

(51) International Patent Classification:
G06F 17/30 (2006.01) **C12Q 1/68** (2006.01)(21) International Application Number:
PCT/KR2008/004735(22) International Filing Date:
14 August 2008 (14.08.2008)

(25) Filing Language: Korean

(26) Publication Language: English

(71) Applicant (for all designated States except US): **KOREA BASIC SCIENCE INSTITUTE** [KR/KR]; 52 Eoeundong Yuseong-gu, Daejeon-shi 305-333 (KR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KWON, Kyung-Hoon** [KR/KR]; #106-1505 Hanareum Apt. Wolpyung-2dong Seo-gu, Daejeon-shi 302-753 (KR). **PARK, Gun Wook** [KR/KR]; #107-206 Sangkoksu Apt. Manbyeon-dong Seo-gu, Daejeon-shi 302-781 (KR). **LEE, Jeong Hwa** [KR/KR]; #105-1509 Lottecastle Venezia 2545, Hwanghak-dong Chung-gu, Seoul 100-440 (KR). **YOO, Jong Shin** [KR/KR]; #1-701 GaepoLG Apt. Dokok-dong Kangnam-gu, Seoul 135-270 (KR).(74) Agent: **LEE, Won-Hee**; 8th Fl. Sung-ji Heights II 642-16, Yoksam-dong Kangnam-ku, Seoul 135-080 (KR).

(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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

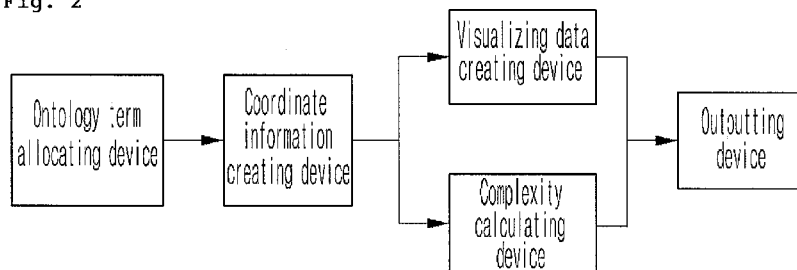
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, 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:

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

(54) Title: APPARATUS FOR VISUALIZING AND ANALYZING GENE EXPRESSION PATTERNS USING GENE ONTOLOGY TREE AND METHOD THEREOF

Fig. 2



(57) Abstract: The present invention relates to an apparatus for visualizing or analyzing gene expression patterns of a biological sample using gene ontology tree and a method of the same, more precisely, an apparatus for visualizing or analyzing gene expression patterns which are regarded as biologically important using gene ontology tree by calculating complexity and a method of the same. This apparatus of the present invention is useful for comparing gene expression patterns in different biological samples using data produced by the said apparatus and the method of the present invention. According to the present invention, gene expression patterns in relation to molecular functions, biological processes or cellular components can be analyzed by investigating protein or RNA expression in biological samples. So, the apparatus and the method of the present invention can be effectively used for analysis of biologically important aspects such as functional changes, evolutionary stages or developmental stages of each biological sample.

【DESCRIPTION】**【Invention Title】**

APPARATUS FOR VISUALIZING AND ANALYZING GENE
EXPRESSION PATTERNS USING GENE ONTOLOGY TREE AND METHOD
5 THEREOF

【Technical Field】

The present invention relates to a method for
analyzing gene expression patterns of biological samples,
10 more precisely an apparatus for visualizing and analyzing
gene expression patterns of biological samples using gene
ontology tree and a method thereof.

【Background Art】

15 Proteome is a combined word of protein and ome,
which is the integrative term for whole proteins. In
general, a cell of a unicellular organism has one type
proteome, while each cell of a multicellular organism has
same genome but different types of proteomes. That is, in
20 a multicellular organism, genome which proteome is
originated from is all equal but aspects of proteome shown
in a specific cell or under specific condition are all
different. Proteomics is the study to identify proteins,
expression levels of proteins, transformation and
25 intracellular locations of proteins, and interactions

between proteins. By proteomics, proteins expressed in cells can be identified and network among these proteins can be disclosed, suggesting that proteomics provides the explanations on all the biological phenomena from genome
5 to protein.

There are two main techniques for the study of proteomics. First is the technique to separate total protein groups from a cell and then regrouping them to
10 multiple structural proteins. Second is the technique to analyze those separated proteins. 2-dimensional electrophoresis most widely used recently is 2 dimensional polyacrylamide gel electrophoresis (2D-PAGE) separating proteins by using isoelectric point (pI) and molecular
15 weight (MW). To identify proteins separated by 2D-PAGE, scientists additionally perform mass analysis with each protein using ultra high speed mass spectrometry (MS). 2D-PAGE proteome database established by Julio Celis et al provides every details of each protein with explanations
20 made by using amino acid sequence, peptide mass fingerprinting (PMF), pI and molecular weight (MW). Based on the studies screening the general information about proteins and investigating expressions, transformations and intracellular locations of proteins as well as
25 interactions among proteins, total proteins expressed in

cells and the network connecting them are disclosed, which paves the way to explanation on the whole biological phenomena from genome to protein.

5 In particular, analysis of expression profile is to investigate general protein expression patterns according to experimental conditions. Precisely, it is originated from gene microarray technique analyzing genes harboring necessary information by integrating on a chip. This
10 analysis facilitates gene expression investigation in a large scale by various statistical analysis methods, overcoming the limit of the conventional method enabling gene expression analysis of only one or two genes at a time.

15

 Profile analysis has been tried in many different ways to analyze protein expression accurately and to overcome the problem of inconsistency between actual mRNA expression and protein expression. For example, protein
20 profile analysis was tried using protein spot intensity information obtained from 2D-PAGE.

 However, protein expression analysis by the conventional proteomics or RNA analysis by microarray
25 could only provide massive information on gene expression

as a whole or a specific target gene expression. So, it was not possible to compare general gene expression quantitatively or to compare expressions of different proteomes by the difference of their functions or to
5 compare expressions over cellular components.

According to the conventional studies on evolution, genes have been generally classified over phenotypes. But with the recent advancement of molecular biology and
10 discovery of gene map, evolution of species is determined based on the similarity of genotypes, which is similarity of gene sequences. Gene expression analysis was not provided for evolution studies in early days. Although gene sequences were different, if the expressed gene
15 sequences or functions were similar, suggesting that evolution or development proceeded similarly, genotypes and phenotypes of those genes had to be analyzed together, which was not facilitated by the conventional method. Therefore, the conventional analysis techniques could not
20 provide information on functional changes according to gene expression in evolutionary relationships or developmental stages.

The term gene ontology was introduced by Gene
25 Ontology Consortium. Ontology herein indicates a system

classifying biological terms or vocabularies. Gene
Ontology Classification System Consortium was established
for standardization of biological terms. To explain
functions of genes in all the species, controlled
vocabularies applied in common were introduced. Again,
gene ontology is a classification system to investigate
relationship among genes or among key words of each gene,
which can be applied to bioinformatics approach.

In gene ontology, terms form a tree structure, in
which they are related hierarchically. Total terms are
classified into three categories, and approximately 10,000
terms form a tree structure with forming hierarchical
relationship. According to gene ontology, genetic
functions are classified into three categories, molecular
function, biological process, and cellular component. And,
controlled vocabulary is established hierarchical in each
category. These categories are not exclusive and only
divided by characteristics to describe a gene.

The present inventors tried to develop an analysis
method facilitating not only simple gene expression
analysis in a biological sample but also biologically
important functional analysis of the biological sample.
As a result, the present inventors completed this

invention by developing a method for analyzing expression distributions of whole genes and the distributions of functions and cellular components of biological samples by introducing gene ontology concept.

5 The present invention provides a novel method for identifying evolutionary relationship or cell developmental stages of biospecies.

【Disclosure】

10 **【Technical Problem】**

 It is an object of the present invention to provide an apparatus for visualizing or analyzing gene expression patterns by taking advantage of gene ontology with the results of proteome analysis or RNA expression analysis,
15 in order to analyze functional changes, evolutionary relationships or developmental stages of biological samples and a method thereof.

【Technical Solution】

20 Terms used in this invention are described as follows.

 Gene Ontology indicates a system classifying biological terms or vocabularies provided by Gene Ontology Consortium. Terms classified according to the system are
25 called 'ontology terms'. These ontology terms build a

tree structure together, in which they are related vertically. These terms are divided into three categories, which are molecular function, biological process and cellular component.

5 Ontology terms used in this invention include not only ontology terms themselves to describe a specific gene but also gene ontology codes corresponding to the ontology terms. Gene ontology code herein indicates a code pre-set up on gene ontology database corresponding to a specific
10 gene ontology term.

 Gene ontology tree indicates a classification system dividing gene ontology terms hierarchically, precisely a tree structure composed of branches connecting nodes of gene ontology terms.

15 Hereinafter, the present invention is described in detail.

 To achieve the above object, the present invention
20 provides an apparatus for visualizing gene expression patterns of a biological sample using gene ontology tree comprising the following devices:

 gene ontology term allocating device which receives the results of the analysis of protein or RNA expression
25 of a biological sample and then applies the results to

biological DB or gene ontology database to allocate gene ontology terms corresponding to the protein or RNA;

coordinate creating device which obtains coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

visualizing device which generates visualizing data on the coordinate information obtained by the above coordinate creating device; and

outputting device which outputs the visualizing data.

The present invention also provides an apparatus for analyzing gene expression patterns of a biological sample using gene ontology tree comprising the following devices:

gene ontology term allocating device which receives the results of the analysis of protein or RNA expression of a biological sample and then applies the results to biological DB or gene ontology database to allocate gene ontology terms corresponding to the protein or RNA;

coordinate creating device which obtains coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

complexity calculating device which calculates complexity data on the coordinate information obtained by

the above coordinate creating device; and

outputting device which outputs the complexity data.

