

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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
7 August 2003 (07.08.2003)

PCT

(10) International Publication Number  
**WO 03/064694 A1**

(51) International Patent Classification<sup>7</sup>: **C12Q 1/68**

(21) International Application Number: PCT/US02/34599

(22) International Filing Date: 30 October 2002 (30.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/352,039 25 January 2002 (25.01.2002) US  
60/352,356 28 January 2002 (28.01.2002) US

(71) Applicant (*for all designated States except US*): **APPLERA CORPORATION** [US/US]; 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **KOEHLER, Ryan, T.** [US/US]; 3175 Cromwell, Hayward, CA 94542 (US). **LIVAK, Kenneth, J.** [US/US]; 1148 Polk Lane, San Jose, CA 95117 (US). **STEVENS, Junko** [US/US]; 1157 Windsor Way, Menlo Park, Ca 94025 (US). **DE LA VEGA, Francisco, M.** [MX/US]; 3456 Edison St., San Mateo, CA 94403 (US). **RHODES, Michael** [GB/US]; 107 Ottawa Street, San Mateo, CA 94401 (US). **BELLON, Laurent, R.** [FR/US]; 1927 Ticonderoga Dr., San Mateo, CA 94402 (US).

(74) Agents: **CASSEL, Alan, L.** et al.; Harness, Dickey & Pierce, P.L.C., P.O. Box 828, Bloomfield Hills, MI 48303 (US).

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

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

Published:

— with international search report

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

(54) Title: METHOD FOR OPERATING A COMPUTER AND/OR COMPUTER NETWORK TO DISTRIBUTE BIOTECHNOLOGY PRODUCTS

(57) Abstract: A method for distributing a product to a consumer includes utilizing a computer network to interact with the consumer to obtain information that includes at least one target nucleic acid sequence; providing a forward primer sequence, a reverse primer sequence and a probe sequence having specified characteristics, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence; manufacturing at least one assay that includes a forward primer in accordance with the forward primer sequence, a reverse primer in accordance with the reverse primer sequence, and a probe in accordance with the probe sequence; validating one or more of the forward primer, the reverse primer, and the probe; and delivering the manufactured assay to the consumer.



WO 03/064694 A1

METHOD FOR OPERATING A COMPUTER AND/OR COMPUTER  
NETWORK TO DISTRIBUTE BIOTECHNOLOGY PRODUCTS

CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims the benefit of U.S. Provisional Application No. 60/352,039, entitled "Method of designing and manufacturing custom polynucleotide sequences for a requestor," filed on January 25, 2002, and of U.S. Provisional Application No. 60/352,356, also entitled "Method of designing and manufacturing custom polynucleotide sequences for a requestor," which was filed on January 28, 2002. The disclosures of the above applications are incorporated herein by reference in their entirety.

FIELD

**[0002]** The present invention relates computer systems and networks, and more particularly to methods for utilizing such computer systems and networks for accepting and filling orders for products and services, including, for example, biotechnology laboratory assays and services.

BACKGROUND

**[0003]** One goal of personalized medicine is to improve medical treatment by predicting disease risk and response to therapies. Currently available genome data and the accompanying single nucleotide polymorphisms (SNPs) can enable large-scale studies to achieve this goal. However, laboratories attempting to perform SNP genotyping and gene

expression studies currently spend substantial time, money, and manual labor to design custom assays.

#### SUMMARY

**[0004]** Some configurations of the present invention therefore provide a method for providing a product to a consumer. The method includes utilizing a computer network to interact with the consumer to obtain information associated with at least one target nucleic acid sequence; providing a forward primer sequence, a reverse primer sequence and a probe sequence, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence; manufacturing at least one assay that includes a forward primer in accordance with the forward primer sequence, a reverse primer in accordance with the reverse primer sequence, and a probe in accordance with the probe sequence; validating one or more of the forward primer, the reverse primer, and the probe; and delivering the manufactured assay to the consumer.

**[0005]** Some configurations of the present invention provide a method for providing a product to a consumer, wherein the method includes: utilizing a computer network to interact with the customer to obtain information associated with at least one target nucleic acid sequence; providing a forward primer sequence, a reverse primer sequence and a probe sequence having specified characteristics, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon

lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence; manufacturing at least one assay comprising a forward primer in accordance with the forward primer sequence, a reverse primer in accordance with the reverse primer sequence, and a probe in accordance with the probe sequence; validating one or more of the forward primer, the reverse primer, and the probe; and delivering the manufactured assay to the customer.

**[0006]** Various configurations of the present invention provide a method for operating a computing system. The method includes: receiving information that includes at least one target nucleic acid sequence via a computer communication network; attempting to design an assay that includes a forward primer sequence, a reverse primer sequence, and a probe sequence, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon sequence lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence; validating at least one of the forward primer sequence, the reverse primer sequence, and the probe sequence against a genome database; recording each design attempt in a log file, including whether each of the design attempts succeeded or failed to meet constraints set by the design metrics and scoring metrics; and utilizing the log file to generate output sequence data for at least one forward primer sequence, reverse primer sequence, and probe sequence.

**[0007]** Some configurations of the present invention provide a method for operating a computing system. The method includes: receiving information comprising at least one target nucleic acid sequence via a

computer communication network; providing assay design metrics and scoring metrics; attempting to design an assay comprising a forward primer sequence, a reverse primer sequence, and a probe sequence, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon sequence lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence; validating at least one of the forward primer sequence, the reverse primer sequence, and the probe sequence against a genome database; recording each design attempt in a log file, including whether each said design attempt succeeded or failed to meet constraints set by the design metrics and scoring metrics; and utilizing the log file to generate a batch of output sequence data for at least one forward primer sequence, reverse primer sequence, and probe sequence.

**[0008]** It will be seen below that the various configurations of the present invention will save consumers substantial time, money, and manual labor that custom assay design requires, thereby enabling a new level of high-throughput, low-cost SNP genotyping or gene-expression research studies. Also, various configurations provide ready-to-use, functionally tested assays that confirm performance.

**[0009]** Further areas of applicability of the present invention will become apparent from the detailed description provided hereinafter. It should be understood that the detailed description and specific examples, while indicating preferred embodiments of the invention, are intended for purposes of illustration only and are not intended to limit the scope of the invention.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0010]** The present invention will become more fully understood from the detailed description and the accompanying drawings, wherein:

**[0011]** Figure 1 is a block diagram representation of various configurations of a computing system of the present invention that is useful for distributing biotechnology products to a consumer.

**[0012]** Figure 2 is a flow chart representative of various method configurations of the present invention that can be performed by computing system configurations such as those represented in Figure 1, or by other computing system configurations.

**[0013]** Figure 3 is a flow chart representative of various configurations of the present invention that provide a method for operating a computer system to distribute a product to a consumer.

**[0014]** Figure 4 is a block diagram representative of components and data flow in various configurations of an assay design system.

**[0015]** Figure 5 is a diagram representative of various configurations of assay design program logic suitable for use in assay design system configurations represented by Figure 4.

**[0016]** Figure 6 is a diagram representative of various configurations of reagent design procedures suitable for use in assay design program logic configurations represented by Figure 5.

**[0017]** Figure 7 is a diagram representative of various configurations of probe placing procedures suitable for use in reagent design procedure configurations represented by Figure 6.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

**[0018]** The following description of the preferred embodiment(s) is merely exemplary in nature and is in no way intended to limit the invention, its application, or uses.

**[0019]** As used herein, the terms "distribute" and "provide" may be used synonymously, and are intended encompass selling, marketing, or otherwise providing a product or service. The terms "distributor" and "provider" thus encompass sellers, marketers, and other providers of such products and services, and the term "consumer" encompasses customers and other users of the products and services. Unless explicitly stated otherwise, it is permitted but not required that configurations of the present invention precondition distribution on receipt of a payment or a promise to pay for the distributed products or services.

**[0020]** Also as used herein, unless otherwise explicitly stated, the terms "a," "an," "the," "said," and "at least one" are not intended to be limited in number to "one," but rather are intended to be read as encompassing "more than one" (i.e., a plurality) as well. Also, "a probe" is intended to include two bi-allelic probes in the case of SNP assays.

