



(86) Date de dépôt PCT/PCT Filing Date: 2011/07/06
(87) Date publication PCT/PCT Publication Date: 2012/01/12
(85) Entrée phase nationale/National Entry: 2012/12/19
(86) N° demande PCT/PCT Application No.: US 2011/001184
(87) N° publication PCT/PCT Publication No.: 2012/005764
(30) Priorité/Priority: 2010/07/08 (US61/362,676)

(51) Cl.Int./Int.Cl. *G06F 19/10* (2011.01)
(71) Demandeur/Applicant:
PRIME GENOMICS, INC., US
(72) Inventeur/Inventor:
SHAW, SANDY C., US
(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L., S.R.L.

(54) Titre : SYSTEME POUR LA QUANTIFICATION D'UNE DYNAMIQUE A L'ECHELLE DU SYSTEME DANS DES RESEAUX COMPLEXES

(54) Title: SYSTEM FOR THE QUANTIFICATION OF SYSTEM-WIDE DYNAMICS IN COMPLEX NETWORKS

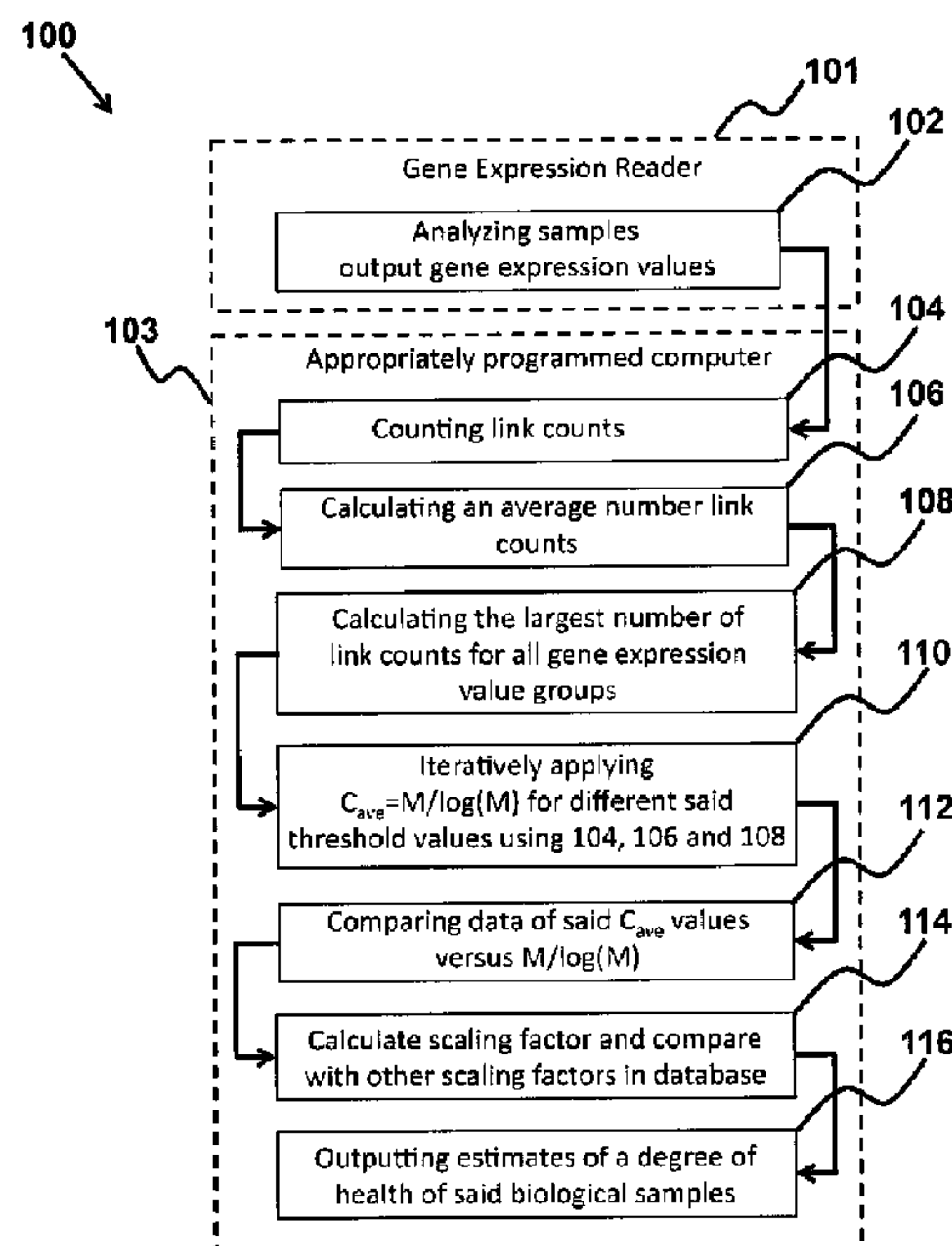


FIG. 1

(57) Abrégé/Abstract:

A device, method and system are provided for diagnosing a disease using a gene expression reader to analyze biological samples and output gene expression values to calculate a scaling factor using a computer by counting a number of link counts C_a for groups of an individual genes' expression values at different times at a threshold value C or for groups of genes' expression values at a single time at the threshold value C , calculating an average number C_{ave} of the link counts C_a , calculating a largest number M of the C_a , iteratively applying a relation $C_{ave} = M/\log(M)$ for different threshold values C , comparing data of the C_{ave} values versus $M/\log(M)$, and calculating a fitting to the compared data to output the scaling factor a . The scaling factor a is compared with other scaling factors a' in a database to output a report of estimates for a degree of health.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 January 2012 (12.01.2012)



(10) International Publication Number
WO 2012/005764 A3

(51) International Patent Classification:
G06F 19/10 (2011.01)

(21) International Application Number:
PCT/US2011/001184

(22) International Filing Date:
6 July 2011 (06.07.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/362,676 8 July 2010 (08.07.2010) US

(71) Applicant (for all designated States except US): **PRIME GENOMICS, INC.** [US/US]; 111 Commonwealth, San Francisco, CA 94118 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **SHAW, Sandy, C.** [US/US]; 111 Commonwealth, San Francisco, CA 94118 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

[Continued on next page]