The present invention also provides an apparatus for
5 comparing gene expression patterns of a biological sample
using gene ontology tree comprising the following devices:

gene ontology term allocating device which receives
the results of the analysis of protein or RNA expression
of a biological sample and then applies the results to
10 biological DB or gene ontology database to allocate gene
ontology terms corresponding to the protein or RNA;

coordinate creating device which obtains coordinate
information of the allocated gene ontology terms
corresponding to gene ontology tree classified in gene
15 ontology database;

comparing device which obtains different coordinate
information by comparing the coordinate information
obtained by the above coordinate creating device and the
coordinate information obtained from another sample;

20 visualizing device which generates visualizing data
on the coordinate information obtained by the comparing
device; and

outputting device which outputs the visualizing data.

25 The present invention also provides a method for

visualizing gene expression patterns of a biological sample using gene ontology tree comprising the following steps:

5 (a) inputting the results of analysis of protein or RNA expression obtained from a sample through computer input system (step 1);

(b) allocating ontology terms corresponding to the input protein or RNA by gene ontology term allocating device (step 2);

10 (c) obtaining coordinate information of the gene ontology terms allocated in step 2 corresponding to gene ontology tree classified in gene ontology database by coordinate creating device (step 3);

15 (d) generating visualizing data of the coordinate information obtained in step 3 by visualizing device (step 4); and

(e) outputting the visualizing data generated in step 4 by computer output system (step 5).

20 The present invention also provides a method for analyzing gene expression patterns of a biological sample using gene ontology tree comprising the following steps:

25 (a) inputting the results of analysis of protein or RNA expression obtained from a sample through computer input system (step 1);

(b) allocating ontology terms corresponding to the input protein or RNA by gene ontology term allocating device (step 2);

5 (c) obtaining coordinate information of the gene ontology terms allocated in step 2 corresponding to gene ontology tree classified in gene ontology database by coordinate creating device (step 3);

10 (d) producing complexity data represented by the following mathematical formula 1 and/or formula 2 on the coordinate information obtained in step 3 by complexity calculating device (step 4); and

(e) outputting the complexity data produced in step 4 by computer output system (step 5).

15 Hereinafter, the present invention is described in more detail.

The present invention provides an apparatus for visualizing gene expression patterns of a biological sample using gene ontology tree comprising the following devices:

20

a) gene ontology term allocating device which receives the results of the analysis of protein or RNA expression of a biological sample and then applies the results to biological DB or gene ontology database to

25

allocate gene ontology terms corresponding to the protein or RNA;

b) coordinate creating device which obtains coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

c) visualizing device which generates visualizing data on the coordinate information obtained by the above coordinate creating device; and

d) outputting device which outputs the visualizing data.

According to the apparatus for visualizing gene expression patterns of the present invention, differences of genes in molecular function, biological process, and cellular component are visualized, leading to general understanding on the gene expression pattern.

In the apparatus for visualizing gene expression patterns, the protein analysis result of step a) is preferably obtained by proteome analysis, but not always limited thereto. The proteome analysis herein is performed preferably by 2-dimensional electrophoresis or mass spectrometry, but not always limited thereto.

In the apparatus for visualizing gene expression patterns, the result of RNA expression analysis in step a)

is preferably performed by microarray analysis, but not always limited thereto.

In the apparatus for visualizing gene expression patterns, the gene ontology term allocating device
5 determines what term among gene ontology terms defining a gene function has to be allocated to the expressed protein or RNA and executes the allocation. If a gene is multi-functional, multiple gene ontology terms can be allocated.

The gene ontology term allocating device identifies
10 a specific gene from biological DB through network and finds out a corresponding term to the gene. In a preferred embodiment of the present invention, accessible biological DB through the network is exemplified by Unigene, LocusLink, Swiss-Prot, MGI, UniProt, EMBL and IPI,
15 but not always limited thereto. Most of the DBs above provide gene ontology terms related to gene functions and if not, they are still able to allocate gene ontology term corresponding to a specific gene using gene ontology database based on the genetic information provided by the
20 above biological databases. In a preferred embodiment of the present invention, gene ontology database accessible through the network is exemplified by GO (Gene Ontology), ChEBI (Chemical Entitles Biological Interest), GOA and NEW, but not always limited thereto.

25 In a preferred embodiment of the present invention,

the apparatus for visualizing gene expression patterns facilitates gene ontology term allocation by gene identification algorithm identifying an expressed gene by screening it through biological DB based on the results of protein or RNA expression analysis and by allocation algorithm allocating gene ontology term corresponding to the identified gene by screening it through gene ontology database.

The gene identification algorithm contains a protein or gene screening tool operable in the network such as Blast and FASTA provided by EBI (European Bioinformatics Institute), PIR (Georgetown University) or ExpASY (Swiss institute of Bioinformatics), but not always limited thereto.

The gene ontology term allocation algorithm contains a gene ontology term screening tool operable in the network such as AmiGO, MGI GO Browser or Ontology Lookup Service, but not always limited thereto.

In a preferred embodiment of the present invention, the coordinate creating device is composed of transform algorithm converting the gene ontology tree classified in gene ontology database into coordinates and coordinate information collection algorithm collecting information of coordinates corresponding to the allocated ontology terms

among the converted gene ontology tree coordinates. The coordinate creating device has functions of matching the ontology term corresponding to the expressed gene to the gene ontology tree classified in gene ontology database.

5 In the ontology term allocating device, if a gene is allocated with multiple terms, multiple coordinates can be obtained.

The gene ontology tree has a tree structure stretching branches of gene ontology terms which are connected with one another by the coordinates (nodes) corresponding to gene ontology terms and the coordinates (nodes) themselves. Figure 1 illustrates an example of gene ontology tree. In the structure of the gene ontology tree shown in Figure 1, the topmost level is the whole gene ontology and the second highest level consists of molecular functions, biological process and cellular components, and levels 3, 4 and 5 are lower levels each forming a tree. As lower goes the level, ontology terms for genes having detailed functions can be provided.

20 In another preferred embodiment of the present invention, the ontology tree contains one of molecular functions, biological processes and cellular components as the topmost level or takes its lower concept as the topmost level.

25 The gene ontology tree can be provided by gene

ontology database or by ontology analysis tool. In a preferred embodiment of the present invention, the gene ontology database and analysis tool are provided by the network at the below internet address, but not always limited thereto.

<http://www.geneontology.org>.

In another preferred embodiment of the present invention, in addition to the database accessible through the above network, gene ontology tree consisting of gene ontology terms can be obtained by using the database established by the present inventors or screening tool.

In a preferred embodiment of the present invention, the coordinate information is preferably 2-dimensional or 3-dimensional coordinate information, but not always limited thereto.

The coordinate information herein indicates the coordinate information for gene ontology term corresponding to a target gene obtained from gene ontology tree. In a preferred embodiment of the present invention, the coordinate information is composed of the information on the shortest path connecting the upmost node of gene ontology tree and the target node, level of the node, location of a branch on the passway among branches stretched from each level, numbers of branches stretched from each level and numbers of genes corresponding thereto,

but not always limited thereto.

In this invention, the visualizing device is composed of algorithm locating coordinates of nodes (gene ontology terms) obtained by the coordinate creating device and visualization algorithm visualizing vertical relationships of those coordinates obtained by a tree structure. The visualizing device facilitates visualization of coordinate data of gene ontology terms by a tree structure.

In a preferred embodiment of the present invention, the outputting device is preferably monitor, printer or plotter, but not always limited thereto.

15

The present invention also provides an apparatus for analyzing gene expression patterns of a biological sample using gene ontology tree comprising the following devices:

a) gene ontology term allocating device which receives the results of the analysis of protein or RNA expression of a biological sample and then applies the results to biological DB or gene ontology database to allocate gene ontology terms corresponding to the protein or RNA;

b) coordinate creating device which obtains

25

coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

5 c) complexity calculating device which calculates complexity data on the coordinate information obtained by the above coordinate creating device; and

d) outputting device which outputs the complexity data.

10 In the apparatus for analyzing gene expression patterns, the protein analysis result of step a) is preferably obtained by proteome analysis, but not always limited thereto. The proteome analysis herein is performed preferably by 2-dimensional electrophoresis or
15 mass spectrometry, but not always limited thereto.

In the apparatus for analyzing gene expression patterns, the result of RNA expression analysis in step a) is preferably performed by microarray analysis, but not always limited thereto.

20 In the apparatus for analyzing gene expression patterns, the gene ontology term allocating device determines what term among gene ontology terms defining a gene function has to be allocated to the expressed protein or RNA and executes the allocation. If a gene is multi-
25 functional, multiple gene ontology terms can be allocated.

The gene ontology term allocating device identifies a specific gene from biological DB through network and finds out a corresponding term to the gene. In a preferred embodiment of the present invention, accessible biological DB through the network is exemplified by Unigene, LocusLink, Swiss-Prot, MGI, UniProt, EMBL and IPI, but not always limited thereto. Most of the DBs above provide gene ontology terms related to gene functions and if not, they are still able to allocate gene ontology term corresponding to a specific gene using gene ontology database based on the genetic information provided by the above biological databases. In a preferred embodiment of the present invention, gene ontology database accessible through the network is exemplified by GO (Gene Ontology), ChEBI (Chemical Entitles Biological Interest), GOA and NEW, but not always limited thereto.

In a preferred embodiment of the present invention, the apparatus for analyzing gene expression patterns facilitates gene ontology term allocation by gene identification algorithm identifying an expressed gene by screening it through biological DB based on the results of protein or RNA expression analysis and by allocation algorithm allocating gene ontology term corresponding to the identified gene by screening it through gene ontology database.

The gene identification algorithm contains a protein or gene screening tool operable in the network such as Blast and FASTA provided by EBI (European Bioinformatics Institute), PIR (Georgetown University) or ExpASy(Swiss
5 institute of Bioinformatics), but not always limited thereto.

The gene ontology term allocation algorithm contains a gene ontology term screening tool operable in the network such as AmiGO, MGI GO Browser or Ontology Lookup
10 Service, but not always limited thereto.

In a preferred embodiment of the present invention, the coordinate creating device is composed of transform algorithm converting the gene ontology tree classified in
15 gene ontology database into coordinates and coordinate information collection algorithm collecting information of coordinates corresponding to the allocated ontology terms among the converted gene ontology tree coordinates. The coordinate creating device has functions of matching the
20 ontology term corresponding to the expressed gene to the gene ontology tree classified in gene ontology database. In the ontology term allocating device, if a gene is allocated with multiple terms, multiple coordinates can be obtained.

