

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 September 2006 (28.09.2006)

PCT

(10) International Publication Number
WO 2006/102401 A2

(51) International Patent Classification:

C12Q 1/68 (2006.01) **G01N 33/50** (2006.01)
G06F 19/00 (2006.01) **C12P 19/34** (2006.01)
G01N 33/48 (2006.01)

(21) International Application Number:

PCT/US2006/010383

(22) International Filing Date: 22 March 2006 (22.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

11/086,253 22 March 2005 (22.03.2005) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SHANNON, Mark** [US/US]; 3131 Arcadia Common #1, Livermore, CA 94550 (US). **OLDHAM, Mark F.** [US/US]; 16500 Soda Springs Road, Los Gatos, CA 95033 (US). **RUFF, David** [US/US]; 2000 Broadway Street, #317, San Francisco, CA 94115 (US).

(74) Agents: **SNYDER, Jeffrey L.** et al.; **HARNESS, DICKY & PIERCE PLC**, P. O. Box 828, Bloomfield Hills, Michigan 48303 (US).

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

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

Published:

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

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

(54) Title: NORMALIZATION OF GENE EXPRESSION DATA

(57) Abstract: A method for determining bias across two domains comprising gene expression data. The method can comprise (a) providing a first domain and a second domain; (b) obtaining information indicative of a bias within the first domain; (c) obtaining information indicative of a bias within the second domain; and (d) using the information indicative of the bias within the first domain and the information indicative of the bias within the second domain to produce an indication of bias across the two domains.



WO 2006/102401 A2

NORMALIZATION OF GENE EXPRESSION DATA

INTRODUCTION

[0001] Currently, genomic analysis, including that of the estimated 30,000 human genes is a major focus of basic and applied biochemical and pharmaceutical research. Such analysis may aid in developing diagnostics, medicines, and therapies for a wide variety of disorders. However, the complexity of the human genome and the interrelated functions of genes often make this task difficult. There is a continuing need for methods and apparatus to aid in such analysis.

DRAWINGS

[0002] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0003] The skilled artisan will understand that the drawings, described herein, are for illustration purposes only. The drawings are not intended to limit the scope of the present teachings in any way.

[0004] FIG. 1 is a flowchart illustrating the use of a database system according to some embodiments;

[0005] FIG. 2 is a flowchart illustrating a process for determining bias.

[0006] FIG. 3 is a graph exemplifying a comparison of amplification with IVT and multiplex preamplification.

[0007] FIG. 4 is a graph exemplifying a $\Delta\Delta C_T$ comparison of a brain and a liver sample.

[0008] FIG. 5 is a graph exemplifying a $\Delta\Delta C_T$ comparison between four different sample inputs.

[0009] FIG. 6 is a graph exemplifying a $\Delta\Delta\Delta C_T$ of a liver and a brain sample with IVT preamplification.

[0010] FIG. 7 is a graph exemplifying a $\Delta\Delta\Delta C_T$ of a liver and a brain sample with multiplex preamplification.

[0011] FIG. 8 is a flowchart illustrating a process for determining bias between two gene expression platforms.

DESCRIPTION OF VARIOUS EMBODIMENTS

[0012] The following description of various embodiments is merely exemplary in nature and is in no way intended to limit the present teachings, applications, or uses. Although the present teachings will be discussed in various embodiments as relating to polynucleotide amplification, such as PCR, such discussion should not be regarded as limiting the present teaching to only such applications.

[0013] All literature and similar materials cited in this application, including but not limited to, patents, patent applications, articles, books, treatises, and internet web pages, regardless of the format of such literature and similar materials, are expressly incorporated by reference in their entirety for any purpose. In the event that one or more of the incorporated literature and similar materials differs from or contradicts this application, including but not limited to defined terms, term usage, described techniques, or the like, this application controls.

[0014] In general, gene expression is a process by which a gene's coded information is converted into the structures present and operating in the cell. Gene expression is a multi-step process that begins with transcription and translation and is followed by folding, post-translational modification and targeting. The amount of protein that a cell expresses depends on the tissue, the developmental stage of the organism and the metabolic or physiologic state of the cell. Expressed genes can include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein. In various embodiments, gene expression can be studied using analytical techniques such as polymerase chain reaction (PCR), Northern blots, serial analysis of gene expression (SAGE) microarrays, hybridization arrays, and high density oligonucleotide arrays.

[0015] Briefly, by way of background, PCR can be used to amplify a sample of target Deoxyribose Nucleic Acid (DNA) for analysis. Typically, the PCR reaction involves copying the strands of the target DNA and then using the copies to generate additional copies in subsequent cycles. Each cycle doubles the amount of the target DNA present, thereby resulting in a geometric progression in the number of copies of the target DNA. The temperature of a double-stranded target DNA is elevated to denature the DNA, and the temperature is then reduced to anneal at least one primer to each strand of the denatured target DNA. In some embodiments, the target DNA can be a cDNA. In some embodiments, primers are used as a pair--a forward primer and a reverse primer--and can be referred to as a primer pair or primer set. In some embodiments, the primer set comprises a 5' upstream

primer that can bind with the 5' end of one strand of the denatured target DNA and a 3' downstream primer that can bind with the 3' end of the other strand of the denatured target DNA. Once a given primer binds to the strand of the denatured target DNA, the primer can be extended by the action of a polymerase. In some embodiments, the polymerase can be a thermostable DNA polymerase, for example, a Taq polymerase. The product of this extension, which sometimes may be referred to as an amplicon, can then be denatured from the resultant strands and the process can be repeated. Temperatures suitable for carrying out the reactions are well known in the art. Certain basic principles of PCR are set forth in U.S. Patent Nos. 4,683,195, 4,683,202, 4,800,159, and 4,965,188, each issued to Mullis et al.

[0016] In some embodiments, PCR can be conducted under conditions allowing for quantitative and/or qualitative analysis of one or more target DNA. Accordingly, detection probes can be used for detecting the presence of the target DNA in an assay. In some embodiments, the detection probes can comprise physical (e.g., fluorescent) or chemical properties that change upon binding of the detection probe to the target DNA. Some embodiments of the present teaching can provide real time fluorescence-based detection and analysis of amplicons as described, for example, in PCT Publication No. WO 95/30139 and U.S. Patent Application No. 08/235,411.

[0017] In some embodiments, assay can be a homogenous polynucleotide amplification assay, for coupled amplification and detection, wherein the process of amplification generates a detectable signal and the need for subsequent sample handling and manipulation to detect the amplified product is minimized or eliminated. Homogeneous assays can provide for amplification that is detectable without opening a sealed well or further processing steps once amplification is initiated. Such homogeneous assays can be suitable for use in conjunction with detection probes. For example, in some embodiments, the use of an oligonucleotide detection probe, specific for detecting a particular target DNA can be included in an amplification reaction in addition to a DNA binding agent of the present teachings. Homogenous assays among those useful herein are described, for example, in commonly assigned U.S. Patent No. 6,814,934.

[0018] In some embodiments, methods are provided for detecting a plurality of targets. Such methods include those comprising forming an initial mixture comprising an analyte sample suspected of comprising the plurality of targets, a polymerase, and a plurality of primer sets. In some embodiments, each primer set comprises a forward primer and a reverse primer and at least one detection probe unique for one of the plurality of primer sets.

In some embodiments, the initial mixture can be formed under conditions in which one primer elongates if hybridized to a target.

[0019] In some embodiments for amplification of a polynucleotide, assay can comprise a preamplification product, wherein one or more polynucleotides in an analyte has been amplified prior to being deposited in at least one of the plurality of wells. In some embodiments, these methods can further comprise forming a plurality of preamplification products by subjecting an initial analyte comprising a plurality of polynucleotides to at least one cycle of PCR to form a detection mixture comprising a plurality of preamplification products. The detection mixture of preamplification products can be then used for further amplification using a PCR. In some embodiments, preamplification comprises the use of isothermal methods.

