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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2007 (07.09.2007)

PCT

(10) International Publication Number
WO 2007/100183 A1

(51) International Patent Classification:
G01N 33/574 (2006.01)

(21) International Application Number:
PCT/KR2006/005836

(22) International Filing Date:
28 December 2006 (28.12.2006)

(25) Filing Language: Korean

(26) Publication Language: English

(30) Priority Data:
10-2006-0019517
28 February 2006 (28.02.2006) KR

(71) Applicant *(for all designated States except US)*: **BIOINFRA INC.** [KR/KR]; 8th Fl., Cancer Research Institute, Seoul National University College of Medicine, 28, Yongon-dong, Chongno-gu, Seoul 110-799 (KR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KIM, Chul Woo** [KR/KR]; San 2-120 Galmae-dong, Guri-si, Gyeonggi-do 471-080 (KR). **OHN, SYNG-YUP** [KR/KR]; #101-2007 Gangchon Apt., Ichon-1dong, Yongsan-gu, Seoul 140-727 (KR). **HAN, MI YOUNG** [KR/KR]; #Da-606 Shinsamho Apt., Bangbae-dong, Seocho-gu, Seoul 137-060 (KR). **SHIN, YONG-SUNG** [KR/KR]; #120-801 SK Apt., 1353, Mia-7dong, Gangbuk-gu, Seoul 142-778 (KR).

(74) Agent: **LEE, Won-Hee**; 8th Fl., Sung-ji Heights II, 642-16 Yoksam-dong, Kangnam-ku, Seoul 135-080 (KR).

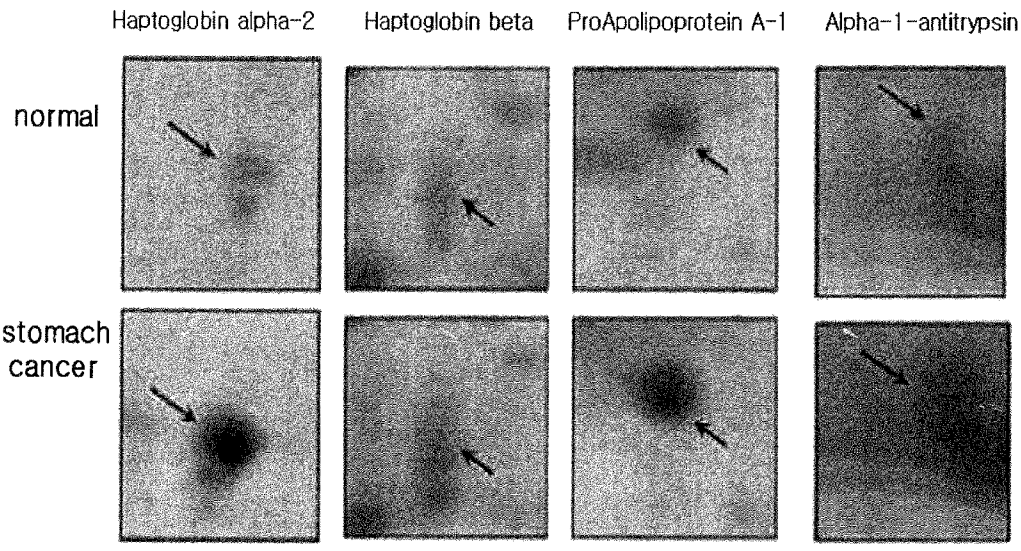
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: PROTEIN MARKERS FOR DIAGNOSING STOMACH CANCER AND THE DIAGNOSTIC KIT USING THEM



(57) Abstract: The present invention relates to protein markers for diagnosing stomach cancer and a diagnostic kit using the same, more precisely protein markers screened by two- dimensional gel electrophoresis and bioinformatics and a diagnostic kit using the same. The markers of the invention can be effectively used for diagnosing stomach cancer and evaluating the extent of progress of the cancer by confirming the expression levels of those marker proteins whose expressions differ in stomach cancer patients from in normal healthy people.

WO 2007/100183 A1

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

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

【Invention Title】

PROTEIN MARKERS FOR DIAGNOSING STOMACH CANCER AND THE DIAGNOSTIC KIT USING THEM

【Technical Field】

The present invention relates to protein markers for diagnosing stomach cancer and the diagnostic kit using the same, more precisely protein markers for diagnosing stomach cancer screened by two-dimensional gel electrophoresis and bioinformatics and the diagnostic kit using the same.

【Background Art】

Cancer is now a leading cause of death by overtaking heart disease in the 21st century. So, the prevention, diagnosis and treatment of cancer are major concerns in the whole world. Korea is not an exception. Since 1988, cancer has been the leading cause of death in Korea. In the case of stomach cancer, it has been reduced in America and European countries but it still exhibits the highest outbreak frequency in Korea and Japan. According to the report of Cancer Registry, Seoul, Korea (1992-95), stomach cancer patients take 23% of total cancer patients (male:

24.7%, female: 17.3%). The average age of outbreak is 54, mostly 40 - 60 but the age of 20s takes approximately 3% and the frequency of the disease in men is double that in women. However, once it is early diagnosed, survival rate is at least 90%, suggesting that early diagnosis of stomach cancer is very important for the National Health Care Surveillance.

Tumorigenesis, progress and malignant change are the result of combined action of both genetic factors and environmental factors to result in proteome changes. The correlation between the level of molecular concentrations of proteins, directly involved in essential metabolism and regulation pathways for the survival of every cells including cancer cells, and the relevant mRNA level is very low (correlation coefficient = 0.48), so the mRNA focused approach based on high-throughput screening of a target gene might lead to misunderstanding on the biological functions of the target protein. The activity, stability, intracellular location and turnover of the protein involved in cell survival, differentiation and death are regulated by diverse post-translational modifications (PTM). Thus, it is more important to analyze proteome than to focus on mRNA.

Cancer is a systemic disease, in which rather many

functions and biological processes of various organs than one or two kinds of cells or tissues are involved. Therefore, such clinical samples, which are able to reflect the whole proteome changes, such as, in serum or body fluid, have to be examined. In fact, general clinical diagnosis has been made through the analysis of the body fluid. So, developing a novel diagnostic method using body fluid that facilitates the effective primary screening might pave the way to significant improvement of survival rate of cancer patients.

The combination of biotechnology (BT) and information technology (IT) gave birth to bioinformatics and bioelectronics that facilitate easy, fast and effective analysis of huge amount of data. Particularly, bioinformatics is a technique to collect, store and analyze huge amount of information that a living body contains, and to use such information in the field of drug development, food production, agriculture and environmental business, etc, resulting in the establishment of biological information S/W, biological information service, and information technology infrastructure. The image mining is a new database applying technique, which is technical combination of image database technique and data mining technique. Unlike the general data mining tried in the

field of bioinformatics, the image mining is a novel technique never tried before in domestic and overseas. General data mining techniques are based on applied statistics, suggesting that it is limited in obtaining precise relevant information. For example, proteome spots, seen on the image made by two-dimensional gel electrophoresis, are expected to include information on the disease condition (normal or ill) and various characteristics of the patient. So, if it is possible to extract relevant information by using the image mining technique, diagnostic method and prediction of prognosis, depending on stage and histopathological classification, will be significantly improved and/or complemented.

Korean Patent No. 2004-0055893 describes the automatic template generation method for constructing protein interaction networks, and Korean Patent No. 2003-0092462 describes the method and kit for diagnosis of progress stage of cancer by investigating the level of p53 protein or Bcl-X L protein or the correlation of the two and Bcl-2 protein. However, these two descriptions do not include the explanations on the stomach cancer marker development by using two-dimensional gel electrophoresis and bioinformatics.