(54) Title: SYSTEM FOR THE QUANTIFICATION OF SYSTEM-WIDE DYNAMICS IN COMPLEX NETWORKS

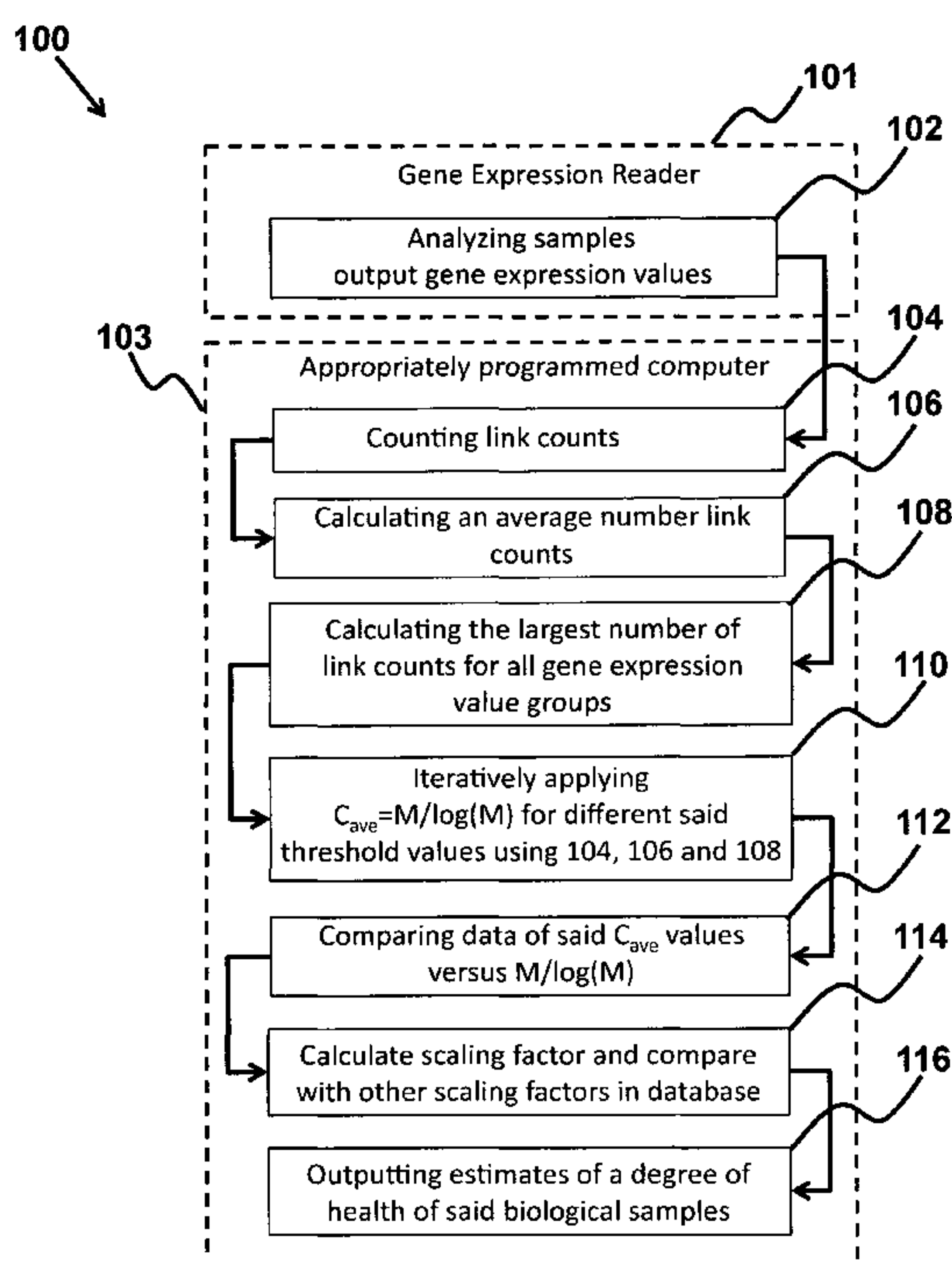


FIG. 1

(57) Abstract: A device, method and system are provided for diagnosing a disease using a gene expression reader to analyze biological samples and output gene expression values to calculate a scaling factor using a computer by counting a number of link counts C_a for groups of an individual genes' expression values at different times at a threshold value C or for groups of genes' expression values at a single time at the threshold value C , calculating an average number C_{ave} of the link counts C_a , calculating a largest number M of the C_a , iteratively applying a relation $C_{ave} = M/\log(M)$ for different threshold values C , comparing data of the C_{ave} values versus $M/\log(M)$, and calculating a fitting to the compared data to output the scaling factor a . The scaling factor a is compared with other scaling factors a' in a database to output a report of estimates for a degree of health.

WO 2012/005764 A3

- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))* (88) **Date of publication of the international search report:**
12 April 2012

SYSTEM FOR THE QUANTIFICATION OF SYSTEM-WIDE DYNAMICS IN COMPLEX NETWORKS

FIELD OF THE INVENTION

5 The present invention relates to diagnosing disease. More particularly, the invention relates to analyzing biological samples for gene expression values to determine a degree of health of the biological sample.

BACKGROUND OF THE INVENTION

10 A large, complex network of interacting components is difficult to describe as a whole dynamic system. In genetics research, scientists examining large numbers of genes, or genetic networks, often focus on identifying one gene or a group of genes that appears to be important to a particular outcome or pathology. What is needed are a low cost and efficient device, method and system for analyzing the interconnections between genes and genetic
15 networks on a large-scale to output a report of a degree of health in a patient.

SUMMARY OF THE INVENTION

To address the needs in the art, a method of diagnosing a disease is provided, according to one embodiment of the invention, that includes a gene expression reader analyzing at least
20 one biological sample and outputting gene expression values from at least two genes based on analyzing the biological samples, calculating a scaling factor a for the biological samples using an appropriately programmed computer, where the scaling factor a is

calculated from the gene expression values by counting a number of link counts C_n for groups of an individual genes' expression values at different times at a threshold value C , or for groups of genes' expression values at a single time at the threshold value C , calculating an average number C_{ave} of the link counts C_n , calculating a largest number M of the C_n , where the M includes the largest of the number of link counts C_n for a given threshold value C for all the gene expression value groups, iteratively applying a relation $C_{ave}=M/\log(M)$ for different threshold values C , comparing data of the C_{ave} values versus $M/\log(M)$, and calculating a fitting to the compared data to output the scaling factor a , where the scaling factor a is the slope of the fitting. The method further includes comparing values of the scaling factor a for the biological samples with other scaling factors a' in a database from analyzed biological samples using the appropriately programmed computer, and outputting a report using the appropriately programmed computer, where the report includes estimates of the at least one biological sample for a degree of health.

15

According to one aspect of the current method embodiment, the at least one biological sample can include saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, or other organic material.

20

In another aspect of the current method embodiment, the gene expression reader includes at least two gene probes.

In a further aspect of the current method embodiment, the number of link counts C_n includes a number of link counts for each of N expression value groups, where each expression value group includes a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value C between the expression value group and the sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other $N-1$ gene expression value groups.

According to another aspect of the current method embodiment, the scaling factor a is calculated by iteratively applying $C_{ave} = M/\log(M)$ for different threshold values C , using the appropriately programmed computer, and comparing C_{ave} values versus $M/\log(M)$, and calculating a linear fitting of the comparison to get the scaling factor a .

In yet another aspect of the current method embodiment, comparing values of a further includes comparing byproducts of the scaling factor a , comparing healthy samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.

According to another aspect of the current method embodiment, the threshold value C is in a range between 0 and 1.