[0020] In some embodiments, a two-step multiplex amplification reaction can be performed wherein the first step truncates a standard multiplex amplification round to boost a copy number of the DNA target by about 100-1000 or more fold. Following the first step, the resulting product can be divided into optimized secondary single amplification reactions, each containing one or more of the primer sets that were used previously in the first or multiplexed booster step. The booster step can occur, for example, using an aqueous target or using a solid phase archived nucleic acid. See, for example, U.S. Patent 6,605,452, Marmaro.

[0021] In some embodiments, preamplification methods can employ in vitro transcription (IVT) comprising amplifying at least one sequence in a collection of nucleic acids sequences. The processes can comprise synthesizing a nucleic acid by hybridizing a primer complex to the sequence and extending the primer to form a first strand complementary to the sequence and a second strand complementary to the first strand. The primer complex can comprise a primer complementary to the sequence and a promoter region in anti-sense orientation with respect to the sequence. Copies of anti-sense RNA can be transcribed off the second strand. The promoter region, which can be single or double stranded, can be capable of inducing transcription from an operably linked DNA sequence in the presence of ribonucleotides and a RNA polymerase under suitable conditions. Suitable promoter regions may be prokaryote viruses, such as from T3 or T7 bacteriophage. In some embodiments, the primer can be a single stranded nucleotide of sufficient length to act as a template for synthesis of extension products under suitable conditions and can be poly (T) or a collection of degenerate sequences. In some embodiments, the methods involve the incorporation of an RNA polymerase promoter into selected cDNA molecule by priming

cDNA synthesis with a primer complex comprising a synthetic oligonucleotide containing the promoter. Following synthesis of double-stranded cDNA, a polymerase generally specific for the promoter can be added, and anti-sense RNA can be transcribed from the cDNA template. The progressive synthesis of multiple RNA molecules from a single cDNA
5 template results in amplified, anti-sense RNA (aRNA) that serves as starting material for cloning procedures by using random primers. The amplification, which will typically be at least about 20-40, typically to 50 to 100 or 250-fold, but can be 500 to 1000-fold or more, can be achieved from nanogram quantities or less of cDNA.

[0022] In some embodiments, a two stage preamplification method can be used
10 to preamplify assay in one vessel by IVT and, for example, this preamplification stage can be 100 X sample. In the second stage, the preamplified product can be divided into aliquots and preamplified by PCR and, for example, this preamplification stage can be 16,000 X sample or more.

[0023] In some embodiments, the preamplification can be a multiplex
15 preamplification, wherein the analyte sample can be divided into a plurality of aliquots. Each aliquot can then be subjected to preamplification using a plurality of primer sets for DNA targets. In some embodiments, the primer sets in at least some of the plurality of aliquots differ from the primer sets in the remaining aliquots. Each resulting preamplification product
20 detection mixture can then be dispersed into at least some of a plurality of wells of a microplate comprising an assay having corresponding primer sets and detection probes for further amplification and detection according to the methods described herein. In some
embodiments, the primer sets of assay in each of the plurality of wells can correspond to the primer sets used in making the preamplification product detection mixture. The resulting
25 assay in each of the plurality of wells thus can comprise a preamplification product and primer sets and detection probes for amplification for DNA targets, which, if present in the analyte sample, have been preamplified.

[0024] Since a plurality of different sequences can be amplified simultaneously in a single reaction, the multiplex preamplification can be used in a variety of contexts to
effectively increase the concentration or quantity of a sample available for downstream
30 analysis and/or assays. In some embodiments, because of the increased concentration or quantity of target DNA, significantly more analyses can be performed with multiplex amplified samples than can be performed with the original sample. In many embodiments, multiplex amplification further permits the ability to perform analyses that require more sample or a higher concentration of sample than was originally available. In such embodiments,

5 multiplex amplification enables downstream analysis for assays that could not have been possible with the original sample due to its limited quantity. In some embodiments, the plurality of aliquots can comprise 16 aliquots with each of the 16 aliquots comprising about 1536 primer sets. In such embodiments, a sample comprising a whole genome for a species, for example a human genome, can be preamplified. In some embodiments, the plurality of aliquots can be greater than 16 aliquots. In some embodiments, the number of primer sets can be greater than 1536 primer sets. In some embodiments, the plurality of aliquots can be less than 16 aliquots and the number of primer sets can be greater than 1536 primer sets. For examples of such embodiments, see PCT Publication No. WO 10 2004/051218 to Andersen and Ruff.

[0025] In some embodiments, multiplex methods are provided wherein assay comprises a first universal primer that binds to a complement of a first target, a second universal primer that binds to a complement of a second target, a first detection probe comprising a sequence that binds to the sequence comprised by the first target, and a 15 second detection probe comprising a sequence that binds to a sequence comprised by the second target. In some embodiments, at least some of a plurality of wells of a microplate comprise a solution operable to perform multiplex PCR. The first and second detection probes can comprise different labels, for example, different fluorophores such as, in non-limiting example, VIC and FAM. Sequences of the first and second detection probes can 20 differ by as little as one nucleotide, two nucleotides, three nucleotides, four nucleotides, or greater, provided that hybridization occurs under conditions that allow each detection probe to hybridize specifically to its corresponding detection probe.

[0026] In some embodiments, multiplex PCR can be used for relative quantification, where one primer set and detection probe amplifies the target DNA and 25 another primer set and detection probe amplifies an endogenous reference. In some embodiments, the present teaching provide for analysis of at least four DNA targets in each of a plurality of wells and/or analysis of a plurality of DNA targets and a reference in each of a plurality of wells.

[0027] In some embodiments, as seen in FIG. 1, a plurality of microplates having 30 assay filled thereon can be analyzed as described herein with sequence detection system, such as a PCR system to generate data. In some embodiments, this data can be stored in a gene expression analysis system database 736. Software can then be used to generate gene expression analysis information 738.

[0028] In some embodiments, a gene expression analysis system can utilize computer software that organizes analysis sessions into studies and stores them in database 738. An analysis session can comprise the results of running microplate in sequence detection system. To analyze session data, one can load an existing study that contains analysis session data or create a new study and attach analysis session data to it. Studies can be opened and reexamined an unlimited number of times to reanalyze the analysis session data or to add other analysis sessions to the analysis.

[0029] In some embodiments, gene expression analysis system database 736 stores the analyzed data for each microplate run on sequence detection system as an analysis session in database 736. The software can identify each analysis session by marking indicia of the associated microplate and the date on which it was created. Once analysis sessions have been assigned to a study, various functions can be performed. These functions comprise, but are not limited to, designating replicates, removing outliers, filtering data out of a particular view or report, correction of preamplification values via stored values, and computation of gene expression values.

[0030] In various embodiments, real time PCR is adapted to perform quantitative real time PCR (qRT-PCR). In various embodiments, two different methods of analyzing data from qRT-PCR experiments can be used: absolute quantification and relative quantification. In some embodiments, absolute quantification can determine an input copy number of the target DNA of interest. This can be accomplished by relating a signal from a detection probe to a standard curve. In various embodiments, relative quantification can describe the change in expression of the target DNA relative to a reference or a group of references such as, for an example, an untreated control, an endogenous control, a passive internal reference, an universal reference RNA, or a sample at time zero in a time course study. When determining absolute quantification, the expression of the target DNA can be compared across many samples, for example, from different individuals, from different tissues, from multiple replicates, and/or serial dilution of standards in one or more matrices. In various embodiments of the present teachings, qRT-PCR can be performed using relative quantification and the use of standard curve is not required. Relative quantification can compare the changes in steady state target DNA levels of two or more genes to each other with one of the genes acting as an endogenous reference which may be used to normalize a signal from a sample gene. In various embodiments, in order to compare between experiments, resulting fold differences from the normalization of sample to the reference can be expressed relative to a calibrator sample. In some embodiments, the calibrator sample is

included in each assay 1000. The gene expression analysis system can determine the amount of target DNA, normalized to a reference, by determining

$$\Delta C_T = C_{Tq} - C_{Tendo}$$

where C_T is the threshold cycle for detection of a fluorophore in real time PCR; C_{Tq} is the threshold cycle for detection of a fluorophore for a target DNA in assay 1000; and C_{Tendo} is the threshold cycle for detection of a fluorophore for an endogenous reference or a passive internal reference in assay.

