

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number
WO 02/24960 A1

(51) International Patent Classification⁷: C12Q 1/68,
C12P 19/34

(21) International Application Number: PCT/US01/29922

(22) International Filing Date:
24 September 2001 (24.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/234,857 22 September 2000 (22.09.2000) US

(71) Applicant (for all designated States except US): FRESH-
GENE, INC. [US/US]; 525 Del Rey Ave. Suite B, Sunny-
vale, CA 94085 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): YANG, Qinghong
[CN/US]; 1901 Rock Avenue, Apt. #108, Mountain View,
CA 94043 (US).

(74) Agents: ABRAMS, Samuel, B. et al.; Pennie & Edmonds
LLP, 1155 Avenue of the Americas, New York, NY 10036
(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, PH, 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, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

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



WO 02/24960 A1

(54) Title: DETECTION OF DNA VARIATION

(57) Abstract: The invention provides methods and reagents for detecting the presence of a difference between two related nucleic acid sequences. In preferred embodiments of the invention, the difference is a mutation, such as a point mutation, deletion or insertion.

Detection of DNA variation

INTRODUCTION

The present invention relates to the field of molecular biology, more particularly nucleic acid hybridization, Holiday junction formation and branch migration. In one aspect, the invention provides methods and reagents for detecting the presence of a difference between two related nucleic acid sequences. In preferred embodiments of the invention, the difference is a mutation, such as a point mutation, deletion or insertion. Practical applications of the invention include, but are not limited to, genotyping, discovery and detection of single nucleotide polymorphisms, characterization and quantitation of polynucleotides, mutation rate detection, gene expression analysis. Furthermore, the invention is capable of distinguishing between homozygous and heterozygous genetic variation

BACKGROUND

The tendency of nucleic acids to bind selectively and specifically to complementary nucleic acid sequence has been exploited in the development of numerous nucleic acid hybridization techniques. Not only are such techniques useful for detecting complementarity and/or identity between nucleic acid sequences (e.g.: quantitating differential gene expression level such as Northern blots, Southern blots and gene expression chip/micro-arrays), but in some cases they are exploited to be used for detecting differences between related nucleic acid sequences. Affymetrix's SNP-chip and various micro-array-based SNP scoring chip (e.g.: Motorola' DNA chip) are examples of a current genotyping technology that is based on allele-specific hybridization. The big problem with current allele-specific-hybridization-based genotyping technology is its poor accuracy due to low specificity. Gene-expression chips/micro-arrays have much better specificity/accuracy than SNP chips/micro-arrays due to the fact that the hybridization between specific cDNAs and their corresponding oligos/DNA fragments immobilized on the chip/micro-array is largely "Non" or "ALL"—or, in other words, highly specific. In contrast, specific DNA strands/oligos containing version 1 of a specific SNP can hybridize not only to its perfectly matched complementary DNA but also to non-perfectly matched ones such as those contain version 2 of the specific SNP. The hybridization is stronger between two perfectly complementary DNA strands than that between two non-perfectly

complementary DNA strands (including those that have either a single or multiple base-pair-mismatch between the two complementary strands). However, the single-base-pair difference is usually too small to render a high enough specificity for SNP scoring. The method disclosed in this patent application addresses the problem by combining highly specific allele-specific holiday structure formation with nucleic acid hybridization techniques (e.g.: gene chip/micro-array or fluorescence-labeled beads). As a result, SNP chips/micro-arrays can achieve the same high level of specificity/accuracy as gene-expression chips/micro-arrays.

In Panyutin IG *et al's* 1993 paper (Panyutin IG, Hsieh P, Formation of a Single Base Mismatch Impedes Spontaneous DNA Branch Migration (1993) *J. Mol. Biol.*, 230:413-24.), a single-stranded oligo that is completely (or partially) complementary to a specific part of single-stranded M13mp18 viral DNA anneals to the viral DNA and form a partial duplex with either 1 (or 2) tail(s) at each end. The partial duplex formed between the oligo and the M13mp18 viral DNA can then form a four-way Holiday-like structure with an invading partial duplex with either 1 (or 2) complementary tails. The four-way Holiday-like structure then undergoes branch migration in the direction away from the tail(s) (It can not branch migrate back towards the tail(s) due to energy barrier: breaking existing H-bonds without forming new ones). For Holiday structures formed between single-tailed partial duplexes, a single (or multiple) base pair difference between the duplex part of oligo/M13mp18 partial duplex and the duplex part of the invading partial duplex poses enough energy barrier (2 H-Bonds \rightarrow 0 H-bond) to impede the branch-migration and prevent the release of the annealed oligo, regardless of the presence or absence of Mg⁺⁺. For Holiday structure formed between double-tailed partial duplexes, a single base pair difference (substitution, deletion or insertion) between the duplex part of oligo/M13mp18 partial duplex and the duplex part of the invading partial duplex poses enough energy barrier to impede the branch-migration and prevent the release of the annealed oligo ONLY in the presence of Mg⁺⁺.

