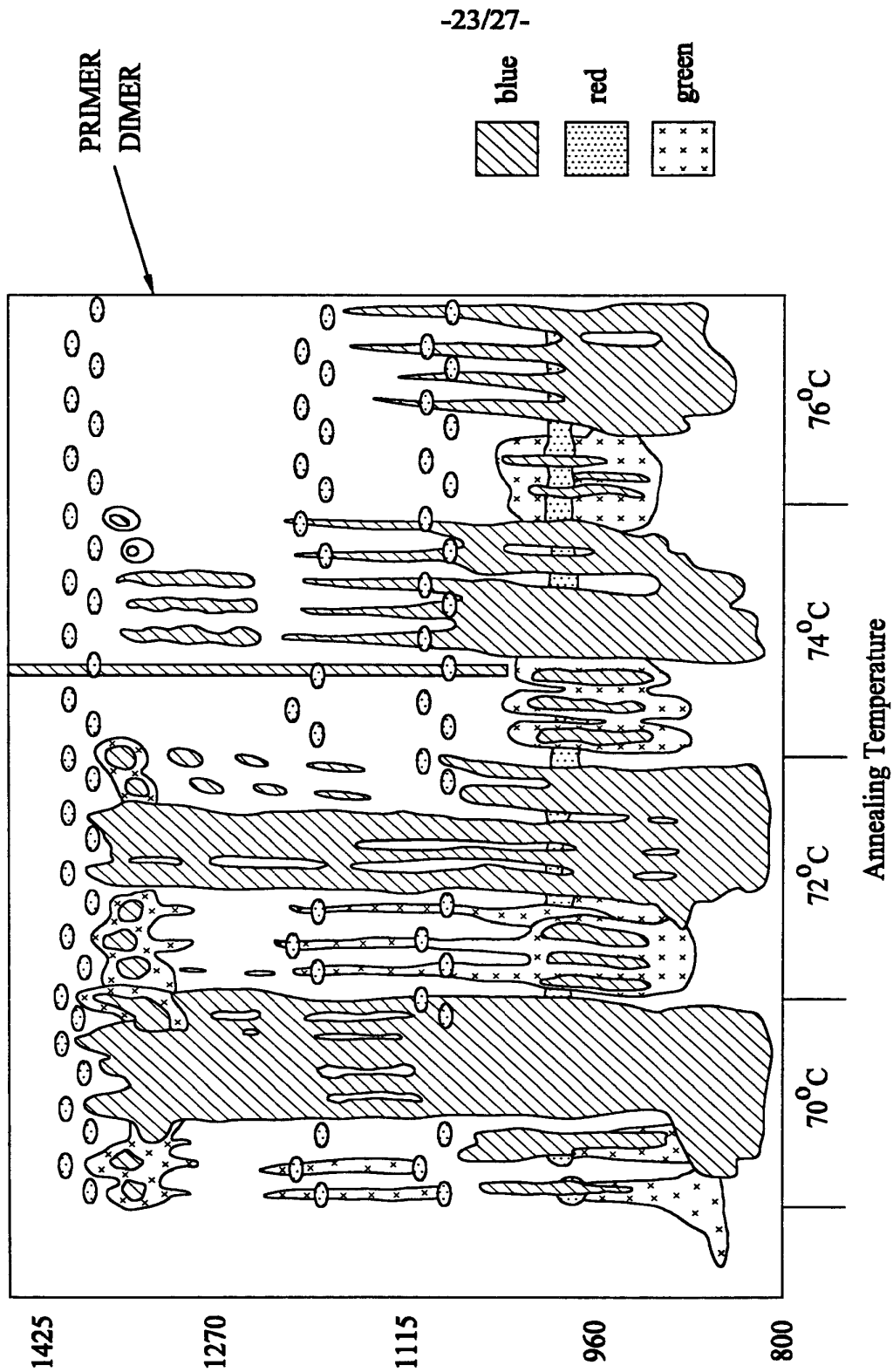


FIG. 24

-24/27-

A) PCR can occur with 5' overhangs

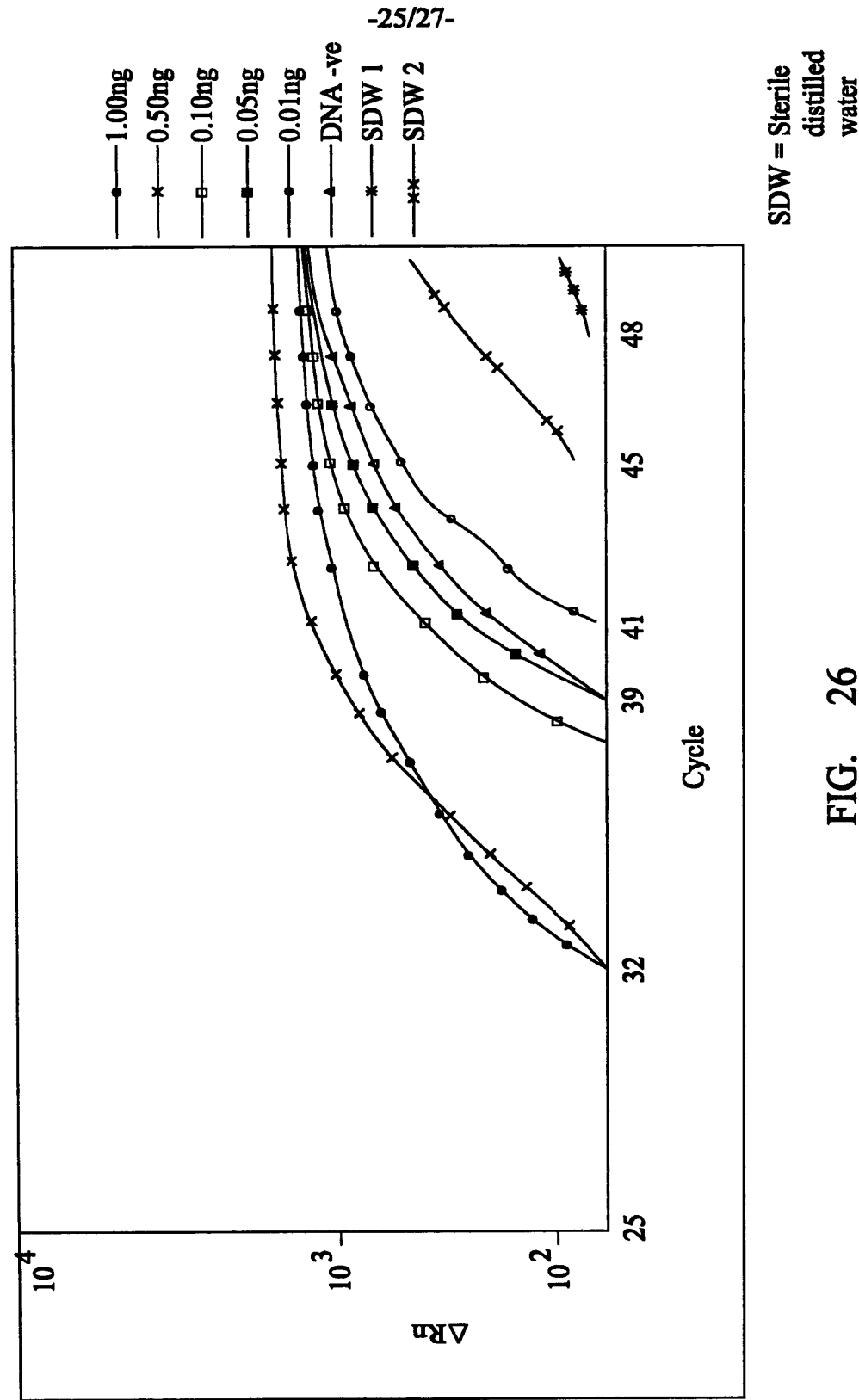
FIG. 25a

B) PCR cannot occur with 3' overhangs