25 The gene ontology tree has a tree structure

stretching branches of gene ontology terms which are connected with one another by the coordinates (nodes) corresponding to gene ontology terms and the coordinates (nodes) themselves. Figure 1 illustrates an example of gene ontology tree. In the structure of the gene ontology tree shown in Figure 1, the topmost level is the whole gene ontology and the second highest level consists of molecular functions, biological process and cellular components, and levels 3, 4 and 5 are lower levels each forming a tree. As lower goes the level, ontology terms for genes having detailed functions can be provided.

In another preferred embodiment of the present invention, the ontology tree contains one of molecular functions, biological processes and cellular components as the topmost level or takes its lower concept as the topmost level.

The gene ontology tree can be provided by gene ontology database or by ontology analysis tool. In a preferred embodiment of the present invention, the gene ontology database and analysis tool are provided by the network at the below internet address, but not always limited thereto.

<http://www.geneontology.org>.

In another preferred embodiment of the present invention, in addition to the database accessible through

the above network, gene ontology tree consisting of gene ontology terms can be obtained by using the database established by the present inventors or screening tool.

5 In a preferred embodiment of the present invention, the coordinate information is preferably 2-dimensional or 3-dimensional coordinate information, but not always limited thereto.

10 The coordinate information herein indicates the coordinate information for gene ontology term corresponding to a target gene obtained from gene ontology tree. In a preferred embodiment of the present invention, the coordinate information is composed of the information on the shortest path connecting the upmost node of gene ontology tree and the target node, level of the node, 15 location of a branch on the passway among branches stretched from each level, numbers of branches stretched from each level and numbers of genes corresponding thereto, but not always limited thereto.

20 In the apparatus for analyzing gene expression patterns, the complexity calculating device is composed of computer arithmetic algorithm calculating complexity defined by the following mathematical formula 1, but not always limited thereto.

25 **【Mathematical formula 1】**

$$\text{Complexity}(S) = \sum_{\alpha=1}^N [-\log(P_{\alpha})]/N$$

N : number of coordinates of genes corresponding to gene ontology tree

P_{α} : probability of selection of the specific gene

5 coordinate α in the gene ontology tree, P_{α} is defined by the following mathematical formula 2.

【Mathematical Formula 2】

$$P_{\alpha} = \prod_{j=1}^K 1/n_{\alpha j}$$

10 $n_{\alpha j}$

Figure 4 is a diagram illustrating an example of calculating complexity from the coordinate information distributed on gene ontology tree.

The coordinate information distributed on the gene ontology tree is represented by identification marks, for
 15 example, 401, 402, 403 and 404. The number of coordinates (N) is 4 and probability of selection of each coordinate (P) is calculated as follows: for example in the case of

identification mark 401, the first coordinate among those on the shortest path passing over 401 stretches three branches and the second coordinate also stretches three branches, so $P_{401} = 1/3 * 1/3$, which is $1/9$. This formula
5 can be applied in other coordinates and as a result, probabilities of 402, 403 and 404 are $1/9$, $1/3$ and $1/3$ respectively. So, complexity S is $-0.25 * (\log(1/9) + \log(1/9) + \log(1/3) + \log(1/3)) = 1.5 * \log 3 = 0.71$.

Complexity S increases as probability P_i reduces.
10 Small P_i indicates that the coordinate is located far from the starting point of ontology tree. So, as coordinates of genes are located far from the starting point of gene ontology tree, their complexities increase.

Quantitative analysis of gene expression patterns of
15 a biological sample can be facilitated by calculating complexity. The complexity of a specific biological sample differs from the kinds and numbers of proteins and differentiation stages thereof. Complexity is increased when expressed proteins are diverse and differentiated
20 proteins are dominant in a sample.

The apparatus for analyzing gene expression patterns of the present invention facilitates digitization of gene expression of a biological sample by using complexity and more specifically digitization of gene expression
25 according to molecular functions, biological processes and

cellular components. Therefore, the apparatus enables the examination of general expression patterns and cell functions as well as identification of developmental stages.

5 In a preferred embodiment of the present invention, the outputting device is preferably monitor, printer or plotter, but not always limited thereto.

10 The present invention also provides an apparatus for comparing gene expression patterns of a biological sample using gene ontology tree comprising the following devices:

 gene ontology term allocating device which receives the results of the analysis of protein or RNA expression of a biological sample and then applies the results to biological DB or gene ontology database to allocate gene ontology terms corresponding to the protein or RNA;

15 coordinate creating device which obtains coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

20 comparing device which obtains different coordinate information by comparing the coordinate information obtained by the above coordinate creating device and the coordinate information obtained from another sample;

25 visualizing device which generates visualizing data

on the coordinate information obtained by the comparing device; and

outputting device which outputs the visualizing data.

The apparatus for comparing gene expression patterns facilitates visualization of information on gene expressions in different biological samples. That is, owing to this apparatus, differences in major cell functions particularly molecular functions, biological processes and cellular components among different samples can be explained. Again, differences in functions and developmental stages among different samples can be screened by this apparatus.

In the study of evolution, it is only explained how far evolution has been progressed by comparing gene sequences without using the apparatus for comparing gene expression patterns of the present invention. Therefore, it is very difficult to compare evolutionary stages between individuals having same or similar gene sequences. However, according to the present invention, gene expression patterns can be analyzed and compared, suggesting that evolutionary stages can be compared more precisely, if individuals have different gene expression patterns despite they have the same or similar gene sequences.

In a preferred embodiment of the present invention,

the comparing device of the apparatus for comparing gene expression patterns consists of comparison algorithm comparing the coordinate information obtained by the coordinate creating device and the coordinate information
5 obtained from other samples to identify their locations and coordinate producing algorithm collecting only those coordinates having different locations by eliminating coordinate information having the same locations, but not always limited thereto.

10 In another preferred embodiment of the present invention, the comparing device can additionally include selection algorithm producing coordinate information corresponding to gene ontology term included in specific hierarchical classification which is operated before
15 applying the comparison algorithm. The apparatus for comparing gene expression patterns comprising the comparing device including the additional selection algorithm facilitates the comparison of not only the whole expressions of samples but also a specific function based
20 expression of a gene among different samples.

The present invention also provides a method for visualizing gene expression patterns of a biological sample using gene ontology tree comprising the following
25 steps using the above visualizing device:

(a) inputting the results of analysis of protein or RNA expression obtained from a sample through computer input system (step 1);

5 (b) allocating ontology terms corresponding to the input protein or RNA by gene ontology term allocating device (step 2);

10 (c) obtaining coordinate information of the gene ontology terms allocated in step 2 corresponding to gene ontology tree classified in gene ontology database by coordinate creating device (step 3);

(d) producing visualizing data of coordinate information obtained in step 3 by visualizing device (step 4); and

15 (e) outputting the visualizing data produced in step 4 by computer output system (step 5).

In this method, the input system is preferably key board, scanner, barcode reader, mouse, tablet, track ball, electronic pen or digital camera, but not always limited thereto.

20 In the above method, the output system is preferably monitor, printer or plotter, but not always limited thereto.

25 In addition, the present invention provides a method for analyzing gene expression patterns of a biological

sample using gene ontology tree comprising the following steps:

(a) inputting the results of analysis of protein or RNA expression obtained from a sample through computer input system (step 1);

(b) allocating ontology terms corresponding to the input protein or RNA by gene ontology term allocating device (step 2);

(c) obtaining coordinate information of the gene ontology terms allocated in step 2 corresponding to gene ontology tree classified in gene ontology database by coordinate creating device (step 3);

(d) producing complexity data represented by the following mathematical formula 1 on the coordinate information obtained in step 3 by complexity calculating device (step 4); and

(e) outputting the complexity data produced in step 4 by computer output system (step 5).

In this method, the input system is preferably keyboard, scanner, barcode reader, mouse, tablet, track ball, electronic pen or digital camera, but not always limited thereto.

In the above method, the output system is preferably monitor, printer or plotter, but not always limited thereto.

【Advantageous Effect】

The apparatus for visualizing or analyzing gene expression patterns and a method using the same are significantly improved ones from the conventional method and apparatus which provide a way to understand molecular functions, biological processes or cellular components in a biological sample based on the results of protein or RNA expression analysis. So, the apparatus and the method of the present invention facilitate analysis of biological aspects such as functional changes, evolutionary stages or developmental stages of each biological sample.

【Description of Drawings】

The application of the preferred embodiments of the present invention is best understood with reference to the accompanying drawings, wherein:

Figure 1 is a diagram illustrating an example of gene ontology tree structure.

Figure 2 is a flow chart illustrating the general workout of the apparatus for visualizing or analyzing gene expression patterns using gene ontology according to a preferred embodiment of the present invention.

Figure 3 is a diagram illustrating the visualization

of gene expression patterns using gene ontology according to a preferred embodiment of the present invention.

Figure 4 is a diagram illustrating the calculation of complexity from coordinate information distributed on gene ontology tree.

N : numbers of coordinates of genes corresponding to gene ontology tree,

P_{α} : probability of selection of the specific gene coordinate α in the gene ontology tree

Figures 5-8 are diagrams illustrating the visualized data of gene expression patterns according to a preferred embodiment of the present invention. In each diagram, gene expression patterns are visualized as the tree structure by taxonomy according to cellular component, molecular function and biological process.

Figure 5 is a set of diagrams illustrating the visualized data of gene expression patterns, taking the group related to extracellular matrix, among cell components, as the first level. (A): visualized data of protein expression pattern of brain tissues, (B): visualized data of protein expression pattern of neural stem cells, (C): visualized data of RNA expression pattern of neural stem cells, (D): visualized data of RNA expression pattern of oligodendrocytes.

Figure 6 is a set of diagrams illustrating the visualized data of gene expression patterns, taking the group related to auxiliary transport, among molecular functions, as the first level. (A): visualized data of protein expression pattern of brain tissues, (B): visualized data of protein expression pattern of neural stem cells, (C): visualized data of RNA expression pattern of oligodendrocytes.