Based on Panyutin IG and Hsieh P's finding, we designed a genotyping method that allows multiplexing of genotyping assays (tens of thousands of SNPs/mutations can be assayed simultaneously in one assay reaction) and eliminate the requirement of individual PCR reactions. As a result, our method has the potential to dramatically reduce the cost and improve the throughput for genotyping.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

1. A DNA sample of interest (or target DNA, e.g.: genomic DNA or other DNA preparations that need to be genotyped) are immobilized on a solid surface. As an illustration rather than limitation, the DNA sample can be immobilized on a piece of nitrocellulose paper, baked and followed by UV cross-linking (standard procedure as in Southern blot). The target DNA can be denatured first and then immobilized on the solid surface, or it can be immobilized on the solid surface first and then denatured.
2. After the (immobilized) target DNA is denatured or during the target DNA is being denatured, a collection of probes are mixed with the immobilized and denatured target DNA. The collection of probes is comprised of n (1-10,000,000) different probes, each targeted at a specific SNP (Figure1, only 4 different probes targeted for 4 SNPs are shown). Each probe is comprised of three parts (figure 1):
 - 1 A 0-80 bp long (preferably 0bp or 10-50 bp) 5' tail **Trn** (**Tr1,Tr2,Tr3...**). When $Trn=0$ bp, the partial duplexes formed between the probes and their target DNA have a single tail at one end. When $Trn \neq 0$ bp, the partial duplexes formed have double-tails at both ends. Sequences for Trn are not found in the target DNA sample and therefore will not hybridize with the target DNA. For example, for human genotyping, sequences for Trn can derive from bacteria specific sequences that have no homology with the human DNA.
 - 2 A middle part that is unique, 1-600 bp long (preferably 5-100bp), and contains one version of a specific SNP. It will anneal to its target position in the target DNA sample.
 - 3 A 3' tail **T** (0-80 bp, preferably 12-50 bp)) that is **universal** for all probes in any one collection of probes used for each assay. Sequences for **T** are not found in the target DNA sample and therefore will not hybridize with the target DNA. For example, for human genotyping, sequences for Trn can derive from bacteria specific sequences that have no homology with the human DNA.
3. Wash away any probes that do not annealed to the immobilized target DNA and therefore are not bound to the nitrocellulose paper with any buffer that will allow the hybridization between the probes and the target DNA. As an illustration rather than limitation, washing buffer used for standard southern blot can be used.

4. Add a collection of n (1-10,000,000) reference DNA/partial duplexes containing one version of each of the SNPs the probes in step 2 targeted at/hybridized with. With certain buffer (e.g.: many commonly used buffers including TES buffer (50mM-Tris-Hcl(PH 7.5), 50mM NaCl, 1mM-EDTA), TSM buffer (50mM-Tris-Hcl(PH 7.5), 25mM NaCl, 10mM MgCl₂, 1mM-EDTA), PCR buffers (with Mg⁺⁺ for double-tailed partial duplexes and PCR buffers with/without Mg⁺⁺ for single-tailed partial duplexes)) at certain temperature (10°C – 75°C, preferably, 37°C – 65°C), the reference partial DNA duplexes form holiday structures with their corresponding partial duplexes formed (in step 2) between the immobilized target DNA and corresponding probes. The formed Holiday structures will undergo branch migration (1minute -240 minutes, in certain buffer (e.g.: many commonly used buffers including TES buffer (50mM-Tris-Hcl(PH 7.5), 50mM NaCl, 1mM-EDTA), TSM buffer (50mM-Tris-Hcl(PH 7.5), 25mM NaCl, 10mM MgCl₂, 1mM-EDTA), PCR buffers (with Mg⁺⁺ for double-tailed partial duplexes and PCR buffers with/without Mg⁺⁺ for single-tailed partial duplexes)) at certain temperature (10°C – 75°C, preferably, 37°C – 65°C)).