[0031] In some embodiments, a gene expression analysis system can determine the amount of target DNA, normalized to a reference and relative to a calibrator, by determining:

$$\Delta\Delta C_T = \Delta C_{Tq} - \Delta C_{Tcb}$$

[0032] where C_{Tq} is the threshold cycle for detection of a fluorophore for the target DNA in assay 1000; C_{Tcb} is the threshold cycle for detection of a fluorophore for a calibrator sample; ΔC_{Tq} is a difference in threshold cycles for the target DNA and an endogenous reference; and ΔC_{Tcb} is a difference in threshold cycles for the calibrator sample and the endogenous reference. If $\Delta\Delta C_T$ is determined, the relative quantity of the target DNA can be determined using a relationship of relative quantity of the target DNA can be equal to $2^{-\Delta\Delta C_T}$. In various embodiments, $\Delta\Delta C_T$ can be about zero. In some embodiments, $\Delta\Delta C_T$ can be less than ± 1 . In various embodiments, the above calculations can be adapted for use in multiplex PCR (See, for example, Livak et al. Applied Biosystems User Bulletin #2, updated October 2001 and Livak and Schmittgen, Methods (25) 402-408 (2001)).

[0033] In some embodiments, assay can be preamplified, as discussed herein, in order to increase the amount of target DNA prior to distribution into a plurality of wells of a microplate. In some embodiments, assay can be collected, for example, via a needle biopsy that typically yields a small amount of sample. Distributing this sample across a large number of wells can result in variances in sample distribution that can affect the veracity of subsequent gene expression computations. In such situations, assay can be preamplified using, for example, a pooled primer set to increase the number of copies of all target DNA simultaneously.

[0034] In various embodiments, preamplification processes can be non-biased, such that all target DNA are amplified similarly and to about the same power. In such embodiments, each target DNA can be amplified reproducibly from one input sample to the next input sample. For example, if target DNA X is initially present in sample A at 100 target molecules, then after 10 cycles of PCR amplification (1000-fold), 100,000 target molecules

should be present. Continuing with the example, if target DNA X is initially present in sample B at 500 target molecules, then after 10 cycles of PCR amplification (1000-fold), 500,000 target molecules should be present. In this example, the ratio of target DNA X in samples A/B remains constant before and after the amplification procedure.

5 **[0035]** In various embodiments, a minor proportion of all target DNA can have an observed preamplification efficiency of less than 100%. In such embodiments, if the amplification bias is reproducible and consistent from one input sample to another, then the ability to accurately compute comparative relative quantitation between any two samples containing different relative amounts of target can be maintained. Continuing the example
10 from above and assuming 50% reproducible amplification efficiency, if target DNA X is initially present in sample A at 100 target molecules, then after 10 cycles of PCR amplification (50% of 1000-fold), 50,000 target molecules should be present. Further continuing the example, if target X is initially present in sample B at 500 target molecules, then after 10 cycles of PCR amplification (50% of 1000-fold), 250,000 target molecules
15 should be present. In this example, the ratio of template X in samples A/B remains constant before and after the amplification procedure and is the same ratio as the 100% efficiency scenario.

[0036] In various embodiments, an unbiased amplification of each target DNA (x, y, z, etc.) can be determined by calculating the difference in CT value of the target DNA
20 (x,y,z, etc.) from the C_T value of a selected endogenous reference, and such calculation is referred to as the ΔC_T value for each given target DNA, as described above. In various embodiments, a reference for a bias calculation can be non-preamplified, amplified target DNA and an experimental sample can be a preamplified amplified target DNA. In some
25 embodiments, the standard sample and experimental sample can originate from the same sample, for example, same tissue, same individual and/or same species. In various embodiments, comparison of ΔC_T values between the non-preamplified amplified target DNA and preamplified amplified target DNA can provide a measure for the bias of the preamplification process between the endogenous reference and the target DNA (x, y, z, etc.).

30 **[0037]** In various embodiments, the difference between the two ΔC_T values ($\Delta\Delta C_T$) can be zero and as such there is no bias from preamplification. This is explained in greater detail below with reference to FIG. 2. In some embodiments, the gene expression analysis system can be calibrated for potential differences in preamplification efficiency that can arise from a variety of sources, such as the effects of multiple primer sets in the same

reaction. In some embodiments, calibration can be performed by computing a reference number that reflects preamplification bias. Reference number similarity for a given target DNA across different samples is indicative that the preamplification reaction ΔC_T s can be used to achieve reliable gene expression computations.

5 **[0038]** In various embodiments of the present teaching, a gene expression analysis system can compute these reference numbers by collecting a sample (designated as Sample A (S_A)) and processing it with one or more protocols. A first protocol comprises running individual PCR gene expression reactions for each target DNA (T_x) relative to an endogenous reference (endo), such as, for example, 18s or GAPDH. These reactions can
10 yield cycle threshold values for each target DNA relative to the endogenous control; as computed by:

$$\Delta C_{T \text{ not preamplified } T_x S_A} = C_{T \text{ not preamplified } T_x S_A} - C_{T \text{ not preamplified endo}}$$

[0039] A second protocol can comprise running a single PCR preamplification step on assay with, for example, a pooled primer set. In various embodiments, the pooled
15 primer set can contain primers for each target DNA. Subsequently, the preamplified product can be distributed among a plurality of wells of a microplate. PCR gene-expression reactions can be run for each preamplified target DNA (T_x) relative to an endogenous reference (endo). These reactions can yield cycle threshold values for each preamplified target DNA relative to the endogenous control, as computed by:

$$\Delta C_{T \text{ preamplified } T_x S_A} = C_{T \text{ preamplified } T_x S_A} - C_{T \text{ preamplified endo } T_x S_A}$$

A difference between these $\Delta C_{T \text{ not preamplified } T_x S_A}$ and $\Delta C_{T \text{ preamplified } T_x S_A}$ can be computed by:

$$\Delta \Delta C_{T_x S_A} = \Delta C_{T \text{ not preamplified } T_x S_A} - \Delta C_{T \text{ preamplified } T_x S_A}$$

[0040] In various embodiments, a value for $\Delta \Delta C_{T_x S_A}$ can be zero or close to zero, which can indicate that there is no bias in the preamplification of target DNA T_x . In
25 various embodiments, a negative $\Delta \Delta C_{T_x S_A}$ value can indicate the preamplification process was less than 100% efficient for a given target DNA (T_x). For example, when using an IVT preamplification process, a percentage of target DNA with a $\Delta \Delta C_T$ of +/- 1 C_T of zero can be ~50%, as shown in FIG. 3. In another example, when using a multiplex preamplification process, a percentage of target DNA with a $\Delta \Delta C_T$ of +/- 1 C_T of zero can be ~90 %, as
30 shown in FIG. 4.

[0041] In various embodiments, an amplification efficiency can be less than 100% for a particular target DNA, therefore $\Delta \Delta C_T$ is less than zero for the particular target DNA. An example, as shown in FIG. 5, can be an evaluation of $\Delta \Delta C_T$ values for a group of target DNA from a 1536-plex for the multiplex preamplification process including four

different human sample input sources: liver, lung, brain and an universal reference tissue composite. In this example, most $\Delta\Delta C_T$ values are near zero, however, some of the target DNA have a negative $\Delta\Delta C_T$ value but these negative values are reproducible from one sample input source to another. In various embodiments, a gene expression analysis system
 5 can determine if a bias exists for target DNA analyzed for different sample inputs.