FIG. 25b

5' CTAGCTGGTGGCTGTGCTAG 3'
3' GATCGTGTCGGTGGTCGATC5'

FIG. 25c



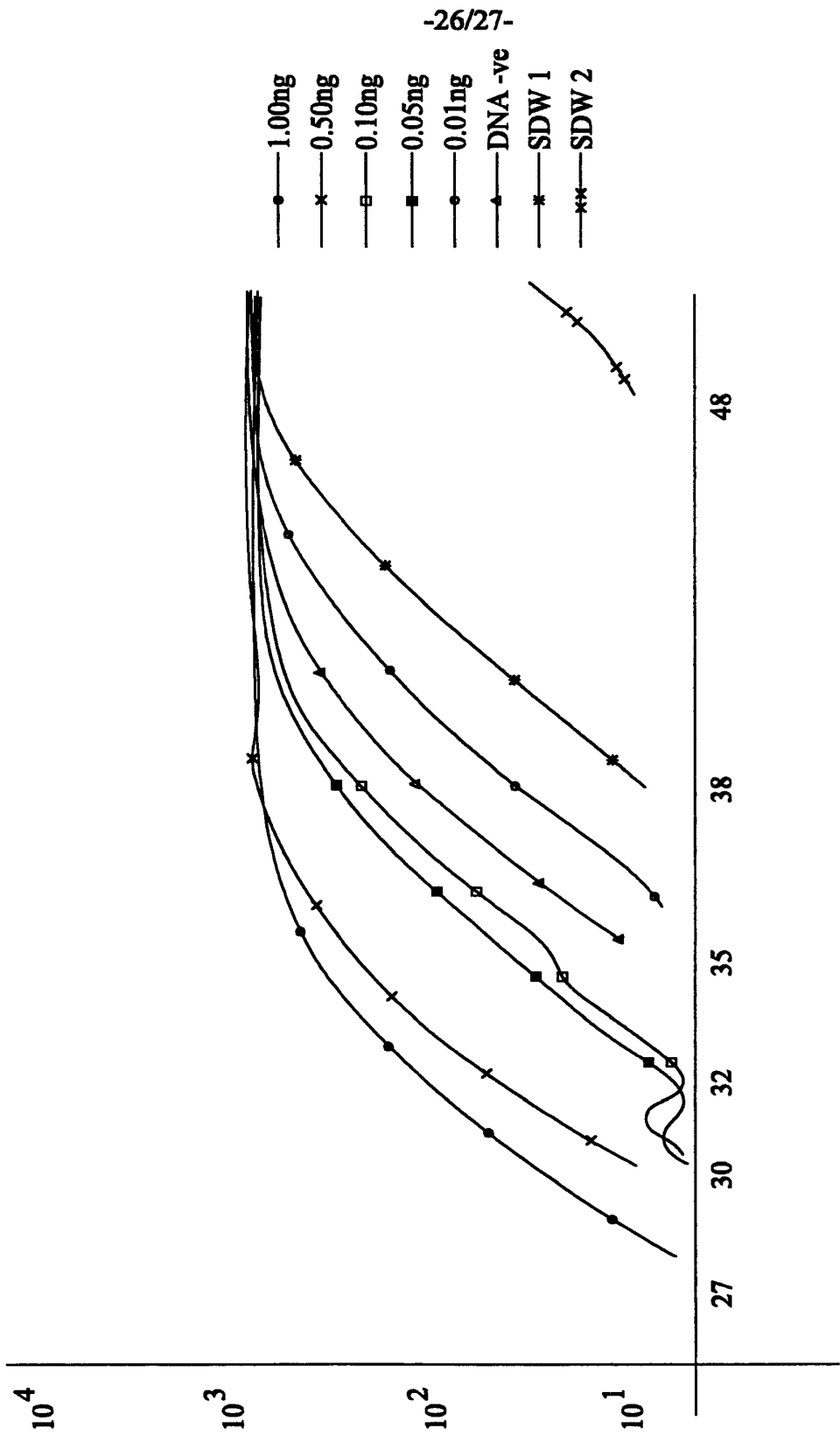


FIG. 27

SUBSTITUTE SHEET (RULE 26)