In another embodiment of the invention, a system for diagnosing disease is provided that includes a gene expression reader for analyzing at least one biological sample and outputting gene expression values of at least two genes, a computer server for receiving from the gene expression reader the gene expression values and for managing and communicating patient information to a user, and a computer program hosted on the

computer server, where the computer program analyzes the gene expression values and outputs a report, where the report includes estimates of the at least one biological sample for a degree of health, where the estimate includes comparing a scaling factor a for the at least one biological sample with other scaling factors a' in a database from previously
5 analyzed biological samples, where the scaling factor a is calculated from the gene expression values using the computer program by counting a number of link counts C_n for groups of an individual genes' expression values at a different times at a threshold value C or for groups of genes' expression values at a single time at the threshold value C , calculating an average number C_{ave} of the link counts C_n , calculating a largest number M of
10 the C_n , where the M includes the largest of the number of link counts C_n for a given threshold value C for all the gene expression value groups, iteratively applying a relation $C_{ave}=M/\log(M)$ for different threshold values C , comparing the C_{ave} data values versus $M/\log(M)$ data, and applying a fitting to the compared data to output the scaling factor a , where the scaling factor a is the slope of the fitting.

15

According to one aspect of the current system embodiment, the at least one biological sample can include saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, or organic material.

20

In another aspect of the current system embodiment, the gene expression reader includes at least two gene probes.

In a further aspect of the current system embodiment, the number of link counts C_n includes a number of link counts for each of N expression value groups, where each expression value group includes a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value C between the expression value group and the sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other $N-1$ gene expression value groups.

According to another aspect of the current system embodiment, the a scaling factor a is calculated by iteratively applying $C_{ave}=M/\log(M)$ for different threshold values C , using the appropriately programmed computer, and comparing C_{ave} values versus $M/\log(M)$ and calculating a linear fitting of the comparison to get the scaling factor a .

In yet another aspect of the current system embodiment, comparing values of a further includes comparing byproducts of the scaling factor a , comparing healthy samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.

In a further aspect of the current system embodiment, the threshold value C is in a range between 0 and 1.

In another embodiment, the invention includes lab-on-a-chip device having a substrate for holding a biological sample receptacle, a gene expression reader and a microprocessor, where biological sample receptacle includes a sample input to the gene expression reader, where the gene expression reader outputs gene expression values of at least two genes based on analyzed the at least one biological sample, where the microprocessor includes a

computer program for analyzing gene expressions in the at least one biological sample, where the computer program compiles the gene expression values, counts a number of link counts C_n for groups of an individual genes' expression values at different times at a threshold value C or for groups of genes' expression values at a single time at the threshold value C , calculates an average number C_{ave} of the link counts C_n , calculates a largest number M of the C_n , where the M includes the largest of the number of link counts C_n for a given the threshold value C for all the gene expression value groups, iteratively applies a relation $C_{ave}=M/\log(M)$ for different threshold values C , compares data of the C_{ave} values versus $M/\log(M)$ data, calculates a fitting to the compared data to output the scaling factor a , where the scaling factor a is the slope of the fitting, compares values of the scaling factor a for the at least one biological sample with other stored scaling factors a' from analyzed biological samples, and outputs a report, where the report includes estimates of the at least one biological sample for a degree of health.

15 According to one aspect of the current device embodiment, the at least one biological sample can include saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, or organic material.

20 In another aspect of the current device embodiment, the gene expression reader includes at least two gene probes.

In a further aspect of the current device embodiment, the number of link counts C_n includes a number of link counts for each of N expression value groups, where each expression

value group includes a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value C between the expression value group and the sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other $N-1$ gene expression value groups.

- 5 According to one aspect of the current device embodiment, the a scaling factor a is calculated by iteratively applying the $C_{ave} = M/\log(M)$ for different threshold values C , using the appropriately programmed computer, and comparing C_{ave} values versus $M/\log(M)$ and calculating a linear fitting the comparison to get the scaling factor a .
- 10 In a further aspect of the current device embodiment, comparing values of a further includes comparing byproducts of the scaling factor a , comparing healthy samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.
- 15 In yet aspect of the current device embodiment, the threshold value C is in a range between 0 and 1.

BRIEF DESCRIPTION OF THE DRAWINGS

- 20 **FIG. 1** shows a flow diagram of a method of one embodiment of the current invention.
- FIG. 2** shows a graphical image of the process used by a computer program to calculate the scaling factor, according to one embodiment of the current invention.

FIG. 3 shows a flow diagram of a system of one embodiment of the current invention.

FIG. 4 shows a schematic drawing of a device of one embodiment of the current invention.

DETAILED DESCRIPTION

5

To address the needs in the art, a method of diagnosing a disease is provided, according to one embodiment of the invention. **FIG. 1** shows a flow diagram of a method **100** of one embodiment of the invention, that includes a gene expression reader **101** analyzing at least one biological sample and outputting gene expression values **102** from at least two genes

10 based on analyzing the at least one biological sample and use this to calculate a scaling factor **a** for the biological sample using an appropriately programmed computer **103**, where the scaling factor **a** is calculated from the gene expression values by counting a number of link counts **C_n** **104** for groups of an individual genes' expression values at different times at a threshold value **C** or for groups of genes' expression values at a single time at the

15 threshold value **C**, calculating an average number **C_{ave}** **106** of the link counts **C_n**, calculating a largest number **M** of the **C_n** **108**, where the **M** includes the largest of the number of link counts **C_n** for a given threshold value **C** for all the gene expression value groups, iteratively applying a relation $C_{ave} = M / \log(M)$ for different threshold values **C** **110**, comparing data of the **C_{ave}** values versus $M / \log(M)$ **112**, and calculating a fitting to the

20 compared data to output the scaling factor **a**, where the scaling factor **a** is the slope of the fitting and comparing values of the scaling factor **a** for the at least one biological sample with other scaling factors **a'** **114** in a database from analyzed biological samples using the appropriately programmed computer, and outputting a report **116** using the appropriately programmed computer, where the report includes estimates of the at least one biological

sample for a degree of health. In one aspect of the current embodiment, the gene expression reader includes at least two gene probes.

According to one embodiment of the method 100, the invention uses gene expression values, for example from a microarray or genechip, for N expression value groups that can include a large number, if not all, the genes in a genome for a given organism, for example. In one embodiment, N does not need to contain all available expression value groups of the microarray data, only a large subset of the microarray data.

10 In one embodiment of the method 100, the gene expression values $\mathbf{n}_T$ can be read from the microarray at multiple time intervals T . The dataset for quantification will include N groups of gene expression values $\mathbf{n}_T$ of the form:

$$\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_T$$

Where $\mathbf{n}$ is the gene expression value of one of N genes taken at T intervals.

15

For the sequence of gene expression values $\mathbf{n}_j$ in the gene expression value group N_i , the absolute value is taken of a correlation between the gene expression value group N_i and every other gene expression value group (the other $N-1$ groups).