[0042] In various embodiments of the present teachings, a gene expression analysis system can use $\Delta\Delta C_T$ values computed for the same target DNA but in different samples (Sample A (S_A) and Sample B (S_B)) in order to determine the accuracy of subsequent relative expression computations. This results in the equation,

$$10 \quad \Delta\Delta\Delta C_T T_x = \Delta\Delta C_T T_x S_A - \Delta\Delta C_T T_x S_B$$

[0043] In various embodiments a value for $\Delta\Delta\Delta C_T T_x$ can be zero or reasonably close to zero which can indicate that the preamplified ΔC_T values for T_x (ΔC_T preamplified $T_x S_A$ and ΔC_T preamplified $T_x S_B$) can be used for relative gene expression computation between different samples via a standard relative gene expression calculation.

[0044] In some embodiments, a standard relative gene expression calculation
 15 can determine the amount of the target DNA. In some embodiments, a standard relative gene expression calculation employs a comparative C_T . In various embodiments, the above methods can be practiced during experimental design and once the conditions have been optimized so that the $\Delta\Delta\Delta C_T T_x$ is reasonably close to zero, subsequent experiments only
 20 require the computation of the ΔC_T value for the preamplified reactions. In various embodiments, $\Delta\Delta C_T T_x S_A$ values can be stored in a database or other storage medium. In such embodiments, these values can then be used to convert $\Delta\Delta C_T$ preamplified $T_x S_A$ values to $\Delta\Delta C_T$ not preamplified $T_x S_A$ values. In such embodiments, the $\Delta\Delta C_T$ preamplified $T_x S_y$ values can be mapped back to a common domain. In various embodiments, a not preamplified domain can
 25 be calculated using other gene expression instrument platforms such as, for example, a microarray. In various embodiments, the $\Delta\Delta C_T T_x S_A$ values need not be stored for all different sample source inputs (S_A) if it can be illustrated that the $\Delta\Delta C_T$ preamplified T_x is reasonably consistent over different sample source inputs. For example, a distribution of $\Delta\Delta\Delta C_T$ for two different sample inputs (liver and brain) are shown in FIG. 6 (IVT preamplification) and FIG. 7 (multiplex preamplification).
 30

[0045] In various embodiments, gene expression can be assessed with microarray technology, which can provide a measure of the cellular concentration of different mRNAs. In some embodiments, a microarray can be a piece of glass or plastic on which single stranded pieces of DNA are affixed in a microscopic array as probes. In some

embodiments, thousands of identical probes can be affixed at each point in the array which can make effective detectors.

[0046] Typically arrays can be used to detect the presence of mRNAs that may have been transcribed from different genes and which encode different proteins. The RNA can be extracted from many cells, ideally from a single cell type, then converted to cDNA. In various embodiments, the cDNA may be amplified in quantity by PCR. Fluorescent tags can be enzymatically incorporated into the or can be chemically attached to strands of cDNA. In various embodiments, a cDNA molecule that contains a sequence complementary to one of the probes will hybridize via base pairing to the point at which the complementary probes are affixed. In such embodiments, the point on the array can then fluoresce when examined using a microarray scanner. In some embodiments, the intensity of the fluorescence can be proportional to the number of copies of a particular mRNA that were present and thus roughly indicates the activity or expression level of that gene.

[0047] In various embodiments, a microarray can be, for example, a cDNA array, a hybridization array, a DNA microchip, a high density sequence oligonucleotide array, or the like. In various embodiments, a microarray can be available from a commercial source such as, for example, Applied Biosystems, Affymetrix, Agilent, Illumina, or Xeotron. In various embodiments, a microarray can be made by any number of technologies including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micromirror devices, or ink-jet printers. The lack of standardization in microarrays can present an interoperability problem in bioinformatics, since it can limit the exchange of array data.

[0048] In various embodiments, microarray output data can be in a format of fluorescence intensity and in other embodiments, microarray output data may be in a format of chemiluminescence intensity. In various embodiments, an intensity value from a microarray output data can be globally normalized. In some embodiments, fold difference values can be determined by subtracting background noise and normalizing the array signal intensity, then dividing experimental sample signal intensity by a control sample signal intensity yielding net sample intensity. In some embodiments, a control sample used to generate the control sample signal intensity can be, for example, Stratagene®, UHR or the like. In some embodiments, a fold difference can be converted to a $\log_2$ by the following equation:

$$2^{\Delta\Delta C_T} = 3.3 \log_{10} (\text{net intensity sample 1} / \text{net intensity sample 2})$$

In such embodiments, microarray output data is in a $\Delta\Delta C_T$ format. In some embodiments, microarray output data can be converted into a $\Delta\Delta C_T$ format by the following equation:

$$R = (1/2)^{\Delta\Delta C_T}$$

where R is the resulting measurement from a microarray. Such calculations are available commercially, such as GeneSpring from Silicon Genetics. Other embodiments include converting microarray output data into a $\Delta\Delta C_T$ format using a Global Pattern Recognition (GPR) algorithm which can convert intensity values generated from microarrays from linear values to algorithmic values and can use transformed intensity cutoffs to effect gene and normalizer filters. In such embodiments, GPR, a software algorithm for gene expression analysis is available from The Jackson Laboratory. In various embodiments, microarray output data can be in a standard language or format such as MAGE-ML (microarray and gene expression markup language), MAML (microarray markup language), or MIAME (minimum information about microarray experiments). In various embodiments, such standardized formats and language can be converted to a $\Delta\Delta C_T$ format.

[0049] In various embodiments, after microarray output data is in a $\Delta\Delta C_T$ format, then real-time PCR data can be directly compared to data from microarray platforms as shown in FIG. 8. In various embodiments, a $\Delta\Delta\Delta C_T$ calculation can be a validation tool to confirm that relative quantitation data can be compared from one amplification/detection process to another. In various embodiments, $\Delta\Delta\Delta C_T$ calculation can be a validation tool to confirm that relative quantitation data can be compared from one sample input source to another sample input source, for example, comparing a sample from liver to a sample from brain in the same individual. In various embodiments, $\Delta\Delta\Delta C_T$ calculation can be a validation tool to confirm that relative quantitation data can be compared from one high-density sequence detector system to another high-density sequence detection system. In various embodiments, $\Delta\Delta\Delta C_T$ calculation can be a validation tool to confirm that relative quantitation data can be compared from one platform to another, for example, data from real time PCR to data from a hybridization array is especially valuable for cross-platform validation. In various embodiments, real time PCR and hybridization array data can be directly compared. In various embodiments, a TaqMan $\Delta\Delta C_T$ can be compared to a microarray output converted to the $\Delta\Delta C_T$ format. In such embodiments, the resultant $\Delta\Delta\Delta C_T$, if within $\pm 1 C_T$ of zero, can determine a high-degree of confidence that the actual fold difference observed within each of the two platforms is correlative.

[0050] In various embodiments, a correction, which can be a quantity added to a calculated or observed value to obtain the true value, may be used so that data generated

on two different platforms can be used together in further calculations and analysis. Such embodiments allow for larger and sometimes more complete data sets to be used in gene expression studies. In some embodiments, the correction can be calculated from a resulting $\Delta\Delta C_T$. In various embodiments, a correction can be a bias correction.