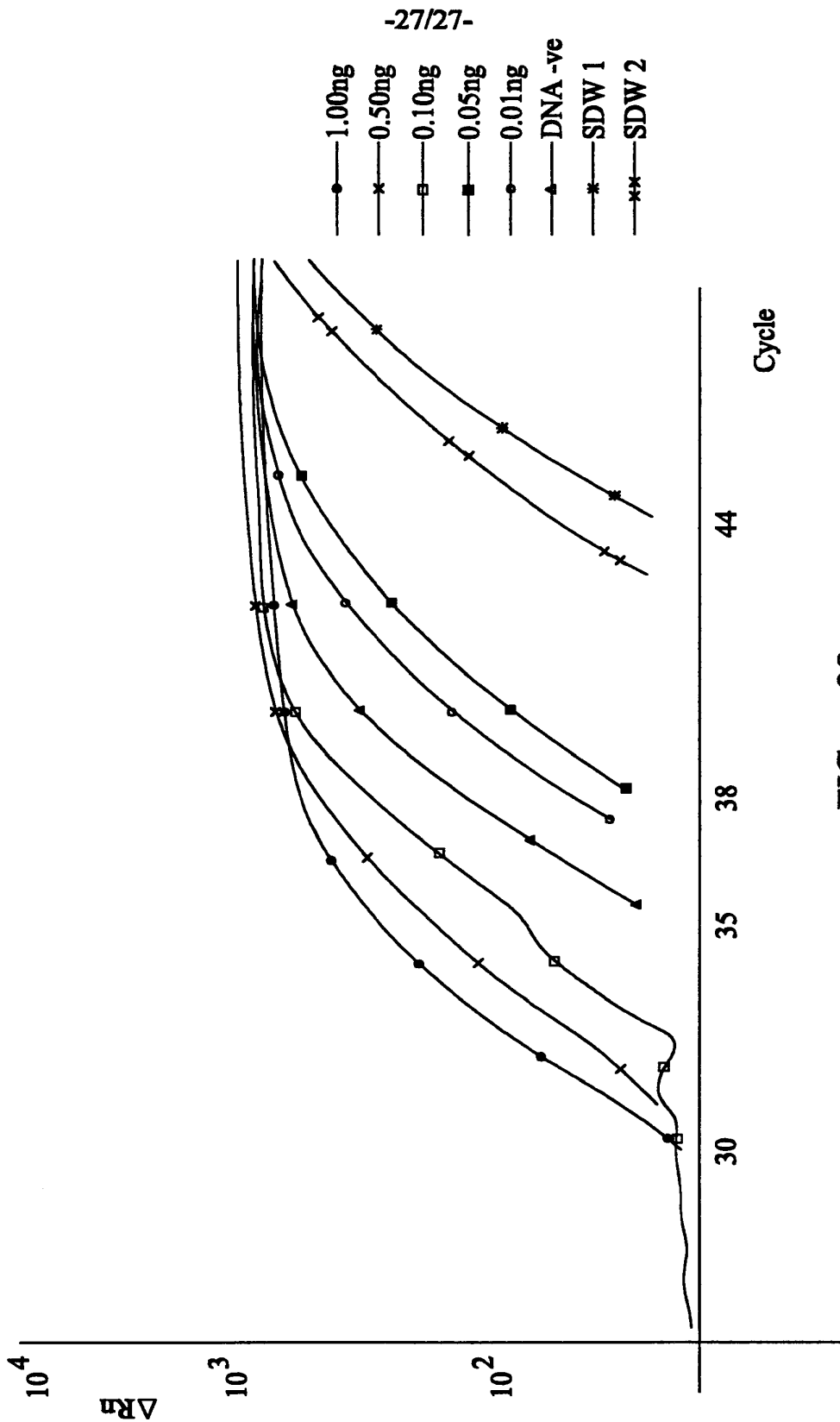


FIG. 28

SUBSTITUTE SHEET (RULE 26)

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 July 2002 (18.07.2002)

PCT

(10) International Publication Number
WO 02/056014 A2

- (51) International Patent Classification⁷: **G01N 33/53**
- (21) International Application Number: PCT/US01/49132
- (22) International Filing Date:
18 December 2001 (18.12.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/747,003 22 December 2000 (22.12.2000) US
- (71) Applicant (for all designated States except US): **GLAXO GROUP LIMITED** [GB/GB]; Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 ONN (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **NELSEN, Anita, J.** [US/US]; GlaxoSmithKline, Five Moore Drive, P.O. Box 13398, Research Triangle Park, NC 27709 (US). **PEPPERS, Lottie, L.** [US/US]; GlaxoSmithKline, Five Moore Drive, P.O. Box 13398, Research Triangle Park, NC 27709 (US). **WEINER, Michael, Phillip** [US/US]; c/o 454 Corporation, 20 Commercial Street, Branford, CT 06405 (US).
- (74) Agents: **LEVY, David, J.** et al.; GlaxoSmithKline, Five Moore Drive, P.O. Box 1398, Research Triangle Park, NC 27709 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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



WO 02/056014 A2

(54) Title: METHODS FOR ENCODING AND DECODING COMPLEX MIXTURES IN ARRAYED ASSAYS

(57) Abstract: The present invention is a method of encoding a complex mixture of assay constituents comprising using combinations of detectable tags and a total number of detectable tags less than the total number of constituents to be encoded. The method comprises determining the total number of constituents to be encoded; determining the number of detectable tags in each combination, wherein the number of detectable tags in each combination is more than one and less than or equal to the number of prime numbers in the number of constituents to be encoded; and determining the total number of detectable tags, wherein the total number of detectable tags equals a sum of a set of factors of the total number of constituents, wherein the number of factors equals the number of detectable tags in each combination. The encoding methods are useful in a multiplexed assay using complex mixtures of assay constituents.

METHODS FOR ENCODING AND DECODING COMPLEX MIXTURES IN ARRAYED ASSAYS

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

This invention relates generally to the fields of molecular biology and chemical analysis. More specifically, the invention relates to methods of encoding and decoding complex mixtures in multiplexed assays in order to minimize time and expense necessary for assaying numerous constituents in a single assay.

BACKGROUND ART

In the past few years the genomes of several organisms have been completely (or nearly completely) sequenced, including those of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Escherichia coli*, *Caenorhabditis elegans* and, most recently, the human genome. To make use of this wealth of available genomic data, rapid, high-throughput methods for analyzing all of the predicted gene products and their roles in the structural and functional organization of the cell were needed. Specifically needed were encoding and decoding means for analyzing the functional information of the thousands of genes in a complete genome.