Figure 7 is a set of diagrams illustrating the visualized data of gene expression patterns, taking the group related to binding, among molecular functions, as the first level. (A): visualized data of protein expression pattern of brain tissues (B): visualized data of protein expression pattern of neural stem cells, (C): visualized data of RNA expression pattern of oligodendrocytes.

Figure 8 is a set of diagrams illustrating the visualized data of gene expression patterns, taking the group related to biological regulation, among biological processes, as the first level. (A): visualized data of protein expression pattern of brain tissues, (B): visualized data of protein expression pattern of neural stem cells, (C): visualized data of RNA expression pattern of oligodendrocytes.

Figure 9 is a diagram illustrating the visualized

data of molecular functions which are observed in neural stem cells but not in oligodendrocytes in a preferred embodiment of the present invention.

Figure 10 is a diagram illustrating the visualized data of molecular functions which are observed in oligodendrocytes but not in neural stem cells in a preferred embodiment of the present invention.

【Mode for Invention】

Practical and presently preferred embodiments of the present invention are illustrative as shown in the following Examples.

However, it will be appreciated that those skilled in the art, on consideration of this disclosure, may make modifications and improvements within the spirit and scope of the present invention.

Example 1: Obtainment of expression analysis results from biological samples and gene ontology terms

<1-1> Obtainment of expression information of biological sample

The present inventors prepared test samples from brain tissues, neural stem cells and oligodendrocytes by the method described below and then expression information was obtained by separation of peptides on 1-dimensional

gel and tandem mass spectrum.

Brain tissues were divided into membrane fraction, soluble fraction and DNA binding protein fraction by the method of Klose. Then, proteins were separated by 1-dimensional electrophoresis. The separated proteins were hydrolyzed by trypsin. Peptides were separated by using SCX, RP column of liquid chromatography, followed by identification of the sequences by tandem mass spectrometer. Quantification was performed by protein abundance index method.

To prepare neural stem cells, the cells separated from the brain of fetus (12-18 weeks old) were cultured. The cells were differentiated into oligodendrocytes by using Olig2 gene (Kim, S. U., Neuropathology, 2004, 24(3), 159-171.) Neural stem cells and oligodendrocytes were labeled with lysine using C12 and C13 by SILAC method, which proceeded to 1-dimensional electrophoresis, hydrolysis with trypsin and liquid chromatography to identify proteins and quantifying thereof (Kwon, K.-H., et al., Proteomics, 2008, 8(6), 1149-61).

<1-2> Obtainment of information on gene expression and gene ontology

The information on expressed proteins obtained in Example <1-1> was applied to IPI (International Protein

Index) database to screen the information on expressed proteins and to identify genes corresponding thereto. Each gene corresponding to each protein was marked with Gene Symbol. In Table 1, some examples of information on the expressed proteins and genes corresponding thereto obtained in Example 1 are shown.

【Table 1】

Feature information of expressed proteins and genes corresponding thereto

<u>annotation</u>	<u>gene-symbol</u>
IPI00759806	ENO1
IPI00759776	ACTN1
IPI00744444	MYL6
IPI00742780	MYH9
IPI00646867	None
IPI00646304	PPIB
IPI00644576	FLNA
IPI00552689	VIM
IPI00514530	ACTA1
IPI00477807	None
IPI00473014	DSTN
IPI00465248	ENO1
IPI00418471	VIM
IPI00413922	MYL6
IPI00412737	WARS
IPI00395772	MYH9

The identified gene information was applied to the gene ontology database GO (Gene Ontology) to allocate gene ontology term information. Some of gene ontology terms allocated to the genes is shown in Table 2.

【Table 2】

Gene ontology terms allocated to expressed genes

Gene symbol	Gene ontology code corresponding to gene ontology term
ALBG	GO:0008150
A2M	GO:0051260,GO:0006886,GO:0051384,GO:0010037,GO:0007584
AADACL1	GO:0008152
AARS	GO:0043039,GO:0006419,GO:0006412,GO:0008033
AASDHPPT	GO:0009059
ABAT	GO:0042135,GO:0009450,GO:0007610,GO:0048148
ABR	GO:0008150,GO:0016337,GO:0032312,GO:0035023,GO:0007166,GO:0007259,GO:0007264,GO:0009615
ACAA1	GO:0006631,GO:0008152,GO:0006091,GO:0006629
ACAD9	GO:0008152,GO:0006118
ACADVL	GO:0006635,GO:0006631,GO:0008152,GO:0006118,GO:0015980,GO:0006629
ACAT1	GO:0008203,GO:0008202,GO:0008152,GO:0006629,GO:0008015
ACAT2	GO:0008203,GO:0008202,GO:0008152,GO:0006629
ACLY	GO:0044262,GO:0006101,GO:0006200,GO:0008610,GO:0015936,GO:0008152
ACO1	GO:0007165,GO:0006101,GO:0006099,GO:0008152
ACO2	GO:0006101,GO:0006099,GO:0008152,GO:0006091
ACOT7	GO:0006629
ACOT9	GO:0006637
ACOX1	GO:0006635,GO:0006693,GO:0006118,GO:0006629,GO:0007283
ACOX3	GO:0006635,GO:0006631,GO:0008206,GO:0008152,GO:0006118,GO:0006629
ACSL3	GO:0006631,GO:0008152,GO:0006629
ACSL4	GO:0006631,GO:0008152,GO:0006629,GO:0007611
ACSL6	GO:0007405,GO:0006637,GO:0008152,GO:0006629
ACSS1	GO:0008152
ACT1	GO:0016055,GO:0008283,GO:0008037,GO:0030198,GO:0008203,GO:0030206,GO:0030210,GO:0050653,GO:0050651,GO:0015014,GO:0019276,GO:0008202,GO:0046398,GO:0009653,GO:0007399,GO:0006629,GO:0008015
ACTA1	GO:0030240,GO:0048741,GO:0006936
ACTG1	GO:0006928,GO:0007605,GO:0051592

ACTN1	GO:0048041,GO:0051271,GO:0042981,GO:0051017
-------	---

The gene ontology code is the pre-selected code corresponding to a specific gene ontology term on gene ontology database, which has been defined by Gene Ontology Consortium. So, the gene ontology code is possibly added when the gene ontology term is added. But, the code itself is not changed and defined in the gene ontology database along with the gene ontology term. For example, it can be confirmed in the gene ontology database such as GO (Gene ontology) database through www.geneontology.org.

Example 2: Obtainment of coordinate information corresponding to gene expression information

Gene ontology tree classified in the gene ontology database GO (Gene ontology) was constructed using gene ontology information corresponding to the expression information obtained in Example 1.

Coordinate information of corresponding gene ontology term was obtained by using the constructed tree. Particularly, AD3 gene corresponded to GO:0016337, which passed on the "first" branch of level 1 branches on the tree, the "first" branch of 22 level 2 branches, the "second" branch of 5 level three branches and the "third" branch of 5 level 4 branches. Thus, it could be

[illegible]

Table 3 shows coordinate information corresponding to gene ontology information about each gene. Coordinate information contains information on the shortest path from the topmost node of gene ontology tree to a target node, precisely 1) level where a target node is located, 2) branch of each level on which a target path passes, 3) number of branches stretched from each level and 4) number of corresponding genes (total number of genes passing through the corresponding nodes which are expressed in relation to cellular component, molecular function, and biological process). For example, if the path passes over the first branch of level=0, the third branch of level=1 and the 29th branch of level=2, it is represented as 1:3:29:0:0:0:..., shown in the second line of the below Table.

【Table 3】

Level	Branch	Number of branches	Number of genes
1			
1	1:0:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:0:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	7945
2			
2	1:5:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	3
2	1:13:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1
2	1:6:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	2
2	1:15:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	148
2	1:1:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	9
2	1:8:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	222
2	1:16:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	170
2	1:2:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	6
2	1:10:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	232
2	1:17:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	19
2	1:3:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	3792
2	1:11:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	31
2	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1217
2	1:4:0:0:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	1:18:0:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	2025
3			
3	1:3:29:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	472
3	1:4:13:0:0:0:0:0:0:0:0 0:0:0:0:0:0:0:0:0	1:18:33:0:0:0:0:0:0:0 :0:0:0:0:0:0:0:0:0	54

3	1:3:35:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:15:9:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:16:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:15:10:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1:18:16:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:3:37:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	12
3	1:16:4:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:15:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	12
3	1:8:15:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:18:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:4:23:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:33:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:3:10:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	3
3	1:10:2:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:2:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	232
3	1:3:43:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1082
3	1:16:6:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:15:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	22
3	1:8:17:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:18:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	19
3	1:16:13:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1:18:15:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	12
3	1:3:12:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	9
3	1:3:45:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:18:10:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1:18:11:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1
3	1:16:8:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:15:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	19
3	1:4:27:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:33:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	17
3	1:4:2:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1:18:33:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	2
3	1:16:15:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1:18:15:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	1
3	1:3:47:0:0:0:0:0:0:0: 0:0:0:0:0:0:0:0:0	1:18:50:0:0:0:0:0:0:0: :0:0:0:0:0:0:0:0:0	11

Example 3: Analysis of expression pattern

<3-1> Visualization of gene expression pattern of a biological sample using gene ontology tree

In the upper part of the Figure, the starting point of tree was located. Then, coordinate information obtained in Example 2 was marked as node on the tree. The marked coordinates were connected from the root of the tree by branches to visualize gene expression patterns.

Some visualized data on the gene expression patterns of biological samples are shown in Figure 5 - Figure 8.

Among ontology terms related to cellular components, the group involved in extracellular matrix was set as the first level branch, and protein and RNA expression patterns of brain tissues, neural stem cells and oligodendrocytes were illustrated in Figure 5. In oligodendrocytes, protein was not detected, so that visualized data on protein expression patterns thereof was not illustrated in the Figure. As shown in Figure 5, protein was expressed in relation to extracellular matrix in brain tissues and neural stem cells, but protein was not expressed in oligodendrocytes. RNA expression in relation to extracellular matrix was detected in oligodendrocytes on microarray, which was different from the protein expression, though.