20 The total number of other gene expression value groups with a correlation above a threshold value C is called C_n and represents the number of links connecting this gene expression value group to all other gene expression value groups in the dataset with a value

of C or greater. The largest of the C_n for a given C for all N gene expression value groups is then taken and called M . The average of all the C_n for a given C is also taken and called C_{avg} . According to one embodiment of the invention, for different values of C , the values of M and C_{avg} form the relation:

$$C_{avg} = (M/\log(M))^a$$

To find the value of the scaling factor a , the method above is repeated by iteratively applying a relation $C_{ave} = M/\log(M)$ for different threshold values C , comparing the C_{ave} data values versus $M/\log(M)$ data, and applying a fitting to the compared data to output the scaling factor a , where the scaling factor a is the slope of the fitting. According to the current embodiment, the threshold value C is in a range between 0 and 1.

In one embodiment of the method 100, shown in FIG. 2 is an exemplary graphical scaling factor representation 200, where the number of values of cutoff value C is nineteen, C is the absolute value of the correlation, for example a Pearson correlation, and C ranges from .95 to .05 at decreasing values of .05 for each point. The slope of the line fitted to a log-log plot of the data is then measured. In this case a is shown to be ~ 1.74 . In FIG. 2, the correlation values measured are between time series of six gene expression values ($T=6$) taken at seven-minute intervals for 3360 genes ($N=3360$) in yeast (*S. cerevisiae*). Although 3360 genes are used in this example, the genes used in other examples can be any number, but are generally in the thousands. In one embodiment, it is possible to apply this method to groups of gene expression values measured at a single time rather than individual gene's expression values at different times. In other words, the correlation values are between N groups made up of gene expression values from T genes taken at a single time.

In one example of this embodiment, given gene expression values for 5 different genes at a single time labeled 1-5, three gene expression value groups ($N=3$) can be made containing three gene expression values each ($T=3$). For example, the gene expression values from genes 1-3, 2-4, 3-5. The invention calculates the absolute values of the Pearson correlation between each group, and the other two ($N-1=2$). Assume that 4 of the correlation values calculated are $> .95$. Then C_{ave} for $C=.95$ and $N=3 = 4/3 = 1.33$. Further, assume that the largest number of absolute Pearson correlation values $> .95$ for any single gene expression value group is 2. Then M for $C=.95$ would be 2.

The essence of both the single-time groups and the time series (time groups) approach is that in each case correlation values are taken between one group and all the other groups. Then it is calculated how many correlation values are greater than the threshold C . The largest number for any single group is M . The total number for all groups divided by the number of groups (N) gives C_{ave} . Though these are two different ways to calculate scaling factors a that could be different values, according to one aspect of the invention, the only requirement is that either method used to generate a must be consistent when comparing values of a between biological samples.

According to one aspect of the method 100, the at least one biological sample can include saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, or other organic material.

In another aspect of the method **100**, comparing values of **a** further includes comparing byproducts of the scaling factor **a**, comparing healthy samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.

5

In another embodiment of the invention, **FIG. 3** shows a system for diagnosing disease **300** that includes a user **302** having a biological sample **304** to input to a gene expression reader **306** for analyzing at least one biological sample **304** and outputting **310** gene expression values of at least two genes, and communicating **310** the gene expression values, for example using the internet, to a computer server **312** for receiving from the gene expression reader **306** the gene expression values and for managing and communicating patient information, where the patient information is then provided to the user **302**. A computer program **314** is hosted on the computer server **312** and analyzes the gene expression values to then output a report **316** that can be viewed on a display **318** that includes estimates of the at least one biological sample for a degree of health. According to the current embodiment, the estimate includes comparing a scaling factor **a** for the at least one biological sample with other scaling factors **a'** in a database from previously analyzed biological samples, where the scaling factor **a** is calculated from the gene expression values using the computer program **314** by counting a number of link counts C_n for groups of an individual genes' expression values at a different times at a threshold value **C** or for groups of genes' expression values at a single time at the threshold value **C**, calculating an average number C_{ave} of the link counts C_n , calculating a largest number **M** of the C_n , where the **M** includes the largest of the number of link counts C_n for a given threshold value **C** for all the gene expression value groups, iteratively applying a relation $C_{ave}=M/\log(M)$ for different threshold values **C**, comparing the C_{ave} data values versus $M/\log(M)$ data, and applying a

10
15
20
25

fitting to the compared data to output the scaling factor **a**, where the scaling factor **a** is the slope of the fitting.

According to one embodiment of the system **300**, the at least one biological sample can
5 include saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, or organic material.

In another aspect of the system **300**, the gene expression reader includes at least two gene
10 probes.

In a further aspect of the system **300**, the number of link counts C_n includes a number of link counts for each of **N** expression value groups, where each expression value group includes a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value **C** between
15 the expression value group and the sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other **N-1** gene expression value groups.

According to another aspect of the system **300**, the a scaling factor **a** is calculated by iteratively applying $C_{ave} = M / \log(M)$ for different threshold values **C**, using the
20 appropriately programmed computer, and comparing C_{ave} values versus $M / \log(M)$ and calculating a linear fitting of the comparison to get the scaling factor **a**.

In yet another aspect of the system **300**, comparing values of **a** further includes comparing byproducts of the scaling factor **a**, comparing healthy samples against disease samples, or

comparing an unknown sample with a database of values from samples with a known condition.

In a further aspect of the system 300, the threshold value C is in a range between 0 and 1.

5

FIG. 4 shows another embodiment of the invention that includes lab-on-a-chip device 400 having a substrate 402 for holding a biological sample receptacle 404, a gene expression reader 406 and a microprocessor 408, where biological sample receptacle 404 includes a sample input 410 to the gene expression reader, where the gene expression reader outputs

10 412 gene expression values of at least two genes based on analyzed the at least one biological sample, where the microprocessor 408 includes a computer program 314 for analyzing gene expressions in the biological sample 304 input by the user 302 to the sample receptacle 404. The computer program 314 compiles the gene expression values, counts a number of link counts C_n for groups of an individual genes' expression values at different

15 times at a threshold value C or for groups of genes' expression values at a single time at the threshold value C , calculates an average number C_{ave} of the link counts C_n , calculates a largest number M of the C_n , where the M includes the largest of the number of link counts C_n for a given the threshold value C for all the gene expression value groups, iteratively applies a relation $C_{ave}=M/\log(M)$ for different threshold values C , compares data of the

20 C_{ave} values versus $M/\log(M)$ data, calculates a fitting to the compared data to output the scaling factor a , where the scaling factor a is the slope of the fitting, compares values of the scaling factor a for the at least one biological sample with other stored scaling factors a' from analyzed biological samples, and outputs a report 316, where the report 316 includes estimates of the at least one biological sample for a degree of health. The report can be

25 communicated to a computer 414 having computer software 416 and a display or printer

418. Further, it is understood that the substrate **402** can be any suitable platform, host or housing and that the computer **414** can be separate or integrated with the substrate **402**.

According to one aspect of the device **400**, the at least one biological sample can include
5 saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, or organic material.

In another aspect of the device **400**, the gene expression reader includes at least two gene
10 probes.

In a further aspect of the device **400**, the number of link counts C_n includes a number of link counts for each of N expression value groups, where each expression value group includes a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value C between
15 the expression value group and the sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other $N-1$ gene expression value groups.