One technology previously introduced used unique "bar-coding" tags for each of thousands of yeast genes and a silicon chip for decoding (Shoemaker et al., 1996). This method was useful for examining differential gene expression of a population of yeast strains. However, it required several thousand tags and a relatively expensive readout platform.

Prior to the present invention, no multiplexed method had been provided that could be scaled to accommodate assays of varying complexity using multiplexed, inexpensive, high-throughput methods.

5

SUMMARY OF THE INVENTION

In accordance with the purpose(s) of this invention, as embodied and broadly described herein, this invention, in one aspect, relates to a method of encoding a complex mixture of assay constituents comprising using combinations of detectable
10 tags and a total number of detectable tags that is less than the total number of constituents to be encoded. More specifically, the invention relates to a method further comprising determining the total number of constituents to be encoded; determining the number of detectable tags in each combination, wherein the number of detectable tags in each combination is more than one and less than or equal to the
15 number of prime numbers in the number of constituents to be encoded; and determining the total number of detectable tags, wherein the total number of detectable tags equals a sum of a set of factors of the total number of constituents, wherein the number of factors equals the number of detectable tags in each combination.

20 In yet another aspect, the invention relates to a method of performing a multiplexed assay using complex mixtures of assay constituents encoded according to the encoding method of the invention. Specifically, the invention relates to a method comprising performing an assay to produce assay constituents using an array of the complex mixtures, wherein each constituent in a single complex mixture is
25 detectably tagged with a unique combination of detectable tags; detecting which complex mixtures of assay constituents in the array have a positive response; and decoding the constituents in the complex mixtures having the positive response to determine which specific constituent or constituents are positive.

In another embodiment, the invention relates to a kit for performing a
30 multiplexed assay using complex mixtures of encoded assay constituents,

comprising a means of detectably tagging assay constituents encoded according to the encoding method of the invention, and an arraying means for a plurality of complex mixtures, and a container therefor.

The advantages of this invention include scalability, throughput and low-cost as compared to the currently available methods. Additional advantages of the invention will be set forth in part in the description that follows, and in part will be obvious from the description, or may be learned by practice of the invention. The advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

15

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments of the invention and, together with the description, serve to explain the principles of the invention.

Figure 1 shows a schematic of the bait vector, pMW101, and the prey vector, pMAR101, used in the Yeast Two Hybrid (Y2H) study.

Figure 2 shows a schematic of the synthesis of the 96 pMAR101 prey vectors used in the Y2H study. Each 5' ZipCode was bracketed by a similar DNA sequence (5'- TGGGCGACTTCTCCAAAC -3', (SEQ ID NO:2) which was labeled the "Watson" sequence). And each 3' sequence was bracketed by a second DNA sequence (5'- CTTGCAGATTCGGCAGTT -3' (SEQ ID NO:3), which was labeled the "NCrick" sequence). PCR amplification was used to generate 96 different fragments of the *Cm^r* gene; each fragment with the following order: 5'-Watson-ZipCode₁₋₁₂-*Cm^r* - ZipCode_{A-H}-NCrick-3'. Fragments were cloned into the pMAR101 vector at a unique *Swa*I site.

Figure 3 shows the method of bead-based genotyping by hybridization to

Luminex beads, which was used to decode the Y2H positive wells. Following the Y2H assay, clones in positive wells were PCR amplified using biotinylated Watson and nCrick primers. For a given fragment, querying which pair of 3' and 5' ZipCodes were contained therein involved hybridizing the fragment to the cZipCodes on the microsphere. Flow cytometry was used to detect the label captured on a particular pair of microspheres.

Figure 4 shows an example of a decode of 20 PCR products hybridized to a set of 20 ZipCode beads. A set of 96 vectors, each encoding a unique region containing two ZipCodes bracketed by a Watson and nCrick was used. DNA sequence served as a PCR template in a reaction containing Watson and nCrick primers. The PCR product was then used in a microsphere-based genotyping method and both of the ZipCodes on either side of the Cmr gene were decoded by hybridization to a set of 20 different beads. Shown are the MFI values obtained from the first twenty PCR products of the 96-member set.

15

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention may be understood more readily by reference to the following detailed description of preferred embodiments of the invention and the Examples included therein and to the Figures and their previous and following description.

Before the present compounds, compositions, articles, devices, kits, and/or methods are disclosed and described, it is to be understood that this invention is not limited to specific assay methods, specific means or methods of detection, or to particular encoding or decoding means, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a microsphere" includes mixtures of various

30

microspheres, reference to “an assay constituent” includes mixtures of two or more constituents, and the like.

Ranges may be expressed herein as from “about” one particular value and/or to “about” another particular value. When such a range is expressed, another
5 embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent “about,” it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint and independently of the other
10 endpoint.

In this specification and in the claims that follow, reference will be made to a number of terms that shall be defined to have the following meanings:

“Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur, and that the description includes instances
15 where said event or circumstance occurs and instances where it does not. For example, the phrase “detectable tags optionally are contained in or coupled to microspheres” means that the detectable tags may or may not be contained in or coupled to microspheres and that the description includes both detectable tags contained in or coupled to microspheres and detectable tags otherwise used to label
20 the desired assay constituents.

The present invention provides a method of encoding a complex mixture of assay constituents comprising using combinations of detectable tags and a total number of detectable tags that is less than the total number of constituents to be encoded. This method offers an advantage over the prior art because it reduces the
25 number of labels necessary to detect a given number of constituents and lends itself to highly complex, multiplexed formats that are useful in high-throughput assays with pooled samples.

As used throughout, “encoding” refers to tagging an assay constituent with one or more detectable tags so that the tag(s) can be detected and the constituent

identified by decoding (i.e., attributing the detectable tag(s) to a specific assay constituent).

An assay constituent can be either a reactant or a product of the assay. The constituents are selected from the group consisting of proteins, peptides, amino
5 acids, small molecules, nucleotides, fatty acids, sugars, cofactors, receptors, receptor ligands, protein domains, oligonucleotides, transcription factors, nucleic acids, and small compounds.