**[0021]** Where examples are recited herein, such examples are intended to be non-limiting.

**[0022]** In various configurations of the present invention and referring to Figure 1, a computing system 10 comprising a plurality of computers 12, 16 is utilized to distribute a product to a consumer 18. A first computer 12 (i.e., a distributor computer) on a computer network 14 (e.g., a public network, such as the Internet) interacts with a consumer 18 using a

second computer 16 (i.e., a consumer computer) to obtain information that is associated with a human or nonhuman target DNA (or RNA) sequence, including SNP and/or exon locations, i.e., the sequence itself, the SNP and/or exon locations themselves, or other information from which these items may be determined such as, for example, a gene name, accession number, etc. In some configurations, this interaction is initiated by consumer 18 typing a uniform resource locator (URL) into a web browser running on consumer computer 16 and downloading a hypertext mark-up language (HTML) or other type of web page from a server 20 running on distributor computer 12. The web page displayed on consumer computer 16 may include various types of introductory and sales information, provide a login for authorized user/purchasers, and solicit the DNA (or RNA) sequence and other information, as is necessary or desirable. In some configurations, the initial web page is one of several web pages provided by server 34 that interact with consumer 18 to obtain information. For example, in some configurations, the initial web page accessed by consumer 18 is a corporate web site that provides information for consumer 18 as well as a form in which consumer 18 types identifying information using consumer computer 16. Distributor computer 12 receives the information entered by consumer 18 and sent by consumer computer 16 via computer network 14. In some configurations, distributor computer 12 verifies the identity of consumer 18 and his or her qualifications to access a sales page and to purchase assays from the distributor. For example, this verification may be performed by a web application server 22 (for example, the IBM® WEBSphere® Application Server available from International Business Machines Corporation, Armonk



NY) running on distributor computer 12 with reference to a consumer database 24 of qualified consumers and consumer identifications. If consumer 18 cannot be verified or is not qualified to make a purchase, this information may be returned by web application server 22 and web page server 20 via computer network 14 to consumer 18, and consumer 18 will not be allowed to complete a purchase and/or to access additional information.

**[0023]** If consumer 18 is verified and qualified (or in configurations in which verification and/or qualification is not required), consumer 18 specifies information including at least one target nucleic acid sequence. In some configurations, the target nucleic acid sequence is a target DNA sequence, and the information provided by consumer 18 and obtained by distributor computer 12 may also include an exon or a portion thereof, and/or a single nucleotide polymorphism (SNP). Other information that may be provided by consumer 18 and obtained by distributor computer 12 may include a SNP location and/or an exon location. Information from consumer 18 may, for example, be provided on a web page form by consumer 18 using consumer computer 16 sent to and received by distributor computer 12 via computer network 14. Upon receiving information from consumer 18, some configurations of the present invention analyze the target nucleic acid sequence provided by the consumer for format errors. A variant configurator 26 (such as SELECTICA<sup>®</sup> Configurator<sup>™</sup>, available from Selectica, Inc., San Jose, CA) interacts with consumer 18 via network 14 to produce a list of specified characteristics, as discussed below. Configurator 26 is essentially an automated decision tree that produces the input for assay design program 28 and that ensures that input parameters to assay design program 28 are

within bounds that can be handled by program 28. If there are no errors, assay design program 28 then uses a lookup process, a design process, or another suitable method to provide a forward primer sequence, a reverse primer sequence, and a probe sequence that have the specified characteristics. The forward primer sequence and the reverse primer sequence define an amplicon sequence, which itself lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence. Some configurations of the present invention utilize distributor computer 12 to validate or "BLAST" the results from assay design program 28 against a genome database 30, for example, a human genome database. The term "BLAST" as used herein is intended to refer to a Basic Local Alignment Search Tool such as was developed by Altschul et al. (*J Mol Biol* 215:403-10, 1990) which comprises a fast search algorithm to search DNA databases based upon sequence similarities. Such a BLAST search tool is useful for various validation methods that can be used for the probe and primer sequences. The validation process is intended to verify that the probe and primer sequences are selective for the target region, i.e. they will hybridize to the target region but not to other regions of the genome.

**[0024]** Upon successful validation, oligo factory 32 accepts the order from consumer 18, manufactures at least one assay having components including a forward primer, a reverse primer and a probe and ships the manufactured assay to the consumer. The manufactured forward primer, reverse primer, and probe are manufactured in accordance with the validated sequences. In some configurations, the assay is shipped as a homogenous assay in a single tube format with a two-dimensional bar code.

**[0025]** In various configurations, the manufactured assay is tested prior to delivering to the consumer to verify that the assay meets the specified characteristics. For example, the testing may include performing mass spectroscopy on the assay to determine that an oligonucleotide sequence is correct, and/or the testing may include performing a functional test to determine that an amplification has occurred and at least one allelic discrimination is confirmed.

**[0026]** In some configurations, the probe in the assay shipped to the consumer includes a non-fluorescent dye that is configured to reduce background fluorescence and increase quenching efficiency. Thus, the assay is particularly suitable and provides a substantial benefit to consumers using PCR sequence detection systems such as the Applied Biosystems PRISM® 7900HT Sequence Detection System, enabling high-throughput SNP genotyping in which approximately 250,000 genotypes per day are analyzed, each needing only a small amount of sample DNA.

**[0027]** Referring to Figures 1 and 2, various configurations of the present invention perform a method 34 for distributing a biotechnology product to a consumer. More particularly, the method includes utilizing a computer network 14 to interact at 36 with a consumer 18 to obtain information associated with (i.e., indicative of) at least one nucleic acid sequence. The target nucleic acid sequence obtained from the consumer is, for example, a target RNA or DNA sequence, which itself may include an exon or a portion thereof, and/or a single nucleotide polymorphism (SNP). The information may further include information associated with a SNP location and/or an exon location. The provided nucleic acid sequence is

analyzed at 38 for format errors. If errors are detected, further interaction at 36 may be performed to correct the format errors. (In some configurations, prior to interacting at 36 with consumer 18 to obtain information comprising a nucleic acid sequence, consumer 18 is required to verify his or her identity via computer network 14, and/or confirm his qualifications to place an order.)

**[0028]** Upon obtaining information from consumer 18, various methods of the present invention provide, at 40, a forward primer sequence, a reverse primer sequence, and a probe sequence having specified characteristics. For example, the specified characteristics may include size limits, target  $T_m$  (melting temperature), minimum  $T_m$ , total matching bases in a hairpin stem, contiguous matching groups in a hairpin stem, combined G and C content, runs of G bases, runs of non-G bases, and/or G content not greater than C content ( $G \leq C$ ).

**[0029]** The forward primer sequence and the reverse primer sequence together define an amplicon sequence. The amplicon lies within the target nucleic acid sequence. The probe sequence is complementary to a portion of the amplicon sequence. Next, in various configurations, one or more of the forward primer sequence, the reverse primer sequence, and the probe sequence are validated at 42, using, for example, a genome database such as database 30. Validation may include BLASTing of one or more of the sequences, as described above. At least one assay is manufactured at 44. The manufactured assay comprises a forward primer in accordance with the forward primer sequence, a reverse primer in accordance with the reverse primer sequence, and a probe in accordance with the probe sequence. In some configurations, the forward primer sequence, the reverse primer

sequence, and/or the probe sequence is a validated sequence from 42. The assay is shipped at 48 to consumer 18. Some configurations of the present invention ship the assay in a single tube format with a two-dimensional bar code. In some configurations, the probe in the manufactured assay comprises a non-fluorescent dye configured to reduce background fluorescence and increase quenching efficiency. The assay itself is suitable for use in a sequence detection system.

**[0030]** Some configurations test, at 46, the manufactured forward primer, the manufactured reverse primer, and/or the manufactured probe before delivery to verify that the assay meets specified characteristics. Tests at 46 may include, for example, performing mass spectroscopy on the manufactured assay to determine that an oligonucleotide sequence is correct, and/or performing a functional test to determine that an amplification has occurred and at least one allelic discrimination is confirmed.

**[0031]** In various configurations of the present invention and referring to Figure 3, a method 50 for operating a computer system is provided. The method includes receiving, at 52, information including at least one target nucleic acid sequence via a computer communication network such as network 14 of Figure 1. Assay design metrics, i.e. specified characteristics of probe and primers as well as scoring values for the specified characteristics are provided at 54, for example, from local storage. The scoring approach involves assessing how closely the probe and primers conform to each of the specified characteristics and, then, arriving at an overall value for conformance which can be used to determine whether the probes and primers are acceptable.

**[0032]** Using the information received at 52, the computer system is used to attempt to design, at 58, an assay that comprises a forward primer sequence, a reverse primer sequence, and a probe sequence. The forward primer sequence and the reverse primer sequence together define an amplicon sequence. The amplicon sequence lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence. Some configurations also attempt to apply, at 60, design metrics or constraints to the resulting assay design, or portions thereof. For example, constraints may include size limits, target  $T_m$ , minimum  $T_m$ , total matching bases in a hairpin stem, contiguous matching groups in a hairpin stem, combined G and C content, runs of G bases, and runs of non-G bases. These constraints may be applied to either the forward primer sequence, the reverse primer sequence, or the probe sequence, or a combination thereof. Attempting to design the assay at 58 may also comprise, in some configurations, attempting to design at least one of the primer sequences in accordance with at least one constraint including a limit on G+C at a 3' end (5 bases) of the primer sequence.