CLAIMS

What is claimed is:

1. A method for determining bias across two domains comprising gene
5 expression data, the method comprising:
 providing a first domain and a second domain;
 obtaining information indicative of a bias within the first domain;
 obtaining information indicative of a bias within the second domain; and
 using the information indicative of the bias within the first domain and the information
10 indicative of the bias within the second domain to produce an indication of bias across the
two domains.
2. The method according to Claim 1 wherein the providing a first domain and a
second domain further comprises providing at least one of the first domain and the second
domain comprising information collected from a polynucleotide amplification instrument.
- 15 3. The method according to Claim 2 wherein the polynucleotide amplification
instrument is capable of running polymerase chain reactions.
4. The method according to Claim 1 wherein the providing a first domain and a
second domain further comprises providing at least one of the first domain and the second
domain comprising information generated by gene expression analysis.
- 20 5. The method according to Claim 4 wherein the providing at least one of the
first domain and the second domain comprise data generated by gene expression analysis
comprising generating gene expression data from an polynucleotide amplification
instrument.
6. The method according to Claim 4 wherein the providing at least one of the
25 first domain and the second domain comprise data generated by gene expression analysis
comprising generating gene expression data from a microarray.
7. The method according to Claim 4 wherein the providing at least one of the
first domain and the second domain comprise data generated by gene expression analysis
comprising generating gene expression data from a hybridization chip.
- 30 8. The method according to Claim 1 wherein the providing a first domain and a
second domain further comprises providing information from a preamplified and amplified
sample and providing information from an amplified sample.
9. The method according to Claim 1 further comprising amplifying a
polynucleotide sample.

10. The method according to Claim 9 further comprising collecting information indicative of the amplifying a polynucleotide sample.

11. The method according to Claim 10 further comprising organizing the information indicative of the amplifying a polynucleotide sample into the first domain.

5 12. The method according to Claim 9 further comprising preamplifying a polynucleotide sample.

13. The method according to Claim 12 further comprising collecting information indicative of the preamplifying a polynucleotide sample and amplifying a polynucleotide sample.

10 14. The method according to Claim 13 further comprising organizing the information indicative of the preamplifying a polynucleotide sample and amplifying a polynucleotide sample into the first domain.

15 15. The method according to Claim 14 wherein the obtaining information indicative of bias within the first domain comprises evaluating a bias between the amplifying a polynucleotide sample and the preamplifying a polynucleotide sample and amplifying a polynucleotide sample.

16. The method according to Claim 15 wherein the obtaining information indicative of bias within the second domain comprises evaluating a bias between the amplifying a second polynucleotide sample and the preamplifying a polynucleotide second sample and amplifying a second polynucleotide sample.

17. The method according to Claim 16 wherein the polynucleotide sample and the second polynucleotide sample are from different tissues.

18. The method according to Claim 17 wherein the different tissues are from the same individual.

25 19. The method according to Claim 1 wherein the obtaining information indicative of bias within the first domain further comprises performing a $\Delta\Delta C_T$ calculation.

20. The method according to Claim 1 wherein the obtaining information indicative of bias within the second domain further comprises performing a $\Delta\Delta C_T$ calculation.

21. The method according to Claim 1 wherein the using the information indicative of bias within the first domain and the information indicative of bias within the second domain to produce an indication of bias across the two domains further comprises performing a $\Delta\Delta\Delta C_T$ calculation.

22. The method according to Claim 1 wherein the obtaining information indicative of bias within the first domain comprises information from a first amplification device and the

obtaining information indicative of bias within the second domain comprises from information a second amplification device.

23. The method according to Claim 1 further comprising converting information from a micro array device into a $\Delta\Delta C_T$ format.

24. The method according to Claim 23 the obtaining information indicative of bias within the first domain comprises information from a first amplification device and the obtaining information indicative of bias within the second domain comprises from information from a microarray device.

25. The method according to Claim 1 wherein the obtaining information indicative of bias within the first domain comprises information from a first tissue type and the obtaining information indicative of bias within the second domain comprises from information a second tissue type.

26. The method according to Claim 1 wherein the obtaining information indicative of bias within the first domain comprises information from gene expression of a first individual and the obtaining information indicative of bias within the second domain comprises from information from gene expression of a second individual.

27. The method according to Claim 26 wherein the first individual and the second individual are from the same species.

28. The method according to Claim 26 wherein the first individual and the second individual are from different species.

29. The method according to Claim 1 further comprising applying a correction to at least one of the first domains and the second domains.

30. The method according to Claim 29 wherein the applying a correction to at least one of the first domains and the second domains corrects for a bias across the first domain and the second domain.

31. The method according to Claim 1 wherein at least one of the first domain and the second domain comprises genomic information.

32. A method for determining bias across two domains comprising gene expression data, the method comprising:

providing a first domain comprising preamplified gene expression data and non-preamplified gene expression data from a first sample input;

determining a bias between the preamplified gene expression data and the non-preamplified gene expression data of the first domain;

providing a second domain comprising preamplified gene expression data and non-preamplified gene expression data from a second sample input.

determining a bias between the preamplified gene expression data and the non-preamplified gene expression data of the second domain.

5 using the bias of the first domain and the bias of the second domain to produce an indication of bias across the first domain and the second domain.

33. The method according to Claim 32 further comprising amplifying the first sample input.

10 34. The method according to Claim 33 further comprising collecting data from the amplifying the first sample input.

35. The method according to Claim 33 further comprising adding a first reference to the first sample input.

36. The method according to Claim 34 further comprising using the first reference to normalize data from the amplifying the first sample input.

15 37. The method according to Claim 34 further comprising using a ΔC_T calculation to normalize data from the amplifying the first sample input.

38. The method according to Claim 32 further comprising preamplifying the first sample input.

20 39. The method according to Claim 38 further comprising amplifying the first sample input.

40. The method according to Claim 39 further comprising collecting data from the amplifying the first sample input.

41. The method according to Claim 39 further comprising adding a first reference to the first sample input.

25 42. The method according to Claim 41 further comprising using the first reference to normalize data from the preamplifying and amplifying the first sample input.

43. The method according to Claim 41 further comprising using a ΔC_T calculation to normalize data from the preamplifying and amplifying the first sample input.

30 44. The method according Claim 43 further comprising using a $\Delta\Delta C_T$ calculation to determine bias.

45. The method according to Claim 32 further comprising amplifying the second sample input.

46. The method according to Claim 45 further comprising collecting data from the amplifying the second sample input.

47. The method according to Claim 45 further comprising adding a second reference to the second sample input.

48. The method according to Claim 46 further comprising using the second reference to normalize data from the amplifying the second sample input.

5 49. The method according to Claim 46 further comprising using a ΔC_T calculation to normalize data from the amplifying the second sample input.

50. The method according to Claim 32 further comprising preamplifying the second sample input.

10 51. The method according to Claim 50 further comprising amplifying the second sample input.

52. The method according to Claim 51 further comprising collecting data from the amplifying the second sample input.

53. The method according to Claim 50 further comprising adding a second reference to the second sample input.

15 54. The method according to Claim 53 further comprising using the second reference to normalize data from the preamplifying and amplifying the second sample input.

55. The method according to Claim 53 further comprising using a ΔC_T calculation to normalize data from the preamplifying and amplifying the second sample input.

20 56. The method according Claims 49 and 55 further comprising using a $\Delta\Delta C_T$ calculation to determine bias.

57. A method according to Claim 32 wherein the using the bias of the first domain and the bias of the second domain to produce an indication of bias across the first domain and the second domain further comprises performing a $\Delta\Delta\Delta C_T$ calculation.

25 58. A system for determining bias across two domains, the system comprising:
a first domain stored on a media;
a second domain stored on a media;
a first algorithm for obtaining information indicative of a bias within the first domain;
a second algorithm for obtaining information indicative of a bias in the second domain;
30 domain;
a third algorithm for using the information indicative of the bias within the first domain and the information indicative of the bias within the second domain to produce an indication of bias across the two domains; and
an output.

59. The system according to Claim 58 further comprising at least one computer.

60. The system according to Claim 58 further comprising at least one PCR device.

61. The system according to Claim 58 further comprising at least one microarray device.

62. The system according to Claim 58 further comprising high density sequence detection system.

63. The system according to Claim 58 further comprising at least one sample.

64. The system according to Claim 58 further comprising a graphical user interface.

65. The system according to Claim 58 further comprising a network.

66. The system according to Claim 65 wherein the first domain and the second domain are at different loci on the network.