As used throughout, "an assay" can be a chemical assay, protein assay, pharmacologic assay, hybrid assay (e.g., yeast two hybrid, prokaryotic two-hybrid,
10 reverse-two hybrid, or three-hybrid assay), display assay (e.g., phage display-, F pilli- and lacI-fusion), protein readout assay, binding assay (ligand, nucleic acid, antibody, small molecule, or small compound binding assay), cell-based assay, genomic assay, read-out assay (transcriptional or protein read-out assay), or the like.

A "detectable tag" refers to any label that can be detected with a detection
15 means and can include the absence of a label. Thus, if one hundred thousand assay constituents are to be encoded, the methods of the present invention provide that less than one hundred thousand detectable tags are used, even if one of those tags is the absence of a label. Preferably, the detectable tags are directly or indirectly coupled to the constituents. The detectable tags as used in the methods of the present
20 invention optionally are contained in or coupled to a solid support that binds the constituents either directly or through an intermediary. Thus, the detectable tags can be coupled to a non-mobile solid support, like a plate or a chip, or a mobile solid support, like microspheres. Optionally, each detectably tagged microsphere used in the methods of the present invention is coupled to a means of specifically binding a
25 constituent. For example, in one embodiment, the coupled means is a nucleic acid (called a "ZipCode"), which is complementary to a nucleic acid in or bound to the constituent to be encoded.

Preferably the detectable tags are selected from the group consisting of radiolabels, dyes, fluorescent labels, Quantum Dot® (Quantum Dot Corp.), and

combinations thereof. "Dyes" include, but are not limited to, chemiluminescent, magnetic, and radiofrequency labels.

Optionally, the method of the present invention further comprises determining the total number of constituents to be encoded; determining the number of detectable tags in each combination, wherein the number of detectable tags in each combination is more than one and less than or equal to the number of prime numbers in the number of constituents to be encoded; and determining the total number of detectable tags, wherein the total number of detectable tags equals a sum of a set of factors of the total number of constituents, wherein the number of factors equals the number of detectable tags in each combination. For example, if 100,000 genes are to be screened using a traditional method of encoding for each assay constituent to be screened, then 100,000 different detectable tags would have to be used in a traditional one dimensional assay (i.e., one detectable tag for each constituent). In the present method, however, the number of different detectable tags can be reduced using a multi-step process. Where the total number of constituents to be screened is 100,000, the number of detectable tags in each combination can be any number between two and ten, because ten is the number of prime numbers that are multiples of 100,000 (i.e., $2 \times 2 \times 2 \times 2 \times 2 \times 5 \times 5 \times 5 \times 5 \times 5 = 10$ prime numbers). Thus, if three detectable tags will be used in each combination to encode a single constituent, the total number of detectable tags needed is calculated by determining three factors of the total number of constituents (e.g., $10 \times 100 \times 100$) and adding those three factors together (i.e., 210) to determine the total number of detectable tags. Using this paradigm, the entire 100,000 genes could be screened using a total of 210 detectable tags.

Table 1: Examples of assay arrays and detectable tags for an assay of 100,000 constituents

Number of detectable tags in each combination	Example number of constituents to be screened (factors)	Example total number of detectable tags needed
1	100,000 (100,000)	100,000
2	100,000 (250 X 400) (500 X 200) (160 X 625) (125 X 800) (100 X 1,000) (80 X 1250) (40 X 2,500) (50 X 2000) (32 X 3125) (25 X 4000) (20 X 5,000) (16 X 6,250) (10 X 10,000) (8 X 12,500) (5 X 20,000) (4 X 25,000) (2 X 50,000)	650 700 785 925 1,100 1,330 2,540 2050 3,157 4,025 5,020 6,266 10,010 12,508 20,005 25,004 50,002
3	100,000 (50 X 25 X 80) (100 X 25 X 40) (32 X 25 X 125) (20 X 125 X 40) (16 X 125 X 50) (10 X 100 X 100) (160 X 25 X 25) (10 X 125 X 80) (8 X 125 X 100) (20 X 25 X 200) (160 X 5 X 125) (16 X 25 X 250) (100 X 5 X 200) (50 X 8 X 250) (4 X 125 X 200) (250 X 5 X 80) (250 X 4 X 100) (10 X 25 X 400) (250 X 2 X 200) (50 X 5 X 400) (400 X 2 X 125) (8 X 25 X 500) (500 X 5 X 40) (50 X 4 X 500) (100 X 2 X 500) (8 X 625 X 20) (16 X 625 X 10)	155 165 182 185 191 210 210 215 233 245 290 291 305 308 329 335 354 435 452 455 527 533 545 554 602 653 651

	(32 X 5 X 625)	662
	(4 X 625 X 40)	669
	(2 X 625 X 80)	707
	(800 X 5 X 25)	830
	(10 X 10 X 1000)	1,020
	(20 X 5 X 1000)	1,025
	(4 X 25 X 1000)	1,029
	(50 X 2 X 1000)	1,052
	(10 X 8 X 1250)	1,268
	(16 X 5 X 1250)	1,271
	(1250 X 4 X 20)	1,274
	(1250 X 2 X 40)	1,292
	(10 X 5 X 2000)	2,015
	(2 X 25 X 2000)	2,027
	(8 X 5 X 2500)	2,513
	(10 X 4 X 2500)	2,514
	(20 X 2 X 2500)	2,522
	(4 X 3125 X 8)	3,137
	(2 X 3125 X 16)	3,143
	(4,000 X 5 X 5)	4,010
	(2 X 125 X 4000)	4,127
	(10 X 2 X 5000)	5,012
	(4 X 5 X 5000)	5,009
	(6,250 X 4 X 4)	6,258
	(6,250 X 2 X 8)	6,260
	(2 X 5 X 10,000)	10,007
	(12,500 X 2 X 4)	12,506
	(25,000 X 2 X 2)	25,004
4	100,000 (e.g., 10 X 10 X 10 X 100)	130
5	100,000 (e.g., 10 X 10 X 10 X 10 X 10)	50
6	100,000 (e.g., 2 X 5 X 10 X 10 X 10 X 10)	47
7	100,000 (e.g., 2 X 2 X 5 X 5 X 10 X 10 X 10)	44
8	100,000 (e.g., 2 X 2 X 2 X 5 X 5 X 5 X 10 X 10)	41
9	100,000 (e.g., 2 X 2 X 2 X 2 X 5 X 5 X 5 X 5 X 10)	38
10	100,000 (2 X 2 X 2 X 2 X 2 X 5 X 5 X 5 X 5 X 5)	35

As demonstrated by the above table, the total number of detectable tags is minimized by selecting factors of the total number of constituents that are equal or approximate. Thus, in one embodiment of the present method, the total number of detectable tags is minimized by selecting factors of the total number of constituents
5 that are equal or approximate.