According to one aspect of the device **400**, the a scaling factor a is calculated by iteratively applying the $C_{ave} = M / \log(M)$ for different threshold values C , using the appropriately
20 programmed computer, and comparing C_{ave} values versus $M / \log(M)$ and calculating a linear fitting the comparison to get the scaling factor a .

In a further aspect of the device **400**, comparing values of a further includes comparing byproducts of the scaling factor a , comparing healthy samples against disease samples, or

comparing an unknown sample with a database of values from samples with a known condition.

In yet aspect of the device 400, the threshold value **C** is in a range between 0 and 1.

5

The present invention has now been described in accordance with several exemplary embodiments, which are intended to be illustrative in all aspects, rather than restrictive. Thus, the present invention is capable of many variations in detailed implementation, which may be derived from the description contained herein by a person of ordinary skill in the art. For example, other complex interconnected networks where a single network component or node in the network can have the degree to which is it switched “on” quantified in a way similar to single gene expression values in a genetic network. Examples could include: numbers characterizing the total energy that each single protein in a protein-protein interaction network acquires from binding with other proteins in the network, other biochemical networks where the interaction between single components and other components can be similarly quantified for each component, numbers reflecting the flow of information to/from each single node in a communication or computer network, and numbers reflecting the flow of traffic through individual intersections in a city traffic network or between individual hubs in a transportation network.

20

All such variations are considered to be within the scope and spirit of the present invention as defined by the following claims and their legal equivalents.

CLAIMS

What is claimed:

1. A method of diagnosing a disease, comprising:
 - a. a gene expression reader analyzing at least one biological sample and outputting
5 gene expression values from at least two genes based on said analyzing said at least one biological sample;
 - b. calculating a scaling factor **a** for said at least one biological sample using an appropriately programmed computer, wherein said scaling factor **a** is calculated from said gene expression values comprising:
 - 10 i. counting a number of link counts C_n for groups of an individual genes' expression values at different times at a threshold value **C** or for groups of genes' expression values at a single time at said threshold value **C**;
 - ii. calculating an average number C_{ave} of said link counts C_n ;
 - iii. calculating a largest number **M** of said C_n , wherein said **M** comprises the
15 largest of said number of link counts C_n for a given said threshold value **C** for all said gene expression value groups;
 - iv. iteratively applying a relation $C_{ave}=M/\log(M)$ for different said threshold values **C**;
 - v. comparing data of said C_{ave} values versus $M/\log(M)$; and
 - 20 vi. calculating a fitting to said compared data to output said scaling factor **a**, wherein said scaling factor **a** is the slope of said fitting;

- c. comparing values of said scaling factor a for said at least one biological sample with other scaling factors a' in a database from analyzed biological samples using said appropriately programmed computer; and
- d. outputting a report using said appropriately programmed computer, wherein said report comprises estimates of said at least one biological sample for a degree of health.
2. The method of claim 1, wherein said at least one biological sample is selected from the group consisting of saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, and other organic material.
3. The method of claim 1, wherein said gene expression reader comprises at least two gene probes.
4. The method of claim 1, wherein said number of link counts C_n comprises a number of link counts for each of N expression value groups, wherein each said expression value group comprises a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value C between said expression value group and said sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other $N-1$ gene expression value groups.

- 5 5. The method of claim 1, wherein said scaling factor a is calculated by iteratively applying said $C_{ave}=M/\log(M)$ for different said threshold values C , using said appropriately programmed computer, and comparing C_{ave} values versus $M/\log(M)$ and calculating a linear fitting of said comparison to get said scaling factor a .
- 10 6. The method of claim 1, wherein said comparing values of said a further comprises comparing byproducts of said scaling factor a , comparing healthy samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.
7. The method of claim 1, wherein said threshold value C is in a range between 0 and 1.
- 15 8. A system for diagnosing disease, comprising:
- a. a gene expression reader for analyzing at least one biological sample and outputting gene expression values of at least two genes;
- b. a computer server for receiving from said gene expression reader said gene expression values and for managing and communicating patient information to a
- 20 user; and
- c. a computer program hosted on said computer server, wherein said computer program analyzes said gene expression values and outputs a report, wherein said report comprises estimates of said at least one biological sample for a degree of

health, wherein said estimate comprises comparing a scaling factor **a** for said at least one biological sample with other scaling factors **a'** in a database from previously analyzed biological samples, wherein said scaling factor **a** is calculated from said gene expression values using said computer program comprising:

- i. counting a number of link counts C_n for groups of an individual genes' expression values at a different times at a threshold value **C** or for groups of genes' expression values at a single time at said threshold value **C**;
- ii. calculating an average number C_{ave} of said link counts C_n ;
- 10 iii. calculating a largest number **M** of said C_n , wherein said **M** comprises the largest of said number of link counts C_n for a given said threshold value **C** for all said gene expression value groups;
- iv. iteratively applying a relation $C_{ave}=M/\log(M)$ for different said threshold values **C**;
- 15 v. comparing said C_{ave} data values versus $M/\log(M)$ data; and
- vi. applying a fitting to said compared data to output said scaling factor **a**, wherein said scaling factor **a** is the slope of said fitting.

9. The system of claim 8, wherein said at least one biological sample is selected from the group consisting of saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, and organic material.

10. The system of claim 8, wherein said gene expression reader comprises at least two gene probes.

5 11. The system of claim 8, wherein said number of link counts C_n comprises a number of link counts for each of N expression value groups, wherein each said expression value group comprises a sequence of gene expression values $n_1, n_2, \dots n_T$, at a threshold value C between said expression value group and said sequence of gene expression values $n_1, n_2, \dots n_T$ for the other $N-1$ gene
10 expression value groups.

12. The system of claim 8, wherein said a scaling factor a is calculated by iteratively applying said $C_{ave}=M/\log(M)$ for different said threshold values C , using said appropriately programmed computer, and comparing C_{ave} values versus
15 $M/\log(M)$ and calculating a linear fitting of said comparison to get said scaling factor a .

13. The system of claim 8, wherein said comparing values of said a further comprises comparing byproducts of said scaling factor a , comparing healthy
20 samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.

14. The system of claim 8, wherein said threshold value C is in a range between 0 and 1.

15. A lab-on-a-chip device, comprising:

- 5 a. a substrate for holding a biological sample receptacle, a gene expression reader and a microprocessor, wherein said biological sample receptacle comprises a sample input to said gene expression reader, wherein said gene expression reader outputs gene expression values of at least two genes based on analyzing at least one biological sample, wherein said microprocessor comprises a computer program for analyzing gene expressions in said at least one biological sample, wherein said computer program:
- 10 i. compiles said gene expression values;
- ii. counts a number of link counts C_n for groups of an individual genes' expression values at different times at a threshold value C or for groups of genes' expression values at a single time at said threshold value C ;
- iii. calculates an average number C_{ave} of said link counts C_n ;
- 15 iv. calculates a largest number M of said C_n , wherein said M comprises the largest of said number of link counts C_n for a given said threshold value C for all said gene expression value groups;
- i. iteratively applies a relation $C_{ave}=M/\log(M)$ for different said threshold values C ;
- 20 ii. compares data of said C_{ave} values versus $M/\log(M)$ data;
- iii. calculates a fitting to said compared data to output said scaling factor a , wherein said scaling factor a is the slope of said fitting;

- iv. compares values of said scaling factor **a** for said at least one biological sample with other stored scaling factors **a'** from analyzed biological samples; and
- v. outputs a report, wherein said report comprises estimates of said at least one biological sample for a degree of health.