Visualized data of gene expression patterns of brain tissues, neural stem cells and oligodendrocytes, taking the group related to auxiliary transport protein activity, among molecular functions, as the first level was illustrated in Figure 6. Visualized data of gene expression patterns, taking the group related to binding, among molecular functions, as the first level was illustrated in Figure 7. According to the data, auxiliary transport protein activity related proteins were expressed in brain tissues and oligodendrocytes at a low level, but not in neural stem cells. In the meantime, binding related proteins were expressed in all the three test samples (brain tissues, neural stem cells and oligodendrocytes) at a high level.

Visualized data of gene expression patterns of brain tissues (A), neural stem cells (B) and oligodendrocytes (C), taking the group related to biological regulation, among biological processes, as the first level was illustrated in Figure 8. According to figure 8, proteins related to biological regulation were expressed particularly in brain tissues at a high level.

<3-2> Analysis of gene expression pattern using complexity

Coordinate information obtained in Example 2 was

applied to the following mathematical formula to calculate complexity of expression patterns of brain tissues, neural stem cells and oligodendrocytes and the results are shown in Table 4.

$$S = \sum_{\alpha=1}^N [-\log(P_{\alpha})] / N$$

5 Complexity

As a result, the complexity was highest in brain tissues, followed by in neuronal stem cells and in oligodendrocytes.

10 【Table 4】

	Neural stem cell	Oligodendrocyte	Brain tissue
Biological process	69.90	68.86	87.88
Molecular function	32.29	30.51	39.99
Cellular component	64.67	64.91	70.14

In the brain tissues, diverse brain cells co-exist and each cell has its own protein expressed therein and the protein has been mainly differentiated, suggesting high complexity. Neural stem cells contain neuron, astrocyte and oligodendrocyte related proteins, but these proteins are not differentiated yet, suggesting that protein expression is limited and complexity is low,

compared with that in the brain tissues. Oligodendrocytes comprise only one kind of cells, suggesting that complexity is the lowest.

5 <3-3> Comparison of gene expression patterns

The visualized data of molecular functions which are observed in neural stem cells but not in oligodendrocytes obtained in Example 2 was illustrated in Figure 9. And the visualized data of molecular functions which are
10 observed in oligodendrocytes but not in neural stem cells obtained in Example 2 was illustrated in Figure 10.

According to Figure 9 and Figure 10, neural stem cells show more complicated distribution than oligodendrocytes. But, there were genes only expressed in
15 oligodendrocytes. As shown in Figure 10, genes involved in nucleotide binding corresponding to the 32nd branch of level 2 passing over the third branch of level 1, protein binding corresponding to the 43rd branch of level 2, and hydrolase activity corresponding to the 8th branch of
20 level 2 passing over the 4th branch of level 1 are not expressed in neural stem cells but expressed in oligodendrocytes. In proteomics, protein which is not detected indicates no expression at all or expression at a very low level.

25

Those skilled in the art will appreciate that the conceptions and specific embodiments disclosed in the foregoing description may be readily utilized as a basis for modifying or designing other embodiments for carrying
5 out the same purposes of the present invention. Those skilled in the art will also appreciate that such equivalent embodiments do not depart from the spirit and scope of the invention as set forth in the appended claims.

【CLAIMS】**【Claim 1】**

An apparatus for visualizing gene expression patterns of a biological sample using gene ontology tree
5 comprising the following devices:

gene ontology term allocating device which receives the results of the analysis of protein or RNA expression of a biological sample and then applies the results to biological DB or gene ontology database to allocate gene
10 ontology terms corresponding to the protein or RNA;

coordinate creating device which obtains coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

15 visualizing device which generates visualizing data on the coordinate information obtained by the above coordinate creating device; and

outputting device which outputs the visualizing data.

【Claim 2】

20 The apparatus according to claim 1, wherein the protein analysis result is obtained by proteome analysis.

【Claim 3】

25 The apparatus according to claim 2, wherein the

proteome analysis is performed by 2-dimensional electrophoresis or mass spectrometry.

【Claim 4】

5 The apparatus according to claim 1, wherein the RNA expression analysis result is obtained by microarray analysis.

【Claim 5】

10 The apparatus according to claim 1, wherein the biological DB is Uni-gene, LucusLink, Swiss-Prot, IPI, MGI, Uniprot or EMBL.

【Claim 6】

15 The apparatus according to claim 1, wherein the gene ontology database is GO (Gene Ontology), ChEBI, GOA or NEW.

【Claim 7】

20 The apparatus according to claim 1, wherein the gene ontology term allocating device is composed of gene identification algorithm identifying an expressed gene by screening thereof through biological DB based on the results of protein or RNA expression analysis and allocation algorithm allocating gene ontology term
25 corresponding to the identified gene by screening thereof

through gene ontology database.

【Claim 8】

The apparatus according to claim 1, wherein the
5 coordinate information is 2-dimensional or 3-dimensional
coordinate information.

【Claim 9】

The apparatus according to claim 1, wherein the gene
10 ontology tree is constructed according to molecular
functions, biological processes or cell components.

【Claim 10】

The apparatus according to claim 1, wherein the
15 outputting device is monitor, printer or plotter.

【Claim 11】

An apparatus for analyzing gene expression patterns
of a biological sample using gene ontology tree comprising
20 the following devices:

gene ontology term allocating device which receives
the results of the analysis of protein or RNA expression
of a biological sample and then applies the results to
biological DB or gene ontology database to allocate gene
25 ontology terms corresponding to the protein or RNA;

coordinate creating device which obtains coordinate information of the allocated gene ontology terms corresponding to gene ontology tree classified in gene ontology database;

5 complexity calculating device which calculates complexity data defined as the following formula on the coordinate information obtained by the above coordinate creating device; and

outputting device which outputs the complexity data.

$$\text{Complexity}(S) = \sum_{\alpha=1}^N [-\log(P_{\alpha})]/N$$

10

N : number of coordinates of genes corresponding to gene ontology tree,

P_{α} : probability of selection of the specific gene coordinate α in the gene ontology tree

15

【Claim 12】

An apparatus for comparing gene expression patterns of a biological sample using gene ontology tree comprising the following devices:

20

gene ontology term allocating device which receives the results of the analysis of protein or RNA expression of a biological sample and then applies the results to

biological DB or gene ontology database to allocate gene ontology terms corresponding to the protein or RNA;

coordinate creating device which obtains coordinate information of the allocated gene ontology terms
5 corresponding to gene ontology tree classified in gene ontology database;

comparing device which obtains different coordinate information by comparing the coordinate information obtained by the above coordinate creating device and the
10 coordinate information obtained from another sample;

visualizing device which generates visualizing data on the coordinate information obtained by the comparing device; and

outputting device which outputs the visualizing data.

15
【Claim 13】

The apparatus for comparing gene expression patterns according to claim 12, wherein the comparing device is composed of comparison algorithm comparing the coordinate
20 information obtained by the coordinate creating device and the coordinate information obtained from other samples, and coordinate producing algorithm producing only those coordinates having different locations by eliminating coordinate information having the same locations.

【Claim 14】

The apparatus for comparing gene expression patterns according to claim 13, wherein the comparing device additionally includes selection algorithm producing
5 coordinate information corresponding to gene ontology term included in specific hierarchical classification which is operated before applying the comparison algorithm.

【Claim 15】

10 A method for visualizing gene expression patterns of a biological sample using gene ontology tree comprising the following steps:

(a) inputting the results of analysis of protein or RNA expression obtained from a sample through computer
15 input system (step 1);

(b) allocating ontology terms corresponding to the input protein or RNA by gene ontology term allocating device (step 2);

(c) obtaining coordinate information of the gene
20 ontology terms allocated in step 2 corresponding to gene ontology tree classified in gene ontology database by coordinate creating device (step 3);

(d) producing visualizing data of coordinate information obtained in step 3 by visualizing device(step
25 4); and

(e) outputting the visualizing data produced in step 4 by computer output system (step 5).

【Claim 16】

5 The method for visualizing gene expression patterns of a biological sample using gene ontology tree according to claim 15, wherein the input system is key board, scanner, barcode reader, mouse, tablet, track ball, electronic pen or digital camera.

【Claim 17】

10 The method for visualizing gene expression patterns of a biological sample using gene ontology tree according to claim 15, wherein the output system is monitor, printer or plotter.

【Claim 18】

15 A method for analyzing gene expression patterns of a biological sample using gene ontology tree comprising the following steps:

20 (a) inputting the results of analysis of protein or RNA expression obtained from a sample through computer input system (step 1);

25 (b) allocating ontology terms corresponding to the input protein or RNA by gene ontology term allocating

device (step 2);

(c) obtaining coordinate information of the gene ontology terms allocated in step 2 corresponding to gene ontology tree classified in gene ontology database by coordinate creating device (step 3);

(d) producing complexity data defined as the formular of claim 11 on the coordinate information obtained in step 3 by complexity calculating device (step 4); and

(e) outputting the complexity data produced in step 4 by computer output system (step 5).