Using the encoding method of the present invention, an assay can be designed based on the total number of detectable tags available, based on the total number of tags in each combination, based on the arraying means (e.g., the number of wells on a plate that can be read using automated readers), or a combination.
10 Thus, if 100,000 constituents are to be assayed and only 35 total detectable tags are available, then there must be 10 detectable tags in each combination. Alternatively, if there are practical limitations to the number of tags that can be detected in combination, then the assay could limit the number of detectable tags in combination to, for example, three and the total number of detectable tags could be,
15 for example, 210.

The invention further provides a method of performing a multiplexed assay using complex mixtures of assay constituents encoded according to the encoding method of the invention. Specifically, the method comprises performing an assay to produce assay constituents using an array of the complex mixtures, wherein each
20 constituent in a single complex mixture is detectably tagged with a unique combination of detectable tags; detecting which complex mixtures of assay constituents in the array have a positive response; and decoding the constituents in the complex mixtures having the positive response to determine which specific constituent or constituents are positive. As used herein, "an array" includes a
25 multiwell plate or any other arraying means. Thus, an array using a multiwell plate can be eight wells in one dimension and twelve wells in another dimension as in a 96 well plate. An array could also be sixteen wells in one dimension and twelve wells in another dimension, using two 96 well plates.

30 The detection means is selected as specific for the detectable tags. For example,

if the detectable tag is fluorescent and is contained in or coupled to microspheres, then flow cytometry with a fluorescence detection device or a FAC sorter can be used to detect and distinguish a tag or combination of tags. When a specific mixture (e.g., a complex mixture in a specific well in an assay plate) has a positive response, then that particular mixture is decoded to identify the positive constituent in that mixture. For example, if the well contained ten genes to be screened, then the decoding method would identify which of the ten genes had a positive response. Thus, the decoding is performed by detecting and distinguishing with a detection device the detectable tags in each complex mixture of constituents.

10 In one embodiment the method of encoding and decoding is used with a complex mixture of arrayed cDNA clones in a yeast two-hybrid analysis. The steps comprise using an array of complex mixtures of yeast host cells comprising an encoded set of cDNAs made by cloning each individual cDNA into a member of a set of vectors, wherein each member of the set of vectors comprises a yeast two-
15 hybrid activation domain and a selected pair of identifying nucleic acid sequences ("ZipCodes"), wherein the selected pair of identifying nucleic acids is specific for each individual cDNA, and wherein the yeast host cells containing the set of vector are combined to create complex mixtures of cDNA clones; mating the arrayed host cells with a yeast expressing a bait protein and one or more reporter genes; detecting
20 an interaction or absence of an interaction between the bait protein and the activation domains in each complex mixture of the array by determining the expression of the reporter gene or genes; performing PCR amplification of each complex mixture that shows an interaction, wherein the PCR amplification is performed using labeled primers; and decoding the PCR products using a genotyping assay. In one
25 embodiment, the first member of the selected pair of identifying nucleic acids is at the 5' end of an antibiotic resistance gene and the second member is at the 3' end of the antibiotic resistance gene. Preferably, each vector in the set has two primer nucleic acid sequences present in each vector, wherein the first primer nucleic acid is at the 5' end of the first member of the identifying pair and the second primer nucleic
30 acid is at the 3' end of the second member of the identifying pair. In one

embodiment, the antibiotic resistance gene is a chloramphenicol gene.

In one embodiment of the yeast two hybrid method, each identifying nucleic acid is 25 bases. In another embodiment, twenty different identifying nucleic acids are used in combinations to form 96 different pairs of identifying nucleic acids.

- 5 Thus, each complex mixture can contain up to about 96 different cDNAs and the array of complex mixtures of cDNA clones in host cells can contain up to about 9,220 different cDNAs.

- In one embodiment of the yeast two hybrid method, the reporter gene is a β -galactosidase gene. In another embodiment the reporter gene is a Leu2 gene. In yet
10 another embodiment, the reporter genes are both the β -galactosidase gene and the Leu2 gene.

- The genotyping assay as used in the yeast two hybrid method comprises contacting, under conditions that allow formation of hybridization products, the labeled PCR products of each mixture with a set of microspheres, wherein each
15 member of the set of microspheres is distinguishably labeled and is coupled with a capture nucleic acid complementary to one of the identifying nucleic acid sequences; detecting the label of the PCR product and the label of the microsphere in two or more hybridization products. The presence of a labeled PCR product in two different hybridization products indicates the cDNA specific to the pair of
20 identifying nucleic acid sequences. Preferably, the distinguishable label of the microsphere is a fluorescent label, wherein the PCR product is fluorescently labeled, and wherein the fluorescent label of the microsphere and the PCR product can be detected in the same reaction product or products.

- In one embodiment of the yeast two hybrid method, the microspheres are
25 carboxylated and amino groups at the 5' end of the capture nucleic acids are coupled to the carboxyl groups.

- Preferably, the capture nucleic acid further comprises a luciferase cDNA. Optionally, the luciferase cDNA has the sequence CAGGCCAAGTAACTTCTTCG (SEQ ID NO:1). The capture oligonucleotide can be directly coupled to the
30 microsphere or can be indirectly coupled to the microsphere by a carbon spacer.

In the yeast two hybrid method, the label of the PCR product and the label of the microsphere in two or more hybridization products is preferably detected using flow cytometry.