5

16. The device of claim 15, wherein said at least one biological sample is selected from the group consisting of saliva, urine, other body fluids, synovial fluid, breast ductal fluid, blood and blood components, tissue, tumors, bone marrow, stem cells, induced pluripotent cells, cell lines, plant material, and organic material.

10

17. The device of claim 15, wherein said gene expression reader comprises at least two gene probes.

15

18. The device of claim 15, wherein said number of link counts C_n comprises a number of link counts for each of N expression value groups, wherein each said expression value group comprises a sequence of gene expression values $n_1, n_2, \dots, n_T$, at a threshold value C between said expression value group and said sequence of gene expression values $n_1, n_2, \dots, n_T$ for the other $N-1$ gene expression value groups.

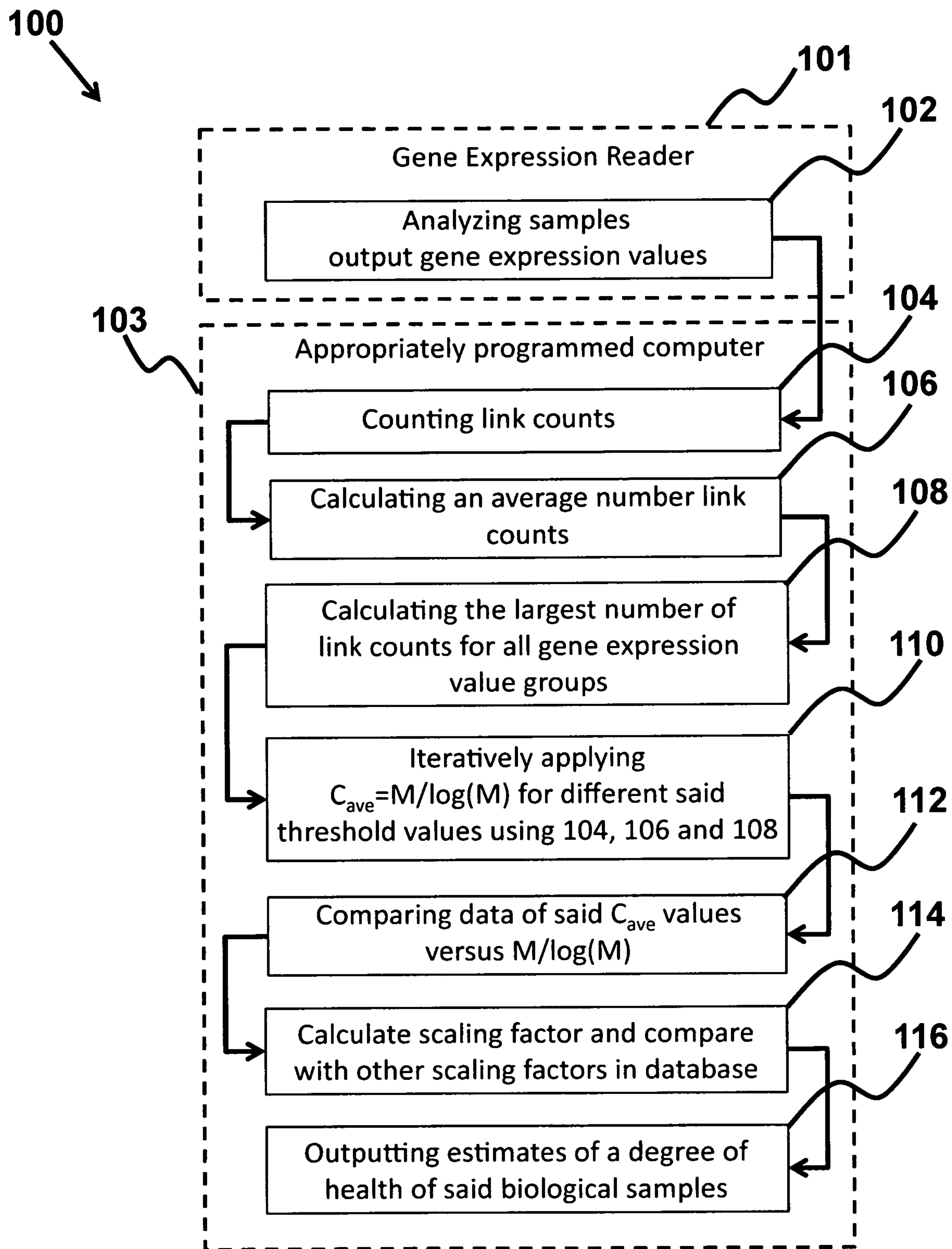
20

19. The device of claim 15, wherein said a scaling factor **a** is calculated by iteratively applying said $C_{ave} = M / \log(M)$ for different said threshold values **C**, using said appropriately programmed computer, and comparing C_{ave} values versus

$M/\log(M)$ and calculating a linear fitting said comparison to get said scaling factor a .

5 20. The device of claim 15, wherein said comparing values of said a further comprises comparing byproducts of said scaling factor a , comparing healthy samples against disease samples, or comparing an unknown sample with a database of values from samples with a known condition.

10 21. The device of claim 15, wherein said threshold value C is in a range between 0 and 1.