【Disclosure】

【Technical Problem】

It is an object of the present invention to provide a method for screening a marker protein for diagnosing stomach cancer, which characteristically differs in expression in stomach cancer, by using two-dimensional gel electrophoresis and bioinformatics and a use of the marker as an early diagnosis marker for stomach cancer.

【Technical Solution】

To achieve the above object, the present invention provides a screening method of stomach cancer high risk group comprising the following steps: (1) investigating the presence or absence of one or more markers for diagnosing stomach cancer in the sample obtained from a patient, selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-IV Transthyretin, ProApolipoprotein A-I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H, and the expression levels or patterns of them; and (2) analyzing the relation of the screening result and stomach cancer progression status.

The present invention also provides a use of the protein selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-IV Transthyretin, ProApolipoprotein A-I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H as a marker for diagnosing stomach cancer.

The present invention further provides a diagnostic kit for stomach cancer containing (1) primary capture reagent binding to one or more markers for diagnosing stomach cancer selected from a group consisting of Leucine-rich alpha-2-glycoprotein (LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-IV Transthyretin, ProApolipoprotein A-I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H; and (2) secondary capture reagent which was not bound to the first capture reagent.

The descriptions on terms used in the present invention are given below to help the understanding of the present invention.

"Marker" indicates the substance, which is distinctly found in the serum sample of a certain disease patient, but not in normal serum samples. The marker or markers comprise a single polypeptide or a combination of polypeptides.

"Proteome pattern" indicates a specific polypeptide group or grouping of polypeptides distinctly found in the serum sample obtained from a patient with a certain disease but not in normal serum samples. For example, a serum protein group, exhibiting specific changes in its level by a disease, and/or two-dimensional location and morphology of the group is included in this criterion.

"Data mining" indicates the procedure designed to find out correlation between relevant data, more precisely, the procedure in which a new data model is extracted from the stored data of database, which has not been disclosed yet, and useful information in the future is taken from the new data model and applied to decision-making. That is, relevant information is extracted by investigating patterns hidden in the data.

Hereinafter, the present invention is described in detail.

The present invention provides a screening method of stomach cancer high risk group comprising the following steps: (1) investigating the presence or absence of one or more markers for diagnosing stomach cancer in the sample obtained from a patient, selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin,

Alpha-1-antitrypsin, Apolipoprotein A-IV Transthyretin, ProApolipoprotein A- I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H, and the expression levels or patterns of them; and (2) analyzing the relation of the screening result and the state of stomach cancer progress.

In step (2), the state of stomach cancer indicates the stomach cancer risk level of an individual, whether stomach cancer has developed or the degree of stomach cancer progression.

The present inventors obtained serum samples from normal healthy individuals and cancer patients, followed by the two-dimensional image treatment for the serum proteome. Then, bioinformatics was applied to analyze the prepared data and as a result the optimum marker proteins being able to distinguish a normal individual from a cancer patient were identified.

Particularly, the present inventors performed two-dimensional gel electrophoresis to establish image data of proteins exhibiting significant expression level changes according to cancer development (herein, stomach cancer), to which bioinformatics techniques were applied, resulting in the establishment of a novel use of the protein markers for diagnosing stomach cancer. First of all, sera were collected from both healthy people and cancer patients.

The stomach cancer patients were at average 58.7 year-old male (age between 33 - 78) and at average 56.2 year-old female (age between 29 - 77) (see Table 1).

Optimum conditions for two-dimensional gel electrophoresis to examine serum proteomes were investigated. As a result, to give the best results, 13 cm strip was used, pH range was 4-7, total volt hour for IEF (isoelectric focusing) was 62,000 Vhr, amount of serum protein loaded was 200 ug, and composition of the rehydration buffer was 8 M urea.

Upon completion of electrophoresis, proteins were detected by the staining patterns. To detect spots on the images of two-dimensional electrophoresis gel, a computer software was used to establish bases for image comparison. Particularly, spots were detected by image filtering, subtracting background, removing vertical and horizontal streaks, and comparing variants using a computer software program. And the database was established with the results.

When samples were measured and thereby database was established, the data was analyzed by bioinformatics approach. The software used in bioinformatics can constitute codes to change the image data into readable codes, which includes the code for applying algorithm to the information on the markers provided by the present

invention. In a preferred embodiment of the present invention, various bioinformatics techniques were used to screen a stomach cancer associated protein marker. The analysis used in the present invention using bioinformatics techniques comprises the steps of generating an example of the proteome harboring disease-specific proteome pattern and training to establish database (training stage); and extracting specific data of target serum proteome and comparing it with the above disease-specific proteome example to determine the disease outbreak in the target serum proteome (testing stage).

"Training" herein indicates that the generation of a classification model by algorithm using established data of known samples from stomach cancer patients and normal healthy people. At this time, the data used to generate a classification model is named "training data set" and the groups providing the data is named "training group". Once training is finished, the classification model recognizes a data pattern from unknown samples and classifies it.

Support vector machine is a learning program facilitating pattern recognition, which is able to treat multiple variants at the same time and classify them. This support vector machine can interpret nonlinear data of the input area into linear data and provide optimal boundary

(optimal separating plane or hyperplane) between each characteristics. The support vector machine is divided largely into training process and evaluating process. In the training process, support vector is generated. And in the evaluating process, judgment is made considering specified characteristics. Sample is composed of n objects, and the i^{th} object is represented as vector xi comprising p variants and then the corresponding class which has been pre-classified is indicated as yi. So, if there are two classes, yi is either '1' or '-1' (cancer or normal). Discriminant function for the random input pattern is as follow.

$$D(x)=\mathrm{sign}(wx+b)=\begin{cases}+1, & \text{if } wx+b\geq 0\\ -1, & \text{else}\end{cases}$$

w: weight vector

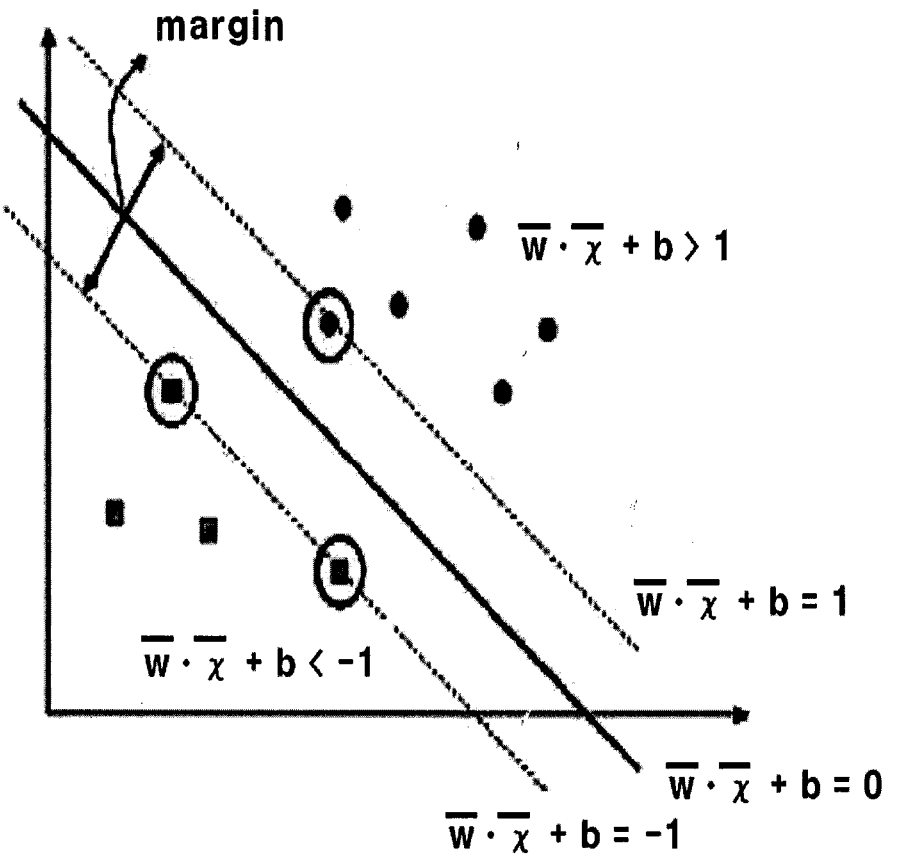
b: threshold

The optimal hyperplane dividing the data into two classes is the dotted line in the following graph. The two

dotted lines and margin (d) are as follows.