The present invention further provides a kit for performing a multiplexed
5 assay using complex mixtures of encoded assay constituents, comprising a means of detectably tagging assay constituents encoded according to the encoding method of the invention, and an arraying means for a plurality of complex mixtures, and a container therefor. The means of detectably tagging assay constituents can include, for example, a set of microspheres, wherein each member of the set of microspheres
10 is detectably tagged and binds selectively to an assay constituent. For example, the kit can comprise a set of detectably tagged microspheres that bind selectively to cDNA clones in a yeast two-hybrid analysis. The kit for performing a yeast two hybrid can further comprise one or more of the following: a set of yeast vectors comprising a reporter gene, a yeast two-hybrid activation domain, and a selected pair
15 of identifying nucleic acid sequences, wherein the selected pair of identifying nucleic acids is specific for each vector; a means for homologously recombining the cDNAs to be encoded into the vectors of the set into yeast host cells; a means for combining the yeast host cells containing the set of vectors to create the complex mixture of cDNA clones; a set of yeast bait cells expressing a bait protein; a means
20 for mating the yeast cells containing the set of vectors and the yeast bait cells; a set of labeled PCR primers; or a set of microspheres, wherein each member of the set of microspheres is detectably labeled and is coupled with a capture nucleic acid complementary to one of the identifying nucleic acid sequences.

25

Experimental

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and
30 evaluated, and are intended to be purely exemplary of the invention and are not

intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient
5 temperature, and pressure is at or near atmospheric.

Example 1: Yeast 2-Hybrid Analysis

Reagents. Restriction and DNA modification enzymes were purchased from
10 various manufacturers and used according to their recommendations. AmpliTaq Gold DNA polymerase and Big Dye Terminator Cycle Sequencing reagent were purchased from Applied Biosystems (Foster City, CA, USA). Unmodified oligonucleotides were purchased from either Keystone Biosource (Camarillo, CA, USA), MWG Research (High Point, NC, USA) or IDT (Coralville, IA, USA). 2-[N-
15 Morpholino]ethanesulfonic acid (MES) and 1-Ethyl-3-(3-Dimethylaminopropyl) carbodiimide Hydrochloride (EDC) were purchased from Sigma (St. Louis, IL, USA) and Pierce (Rockford, IL, USA), respectively. Streptavidin phycoerythrin was purchased from Becton Dickinson (San Jose, CA, USA). Yeast cell preparation and transformation reagents were purchased from Zymo Research (Orange, CA, USA).

20

Preparation of microspheres. Carboxylated fluorescent polystyrene microspheres (5.5 µm in diameter) were purchased from the Luminex Corp. (Austin, TX, USA). Oligonucleotides, synthesized to contain a 5' amino group, C₍₁₅₋₁₈₎ spacer, 20 base luciferase sequence and 25 base Zipcode-complimentary sequence, were
25 ordered from Oligos, Etc. (Wilsonville, OR, USA) or from Applied Biosystems. Oligonucleotides were covalently coupled to the microspheres as described by Iannone and co-workers (Chen et al. (2000); Iannone et al (2000)).

Yeast strains, plasmids, and media. Yeast strains EGY48 and L40 have
30 been described (Finley & Brent (1994)). Plasmids pHybLex/Zeo, pYesTrp2 and

pMW101 have also been described (Finley & Brent (1994); Gyuris et al (1993); Watson et al (1996)). Selective yeast media were prepared as described in Gyuris et al. (1993). The plasmid pBC SK+ was purchased from Stratagene (La Jolla, CA, USA).

5

Construction of pMAR101 and pMAR101 derivatives. To construct pMAR101, a PCR product derived from the amplification of pMW101 with forward primer (5'-GCCGAAGCTTGCGGTTGGGGTATTCGCAACGGCGACTGG-3') (SEQ ID NO:28) and reverse primer (5'-

10 ATACGCATGCAATTCGCCCGGAATTAGCTT GGCTGCAGGT-3') (SEQ ID NO:29) was digested with restriction endonucleases *HindIII* and *SphI* and ligated overnight at 16 °C with *HindIII*, *SphI*-digested, agarose gel purified plasmid pYesTrp2 (Invitrogen, Carlsbad, CA). Addition of this PCR product incorporates regions of approximately 23 bases adjacent to and on either side of the multiple

15 cloning site. The resulting plasmid enables simultaneous homologous recombination of amplified genes into both bait and prey vectors (Figure 1). To construct pMAR101.1-pMAR101.96 twenty primers (Table 2) containing a common sequence for amplification, a 25 base zipcode and a terminal end of the chloramphenicol (*Cm^r*) gene were used to amplify the *Cm^r* marker from the plasmid

20 pBC SK+. Following amplification under standard conditions in Optiprime buffer #5 (Stratagene), 2 units of *Pfu* polymerase were added to each of the 50 µl reactions and incubated at 72 °C for 20 minutes. The resulting 96 unique blunt-ended fragments were ligated with *SwaI* digested pMAR101 for 16 hours at 16 °C. Ligation products were transformed into electrocompetent DH10B cells (LTI,

25 Gaithersburg, MD, USA) and clones were selected on LB agar plates containing carbenicillin (50 µg/ml) and chloramphenicol (12.5 µg/ml) (Figure 2). Colonies were screened by PCR to confirm incorporation of the cassette as well as to select for uniform orientation of the cassette. Common sequence primers, "Watson" (5'-TGGGCGACTTCTCCAAAC-3') (SEQ ID NO:2) and "nCrick" (5'-

30 CTTGCAGATTCGGCAGTT-3') (SEQ ID NO:3), were used to confirm that the

1241 bp fragment was incorporated into the plasmid. Primers Watson and a plasmid-specific oligo, "pYesTrp Forward" (Invitrogen) were used to screen for the orientation of the Cm^r gene.