【Claim 19】

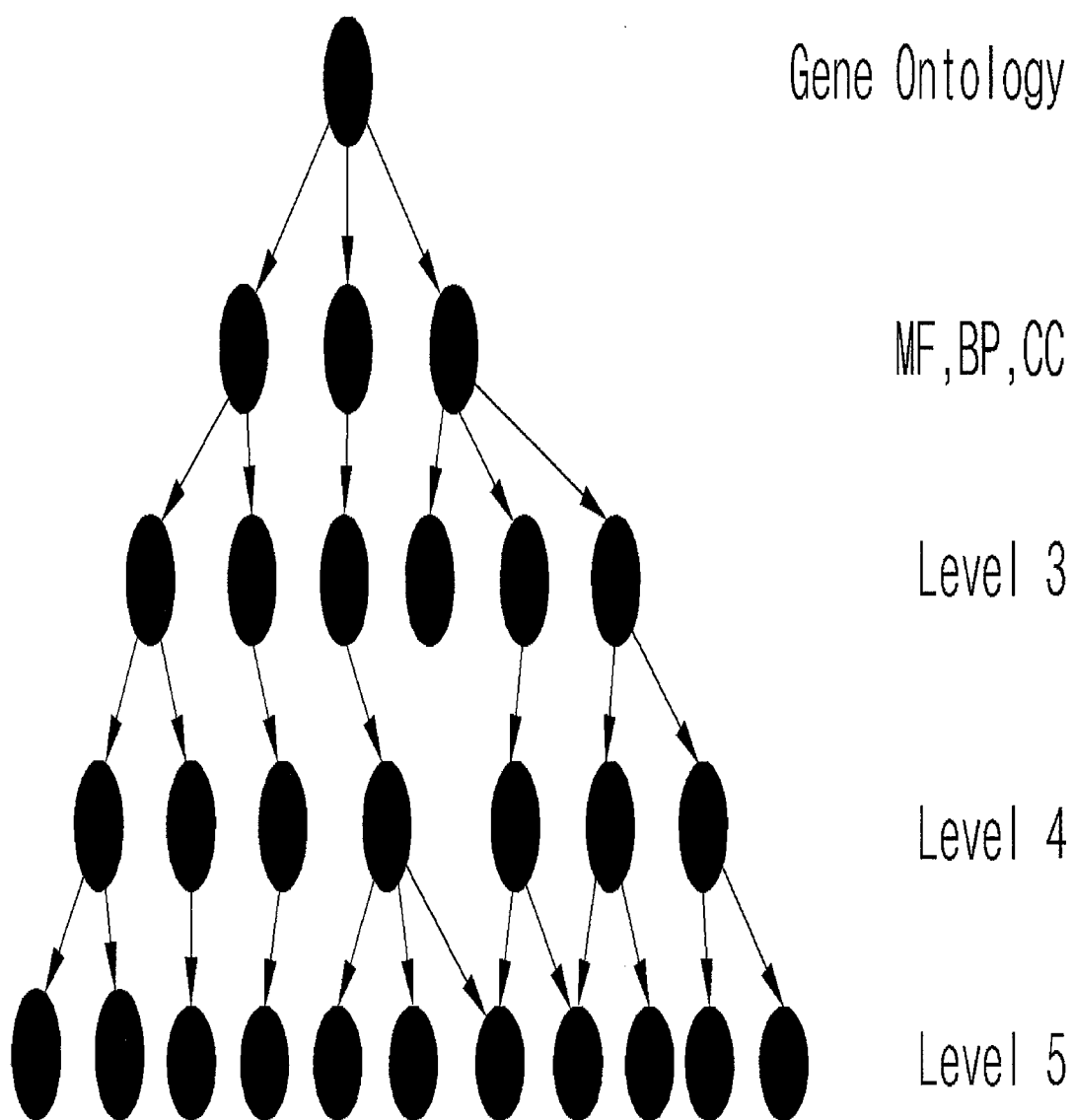
The method for analyzing gene expression patterns of a biological sample using gene ontology tree according to claim 18, wherein the input system is key board, scanner, barcode reader, mouse, tablet, track ball, electronic pen or digital camera.

【Claim 20】

The method for analyzing gene expression patterns of a biological sample using gene ontology tree according to claim 18, wherein the output system is monitor, printer or plotter.