$$(w \cdot x) + b = \pm 1,\ d = 2 / ||w||$$

To design a support vector adequate for the purpose, weight vector(w) and boundary(b) have to be determined. That is, according to the formula: $y[(w \cdot x) + b] \geq 1$, w and b which are able to minimize ||w|| are screened.

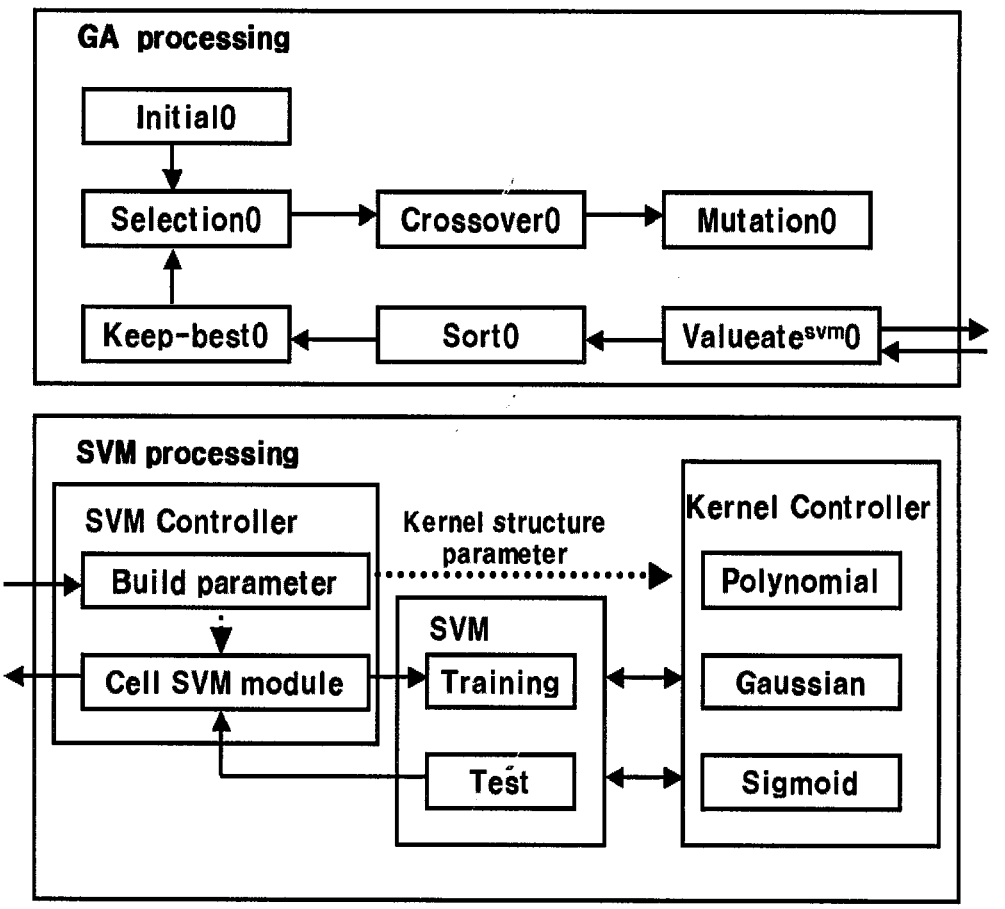


Genetic algorithm is making an engineering model from genetic and evolutionary systems of natural world, which is

dealing with adaptation capacity of a life to environment. The possible solutions for a question are expressed in a certain form and then gradually modified to produce more valuable solutions. Particularly, genetic algorithm is a kind of optimizing algorithm for searching x value calculating the maximum value or the minimum value of the function f(x) against variant x defined in a specific area, as fast as possible. Genetic algorithm is composed of the following steps: determining genotype which encodes and converts genetic factors into symbol string; determining early genetic group to generate various individuals having different characteristics from the defined genotype above; evaluating adaptation of each individual to calculate the adaptation according to the pre-defined method; selecting individuals based on the evaluated adaptation in order to determine survival distribution of individuals; mating by substituting a gene between two chromosomes to generate a new individual; mutating by modifying a part of a gene and maximizing diversity of the genetic group so as to prepare individuals providing better solutions; and returning to the step of evaluating adaptation of each individual. Genetic algorithm facilitates searching for solutions according to cooperative genetic manipulation on selection, mating, etc, among multiple individuals. Thus,

compared with the conventional parallel screening method for solutions, genetic algorithm is easy and simple to screen solutions.

The process of interlocking of GA with SVM is as follows.



Among proteome data obtained from 311 volunteers (stomach cancer patients; 143, normal healthy people; 168),

100 of each stomach cancer patients and normal people were randomly selected and form a training group and evaluating group (see Table 2). The training group data of 200 people was applied to genetic algorithm (GA) interlocked to support vector machine (SVM) (see Korean Patent No. 10-2002-0067298 "Method and System for Analysis of Cancer BiomarkersUsing Proteome Image Mining") and as a result 9 markers for diagnosing stomach cancer were found and database was constructed. Mean values of cancer patients and normal people were calculated with each spot on database, which were then compared. T-test was performed to calculate p-value of each group and 9 spots exhibiting significant difference in expression were identified (see Table 3).

Cross validation and random forest, other bioinformatics techniques, were performed with the proteins of the training group. As a result, stomach cancer decision rate (sensitivity - diagnose a cancer patient as cancer; specificity - diagnose a normal person as normal) was high according to the method of the invention. Random forest was performed again with the data of 111 remaining people who were not included in the training group (stomach cancer patients:43, normal healthy people: 68), and as a result two algorithm exhibited up to 80% diagnostic

accuracy. Therefore, 9 marker candidate proteins for diagnosing stomach cancer were confirmed to be very useful. To support the result, immunoblotting was performed and the result was consistent, that is, marker protein candidates exhibited significant difference in expression level between normal sera and stomach cancer patient sera.

The data of total 143 stomach cancer patients (training group: 100, evaluating group: 43) are divided by cancer stage and the expressions of those 9 proteins were investigated according to the stage. As shown in Fig. 3, each protein exhibited different expressions according to the stage. The expressions of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, ProApolipoprotein A- I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H were higher in stomach cancer patients than in normal healthy group, whereas the expressions of Apolipoprotein A-IV and Transthyretin were reduced in stomach cancer patients. Thus, those 9 marker proteins are presumably associated with stomach cancer progress, supporting the usability of those proteins as biomarkers for diagnosing stomach cancer.