Table 2. DNA Primers and Associated ZipCode Sequences

DNA Sequence ^a			
Primer	Watson/NCrick	ZipCode	Cam gene
1	TGGGCGACTTCTCCAAACGATGATCGACGAGACACTCTCGCCACTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:4)			
2	TGGGCGACTTCTCCAAACCGGTGACGAGCTGCCGCGCAAGATCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:5)			
3	TGGGCGACTTCTCCAAACGACATTGCGGATCGCCGCCCCGCTTTCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:6)			
4	TGGGCGACTTCTCCAAACCGGTATCGCGACCGCATCCCAATCTCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:7)			
5	TGGGCGACTTCTCCAAACGCTCGAAGAGGCGCTACAGATCCTCCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:8)			
6	TGGGCGACTTCTCCAAACACCGCCAGCTCGGCTTCGAGTTCGCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:9)			
7	TGGGCGACTTCTCCAAACGTAATCTCCAGCGGAAGGGTACGGCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:10)			
8	TGGGCGACTTCTCCAAACCTTTTCCCGTCCGTCATCGCTCAAGCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:11)			
9	TGGGCGACTTCTCCAAACGGCTGGGTCTACAGATCCCCAATTCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:12)			
10	TGGGCGACTTCTCCAAACGAACCTTTCGCTTCACCGGCCGATCCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:13)			
11	TGGGCGACTTCTCCAAACTTTCGGGCACGCGCGGGATCACCATCCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:14)			
12	TGGGCGACTTCTCCAAACCTCGGTGGTGCTGACGGTGCAATCCCTGTGACGGAAGATCACTTCGC		
(SEQ ID NO:15)			
A	CTTGCAGATTCGGCAGTTTCAACGTGCCAGCGCCGCTCTGGGACTCCACGGGGAGAGCCTGAGCA		
(SEQ ID NO:16)			
B	CTTGCAGATTCGGCAGTTGCGAAGGAACTCGACGTGGACGCCGCTCCACGGGGAGAGCCTGAGCA		
(SEQ ID NO:17)			
C	CTTGCAGATTCGGCAGTTGGGGGATACCGATCTCGGGCGCACACTCCACGGGGAGAGCCTGAGCA		
(SEQ ID NO:18)			
D	CTTGCAGATTCGGCAGTTGGAGCTTACGCCATCACGATGCGATCTCCACGGGGAGAGCCTGAGCA		
(SEQ ID NO:19)			

E CTTGCAGATTCGGCAGTT**CTGTGGCGGTGCGGAGTTTCCCCGAACTCCACGGGGAGAGCCTGAGCA**
 (SEQ ID NO:20)
 F CTTGCAGATTCGGCAGTT**CGATCCAACGCACTGGCCAAACCTACTCCACGGGGAGAGCCTGAGCA**
 (SEQ ID NO:21)
 G CTTGCAGATTCGGCAGTT**CTGAATCCTCCAACCGGGTTGTCGACTCCACGGGGAGAGCCTGAGCA**
 (SEQ ID NO:22)
 H CTTGCAGATTCGGCAGTT**TCGGCGCTGGCGTAAAGCTTTTGGCTCCACGGGGAGAGCCTGAGCA**
 (SEQ ID NO:23)

^aDNA Sequences are as follows: 3' cam gene, TCCACGGGGAGAGCCTGAGCA (SEQ ID NO:24); 5' cam gene, CTGTGACGGAAGATCACTTCGC (SEQ ID NO:25); Watson, TGGGCGACTTCTCCAAAC (SEQ ID NO:2); NCrick, CTTGCAGATTCGGCAGTT (SEQ ID NO:3). The ZipCode sequences (in bold) are shown between the Watson/NCrick and Cam sequences.

Preparation of pMAR101.1-pMAR101.96. pMAR101.x plasmid DNAs were purified using Qiatip-500 columns (Qiagen, Valencia, CA, USA). The resulting DNAs were digested with *EcoRI* and *XhoI* restriction enzymes and purified on 1% preparative agarose gels. Digested plasmid was transformed into competent EGY48 cells and plated on agar plates containing YNB –Trp + glucose to determine background.

Cloning by Homologous Recombination. Plasmids were constructed *in vivo* in yeast as described by Oldenburg et al. (1997). Briefly, genes of interest were amplified from plasmid DNAs isolated from commercially available cDNA libraries. Primers for amplification were designed to include portions of both the pMAR101 plasmid as well as portions of the gene of interest. The forward primer contained 23 bases of vector sequence immediately adjacent to and 5' of the *EcoRI* restriction site (GCAACGGCGACTGGCTGGAATTC) (SEQ ID NO:26) fused to approximately 25 bases of the 5' end of the gene to be amplified. This primer does not require a start codon, but does require the gene to be in-frame with the *EcoRI* site. The reverse primer contained 23 bases of vector sequence adjacent to and 3' of the *XhoI* site (GCTTGGCTGCAGGTCGACTCGAG) (SEQ ID NO:27) fused to

approximately 25 bases of the 3' end of the gene to be amplified. The 3' primer does require a termination codon. Amplification was carried out in 25 µl reactions, each containing 100 ng cDNA template, 200 nm primers, 1X RedTaq buffer (Sigma), 200 µM dNTP mix (Sigma), 0.75 units RedTaq polymerase (Sigma) and
5 0.125 units *Pfu* polymerase. The MgCl₂ concentration was adjusted to a final concentration of 3.0 mM. Reactions were amplified for 30 cycles (94°C for 30 seconds, 56°C for 45 seconds and 72 °C for 2 minutes) followed by a final extension of 72°C for 7 minutes. Products were verified by electrophoresis of 5 µl on analytical agarose gels. PCR products were then cloned into pMAR101.x by
10 homologous recombination and co-transformed into yeast. Competent EGY48 yeast cells (10 µ) were combined with 50 ng of *Eco*RI, *Xho*I digested vector, 100 ng PCR products and 100 µl EZ3 solution, vortex mixed, and incubated at 30°C. After at least 30 minutes, the entire reaction was plated on agar plates containing YNB – Trp+glucose and incubated at 30°C for 72 hours. Colonies were screened for inserts
15 by PCR with the vector-specific primers pYesTrp Forward and Reverse (Invitrogen). Ninety-six different genes were cloned into the 96 unique pMAR101.x vectors and pooled for analysis against the bait clones. Using a similar method, baits for Y2H were cloned and co-transformed into either pMW101/RFY206/pSH1834T or pHybLex/L40. Bait clones were selected on YNB+Ura +His+glucose or
20 YPD+Zeocin(300 µg/ml), respectively.

Preparation of yeast bait and prey cultures. For validation experiments, several colonies from the bait transformation plates were inoculated into 15 mL of selection media and grown 48 hours at 30°C. For high-throughput experiments, 96
25 different baits were arrayed in 96-well V-bottom microplates containing 200 µl of selection media and grown 48 hours at 30 °C. For both applications, yeast library plates (containing genes cloned into the 96 pMAR101.x vectors) were thawed and 200 µl of selection media (YNB–Trp +glucose with antibiotics) was added to each well. These plates were also incubated at 30°C for 48 hours.