**FIG. 1**

200

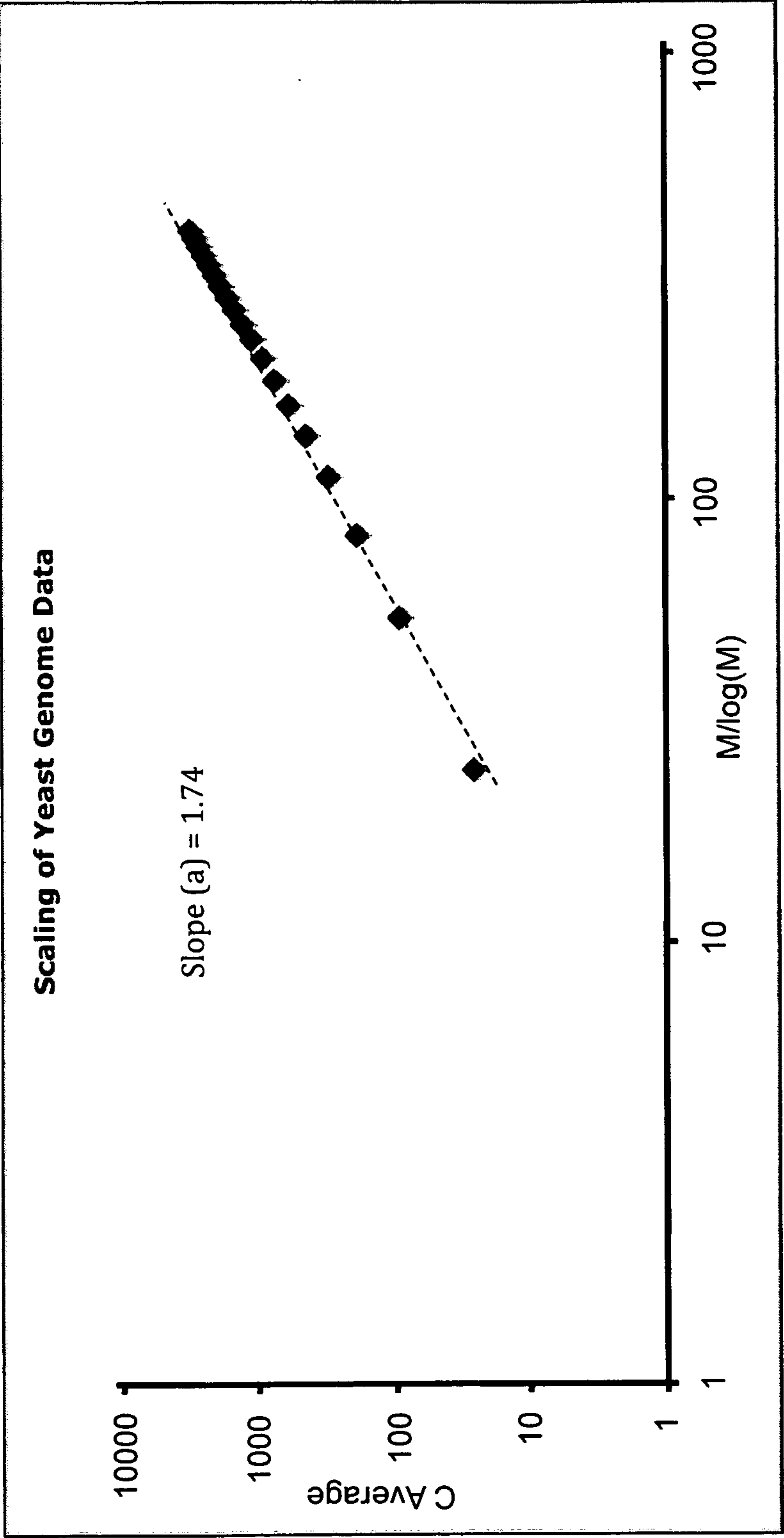
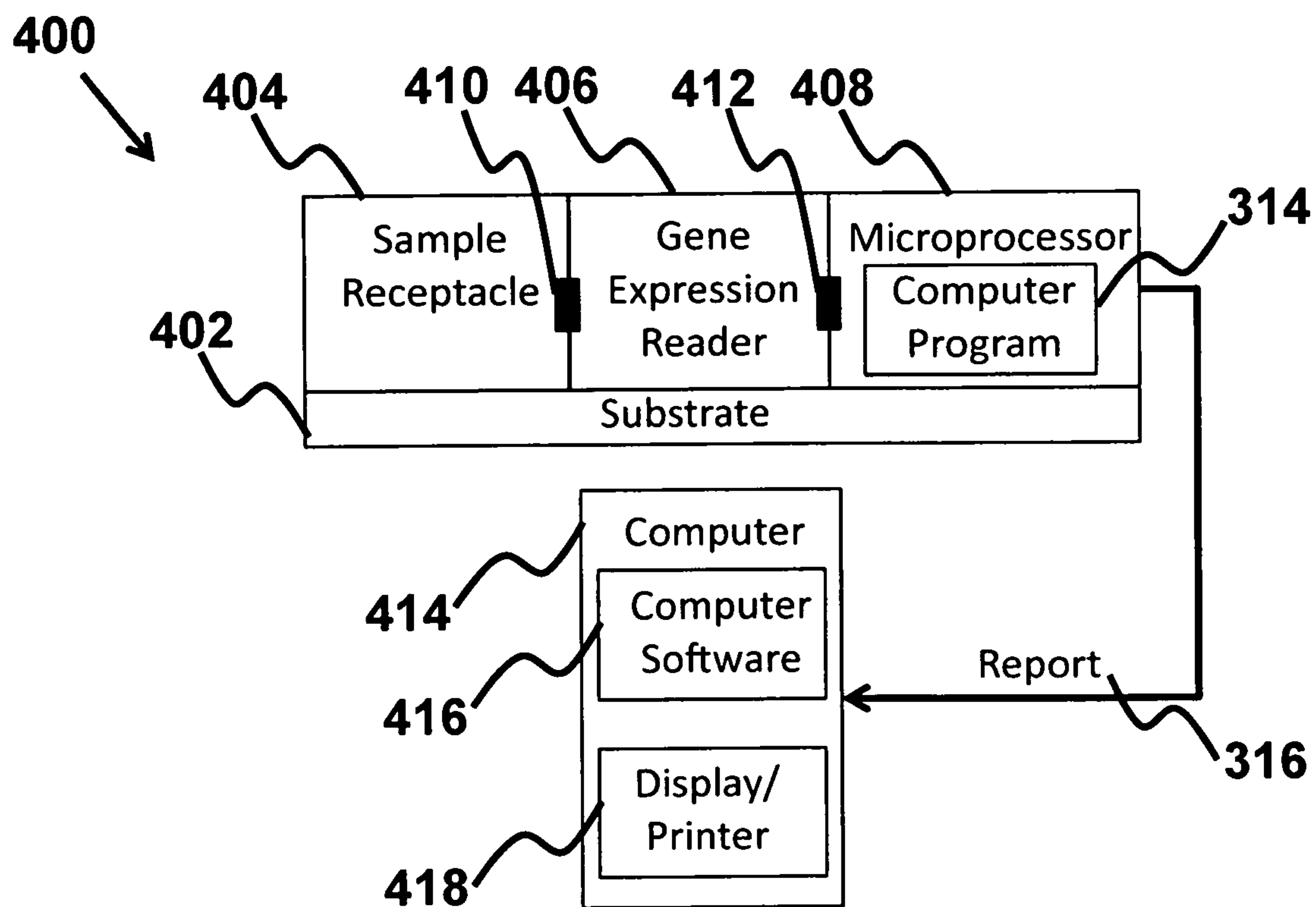
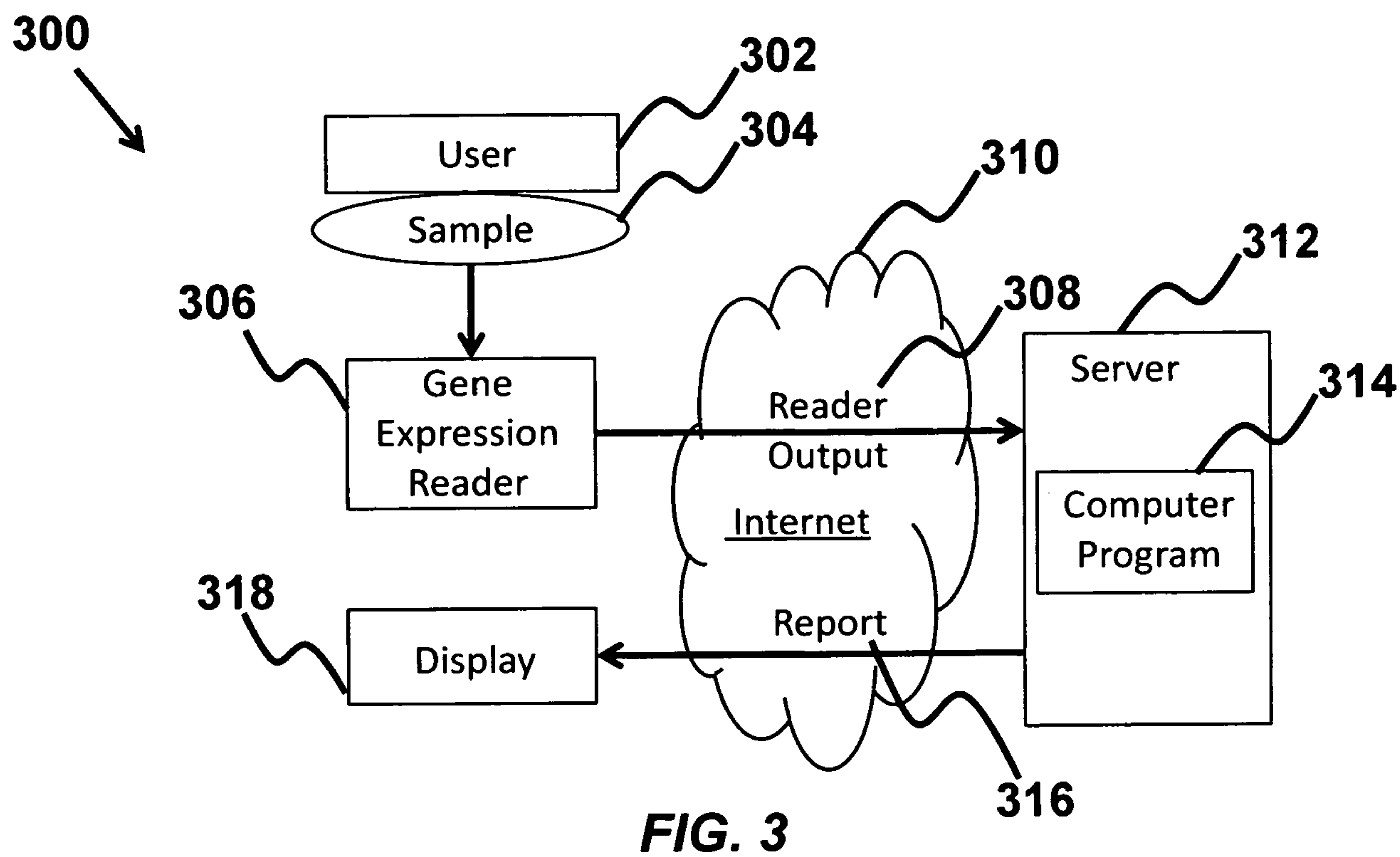