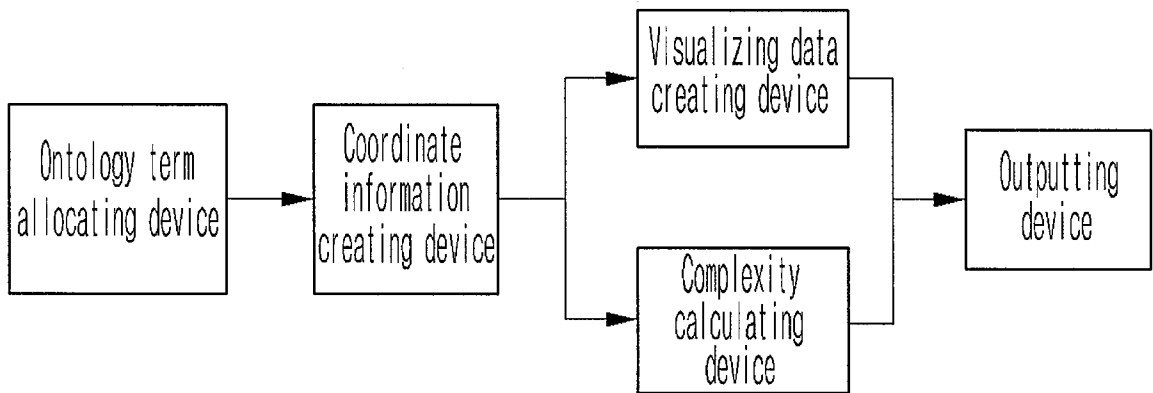
1 / 10

Fig. 1



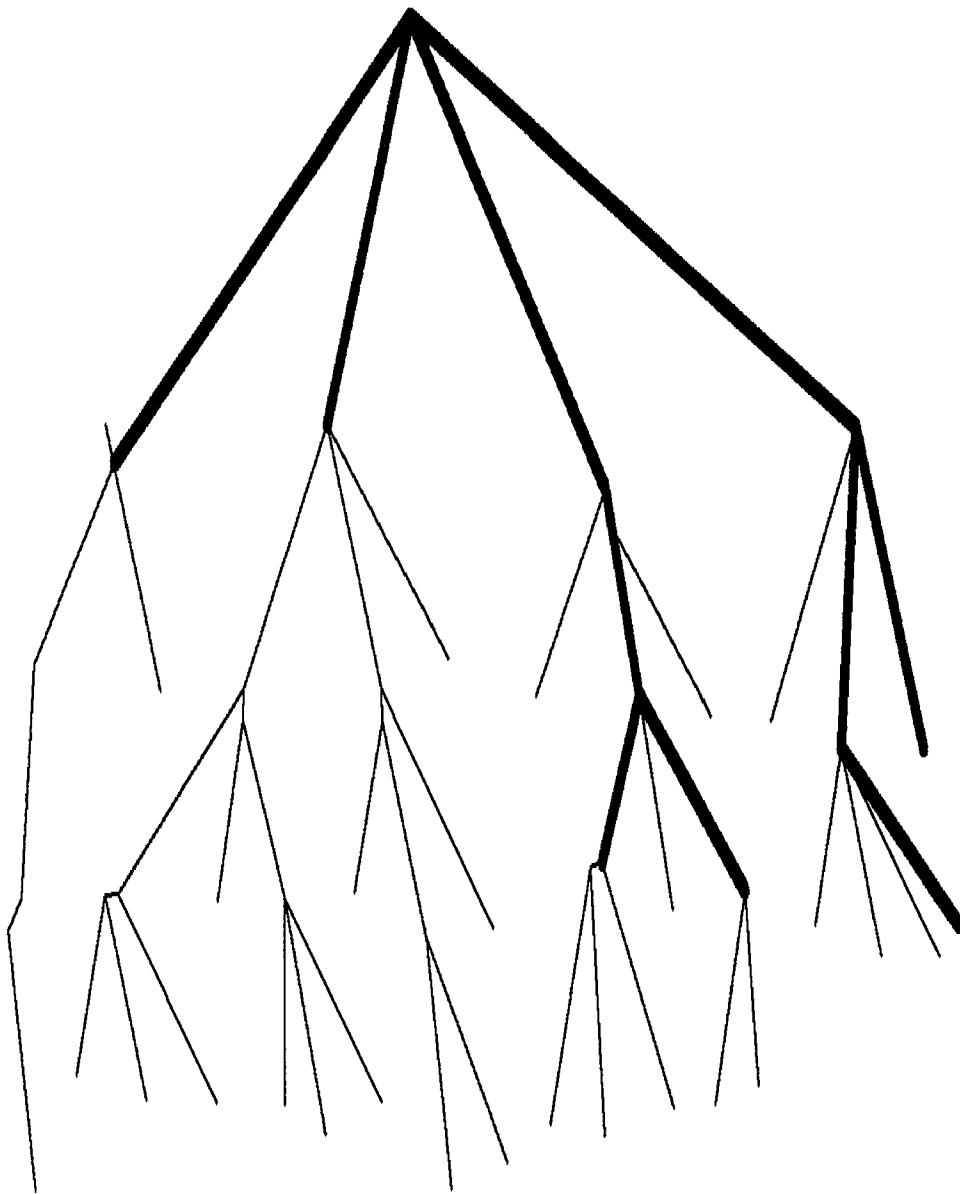
2/10

Fig. 2



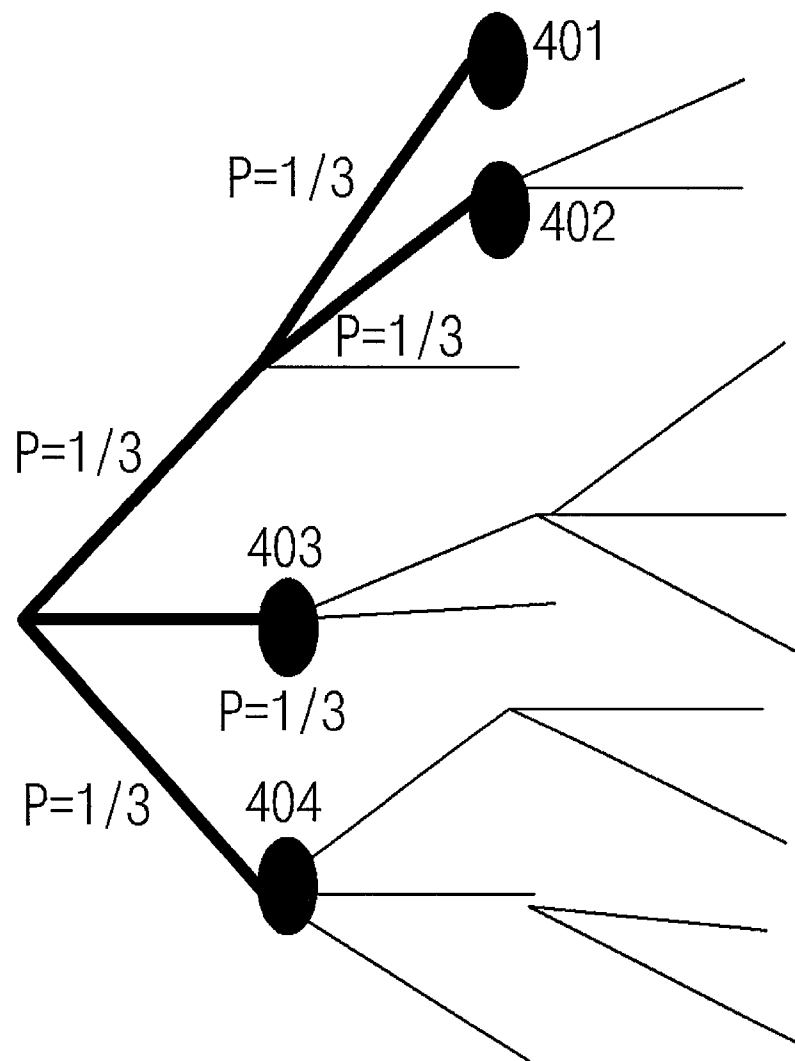
3/10

Fig. 3



4/10

Fig. 4

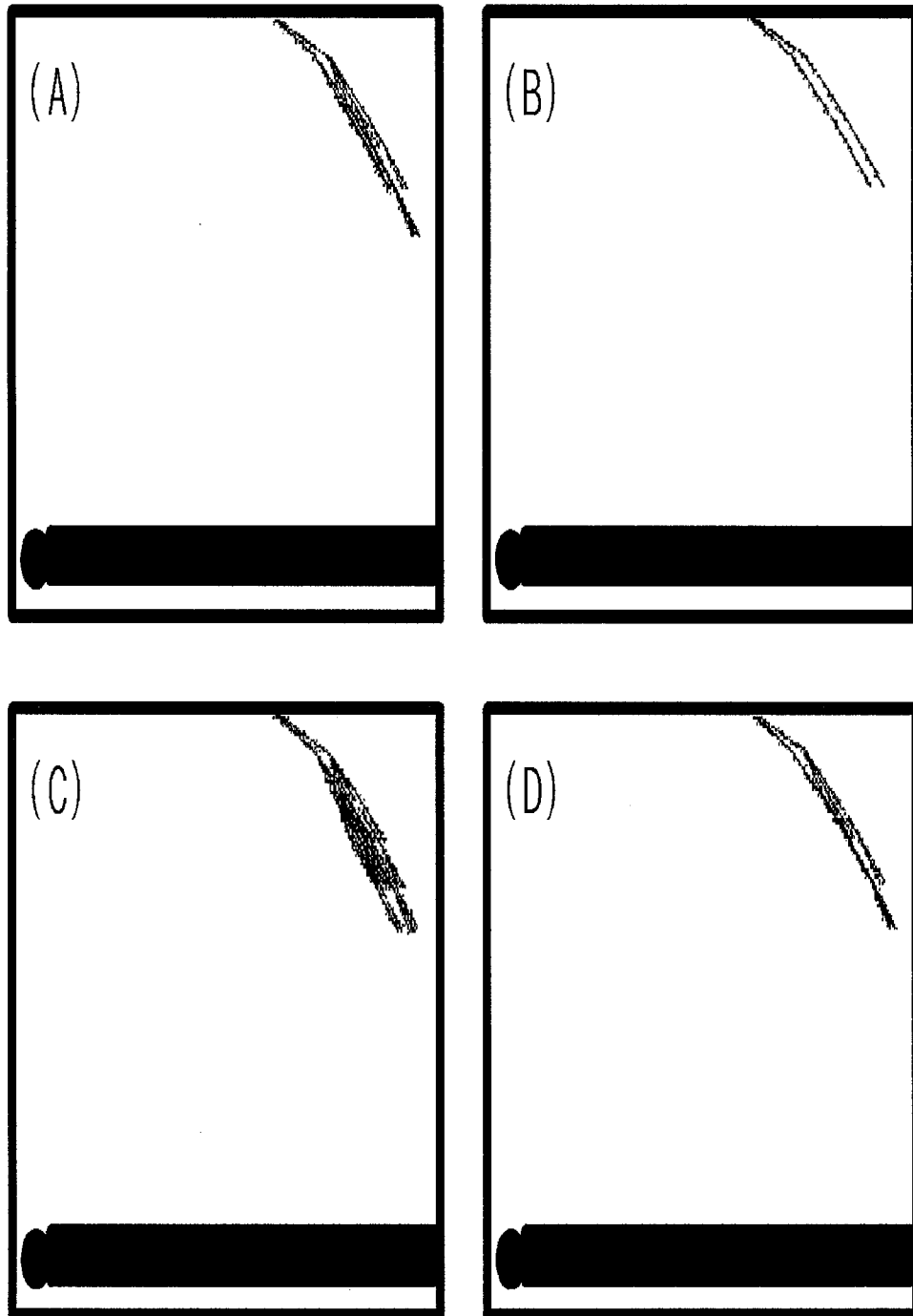


$$S = -0.25^*$$

$$(\log(1/9) + \log(1/9) + \log(1/3) + \log(1/3))$$

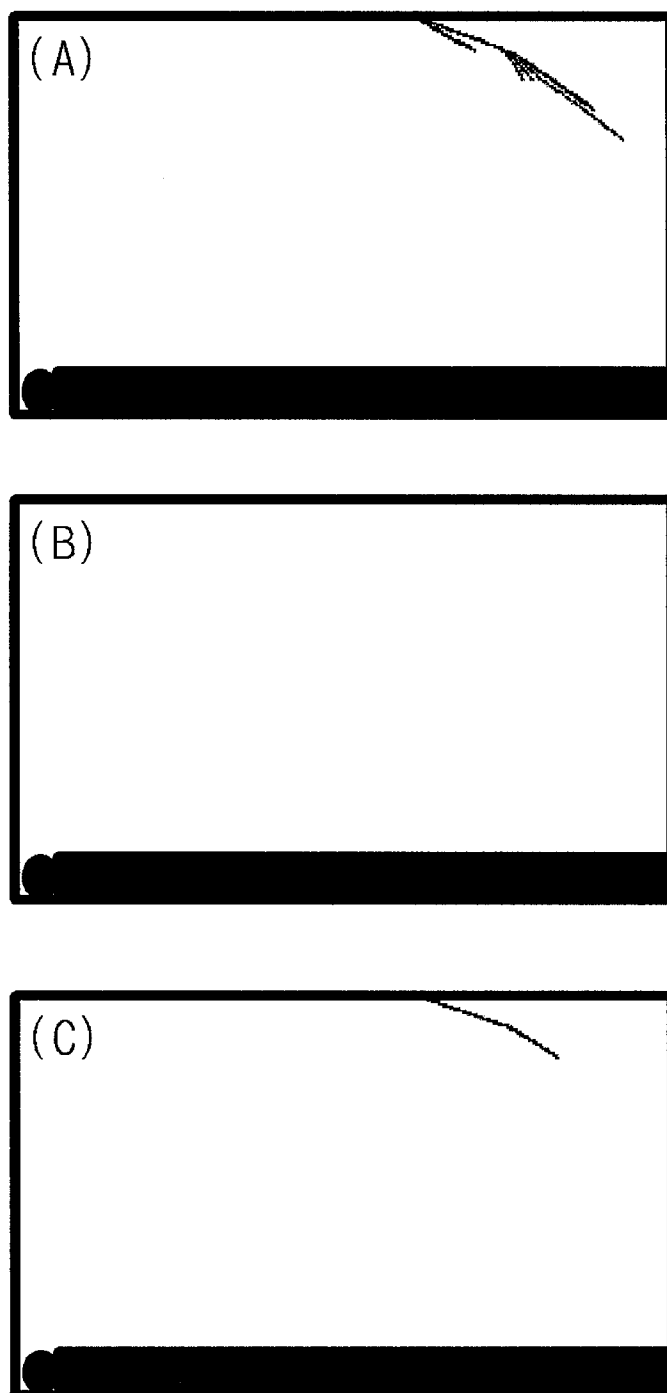
5/10

Fig. 5



6/10

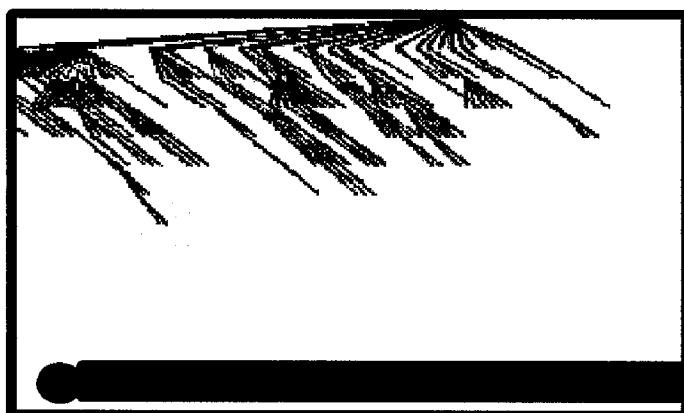
Fig. 6



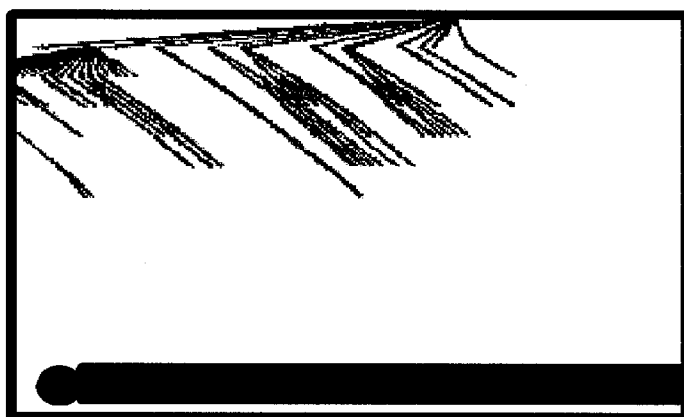
7/10

Fig. 7

(A)



(B)