When a partial duplex formed (in step 2) between the immobilized target DNA and corresponding probes contains a homo-duplex version (e.g: target DNA version 1 anneal with probe DNA version 1, or alternatively, target DNA version 2 anneal with probe version 2) that is different from the homo-duplex version of the reference partial DNA duplex it forms a Holiday junction with, branch migration of that Holiday junction will stop and the probe will not be release from the immobilized target DNA (2 H-bonds → 0 H-bonds, energy barrier).

On the contrary, when a partial duplex formed (in step 2) between the immobilized target DNA and corresponding probes contains a homo-duplex version (e.g: target DNA version 1 anneal with probe DNA version 1, or alternatively, target DNA version 2 anneal with probe version 2) that is the same as the homoduplex version of the reference partial DNA duplex it forms a Holiday junction with, branch migration of that Holiday junction will proceed all the way through and the probe will be release from the immobilized target DNA due to complete strand exchange (2 H-bonds → 2 H-bonds, no energy barrier).

In the case that a partial duplex formed (in step 2) between the immobilized target DNA and corresponding probes contains a hetero-duplex version (e.g: target DNA version 1 anneal with probe DNA version 2, or alternatively, target

DNA version 2 anneal with probe version 1), the Holiday junction it form with either of the two homo-duplex versions of the reference partial DNA duplex will be resolved due to complete branch migration and the probe will be release from the immobilized target DNA due to complete strand exchange (1 H-bond \rightarrow 1 H-bond, no energy barrier).

In the case that a partial duplex formed (in step 2) between the immobilized target DNA and corresponding probes contains a hetero-duplex version (e.g: target DNA version 1 anneal with probe DNA version 2, or alternatively, target DNA version 2 anneal with probe version 1), the Holiday junction it form with either of the two hetero-duplex versions of the reference partial DNA duplex will be resolved due to complete branch migration and the probe will be release from the immobilized target DNA due to complete strand exchange (either 2 H-bonds \rightarrow 2 H-bonds or 0 H-bond \rightarrow 2 H-bonds, both cases no energy barrier).

Each reference partial duplex is comprised of (figure 1, page 1) two strands:

1. First strand is completely complementary to the target DNA at a specific SNP position. This strand is comprised of (one version of, same or different) the middle part of its corresponding probe in step 2 plus the sequences flanking that middle part at both the left and the right side (Figure 1, page 1).
2. 2nd/The other strand is comprised of 3 parts (Figure 1, page 1):
 - a. A middle part that is perfectly complementary to the middle part of the first strand.
 - b. A 0-80 bp long (preferably 0bp or 10-50 bp) 5' tail **Un** (**U1,U2,U3...**). Sequences for Trn are not found in the target DNA sample and therefore will not hybridize with the target DNA. For example, for human genotyping, sequences for Trn can derive from bacteria specific sequences that have no homology with the human DNA.
 - c. A 0-80 bp long (preferably 0bp or 10-50 bp) 3' tail **Trn'** (**Tr1',Tr2',Tr3'...**) that is complementary to and can anneal with tail **Trn** in the corresponding probe. When Trn'=0 bp, the reference partial duplexes have a single tail at one end. When Trn \neq 0 bp, the partial duplexes formed have double-tails at both ends. Sequences for Trn' are not found in the target DNA sample

and therefore will not hybridize with the target DNA. For example, for human genotyping, sequences for Trn' can derive from bacteria specific sequences that have no homology with the human DNA.

5. Collect (and concentrate) all DNA that are not bound to the nitrocellulose after branch migration by washing with minimum amount of buffer (either the same buffer used for Holiday junction formation and branch migration or other appropriate buffers). This collection of DNA includes excess reference partial DNA duplexes and the released probes due to complete strand exchange.
6. Any method that allows the determination of the identity of the released probes in the above collection can be used for SNP scoring.

As an illustration rather than limitation,

1. amplify the probes that are released from the nitrocellulose due to complete strand exchange via using (Figure 1, page 1):

- a. Labeled primer T'--which is complementary to T-- and various DNA polymerases, including Taq polymerase, Taq Gold..., multiple cycles of DNA replication.
- b. When $Tr1=Tr2=Tr3=Trn=Tr$, use primer pair labeled T' (which is complementary to T) and unlabeled Tr to do PCR amplification.
- c. When $Trn=0$ or $Tr1\neq Tr2\neq Tr3\neq \dots\neq Trn$, used labeled primer T' (which is complementary to T) and a pool of random primers to do PCR amplification.