Markers of the present invention facilitate the screening methods, precisely proteomes of target sera to be

investigated to detect stomach cancer were inputted and changed into two-dimensional image, which was then compared with the sample having the pattern of disease-specific marker protein. Or the expressions of those marker proteins in target sera were compared with those in normal sera, and those expressions were turned into numerical values. From the comparison of those values, it could be judged whether the target serum is normal or with cancer.

The present invention also provides a use of the protein selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-IV Transthyretin, ProApolipoprotein A-I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H as a marker for diagnosing stomach cancer.

The above 9 proteins can be used as markers for diagnosing stomach cancer since they were confirmed to be associated with stomach cancer progress.

The present invention further provides a diagnostic kit for stomach cancer containing primary capture reagent binding to one or more markers for diagnosing stomach cancer selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin,

Apolipoprotein A-IV Transthyretin, ProApolipoprotein A-I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H; and secondary capture reagent which was not bound to the first capture reagent.

The diagnostic kit can be used to detect one or more markers of the invention that exhibit different expressions in stomach cancer patients. The diagnostic kit of the present invention facilitates not only the diagnosis of cancer by a doctor but also the monitoring of the after-care response of a patient, in order to modify the treatment. The kit can also be used to identify a compound regulating *in vivo* or *ex vivo* expression of one or more markers in stomach cancer models (for example: animal models such as mice, rats, etc).

The primary capture reagent is an antibody or a metal chelate, more preferably an antibody. The secondary capture reagent is a conjugate labeled with a coloring enzyme, a fluorescein, a radio-isotope or a colloid, which acts as a secondary antibody. The coloring enzyme can be peroxidase, alkaline phosphatase or acid phosphatase (ex: horseradish peroxidase). The fluorescein can be fluorescein carboxylic acid (FCA), fluorescein isothiocyanate (FITC), fluorescein thiourea (FTH), 7-acetoxycoumarin-3-yl, fluorescein-5-yl, fluorescein-6-yl,

2',7'-dichlorofluorescein-5-yl, 2',7'-dichlorofluorescein-6-yl, dihydrotetramethylrosamine-4-yl, tetramethylrodamine-5-yl, tetramethylrodamine-6-yl, 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-ethyl or 4,4-difluoro-5,7-diphenyl-4-bora-3a,4a-diaza-s-indacene-3-ethyl.

When a sample of a patient is exposed on the primary capture reagent, preferably a marker-specific antibody, the sample can be diluted before exposure on the antibody and the antibody can be fixed in a solid phase in order to be used in the next phases including washing or complex separation, etc. The solid phase can be glass or plastic such as microtiter plate, rod, bead or microbead, etc. The antibody can be bound to probe substrate or protein chip. After incubating the sample with the antibody, the sample was washed and incubated with the secondary capture reagent, preferably secondary antibody to measure antibody-marker complex. The measurement or detection of the antibody-marker complex can be performed by one of the processes of fluorescence, luminescence, chemiluminescence, optical density, reflection and transmission. In addition to those methods above, markers in the sample can be detected by indirect methods such as competition or inhibition test with a monoclonal antibody binding to another epitope of the marker.

The kit excludes a substrate to react with an enzyme and a non-bound protein, but includes washed solution or eluent containing bound biomarkers only. Samples for the analysis include serum, urine, tear, saliva and other biomaterials containing disease-specific polypeptide. Preferably, the samples are biological liquid samples such as blood, serum, and plasma, and more preferably serum. Samples can be prepared in order to increase the sensitivity of marker detection. For example, sera obtained from patients can be pre-treated by anion exchange chromatography, affinity chromatography, size exclusion chromatography, liquid chromatography, sequential extraction or gel electrophoresis.

【Description of Drawings】

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

Fig. 1 is a set of photographs illustrating the whole image of the representative serum proteome seen on two-dimensional gel electrophoresis and the optimal urea concentration for the two-dimensional gel electrophoresis.

1: whole image of two-dimensional gel electrophoresis

2: 8 M urea

3: 2 M thiourea/7 urea

Fig. 2 is a set of photographs illustrating the proteins exhibiting different expressions in cancer patients, compared with in normal healthy people, detected by two-dimensional gel electrophoresis.

Fig. 3 is a set of graphs illustrating the stomach cancer stage dependent protein expression.

A: Haptoglobin beta

B: Haptoglobin alpha

C: Leucine-rich alpha-2-glycoprotein(LRG)

D: A1 antitrypsin

E: ProApolipoprotein A- I

F: Apolipoprotein H

G: Clusterin

H: Apolipoprotein A-IV

I: Transthyretin

Fig. 4 is a set of photographs illustrating the expression levels of marker proteins in stomach cancer patients which are different from those in normal healthy people, detected by immunoblotting.

NS: normal serum

SC: stomach cancer serum

Haptoglobin beta

1. Marker
2. NS 0.08 μl
3. SC 0.08 μl

Haptoglobin alpha

1. his-Hp α2 50 ng
2. NS 0.04 μl
3. SC 0.04 μl
4. NS 0.08 μl
5. SC 0.08 μl
6. Haptoglobin 50 ng
7. Marker

Transthyretin

1. Marker
2. NS 0.0016 μl
3. SC 0.0016 μl
4. NS 0.0032 μl
5. SC 0.0032 μl
6. NS 0.0064 μl
7. SC 0.0064 μl
8. NS 0.0128 μl
9. SC 0.0128 μl

Alpha-1-antitrypsin

1. Marker

2. NS 0.0064 ㎕

3. SC 0.0064 ㎕

4. NS 0.0096 ㎕

5. SC 0.0096 ㎕

6. a1 antitrypsin 48 ng

【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: Two-dimensional gel electrophoresis

<1-1> Serum obtainment

Peripheral blood was obtained from stomach cancer patients (143 people) at Department of Surgery, Seoul National University College of Medicine, for two and a half years before the patients got surgery and normal peripheral blood was also obtained from normal healthy people (168 people) proved not to have stomach cancer by medical

examination at the Green Cross Reference Lab. Blood was taken by using vacutainer SST II tube (Becton Dickinson) and sera were separated by centrifugation.

【Table 1】

Serum obtein

sex	
male	female
186	125

	age						
	20	30	40	50	60	70	
cancer	1	11	21	40	45	25	143
normal	0	14	38	55	46	15	168

stage					
	1	2	3	4	
cancer	53	36	34	20	143

total 311 people

<1-2> Optimization of conditions for two-dimensional gel electrophoresis

<1-2-1> Establishment of method for two-dimensional gel electrophoresis

Various experimental conditions were checked to exhibit serum proteome. First, immobilized pH gradient (IPG) strip (Amersham Bioscience) for IEF (isoelectric focusing) was 13 cm long and pH was set at the range between 4 and 7 considering experimental time and efficient

detection of the serum proteome. During IEF, the total volt hour was 62,000 Vhr, which was proved to be optimal volt hour. The amount of each sample for loading was determined to be 200 ug and the whole image generated on two-dimensional gel electrophoresis is shown in Fig. 1-1.

<1-2-2> Establishment of a method for sample treatment

To determine the optimal composition of rehydration buffer used for the analysis of serum proteome, the condition with 8 M urea (Sigma) (Fig. 1-2) and the condition with 7 M urea/2 M thiourea (Sigma) (Fig. 1-3) were compared. As a result, when thiourea was used, vertical streaks were often seen, which made spot analysis difficult. So, the optimal condition for the buffer was determined 8 M urea (Fig. 1-2).