FIG. 2



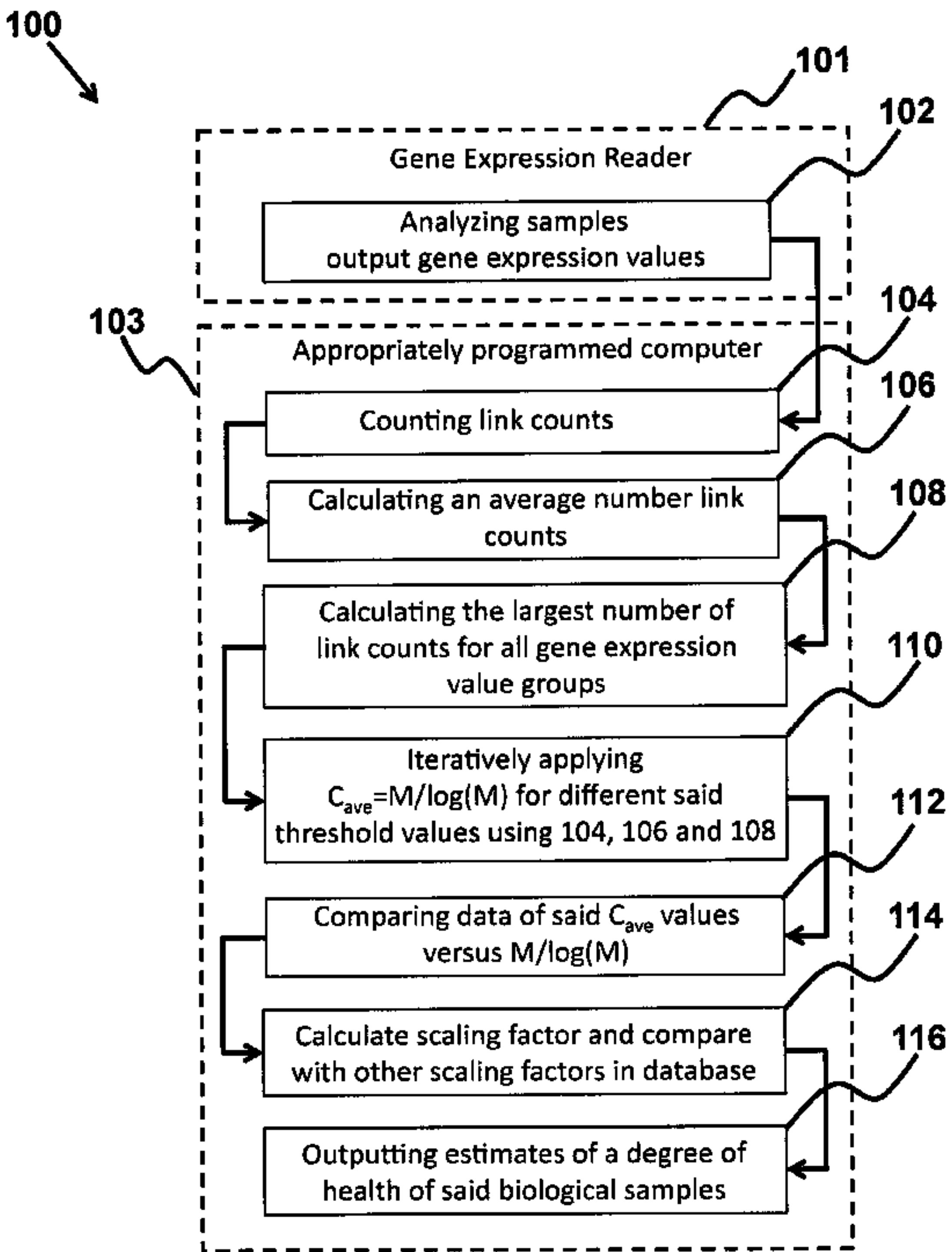


FIG. 1