2. alternatively, selectively amplify the probes that are released from the nitrocellulose due to complete strand exchange via using (figure1, page 2):

- a. universal primer T'--which is complementary to T-- and various DNA polymerases, including Taq polymerase, Taq Gold..., multiple cycles of DNA replication (either PCR or Multiple Displacement Amplification, etc),
- b. Universal primer UR that is not complementary to any DNA sequences in the genotyping assay,
- c. mixture of different oligos, each oligo of the mixture is comprised of a universal 5' tail UR' that is complementary to UR and a 3'

portion that is part of or the whole middle portion of its corresponding probe and is therefore unique for each SNP tested

d. at least one of primer T' or primer UR is labeled for detection

3. For the identification of multiple released probes (multiplexing genotyping), the released probes selectively amplified are hybridized with DNA chip/micro-array or (fluorescent) beads that are immobilized/coated with the DNA sequences between T' and UR for each of the n SNPs of interest.
4. For the identification of released probes one by one (one single SNP at a time), the presence or absence of the released oligos for a specific SNP can be detected by monitoring the amplification of the released oligos (e.g., using fluorescent dyes such as PicoGreen or Ethidium Bromide).

7. In order to eliminate target DNA (released oligos) amplification step (by either PCR or MDA) completely, the probes can be labeled, for an example, with magnet beads and the released probes due to complete strand exchange can then be separated from reference DNA and isolated via magnet. The identification of the isolated released probes can be obtained by using hybridization with DNA chip/micro-array or (fluorescent) beads that are immobilized/coated with the n SNPs of interest.

Tabel 1 shows different bar code for three different genotypes at a specific SNP position by using the above scoring method.

An important issue for genotyping is accuracy. Fluctuation in the amplification step can lead to false or difficult to interpret signals. Therefore, it is important to have controls:

1. External control:
 - a. Genotyping a few well-studied/characterized/genotyped, (e.g: 1-10)
SNPs with highly accurate but expensive genotyping assays for the target DNA and then use these SNPs as external control.

- b. Mix some control DNA (with known sequences that are not found in target DNA) at comparable concentration with the target DNA before immobilization. Add control probes at comparable concentration in the probe pool before hybridization with the immobilized DNA. Also add control reference partial DNA duplexes in the reference partial DNA duplex pool. Correct or wrong scoring of these control DNA can give you a good estimation about the quality of each assay performed.
2. Internal control:
 - a. The collected DNA pools containing released probes from different combination of probe and reference DNA partial duplexes can be labeled differently. For example: The probes released from version 1 probe with version 1 reference partial DNA duplex are labeled with both label X (e.g. red label) and label Y (e.g.: green label), one label at a time. The probes released from version 1 probe with version 2 reference partial DNA duplex are also labeled with both label X (e.g.: red label) and label Y (e.g. green label), one at a time. The X-labeled/amplified pool of probes released from version 1 probe with version 1 reference partial DNA duplex are mixed with the Y-labeled/amplified pool of probes released from version 1 probe with version 2 reference partial DNA duplex before hybridization with the chip/micro-array. In addition, the Y-labeled/amplified pool of probes released from version 1 probe with version 1 reference partial DNA duplex are mixed with the X-labeled/amplified pool of probes released from version 1 probe with version 2 reference partial DNA duplex before hybridization with the chip/micro-array. In this case, the ratio the intensity of red signal vs. green signal will be scored instead of the absolute intensity of red or green signal. Along the same line, many different schemes of scoring can be apparent to a skilled researcher in molecular biology.

To achieve optimal accuracy, the internal controls and external controls can be combined to produce many different schemes for SNP scoring based on the method disclosed here.

CLAIMS

What is claimed is:

1. A method for detecting the sequence identity of a first nucleic acid and a second nucleic acid said method comprising:
 - a. contacting the first nucleic acid with the second nucleic acid under conditions in which the first nucleic acid and second nucleic acid are capable of forming a four-way complex, wherein one strand of the first nucleic acid is immobilized;
 - b. subjecting said four-way complex to branch migration conditions wherein, if a difference between said two related nucleic acid sequences is present, branch migration in said four-way complex ceases and said four-way complex is stabilized, and wherein, if no difference between said two related nucleic acid sequences is present, branch migration in said four-way complex continues until complete strand exchange occurs and said four-way complex resolves thereby releasing the non-immobilized strand of the first nucleic acid;
 - c. detecting the released strand of the first nucleic acid as an indication of a sequence identity of the first nucleic acid and the second nucleic acid.