**[0033]** Each design attempt is recorded at 62 in a log file, including whether each design attempt succeeded or failed to meet constraints set by the design metrics and scoring metrics. If the design at 64 fails to meet these constraints, another attempt to design the assay at 58 may be made. Otherwise, if the design is adequate, the log file is used to generate output sequence data at 66 for at least one forward primer sequence, reverse primer sequence, and probe sequence. Some configurations also assign, at 68, a serial number to each batch of output sequence data.

**[0034]** In some configurations and referring to Figure 4, an assay design system 70 is provided as computer software that allows automated, high-throughput design of TAQMAN<sup>®</sup> primers and probes for allelic discrimination and gene expression assays in a batch format. This computer software is particularly useful when designing hundreds or thousands of assays. Assay design program 28 is a non-interactive pipeline of algorithms for the design of TAQMAN<sup>®</sup> probe and primer reagents. In some configurations, heuristic rules are utilized in assay design program 28. Pre- and post-processing utility programs and wrapper scripts are utilized as components of the complete assay design system 70.

**[0035]** Components and dataflow of assay design system 70 include two data files 72 and 73. A sequence input file 72 contains formatted and annotated sequence data. A parameter file 73 contains keyword-associated settings that govern rules and scoring applied during designs. Prior to attempting any designs, the format of supplied sequence data is checked at 74 for errors. If errors are found at 76 in the sequence data from input file 72, they are reported to an error log 88 and the process terminates.

**[0036]** Assay design program 28 starts by parsing a parameter file 73 to set up rules and scoring schemes. If there are any errors encountered during this initialization phase, they are reported to log file 88 and assay design program 28 stops. Initialization errors may be caused by conflicting options or incorrect file names or formats. Following successful initialization, assay design program 28 sequentially attempts to design assay sets for each target site in each sequence listed in the input sequence data from parameter file 73. As designs are processed, they are recorded in a design log file 80.

Design attempts that fail are also recorded in log file 80. Design failure occurs when no acceptable set of reagents satisfying all rules and scores is found for a sequence target.

**[0037]** If, at 82, there are no valid designs present in design log file 80, this fact is reported in error report 88. Otherwise, following the core design process, design log file 80 may be used to generate output sequence data in a number of different formats. Log pick program 84 performs this post-processing of design log 80 data to produce formatted outputs 86. A script can be implemented in the UNIX operating system to integrate the whole system by tying together all of the processes shown in Figure 4. The script also logs each runs of process 70 and assigns each output batch a serial number for tracking purposes.

**[0038]** Input to assay design program 28 includes a parameter file 73 that specifies design rules and one or more sequence data files 72. Output includes a log file 80 that reports system settings and attributes describing each successful reagent design (including probe, primer, and amplicon sequences). Additional output indicating a system status is reported to a display screen as the program is running, in some configurations.

**[0039]** Separate design rules and constraints are applied to potential probes, primers, and "amplicons," i.e., a region defined by pairing of forward and reverse primers around the target site. All designs resulting from a given run share a common set of rules. Probe constraints include limits on size (i.e., probe length),  $T_m$  (target, minimum, and maximum temperatures), internal loops (total and contiguous matching bases in a "hairpin stem"), G+C content (i.e., combined G and C percentage), and runs of a given base, such



as G. Analogous constraints are also separately applied to primers, which have an additional limit on G+C at the 3' end (5 bases) of the primers. Constraints applied to amplicons include length (including primers), G+C content, and the number of ambiguous bases (note that ambiguous bases are never allowed within probes or primers). In addition, the primers defining amplicons are constrained to limit the maximal size of internal priming sites (i.e., the number of contiguous matching bases starting at the 3' end of one primer that complements any part of the other primer).

**[0040]** For many of the constraints listed above, system 70 may apply either a filter or a score. When applied as a filter, a constraint will be either satisfied or not with the corresponding design being either accepted or rejected. When applied as a score, attributes may be given a graded value that reflects how "optimal" a given design is. For example, a design with all constrained attributes near optimum values will be favored over one with attributes deviating from the optimum values. Scoring provides finer tuning of the constraints that system 70 will use to evaluate and select designs.

**[0041]** The logic of assay design program 28 in various configurations of the present invention is shown in more detail in Figure 5. Upon starting program 28 at 92, an initialization phase 98 reads parameter data from parameter file 73 and sequence data from sequence data file 72. (As shown in Figure 4, sequence data 72 may be checked for errors at 74 and 76 before being read by assay design program 28.) Initialization 98 includes parsing parameter file 73 and setting up for subsequent probe design. If any problems are encountered at 100 as a result of initialization 98, assay design program 28 reports a diagnostic message and stops at 104. Otherwise,

processing continues. In some configurations of the present invention, most parameter file 73 options have default values and may be superceded by command line options. Options actually used during design are reported in log file header 102, which is or becomes part of design log 80 of Figure 4.

**[0042]** Assay design program 28 attempts to acceptably design assay sets for each target site at 110. These designs are logged at 112. An attempt is made to identify acceptable designs at 114 for each input sequence record from sequence data file 72. When records are exhausted at 106, assay design program 28 is done at 104. Otherwise, for each record, each target is tried at 110 in the order listed. If no target information is supplied, the sequence midpoint (if the sequence contains no SNP annotations) or the first SNP (if annotated) is used as a target. When no targets are left for a given record at 108, assay design program 28 progresses at 106 to the next record.

**[0043]** For target sites, assay design program 28 identifies, at 114, successful and unsuccessful designs, according to the design metrics and scoring metrics. If it fails to design for a target, this fact along with the corresponding unsuccessful design is reported to log file 80 and the program progresses at 108 to the next target associated with the record. If it succeeds to design for a target, the details of the chosen record are reported to log file 80. Normally, a single successful design causes assay design program 28 to move to the next record at 106. However, in some configurations, if an option to evaluate all targets listed for each record is enabled, assay design program 28 progresses, at 116, to the next target at 108 rather than to the next record at 106 following a successful design.

**[0044]** The logic of procedure 110 for designing reagents for a simple target in various configurations of the present invention is shown in more detail in Figure 6. Upon starting at 122, design for record/target program 110 extracts design "windows" at 124, e.g., one or two subsequences around the target are extracted. For SNP targets, two separate windows are extracted at 124 around the SNP target site, one for each allele. In addition, any other SNP that is known to be within the sequence of the window is masked by converting it to an N, which represents any nucleotide. For non-SNP targets, a single subsequence window is extracted at 124. Windows are limited in size by the supplied input sequence length or by the maximum allowable amplicon size. Problems encountered at 126 during window extraction (for example, an incorrectly formatted SNP annotation) cause a failure at 146. (In general, failures in this and other procedures or functions may be reported to the consumer and result in no product being shipped. Failures resulting from data that is improper, inconsistent, out-of-bounds, etc. are not fatal, as, in various configurations, the software is configured to reset itself after an order or failure therefore, to be ready for the next order.)

**[0045]** If no problems are encountered, placement of probes is normally attempted next at 130, unless an option to design only primers is enabled at 128, in which case, execution continues at 134. (A primer-only option may be enabled, for example, by a command line option, such as "-op".) Probe placement at 130 yields either one or two acceptable probes (non-SNP and SNP cases, respectively), or not. If acceptable probes are not

identified at 132, target design process 110 fails at 146. Otherwise, bounds are set for primers at 134.

**[0046]** To set primer bounds at 134, three sub-regions within the design window are defined. In cases in which probes are designed (e.g., cases in which not only primers are designed), a central mask region corresponding to coordinates of the probes is defined. Bounds for the mask region may be explicitly designed relative to target site coordinates. For example, in some configurations, a command line option (such as "-pm") is used to specify that the mask region is to be explicitly designed relative to target site coordinates. In this case, the actual mask region is the larger of the specified bounds or the mask formed by the probes. Fixing the central mask region determines the two sub regions where primers may be designed. The "upstrand" sub-region begins at the start of the window and extends to the start of the mask region. The "downstrand" sub-region follows the mask and extends to the end of the window. The three sub-regions of the window (i.e., upstrand, mask, and downstrand) do not overlap.

**[0047]** With the upstrand and downstrand sub-regions determined, design procedure 110 attempts to collect a number of primers in each sub-region at 136. Forward primers are taken from the upstrand region and reverse primers are taken from the downstrand region. Potential primers are evaluated at each nucleotide position starting from the coordinates closest to the mask (i.e., the end and start coordinates of the upstrand and downstrand regions, respectively). Such evaluation may, for example, determine whether a potential primer is acceptable according to standards known and recognized in the art. In some configurations, design procedure 110 collects up to ten

forward and ten reverse primers, but by setting a command line option (such as "-np"), the limit of ten can be changed to another number.