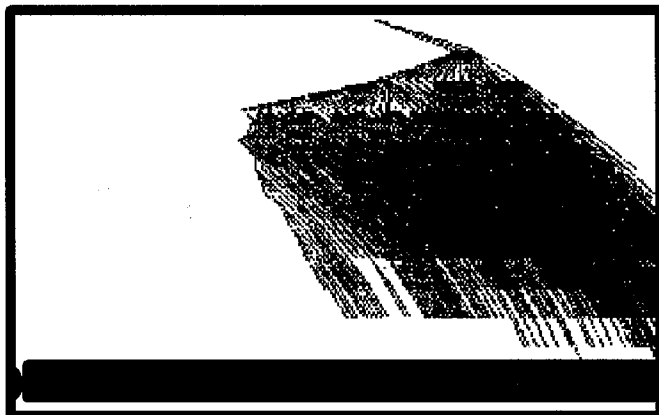
(C)



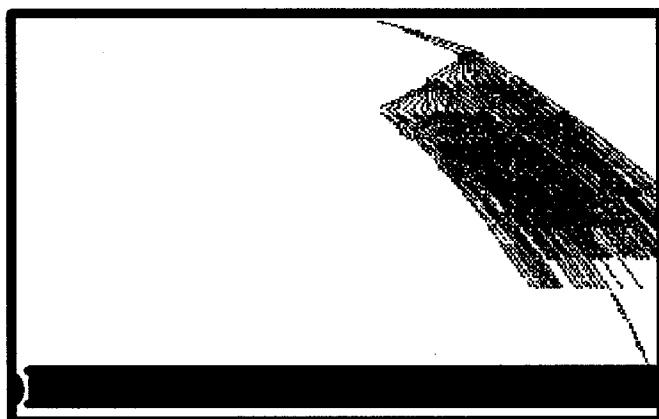
8/10

Fig. 8

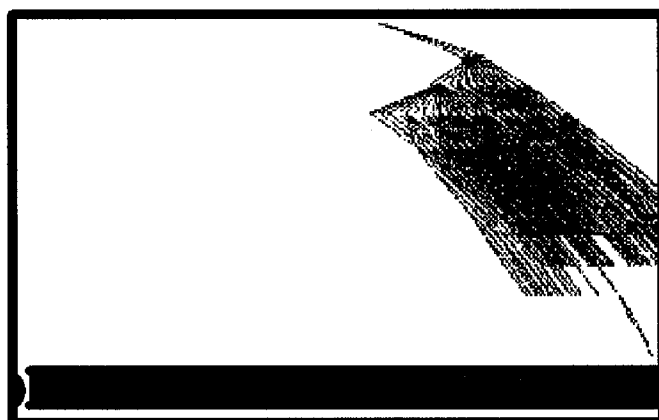
(A)



(B)

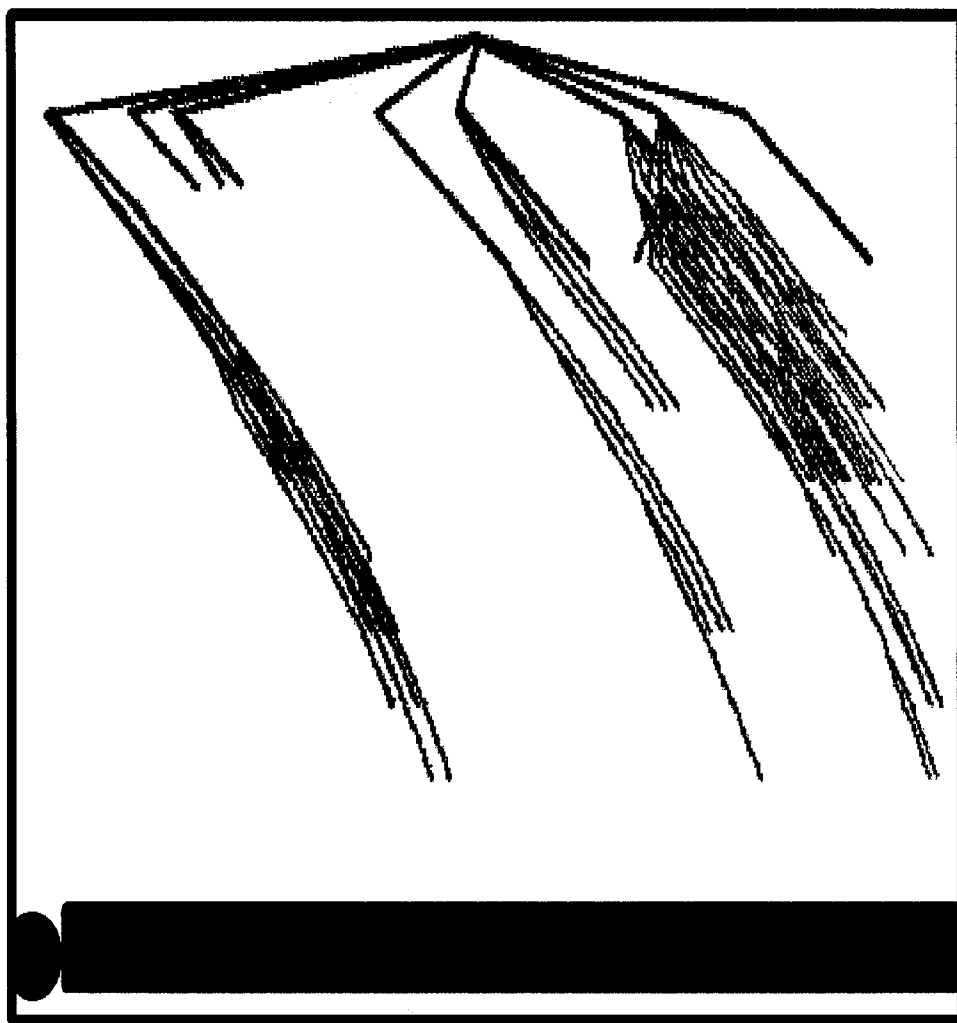


(C)



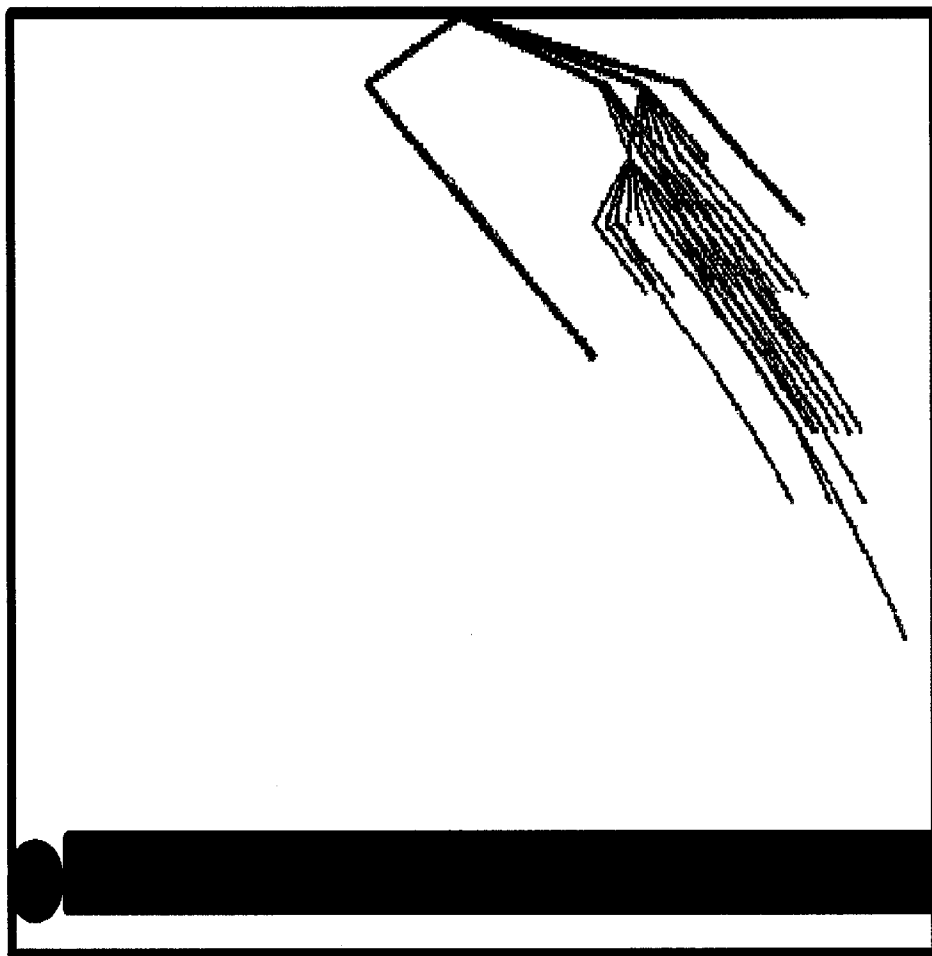
9/10

Fig. 9



10/10

Fig. 10



A. CLASSIFICATION OF SUBJECT MATTER**G06F 17/30(2006.01)i, C12Q 1/68(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F 17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models since 1975

Japanese Utility models and applications for Utility models since 1975

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

eKIPASS(KIPO internal), keywords: search, query, gent, ontology, DNA, tree

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2005/022412 A1 (Kim Yang Suk et al.) 10 March 2005 see whole reference.	1-20
A	US 2005/0137808 A1 (Choi et al.) 23 June 2005 see abstract, paragraph [0018]-paragraph[0050], and figures 6, 7	1-20
A	US 2003/0009294 A1 (Cheng) 09 January 2003 see abstract, paragraph[0005]-paragraph[0033], and figures 4, 5	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

07 MAY 2009 (07.05.2009)

Date of mailing of the international search report

08 MAY 2009 (08.05.2009)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-
gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Lee, Soo Cheol

Telephone No. 82-42-481-8120



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/KR2008/004735Patent document
cited in search reportPublication
datePatent family
member(s)Publication
date

WO 2005-022412 A1

10.03.2005

KR 20050022798 A

08.03.2005

US 2006-0234244 A1

19.10.2006

US 2005-137808 A1

23.06.2005

None

US 2003-009294 A1

09.01.2003

None