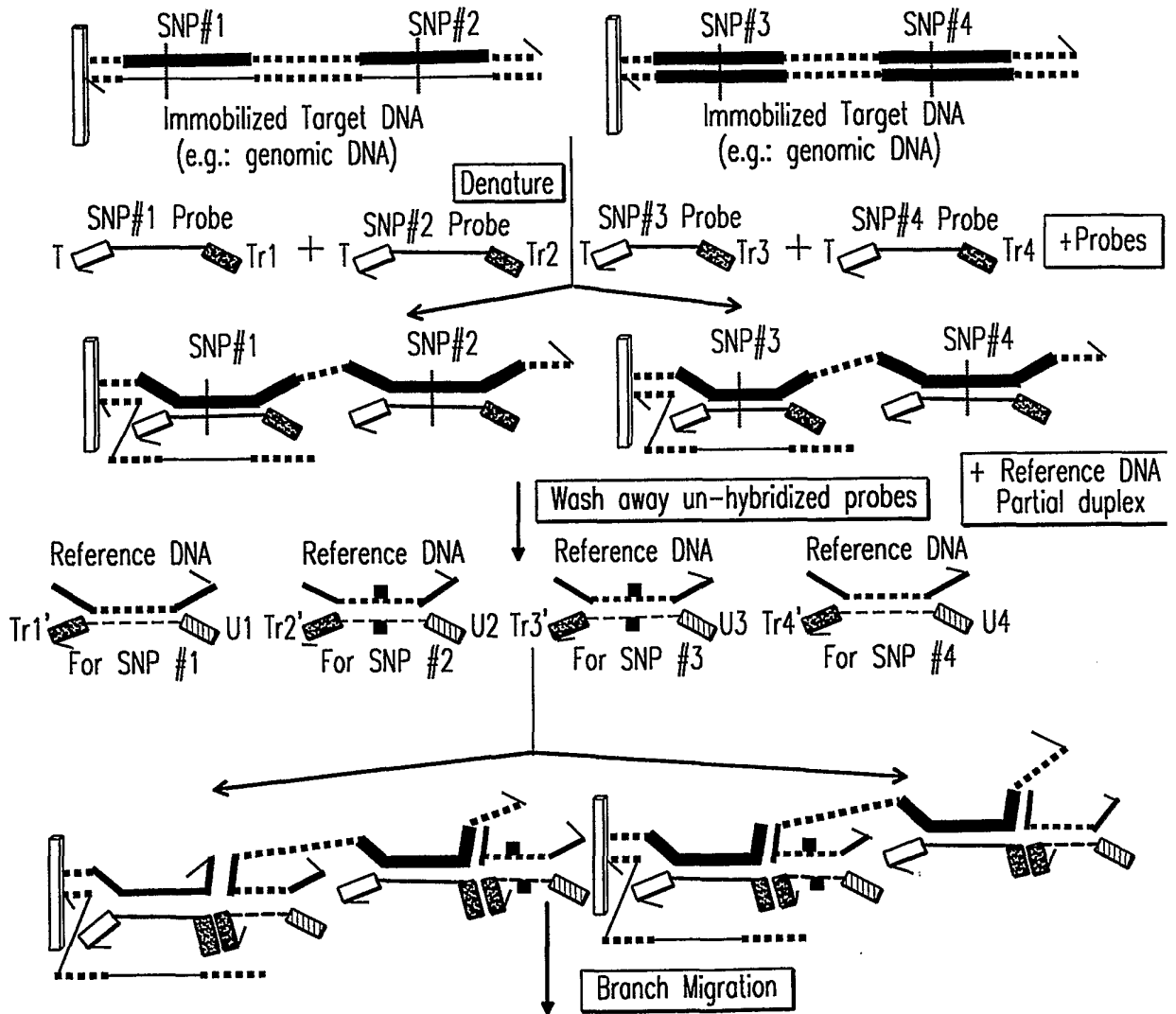


FIG.1A

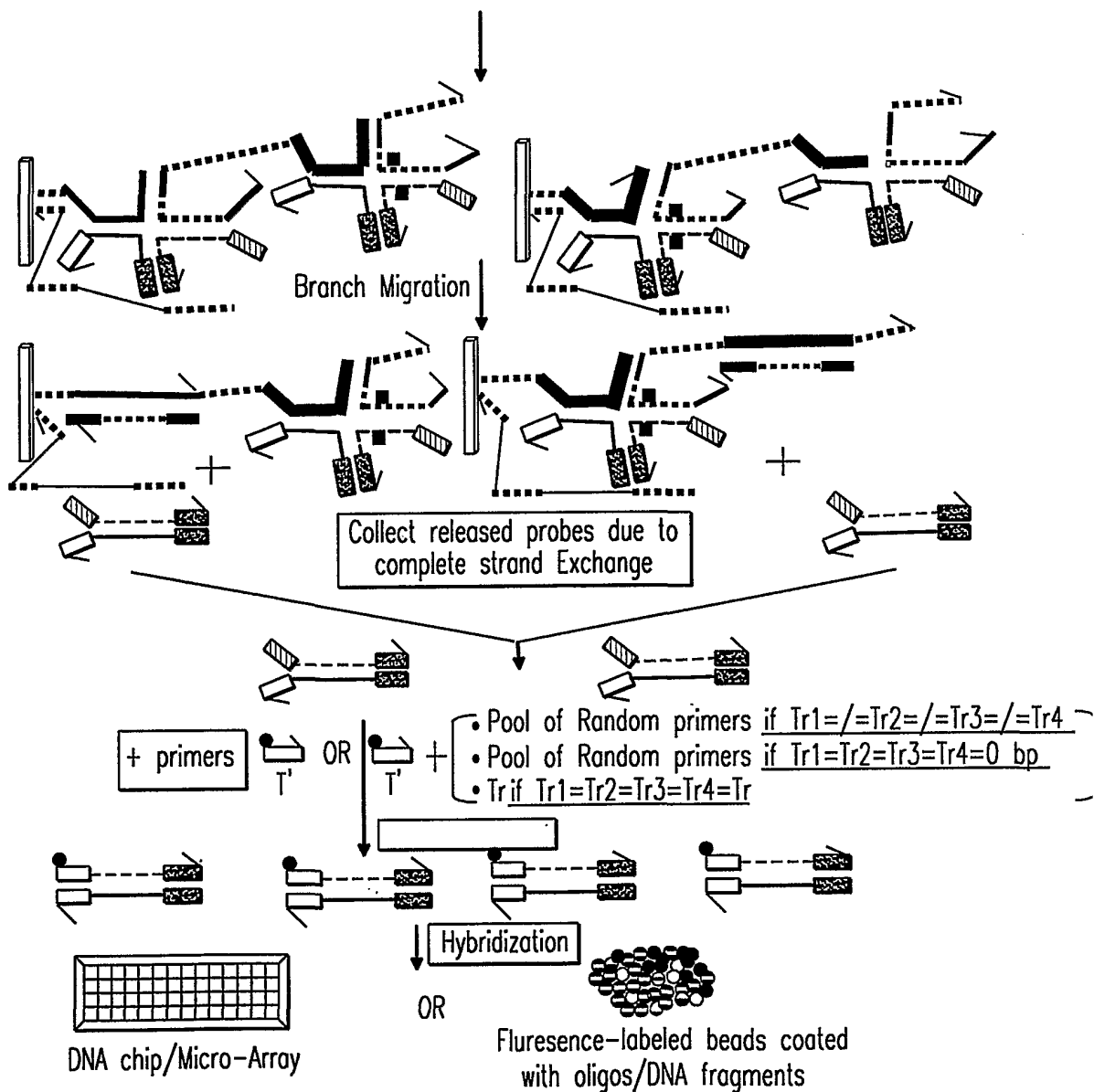


FIG.1B

OR ALTERNATIVELY: 3/3

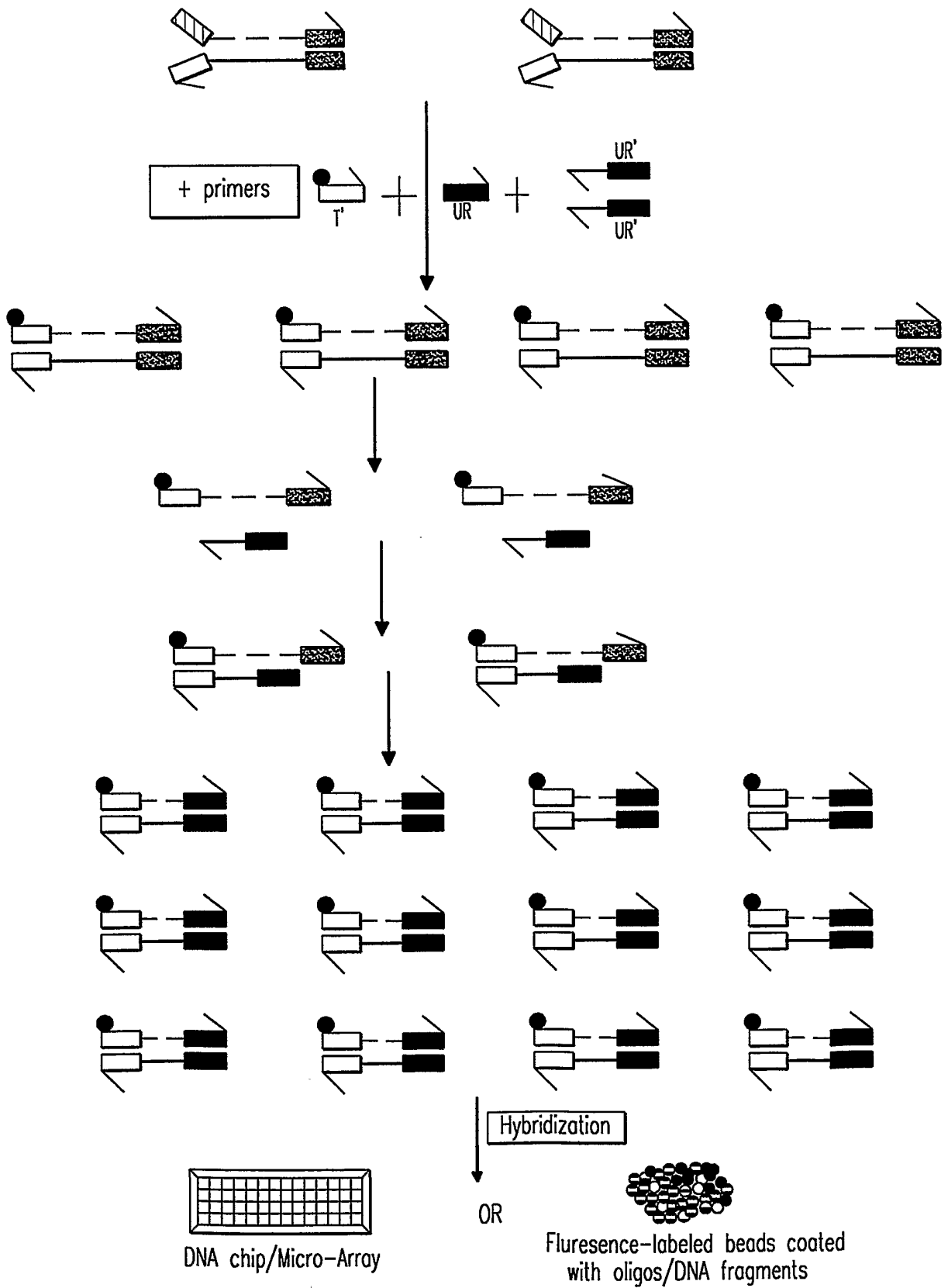


FIG.1C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/29922

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(7) : C12Q 1/68; C12P 19/34
 US CL : 435/6, 91.1, 91.2
 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)
 U.S. : 435/6, 91.1, 91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
 NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,013,439 A (LISHANSKI et al) 11 January 2000, columns 9 and 25.	1
Y	WO 00/20643 A1 (MOSAIC TECHNOLOGIES) 13 April 2000, page 11.	1

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	"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
"E"	earlier application or patent published on or after the international filing date	"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
"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)	"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
"O"	document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 29 October 2001 (29.10.2001)	Date of mailing of the international search report 13 FEB 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Carla Myers <i>Janice Foul for</i> Telephone No. 703-308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29922

Continuation of B. FIELDS SEARCHED Item 3:

DIALOG: Medline, CA, BIOSIS, EMBASE, SciSearch; WEST: US, EP, WO, JP Patents

serach terms: heteroduplex, mutation, polymorphism, variant, Holliday, branch migration, displace