<1-2-3> Establishment of a method for protein staining

Silver stain is a staining method for detecting small amount of protein, which has been often used in laboratories. However, because of the difficulty in reproducibility and the covalent bond between glutaraldehyde of a reagent with a target protein, this

method cannot be used for MALDI-TOF Mass Spectrometry. To overcome the above problems, the present inventors minimized deviations of staining process by keeping recommended volume of a solution and reaction time strictly. To construct image database, the present inventors stained the protein with SyproRuby (Molecular Probe) exhibiting excellent reproducibility and quantitative results.

<1-3> Two-dimensional gel electrophoresis

200 μg of serum protein was loaded in 2% SDS(Sigma)/100 mM Dithiothreitol solution (DTT, Sigma), followed by heating at 95℃for 5 minutes. The solution was loaded in rehydration solution (8 M Urea, 4% CHAPS (Sigma), 50 mM DTT, 0.5% IPG buffer (Amersham Bioscience)), followed by stirring. Centrifugation was performed at room temperature to separate supernatant. IEF (isoelectric focusing) was performed with IPGphor system (Amersham Biosciences) using Immobilized pH Gradient (IPG) strip (pH4-7, 13 cm Amersham Bioscience). SDS-PAGE was performed by vertical electrophoresis using 12.5% polyacrylamide gel. Upon completion of electrophoresis, proteins were stained with SyproRuby (Molecular Probe), followed by detection. Gel images of the two-dimensional gel electrophoresis were analyzed by PDQuest software (Bio-Rad).

Example 2: Construction of image database

To determine optimal conditions, various factors were regulated such as image filtering, background eliminating, uneven vertical streak eliminating, uneven horizontal streak eliminating and spot variant detecting. Spots on image were detected and then specifically distinguishable 110 spots were precisely analyzed. Deviations generated during sample measurement or mechanical deviations generated during image obtainment were normalized by dividing the amount of each spot by the sum of all the spots, in order to regulate the darkness of two-dimensional gel electrophoresis image caused not by the difference of protein expression but by other factors.

Example 3: Bioinformatics analysis

To screen a marker protein in serum, which is associated with stomach cancer, bioinformatics analysis was performed.

As shown in Table 1, 100 people from each stomach cancer patient group and normal healthy people group were randomly selected out of the proteome data of total 311 people (stomach cancer patients 143, and normal healthy people 168) to form a training group. The training group

data of the selected 200 people was interlocked with support vector machine (SVM) and genetic algorithm (GA) (see Korean Patent No. 10-2002-0067298 "Method and System for Analysis of Cancer Biomarkers Using Proteome Image Mining") to screen marker protein candidates for diagnosing stomach cancer (Table 3).

The data of the remaining 111 people not included in the training group (stomach cancer patient 43, and normal healthy people 68) was tested to confirm the usability of those marker protein candidates selected from the training as a diagnostic marker for stomach cancer. Each experimental stage is described in more detail hereinafter.

【Table 2】

Distribution of sex, age and stage

(A)

training group	sex	
	male	female
	121	79

	age						
	20	30	40	50	60	70	
cancer	1	6	12	24	36	21	100
normal	0	7	27	27	27	12	100

	stage				
	1	2	3	4	
cancer	42	23	23	12	100

(B)

evaluating group	sex	
	male	female
	65	46

	age						
	20	30	40	50	60	70	
cancer	0	5	9	16	9	4	43
normal	0	7	11	28	19	3	68

	stage				
	1	2	3	4	
cancer	11	13	11	8	43

total 311 people

*Gender and age distributions of normal healthy people and stomach cancer patients volunteered for the above data analysis were presented according to the training group (A, 200 people) and the evaluating group (B, 111 people).

<3-1> Marker protein selection by SVM/GA

To screen marker proteins that facilitate the distinguishment of stomach cancer group from normal healthy group, proteome data of 100 stomach cancer patients and 100

normal healthy people were interlocked with support vector machine (V. N. Vapnik et. al., Theory of Support Vector Machines, Technical Report CSD-TR-96-17, Univ. of London, 1996.) and genetic algorithm followed by training. As a result, optimal spots (9) which would be markers for stomach cancer owing to their high training result (sensitivity: 91% and specificity: 97%) were screened.

<3-2> t-test

T-test was performed to determine whether the difference of expression of the marker protein for diagnosing stomach cancer screened in the above Example between normal healthy people and stomach cancer patients was significant enough to determine the protein as a marker. Particularly, mean values of expressions were compared between stomach cancer patients and normal healthy people by using SAS program (Statistical Analysis System Institute Inc.) and p-values were obtained. When $p<0.05$, it was judged that the difference was statistically significant. As shown in Table 3, t-test confirmed that 9 spots were screened with suggestion that they were the proteins exhibiting significant differences in expression levels between normal healthy people and stomach cancer patients.

【Table 3】

Marker protein	LRG	Clusterin	Alpha-1-antitrypsin	Apolipoprotein A-IV	Transthyretin
t-test	1.32E-07	4,24E-02	3.12E-05	3.82E-05	6.41E-07
	ProApolipoprotein A-I	Haptoglobin beta	Haptoglobin alpha-2	Apolipoprotein H	
	3.85E-08	1.16E-07	3.85E-14	3.32E-04	

<3-3> Diagnosis by Random Forest

The marker protein candidates for diagnosing stomach cancer, screened by SVM/GA in the above Example, were applied to Random Forest (results of multiple tree-classification determinants were integrated according to majority rule, which would be the final classified result, L. Breiman, "Random forests", Machine Learning, Vol. 45. Issue 1, October 2001), another bioinformatics algorithm. The result was consistent with that obtained by SVM/GA analysis above, that is the equal training group of 200 people also exhibited high decision rate of 80.0% (sensitivity 78%, and specificity 82%).

<3-4> Cross-validation

To investigate errors of the training group selected at random, cross-validation was performed by leave-one-out

(a method to presume generalized errors, in which data were divided into same sized k subsets and k-1 groups were determined to be a training group and the remaining one is determined to be an evaluating group and this classification experiment was performed k times and the average error according to k times was calculated). As a result, decision rate was 81.5% (sensitivity 78%, and specificity 85%).

<3-5> Identification of the selected proteins

Each spot on two-dimensional electrophoresis gel was picked and put in a tube containing distilled water, followed by trypsin digestion in gel, resulting in peptides. The peptides were tested at Korea Basic Science Institute, Taejon, Korea. The proteins were analyzed by MALDI-TOF-TOF (ABI 4700 Proteomics Analyzer) and the spectrums were screened by peptide mass finger print and MS/MS ion search from the database to identify the proteins (Michael O. Glocker et al., Proteomics, 4: 3921-3032, 2004).

<3-6> Verification of the selected proteins

To confirm the competence of the selected marker protein candidates for diagnosing stomach cancer to distinguish cancer from normal, random forest was performed

this time not with the training group proteome but with the evaluating group proteome (111 people who were not included in the training group; 43 stomach cancer patients, 68 normal healthy people). From the result of support vector machine algorithm was confirmed that the marker protein candidates had sensitivity of 81.39% and specificity of 77.94%. Random forest algorithm gave the result that the marker protein candidates had sensitivity of 86.05% and specificity of 82.35%. Two different algorithms consistently proved that the candidates had high sensitivity around 80%, suggesting that these candidates are very useful as marker proteins.

Example 4: Verification of marker proteins for diagnosing stomach cancer - immunoblotting

To investigate expression levels of marker proteins for diagnosing stomach cancer in normal sera and in stomach cancer patient sera, pooled serum samples were prepared from 50 people of each group: normal healthy people and stomach cancer patients, followed by immunoblotting using an antibody against each protein. As a result, as shown in Fig. 4, the marker protein levels in sera of stomach cancer patients were increased or reduced, compared with those in normal sera (Fig. 4).