**[0048]** If at least one potential forward and one potential reverse primer is not found at 138, design process 110 fails at 146. With two lists of primers, design process 110 next attempts to identify an acceptable forward/reverse pair at 140. If no acceptable primer pair is identified at 142, design process 110 fails at 146. Otherwise, a complete design has been found at 144.

**[0049]** The logic of procedure 130 for placing probes in various configurations of the present invention is shown in more detail in Figure 7. When a design attempt is begun at 150, a probe placement is attempted. The logic follows slightly different but similar paths depending upon a determination of whether the target is a SNP site or a non-SNP site at 152. For SNP sites, sample sequences on both alleles of both strands are considered at 156. For non-SNP sites, a determination is made at 154 as to whether both strands or only a single strand is used at 156, 158, or 160. Explicit strands are then determined at 162 or 166 and non-target strand probes are eliminated at 164 or 168 to pick the best probe at 170 or 172. Features that are used to pick the best probe are determined on the basis of  $T_m$  value and filter or score values. Minimal target overlap is an input parameter. For non-SNP targets this value may be negative, allowing a larger sequence region to be sampled. For SNP targets, the minimum value of target overlap is two bases, but the overlap may be increased. Probes targeting both forward and reverse strands are evaluated. Probes may not start with G and normally the requirement that G content does not exceed C

content is applied, but an option is provided to eliminate the  $G \leq C$  rule. In some configurations, and in some cases, both forward and reverse strands are considered explicitly for  $T_m$  delineation. If probe scoring is being applied, the best scoring probe is selected. Otherwise, the constraint satisfying probe most overlapping the target site is selected. From this determination of whether a probe is acceptable or not at 174 or 176, probes are selected that pass, at 178, or alternatively are selected against, at 180, for failing the criteria.

**[0050]** For SNP target sites, sequences corresponding to both alleles (only bi-allelic SNP sites are supported in some configurations) are explicitly constructed and the best probes for both strands of both allele sequences are identified as described above. An acceptable pair of SNP probes must target the same sequence strand. If acceptable probe pairs are found for both strands, the strand yielding the pair with the largest total score is selected. When input sequences have multiple SNP sites denoted, the non-targeted SNP sites are masked (i.e., set to base N) when the sequences for each explicitly targeted allele are constructed.

**[0051]** If no acceptable probe (or, for SNPs, probe pair) is found for a given target, the system reports this fact and attempts to continue, depending upon the number and format of sequence targets supplied. If a single sequence is supplied as input, failure to select a probe (or pair) results in a program termination. If multiple target coordinates (or SNPs) are listed for a given sequence, failure to place a probe at one target coordinate causes probe placement process 130 to consider the next listed coordinate until all listed targets are exhausted. For multiple sequence input, failure to place a

probe at any target coordinate leads the program to address the next listed sequence until all input sequences are exhausted. If there are multiple targets for a given sequence, whether or not a probe can be placed on any one individual target, all targets will be tested and the best design chosen.

**[0052]** Once a probe (or probe pair) sequence is selected, a list of upstream (forward) and downstream (reverse) primers are delineated starting immediately before and after the probe position. These are delineated via  $T_m$  (in some configurations using a different algorithm than used for probe design), and filtered or scored. If SNP probe pairs are being designed, primers are delineated starting immediately before and after the footprint corresponding to both SNP-targeting probe positions. At least one forward and one reverse primer must be identified. By default, up to ten forward and ten reverse primers are collected, but the number of upstream and downstream primers may be changed, such as by using a command line switch. Failure to identify any forward or any reverse primers results in probe placement process 130 to report the problem and continue with the next target coordinate or next sequence as described above.

**[0053]** Forward and reverse primers are checked for pair-wise compatibility and the corresponding amplicons are filtered or scored. The compatibility check includes screening the 3' ends of the primers across the amplicon associated with a given primer pairing. If too great a 3' match is identified, the primers may not be paired. The pair of primers with the best score, by default, the shortest amplicon, is chosen. Failure to select an acceptable primer pair results in probe placement process 130 reporting the problem and continuing as described above.

**[0054]** Acceptable designs comprising one or two probe sequences (such as, for example, probe sequences that can be used to make TAQMAN<sup>®</sup> probes) together with corresponding forward and reverse primer sequences are recorded in the log file. Along with the sequences, the coordinates,  $T_m$  values, and scores are reported for each probe and primer. Any associated auxiliary data (e.g., tracking information) loaded during sequence and target input is also reported to the log file when a successful design is obtained. If no acceptable designs are found for a target sequence, only the target name is recorded in the log file.

**[0055]** It will thus be appreciated that various method configurations of the present invention provide a method for operating a computer system and for using a computer system to distribute products to consumers. These products can include customized and quality-validated assays for single nucleotide polymorphism (SNP) genotyping and gene expression studies. This service can save consumers substantial time, money, and manual labor that home-brew assay design requires, thereby enabling a new level of high-throughput, low-cost SNP genotyping or gene-expression research studies. In some configurations, human or nonhuman target DNA sequences can be provided by the consumer as starting information. Various configurations provide ready-to-use, functionally tested assays that confirm performance, and some configurations provide probes in a single tube format with a two-dimensional bar code for easy sample tracking. The use of a non-fluorescent dye in some configurations eliminates background fluorescence, provides superior quenching efficiency, and increases signal-to-noise ratio, which provides a substantial benefit to consumers using sequence detection



systems such as the Applied Biosystems, Inc. PRISM<sup>®</sup> 7900HT Sequence Detection System, thereby enabling high-throughput SNP genotyping where approximately 250,000 genotypes per day can be analyzed, each requiring only a small quantity of sample DNA.

**[0056]** The description of the invention is merely exemplary in nature and, thus, variations that do not depart from the gist of the invention are intended to be within the scope of the invention. Such variations are not to be regarded as a departure from the spirit and scope of the invention.

## CLAIMS

What is claimed is:

1. A method for providing a product to a consumer, said method comprising;

utilizing a computer network to interact with the consumer to obtain information associated with at least one target nucleic acid sequence;

providing a forward primer sequence, a reverse primer sequence and a probe sequence, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence;

manufacturing at least one assay comprising a forward primer in accordance with the forward primer sequence, a reverse primer in accordance with the reverse primer sequence, and a probe in accordance with the probe sequence;

validating one or more of the forward primer, the reverse primer and the probe; and

delivering the at least one manufactured assay to the consumer.

2. A method in accordance with Claim 1 wherein said validating one or more of the forward primer, the reverse primer and the probe comprises validating one or more of the forward primer sequence, the reverse primer sequence, and the probe sequence against a genome database.

3. A method in accordance with Claim 1 wherein said validating comprises performing mass spectroscopy on the at least one manufactured assay to determine that an oligonucleotide sequence is correct.

4. A method in accordance with Claim 1 wherein said validating comprises performing a functional test to determine that an amplification has occurred.

5. A method in accordance with Claim 4 wherein said performing a functional test further comprises confirming at least one allelic discrimination.

6. A method in accordance with Claim 1 wherein the information associated with the target nucleic acid sequence obtained from the consumer is information associated with a target DNA sequence.

7. A method in accordance with Claim 1 wherein the information associated with the target nucleic acid sequence obtained from the consumer includes information associated with at least one member of the group consisting of an exon or a portion thereof, and a single nucleotide polymorphism (SNP).

8. A method in accordance with Claim 7 wherein the target nucleic acid sequence includes a SNP, and two bi-allelic probes are provided.

9. A method in accordance with Claim 1 wherein the information obtained from the consumer further includes information associated with at least one member of the group consisting of an SNP location and an exon location.

10. A method in accordance with Claim 1 wherein said delivering the at least one manufactured assay to the consumer comprises delivering the at least one manufactured assay in a single tube format.

11. A method in accordance with Claim 1 wherein said delivering the at least one manufactured assay to the consumer comprises delivering the at least one manufactured assay with a two dimensional bar code identification.

12. A method in accordance with Claim 1 wherein said probe further comprises a non-fluorescent dye.

13. A method in accordance with Claim 1 wherein the at least one manufactured assay is suitable for use in a sequence detection system.

14. A method in accordance with Claim 1 wherein said interaction with the consumer further comprises analyzing a target nucleic acid sequence provided by the consumer for format errors.

15. A method in accordance with Claim 1 wherein the target nucleic acid sequence is a human sequence.

16. A method for operating a computing system comprising:

receiving information comprising at least one target nucleic acid sequence via a computer communication network;

attempting to design an assay comprising a forward primer sequence, a reverse primer sequence, and a probe sequence, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon sequence lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence;

validating at least one of the forward primer sequence, the reverse primer sequence, and the probe sequence against a genome database;

recording each design attempt in a log file, including whether each said design attempt succeeded or failed said validation; and

utilizing the log file to generate output sequence data for at least one forward primer sequence, reverse primer sequence, and probe sequence.