67. The system according to Claim 58 further comprising a hybridization chip device.

68. The system according to Claim 58 wherein at least one of the first algorithm, the second algorithm, and the third algorithm is a comparative method.

69. The system according to Claim 58 wherein at least one of the first algorithm and the second algorithm is a $\Delta\Delta C_T$ calculation.

70. The system according to Claim 58 wherein the third algorithm is a $\Delta\Delta\Delta C_T$ calculation.

71. The system of Claim 58 wherein at least one of the first domain and the second domain comprises polynucleotide information.

72. The system of Claim 58 wherein at least one of the first domain and the second domain comprises gene expression information.

73. The system of Claim 58 wherein at least one of the first domain and the second domain comprises polynucleotide amplification information.

74. The system of Claim 58 wherein at least one of the first domain and the second domain comprises genomic information.

75. The system of Claim 58 wherein the first domain comprises information on a first tissue and the second domain comprises information on a second tissue type.

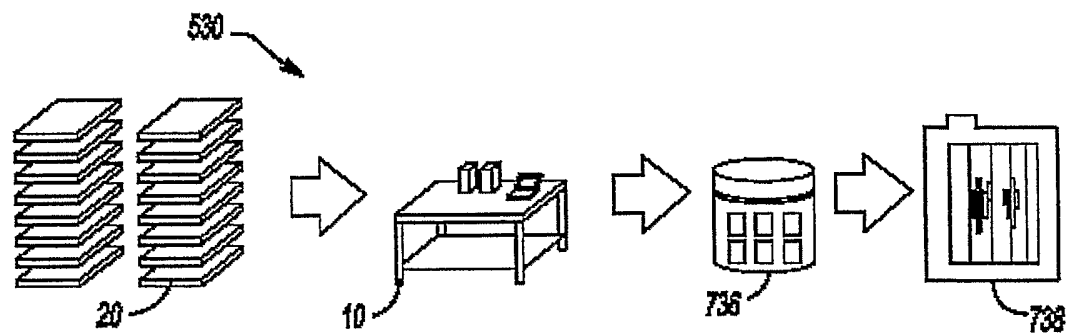
76. The system according to Claim 58 wherein at least one of the first domain and the second domain comprises information from a preamplified and amplified sample and providing information from an amplified sample.

77. The system according the Claim 63 wherein the at least one sample is analyzed for gene expression.

78. The system according to Claim 58 further comprising a data bank of reference information.

5 79. The system according to Claim 78 further comprising obtaining information indicative of a bias within at least one of the first domain and the second domain using the data bank of reference information.

Figure 1



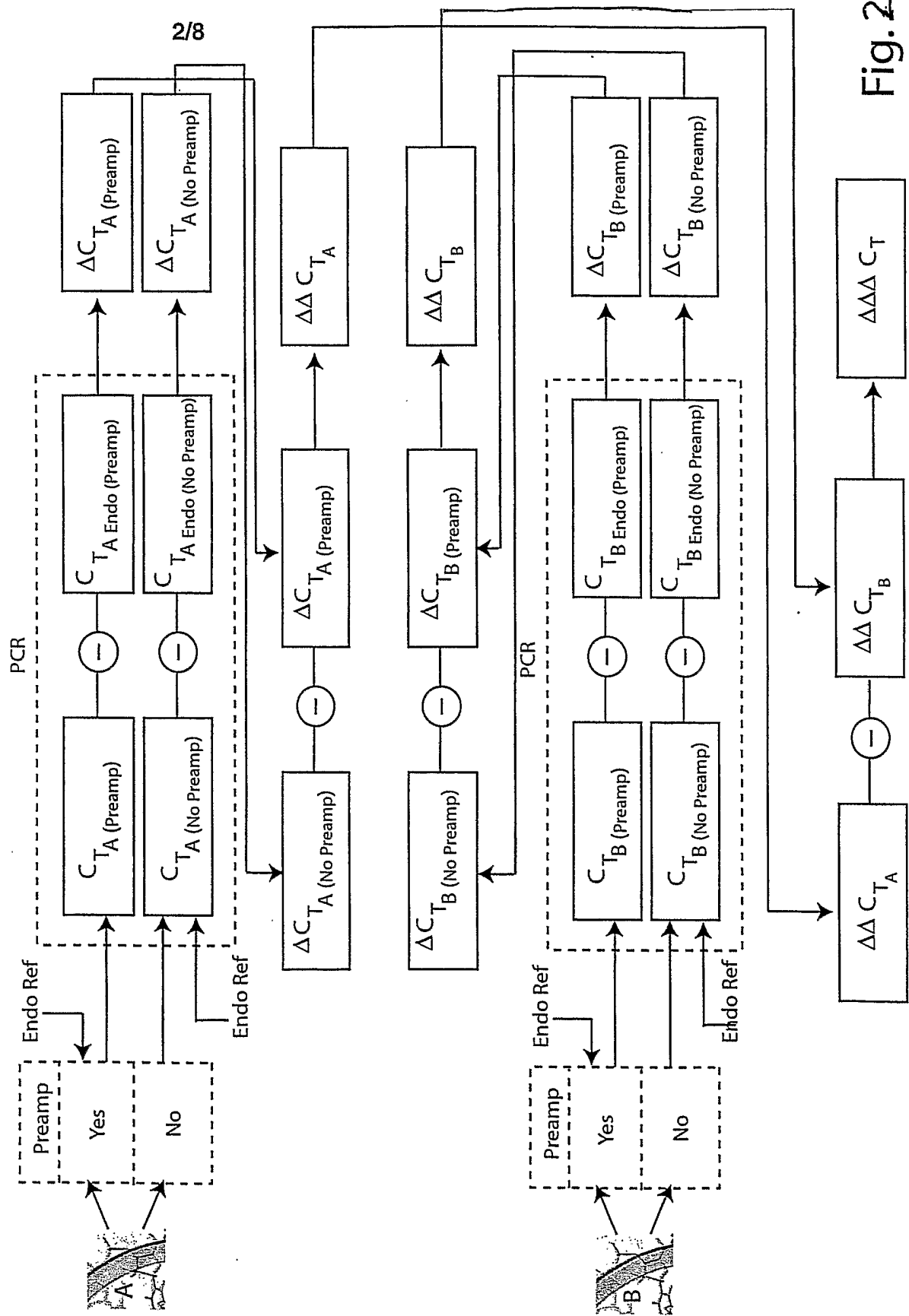
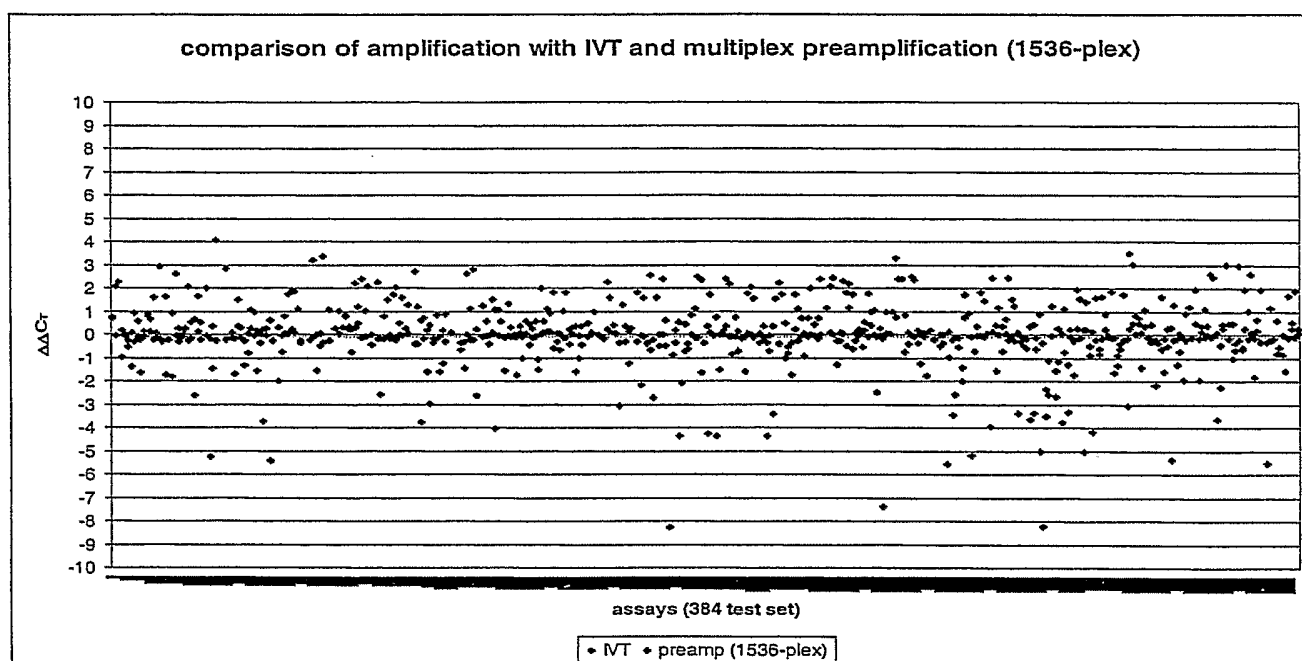