1) Alpha-1-antitrypsin

Human serum was diluted with distilled water and 5x sample buffer (0.0016 ul serum/1 ul sample), which was loaded on 12% SDS-PAGE at different volumes of 1, 2, 4, and 6 ul. 48 ng of alpha-1-antitrypsin (Sigma) was loaded as a positive control, followed by electrophoresis with 25 mA. Western blotting was performed. Particularly, the gel was transferred onto PVDF membrane, followed by blocking with 5% skimmed milk/PBST (0.05% Tween 20). The membrane was treated with alpha-1-antitrypsin antibody (chicken IgY, Abcam) diluted (1:10000) in 5% skim milk/PBST (primary treatment) and anti-chicken IgY-HRP (1:20000, Abcam) (secondary treatment).

2) Haptoglobin alpha

Human serum was diluted (0.04 ul serum/1 ul sample), which was loaded on SDS-PAGE. Purified His-Hp α2 and 50 ng of haptoglobin (Sigma) were loaded as a positive control. Polyclonal anti-Haptoglobin alpha (1:1000) was treated as a primary antibody and anti-Rabbit IgG-HRP (1:20000, Sigma) was treated as a secondary antibody. The polyclonal anti-Haptoglobin alpha used as a primary antibody was prepared by injecting recombinant his-haptoglobin alpha 2 into a

rabbit, which was provided by Asan Medical Center, Seoul, Korea. (his-Hp: recombinant fusion protein labeled with haptoglobin histidine)

3) Haptoglobin beta

Human serum was diluted (0.04 ul serum/1 ul sample), which was loaded on SDS-PAGE. Polyclonal anti-Haptoglobin (1:5000, Sigma) was treated as a primary antibody and anti-mouse IgG-HRP (1:20000, Sigma) was treated as a secondary antibody.

4) Transthyretin

Human serum was diluted (0.0016 ul serum/1 ul sample), which was loaded on SDS-PAGE. Polyclonal anti-transthyretin (1:1000, Dakocytomation, Inc.) was treated as a primary antibody and anti-rabbit IgG-HRP (1:20000, Sigma) was treated as a secondary antibody.

5) ProApolipoprotein A-I

ProApolipoprotein A-I is the precursor of Apolipoprotein A-I in which 7 amino acids at N-terminal were fallen apart. An antibody that is able to distinguish ProApolipoprotein A-1 from Apolipoprotein A-1 could not be obtained or constructed, and thus it was impossible to

detect by Western blotting.

Example 5: Expression of the selected marker protein according to cancer stage

Total 143 stomach cancer patients (training group-100, evaluating group-43) were divided by stage (1-4) and expressions of 9 marker proteins were investigated according to the stage. Particularly, stomach cancer patient data were divided into 4 groups according to the stage (stage 1-53, stage 2-36, stage 3-34 and stage 4-20). Mean value of those 9 protein expressions was calculated for each group, which was divided by mean value of those 9 protein expressions in normal healthy 168 people (Fig. 3). The resultant numerical value higher than 1 indicates that the expressions of those marker proteins are increased in stomach cancer patients, while the numerical value lower than 1 indicates vise-versa. As a result, expressions of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, ProApolipoprotein A-I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H were increased in stomach cancer patients as the cancer progressed. In the meantime, expressions of Apolipoprotein A-IV and Transthyretin were reduced in stomach cancer patients as the cancer progressed. That is, those 9 marker protein

expressions differ between normal and stomach cancer and according to the stage, indicating that these markers exhibit characteristic expression patterns, increase or decrease according to the cancer stage. In conclusion, markers for diagnosing stomach cancer screened above show different expression patterns between normal and stomach cancer status and according to the cancer stage.

【Industrial Applicability】

The marker proteins for diagnosing stomach cancer of the present invention and a diagnostic kit using the same facilitate early diagnosis of stomach cancer simply by checking their expressions in body fluid and thereby facilitate quick response for the treatment. Therefore, it is expected that the marker proteins can contribute to improvement of survival rate of stomach cancer patients developed approximately at least 20,000 patients/year and to reduction of national health care costs by cancer treatment.

【Sequence List Text】

SEQ. ID. NO: 1 is a polypeptide sequence of Leucine-

rich alpha-2-glycoprotein(LRG).

SEQ. ID. NO: 2 is a polypeptide sequence of Clusterin.

SEQ. ID. NO: 3 is a polypeptide sequence of Alpha-1-antitrypsin.

SEQ. ID. NO: 4 is a polypeptide sequence of Apolipoprotein A-IV.

SEQ. ID. NO: 5 is a polypeptide sequence of Transthyretin (Prealbumin).

SEQ. ID. NO: 6 is a polypeptide sequence of ProApolipoprotein A-I.

SEQ. ID. NO: 7 is a polypeptide sequence of Haptoglobin beta.

SEQ. ID. NO: 8 is a polypeptide sequence of Haptoglobin alpha-2.

SEQ. ID. NO: 9 is a polypeptide sequence of Apolipoprotein H.

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 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】

A screening method of stomach cancer high risk group comprising the following steps:

(1) Investigating the presence or absence of one or more markers for diagnosing stomach cancer in the sample obtained from a patient, selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-Ⅳ Transthyretin, ProApolipoprotein A-Ⅰ, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H, and the expression levels or patterns of them; and

(2) Analyzing the relation of the screening result and stomach cancer progress.

【Claim 2】

The screening method of stomach cancer high risk group according to claim 1, wherein the stomach cancer status is selected from a group consisting of potential for stomach cancer, development of stomach cancer and progress of stomach cancer.

【Claim 3】

The screening method of stomach cancer high risk

group according to claim 1, wherein the markers in samples of step (1) are investigated by two-dimensional gel electrophoresis or biochip array.

【Claim 4】

The screening method of stomach cancer high risk group according to claim 3, wherein the biochip array is protein chip array or nucleic acid chip array.

【Claim 5】

The screening method of stomach cancer high risk group according to claim 1, wherein the analyzing the relation is performed by bioinformatics and statistical analysis.

【Claim 6】

The screening method of stomach cancer high risk group according to claim 1, wherein the samples are selected from a group consisting of blood, serum and blood plasma.

【Claim 7】

The screening method of stomach cancer high risk group according to claim 1, wherein the expressions of such

markers for diagnosing stomach cancer as Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, ProApolipoprotein A-Ⅰ, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H increase in stomach cancer patients, compared with in normal healthy people, but the expressions of Apolipoprotein A-Ⅳand Transthyretin decrease in stomach cancer patients.

【Claim 8】

A use of the proteins selected from a group consisting of Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-Ⅳ Transthyretin, ProApolipoprotein A-Ⅰ, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H as markers for diagnosing stomach cancer.

【Claim 9】

The use of the proteins as markers for diagnosing stomach cancer according to claim 8, wherein the markers are used independently or one or more markers are used together.

【Claim 10】

The use of the proteins as markers for diagnosing

stomach cancer according to claim 8, wherein the expressions of such markers as Leucine-rich alpha-2-glycoprotein(LRG), Clusterin, Alpha-1-antitrypsin, ProApolipoprotein A- I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H increase in stomach cancer patients, compared with in normal healthy people, but the expressions of Apolipoprotein A-IVand Transthyretin decrease in stomach cancer patients.