17. A method in accordance with Claim 16 further comprising assigning a serial number to each batch of output sequence data.

18. A method for providing a product to a consumer, said method comprising;

utilizing a computer network to interact with the customer to obtain information associated with at least one target nucleic acid sequence;

providing a forward primer sequence, a reverse primer sequence and a probe sequence having specified characteristics, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence;

manufacturing at least one assay comprising a forward primer in accordance with the forward primer sequence, a reverse primer in accordance with the reverse primer sequence, and a probe in accordance with the probe sequence;

validating one or more of the forward primer, the reverse primer, and the probe; and

delivering the at least one manufactured assay to the customer.

19. A method in accordance with Claim 18 wherein said validating one or more of the forward primer, the reverse primer and the probe comprises validating one or more of the forward primer sequence, the reverse primer sequence, and the probe sequence against a genome database.

20. A method in accordance with Claim 18 wherein said validating comprises performing mass spectroscopy on the at least one manufactured assay to determine that an oligonucleotide sequence is correct.

21. A method in accordance with Claim 18 wherein said validating comprises performing a functional test to determine that an amplification has occurred.

22. A method in accordance with Claim 21 wherein said performing a functional test further comprises confirming at least one allelic discrimination.

23. A method in accordance with Claim 18 wherein the target nucleic acid sequence obtained from the consumer is a target DNA sequence.

24. A method in accordance with Claim 18 wherein the target nucleic acid sequence obtained from the consumer includes at least one member of the group consisting of an exon or a portion thereof, and a single nucleotide polymorphism (SNP).

25. A method in accordance with Claim 24 wherein the target nucleic acid sequence includes a SNP and two bi-allelic probes are provided.

26. A method in accordance with Claim 18 wherein the information obtained from the customer further includes at least one member of the group consisting of an SNP location and an exon location.

27. A method in accordance with Claim 18 further comprising testing the at least one manufactured assay to verify that the at least one manufactured assay meets the specified characteristics.

28. A method in accordance with Claim 27 wherein said testing the at least one manufactured assay includes at least one test selected from the

group consisting of performing mass spectroscopy on the at least one manufactured assay to determine that an oligonucleotide sequence is correct and performing a functional test to determine that an amplification has occurred and at least one allelic discrimination is confirmed.

29. A method in accordance with Claim 18 wherein said delivering the at least one manufactured assay to the consumer comprises delivering the at least one manufactured assay in a single tube format.

30. A method in accordance with Claim 18 wherein said delivering the at least one manufactured assay to the consumer comprises delivering the at least one manufactured assay with a two dimensional bar code.

31. A method in accordance with Claim 18 wherein said probe further comprises a non-fluorescent dye.

32. A method in accordance with Claim 18 wherein the at least one manufactured assay is suitable for use in a sequence detection system.

33. A method in accordance with Claim 18 wherein said interaction with the customer further comprises analyzing the target nucleic acid sequence provided by the customer for format errors.

34. A method in accordance with Claim 18 wherein the specified design constraints are selected from the group consisting of size limits, target



T<sub>m</sub>, minimum T<sub>m</sub>, total matching bases in a hairpin stem, contiguous matching groups in a hairpin stem, combined G and C content, runs of G bases, runs of non-G bases, and G content not greater than C content.

35. A method in accordance with Claim 18 wherein the target nucleic acid sequence is a human sequence.

36. A method for operating a computing system comprising:

- receiving information comprising at least one target nucleic acid sequence via a computer communication network;
- providing assay design metrics and scoring metrics;
- attempting to design an assay comprising a forward primer sequence, a reverse primer sequence, and a probe sequence, wherein the forward primer sequence and the reverse primer sequence together define an amplicon sequence, the amplicon sequence lies within the target nucleic acid sequence, and the probe sequence is complementary to a portion of the amplicon sequence;
- validating at least one of the forward primer sequence, the reverse primer sequence, and the probe sequence against a genome database;
- recording each design attempt in a log file, including whether each said design attempt succeeded or failed to meet constraints set by the design metrics and scoring metrics; and
- utilizing the log file to generate a batch of output sequence data for at least one forward primer sequence, reverse primer sequence, and probe sequence.

37. A method in accordance with Claim 36 further comprising assigning a serial number to each batch of output sequence data.

38. A method in accordance with Claim 36 wherein attempting to design an assay comprises attempting to design the probe sequence in accordance with at least one constraint selected from the constraint group consisting of size limits, target  $T_m$ , minimum  $T_m$ , total matching bases in a hairpin stem, contiguous matching groups in a hairpin stem, combined G and C content, runs of G bases, runs of non-G bases, and G content not greater than C content.

39. A method in accordance with Claim 38 wherein attempting to design the probe sequence in accordance with at least one constraint further comprises applying the at least one constraint as a filter, according to which the forward primer sequence is accepted or rejected in accordance with whether the constraint is satisfied.

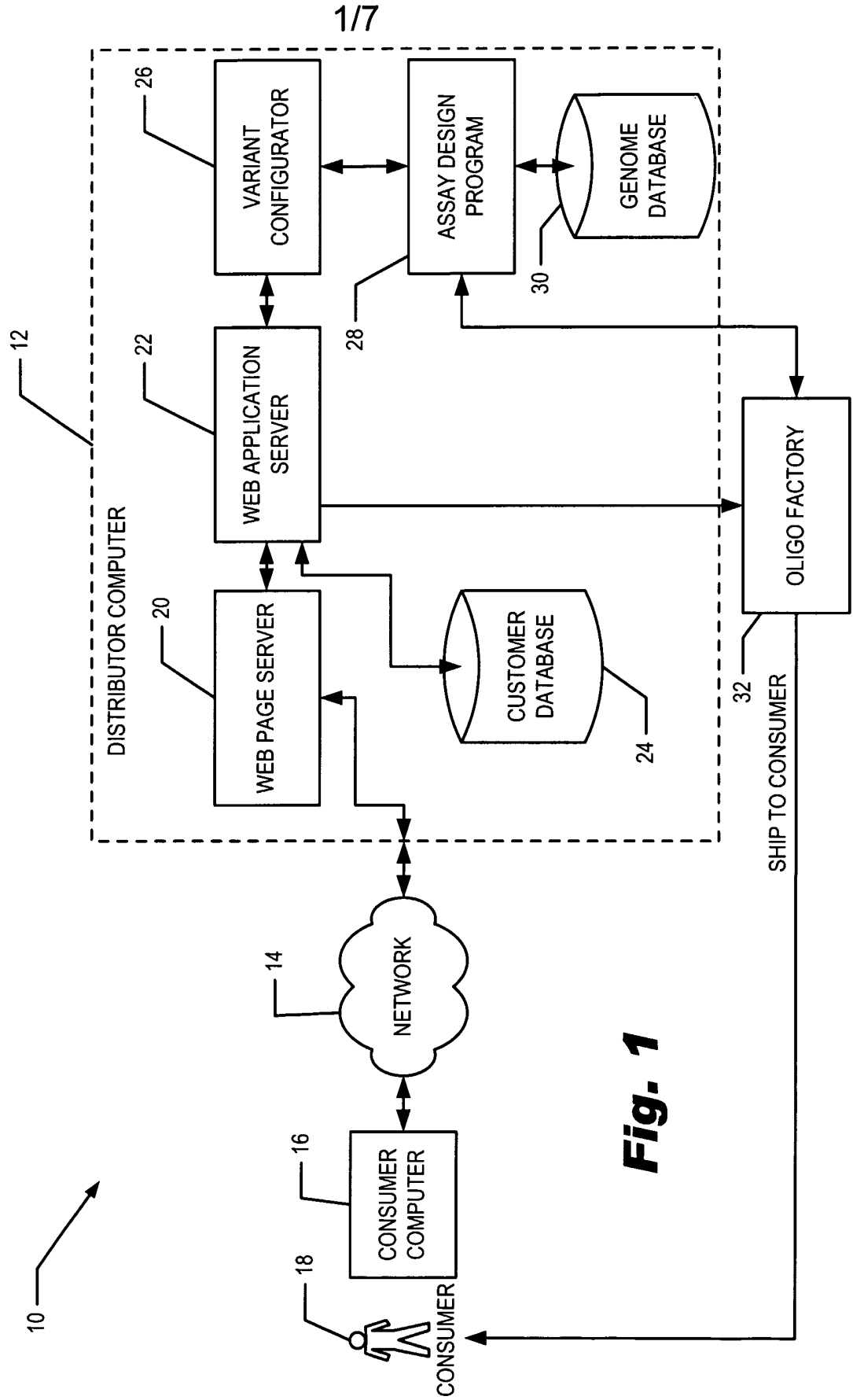
40. A method in accordance with Claim 38 wherein attempting to design the probe sequence in accordance with one constraint further comprises applying the at least one constraint as a score, wherein the forward probe sequence is favored or disfavored according to a comparison with a score of a different probe sequence.