Fig. 2

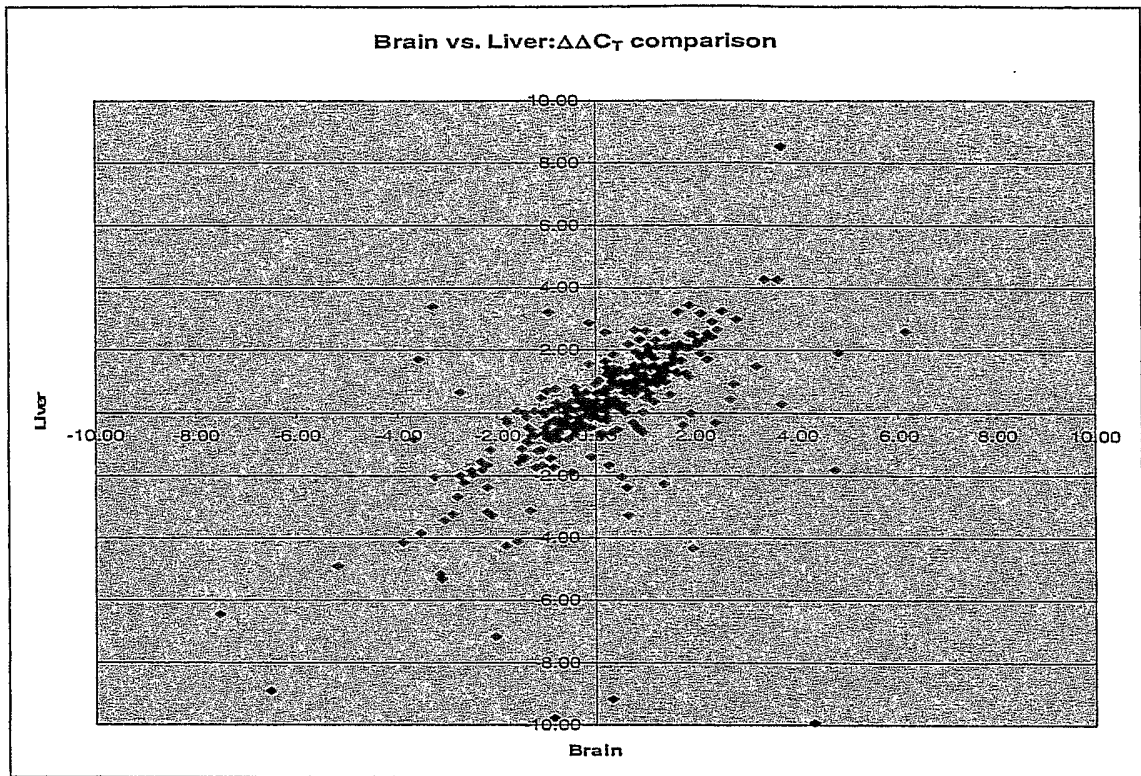
3/8

Figure 3



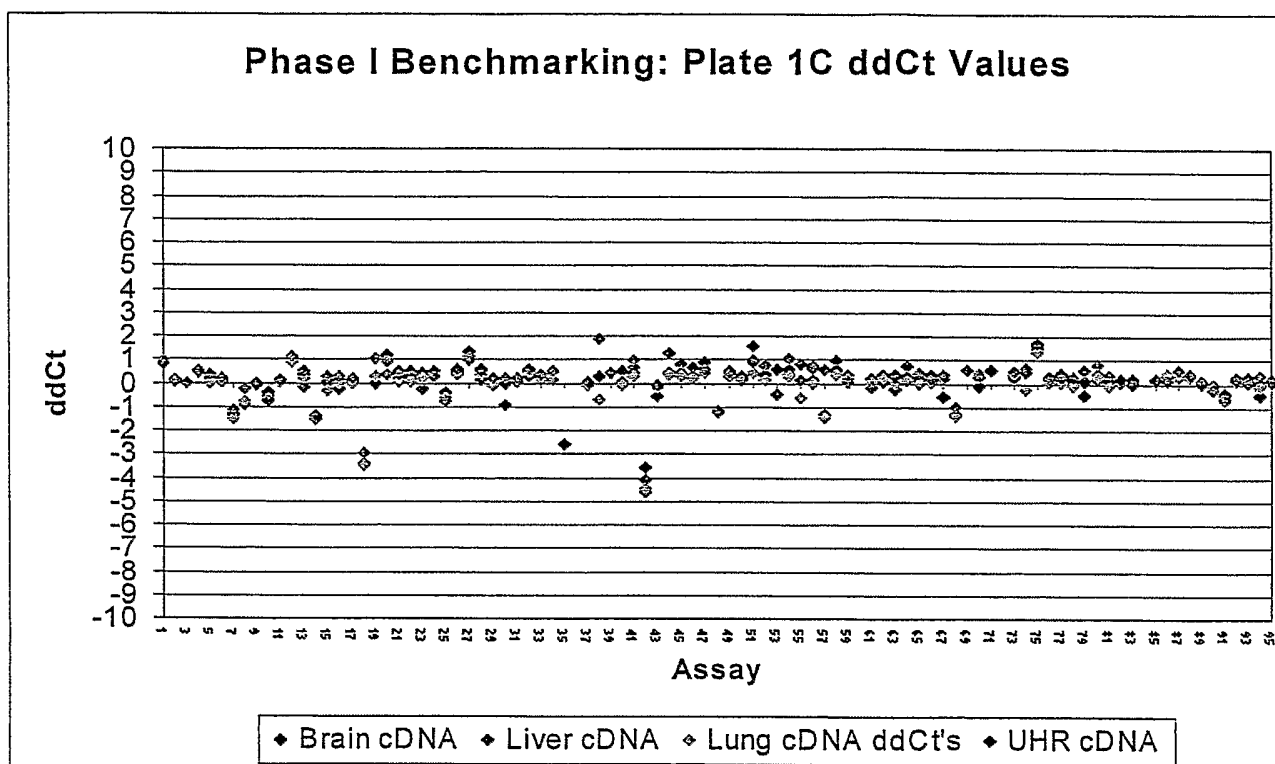
4/8

Figure 4



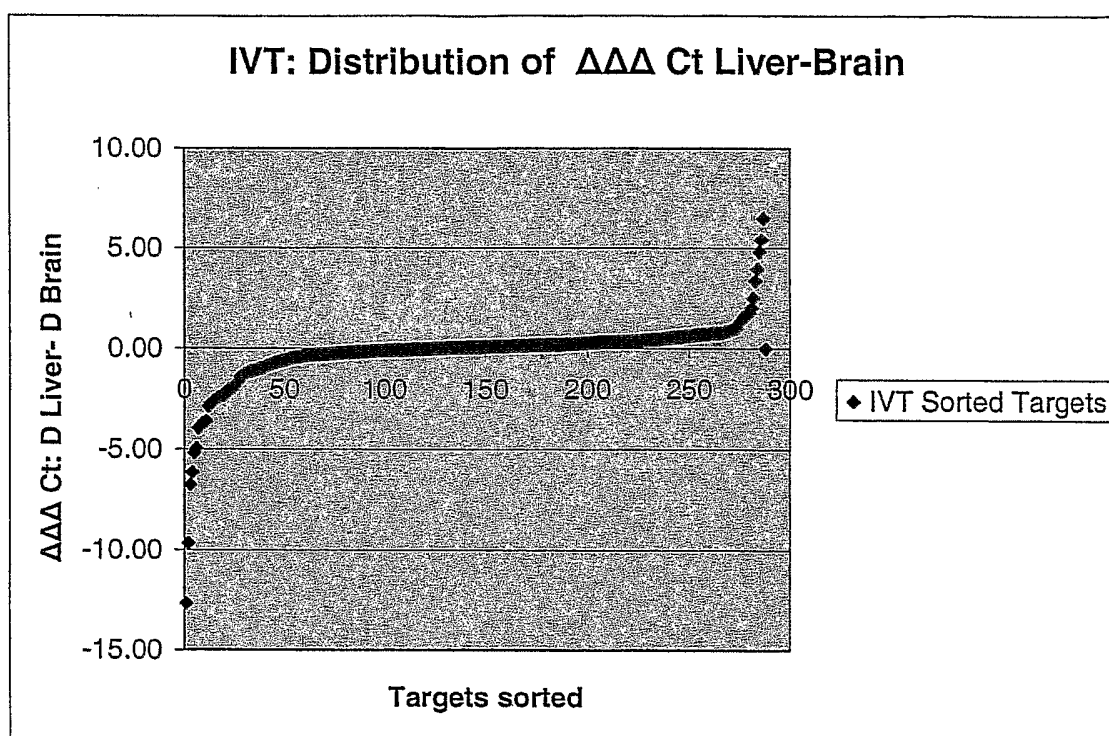