【Claim 11】

A diagnostic kit for stomach cancer containing (1) primary capture reagent binding to one or more markers for diagnosing stomach cancer selected from a group consisting of Leucine-rich alpha-2-glycoprotein (LRG), Clusterin, Alpha-1-antitrypsin, Apolipoprotein A-IV Transthyretin, ProApolipoprotein A- I, Haptoglobin beta, Haptoglobin alpha-2 and Apolipoprotein H; and (2) secondary capture reagent which was not bound to the first capture reagent.

【Claim 12】

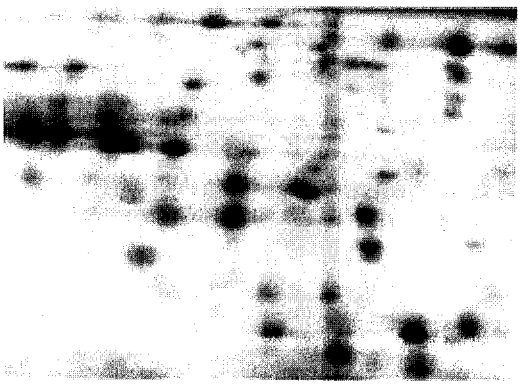
The diagnostic kit for stomach cancer according to claim 11, wherein the primary capture reagent is an antibody or a metal chelate.

【Claim 13】

The diagnostic kit for stomach cancer according to claim 11, wherein the secondary capture reagent is a conjugate labeled with a coloring enzyme, a fluorescein, a radio-isotope or a colloid.

[Fig. 1]