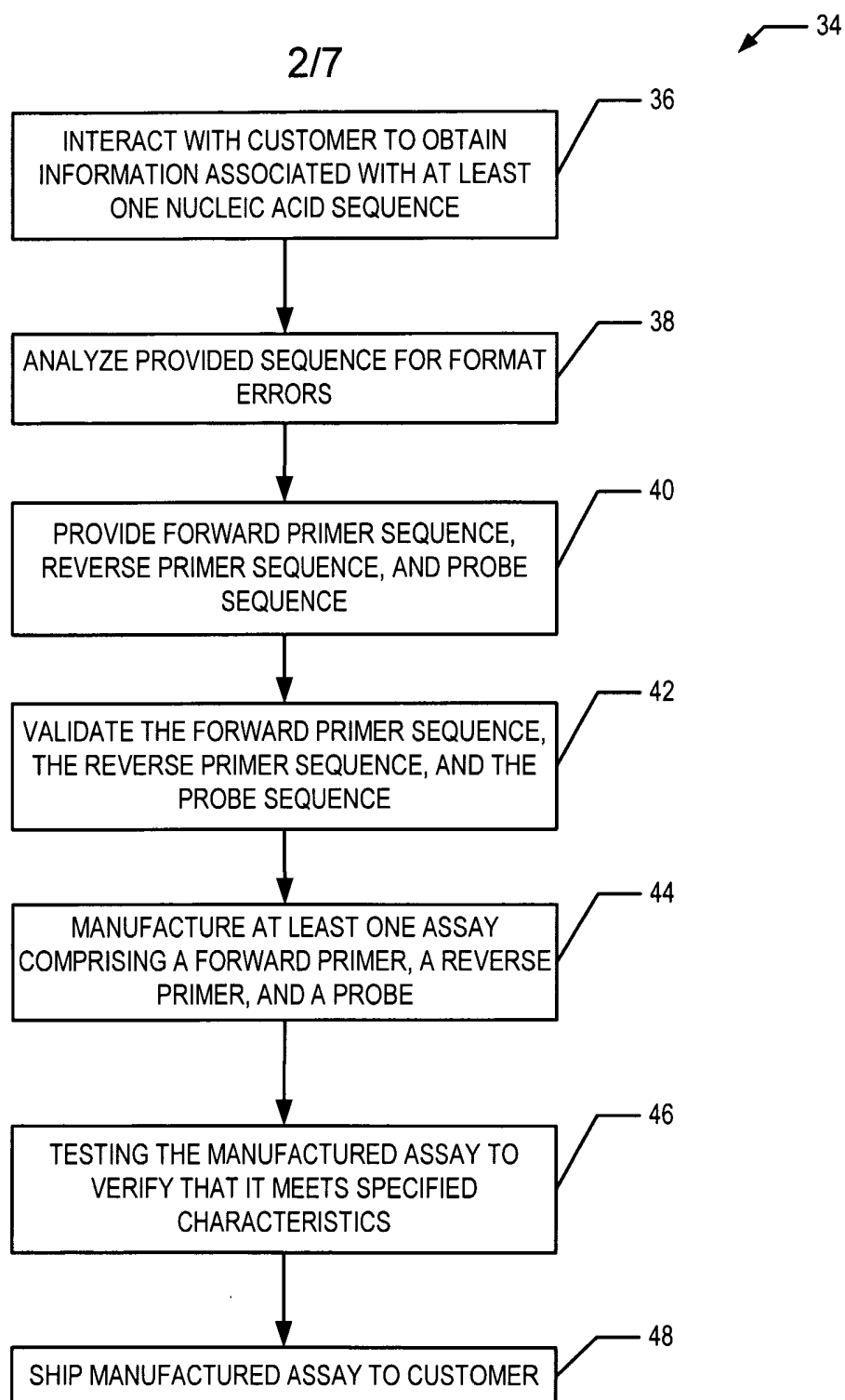
41. A method in accordance with Claim 36 wherein attempting to design an assay comprises attempting to design at least one said primer sequence in accordance with at least one constraint including a limit on G+C at a 3' end (5 bases) of the primer sequence.

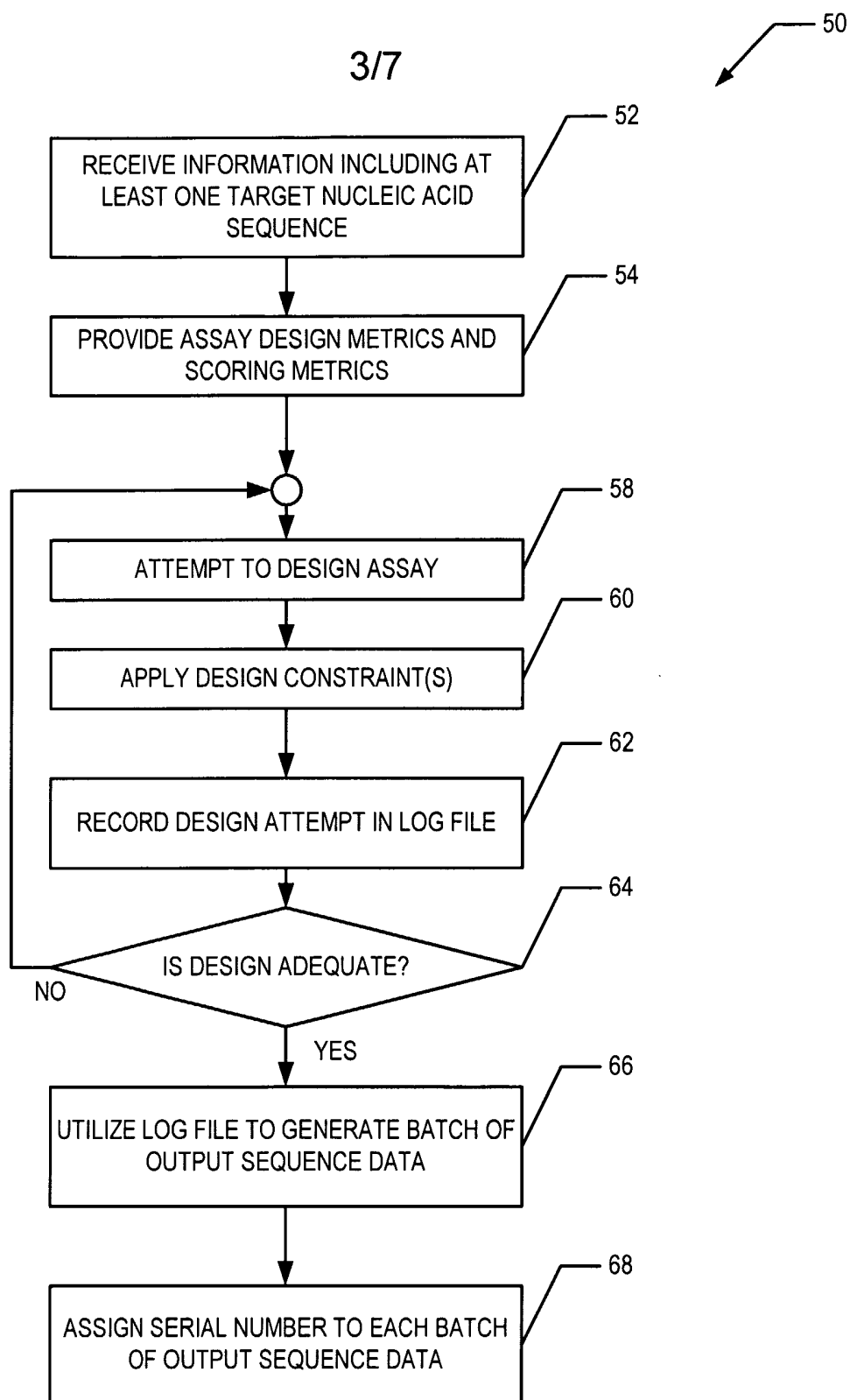
42. A method in accordance with Claim 36 wherein attempting to design an assay comprises attempting to design at least one said primer sequence in accordance with at least one constraint selected from the group consisting of size limits, target  $T_m$ , minimum  $T_m$ , total matching bases in a hairpin stem, contiguous matching groups in a hairpin stem, combined G and C content, runs of G bases, runs of non-G bases, and G content not greater than C content.