5/8

Figure 5



6/8

Figure 6



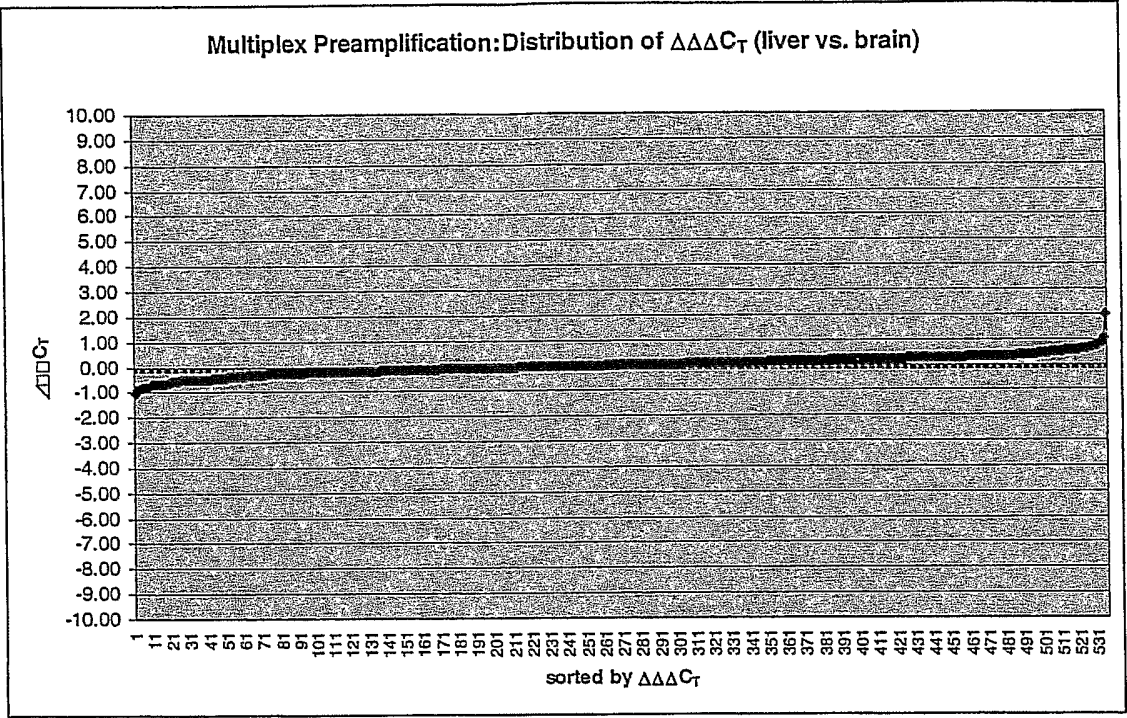


Figure 7

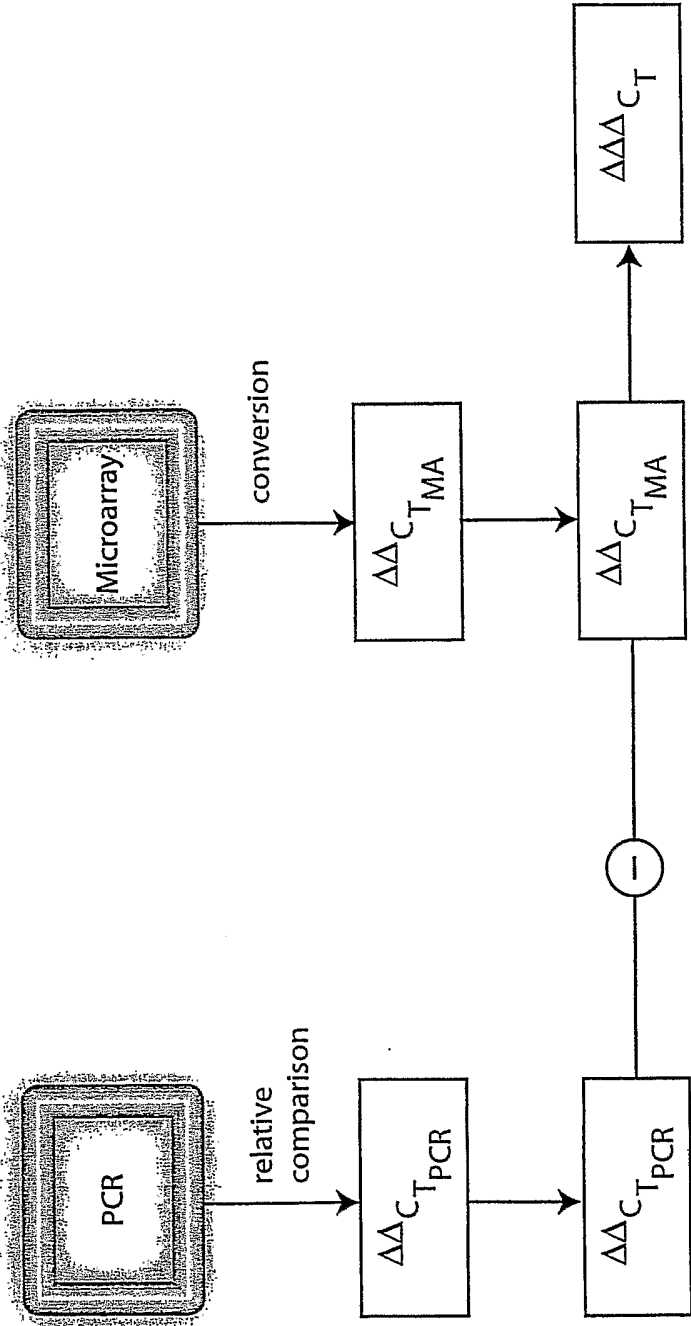


Fig. 8