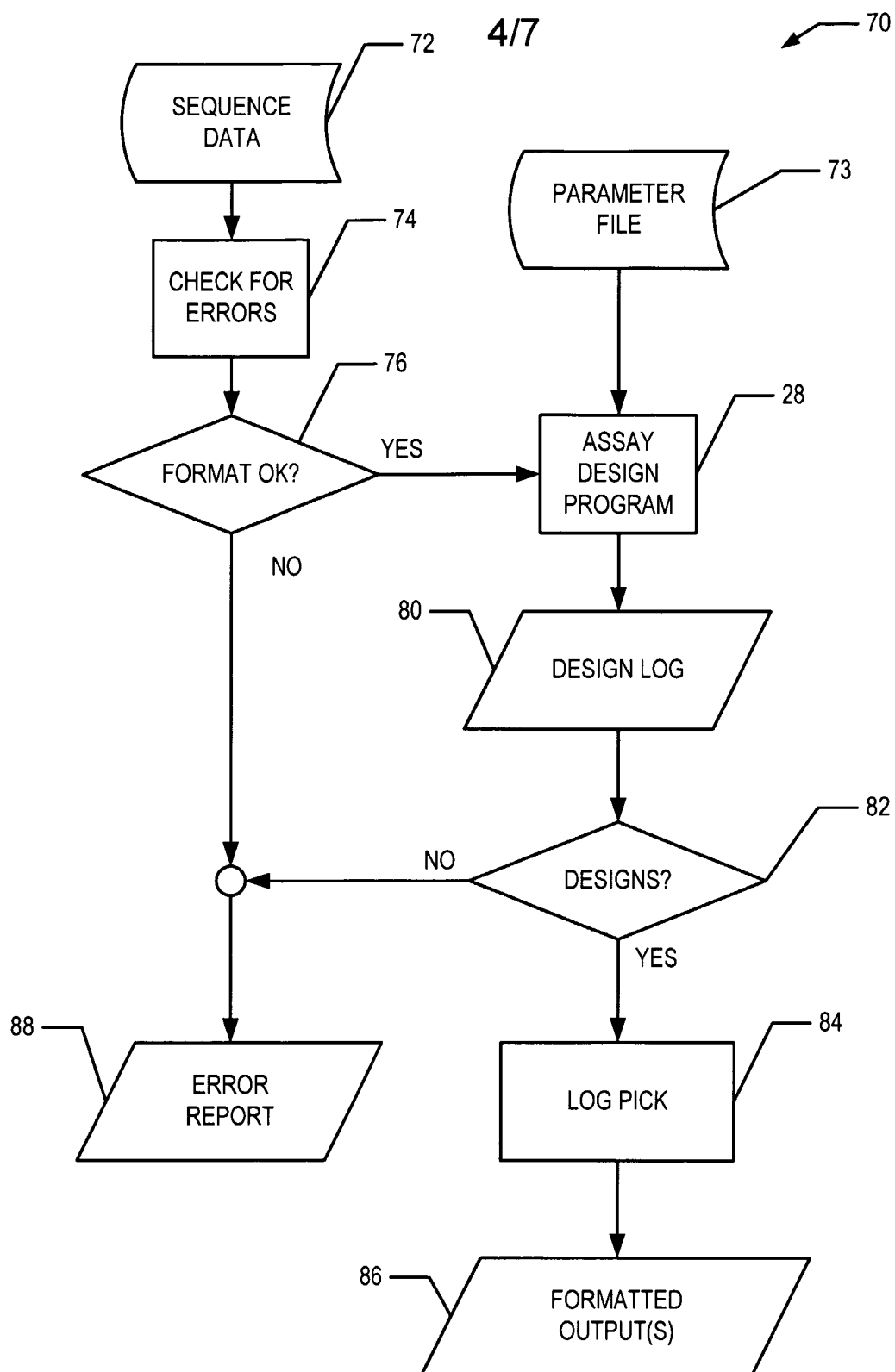
43. A method in accordance with Claim 42 wherein attempting to design said at least one primer sequence in accordance with at least one constraint further comprises applying the at least one constraint as a filter, according to which the primer sequence is accepted or rejected in accordance with whether the constraint is satisfied.

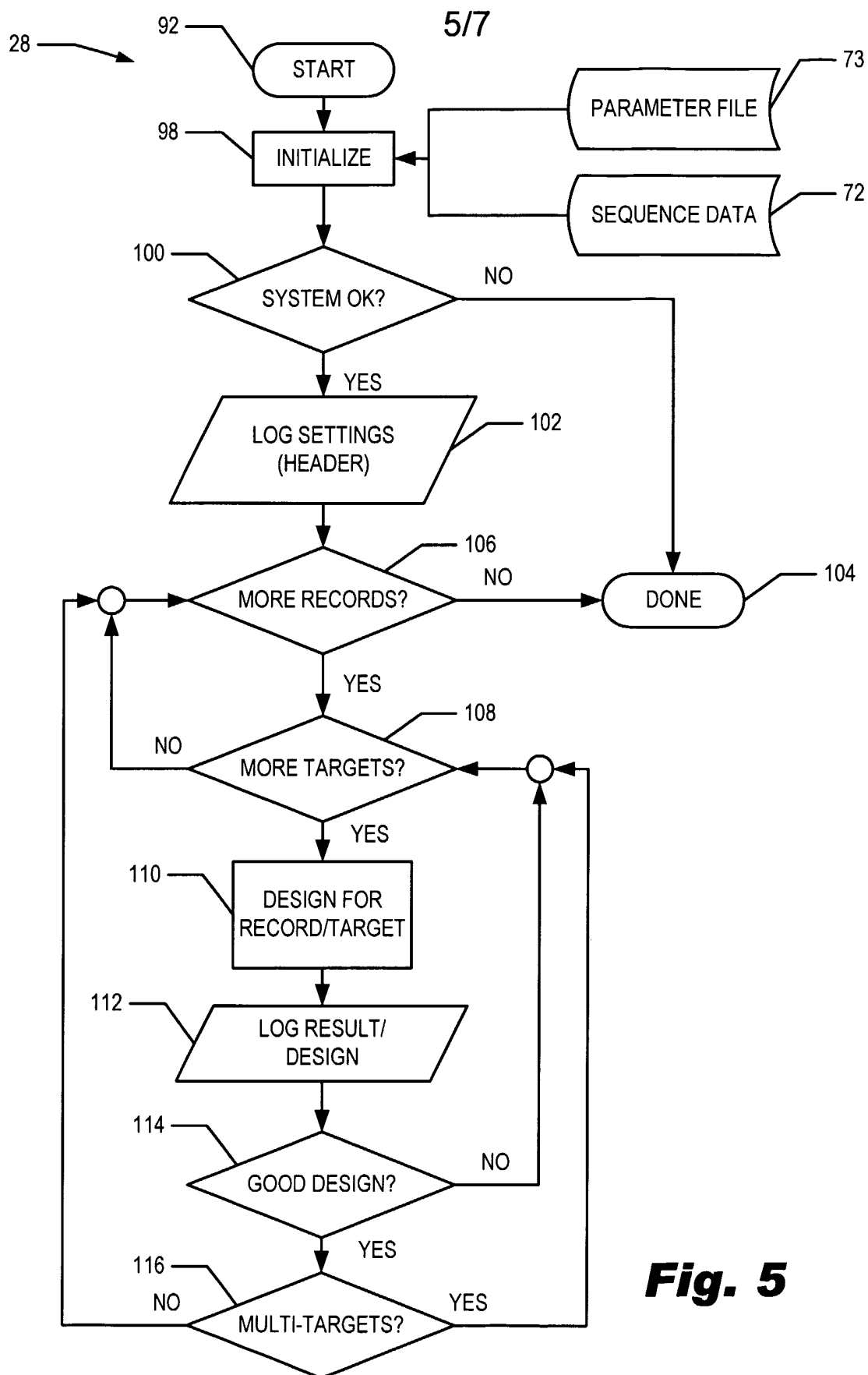
44. A method in accordance with Claim 42 wherein attempting to design said at least one primer sequence in accordance with at least one constraint further comprises applying the at least one constraint as a score, wherein the primer sequence is favored or disfavored according to a comparison with a score of a different primer sequence.



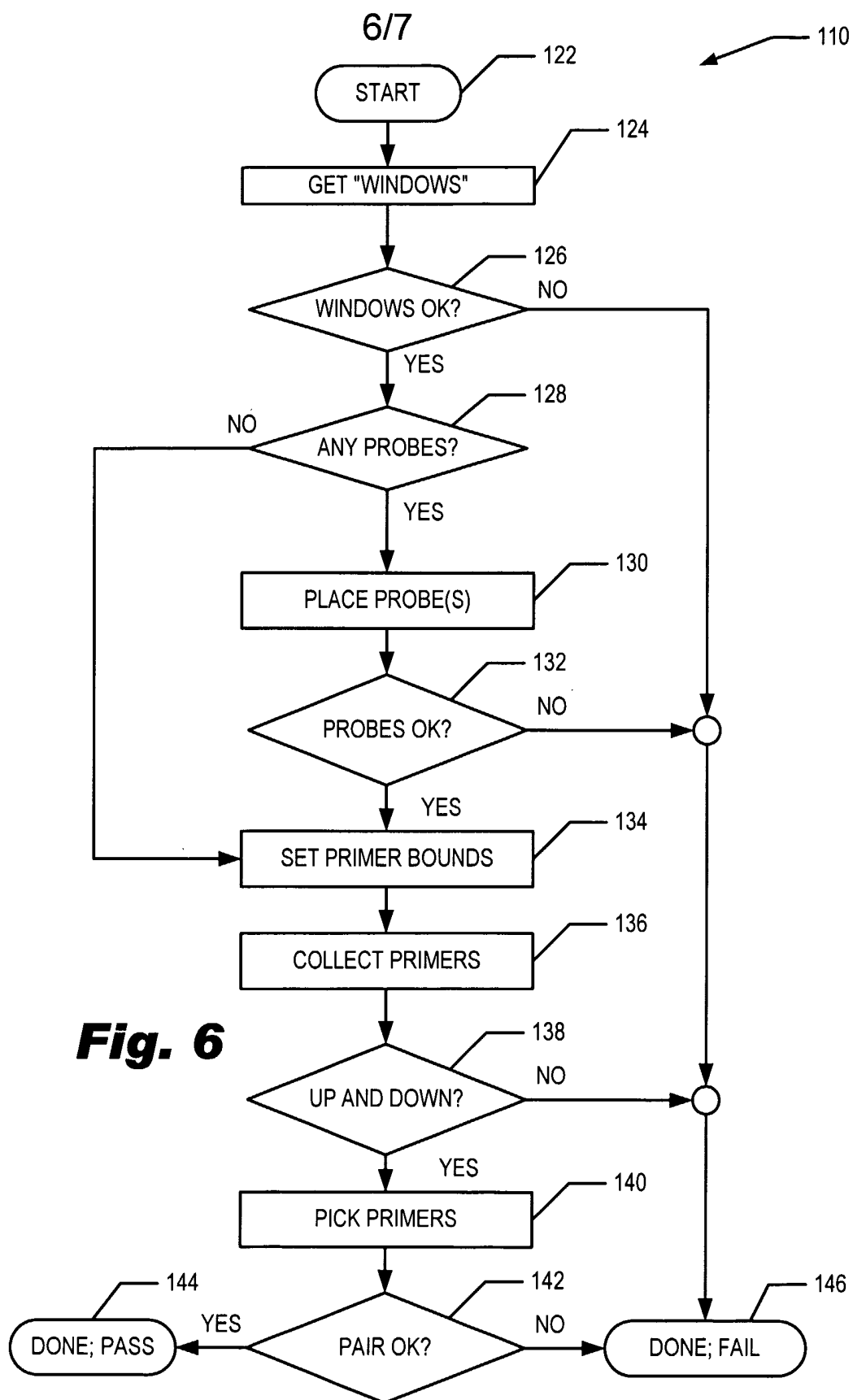
**Fig. 2**

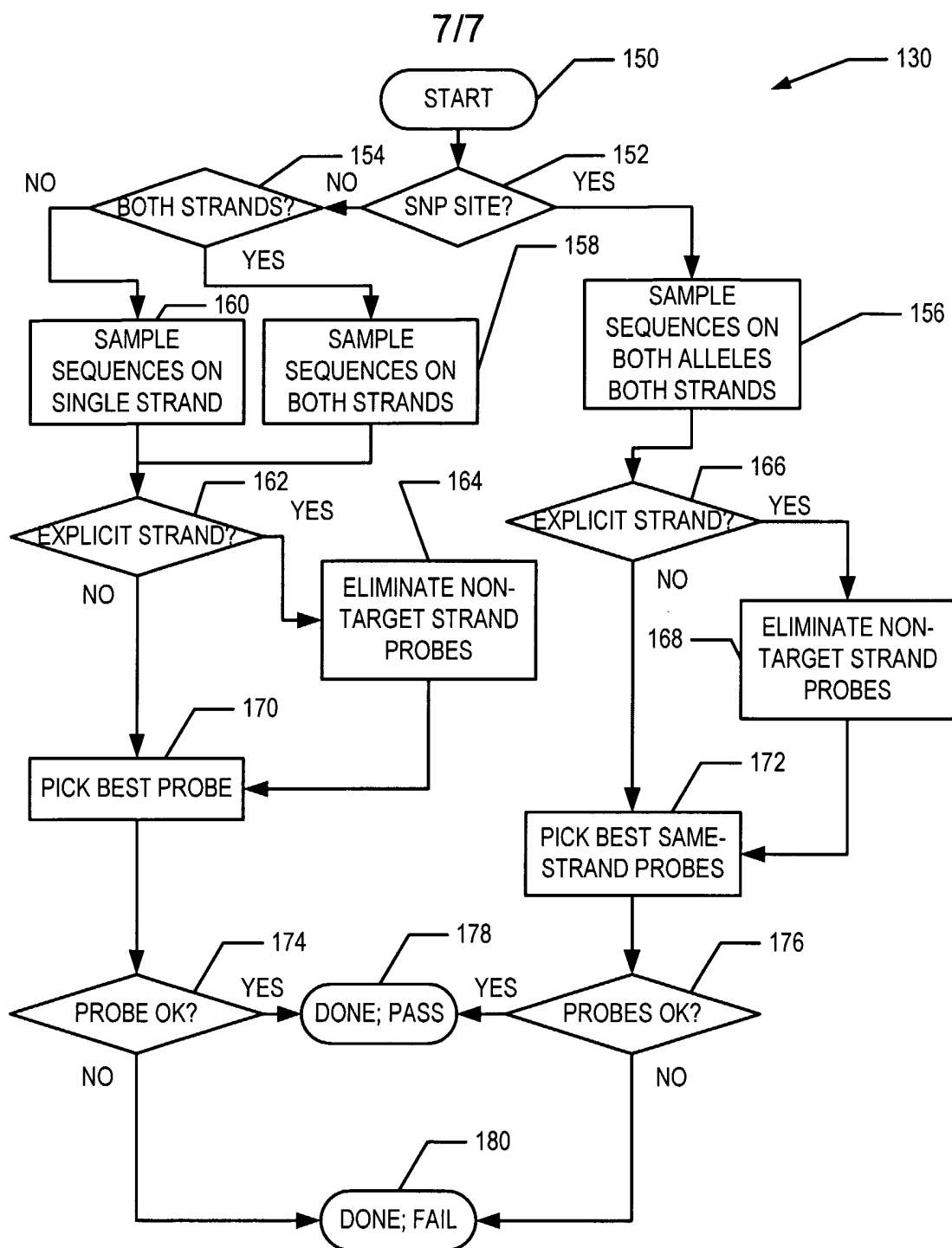
**Fig. 3**

**Fig. 4**

**Fig. 5**





**Fig. 7**

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/34599

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68

US CL : 435/6, 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, 91.2

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)  
Medline, Biosis, Caplus, Internet, primer, manufacture, computer

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ESTRUCH et al. Oligo. In: Methods in Molecular Biology, Vol. 70: Sequence Data Analysis Guidebook. Edited by S. R. Swindell. New Jersey: Humana Press Inc., 1997, pages 279-285, see especially Fig. 3, and p. 285	1-44
X,P ---	Applied Biosystems. Oligo Factory Catalog. August 2002. Especially pages 6-8.	1-15, 18-35 -----
Y,P		16-17, 36-44
A	CFAR. Proposed Strategy for Using the TaqMan. Undated. <a href="http://genomics.ucsd.edu/taqman_strat">http://genomics.ucsd.edu/taqman_strat</a> .	1-44
Y	Beacon Designer 2.0 Manual. PREMIER Biosoft International. 2001.	1-44
Y	ROZEN et al. Primer3 on teh WWW for general Users and for Biologist Programmers. In: Methods in Molecular Biology, Vol. 132: Bioinformatics Methods and Protocols. Edited by S. Misner et al. New Jersey: Humana Press Inc., 2000, pages 365-386	1-44

☒ 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

24 March 2003 (24.03.2003)

Date of mailing of the international search report

29 APR 2003

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

Julia C. Einsmann

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	Press Release. Applied Biosystems Launches "Assays-By-Design" Genomic Assays. 25 October 2001. www.appliedbiosystems.com.	1-15, 18-35 ----- 16-17, 36-44
Y	LOWE et al. A computer program for selection of oligonucleotide primers for polymerase chain reactions. Nucleic Acids Research. 1990, Vol. 18, No. 7, pages 1757-1761.	1-44
Y	US 6,300, 058 B1 (AKITAYA et al.) 09 October 2001 (09.10.2001), col. 61-76	1-44