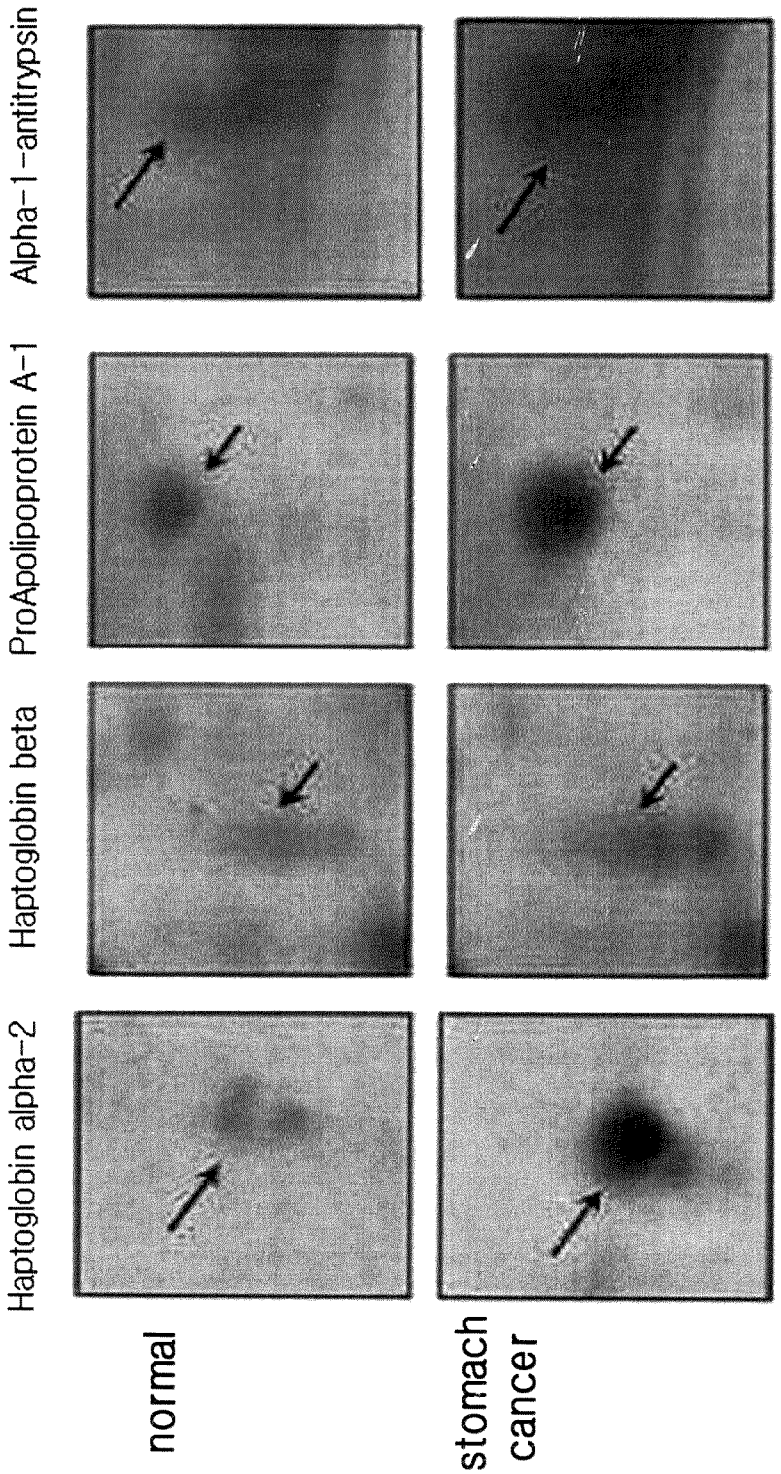
1



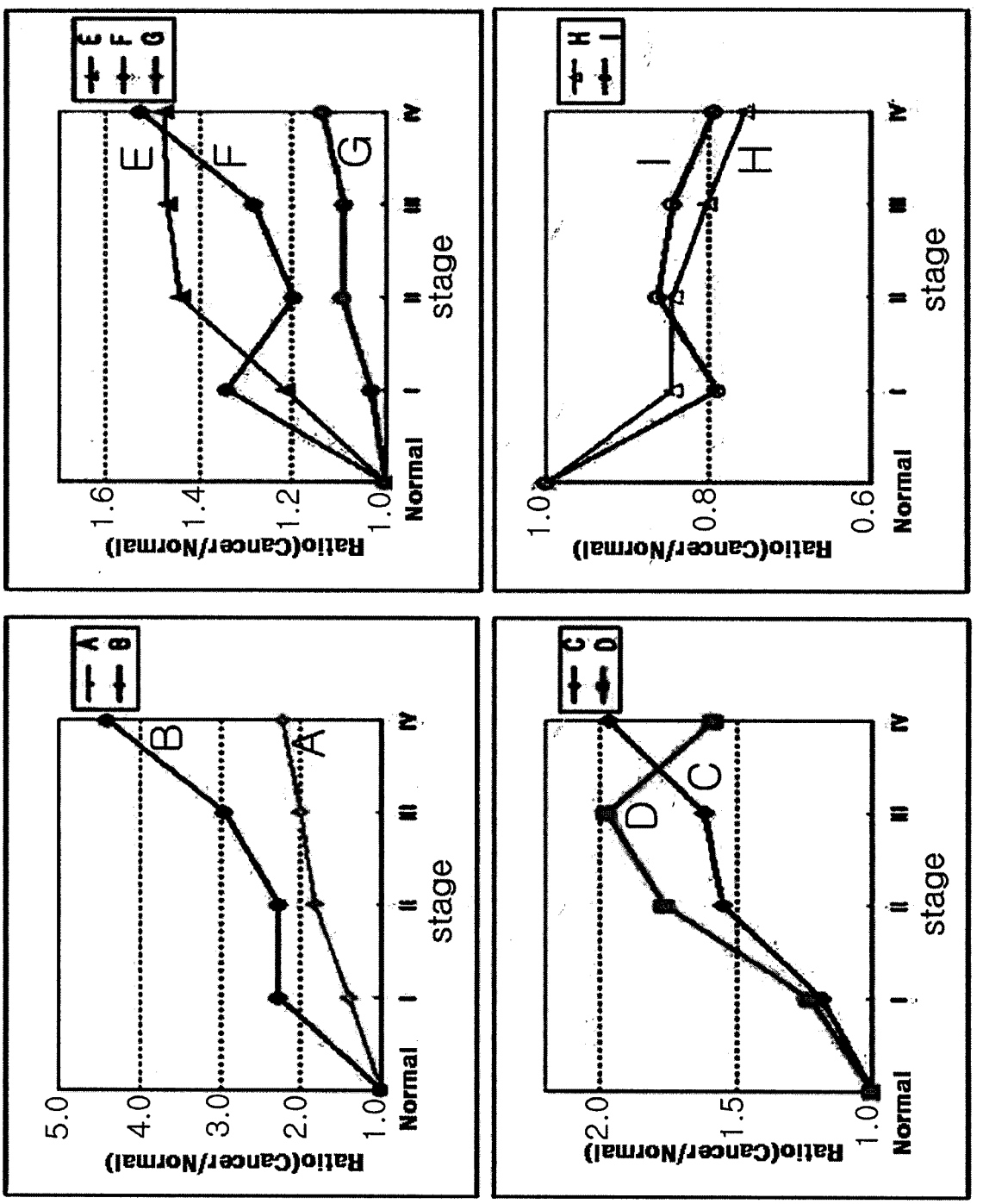
2

3

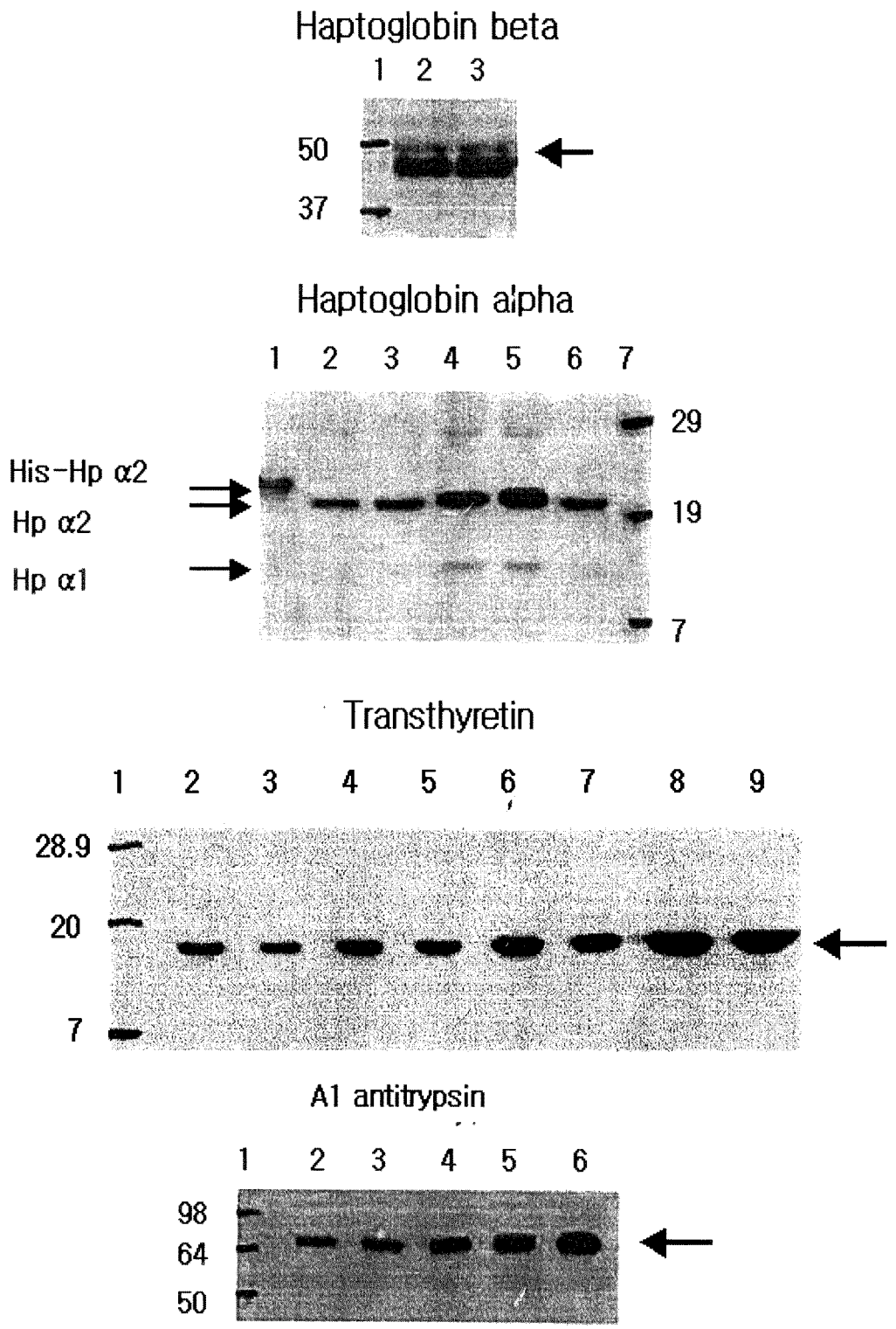
[Fig. 2]


Alpha-1-antitrypsin
ProApolipoprotein A-1
Haptoglobin beta
Haptoglobin alpha-2
normal
stomach cancer

[Fig. 3]


E
F
G
1.6
1.4
1.2
1.0
Ratio(Cancer/Normal)
Normal I II III IV
stage
H
I
1.0
0.8
0.6
Ratio(Cancer/Normal)
Normal I II III IV
stage
A
B
5.0
4.0
3.0
2.0
1.0
Ratio(Cancer/Normal)
Normal I II III IV
stage
C
D
2.0
1.5
1.0
Ratio(Cancer/Normal)
Normal I II III IV
stage

[Fig. 4]


Haptoglobin beta
1 2 3
50
37
Haptoglobin alpha
1 2 3 4 5 6 7
His-Hp α2
Hp α2
Hp α1
29
19
7
Transthyretin
1 2 3 4 5 6 7 8 9
28.9
20
7
A1 antitrypsin
1 2 3 4 5 6
98
64
50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2006/005836

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of :

   a. type of material
   - [x] a sequence listing
   - [ ] table(s) related to the sequence listing

   b. format of material
   - [ ] on paper
   - [x] in electronic form

   c. time of filing/furnishing
   - [ ] contained in the international application as filed
   - [x] filed together with the international application in electronic form
   - [ ] furnished subsequently to this Authority for the purposes of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2006/005836

A. CLASSIFICATION OF SUBJECT MATTER

G01N 33/574(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)
IPC8 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
yesKISTI(Korean Science Technology Information DB)

Electronic data base consulted during the intertnational search (name of data base and, where practicable, search terms used)
e-KIPASS, Goolgle Scholar, NCBI ("gastric cancer, diagnosis, marker, proteome etc)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1394182 A1(Europroteome AG) 03. 03. 2004 See page 6 lines 35-53	1 -13
X	J. S. Kim "Identification of a possible marker protein for gastric cancer by proteomic analysis" in Master Thesis of College of Pharmacy, Choong-Ang University (Korea), 2002 See Abstract, Fig. 3 and Fig.5	1 -13
X	J. W. Ryu et al., "The proteomics approach to find biomarkers in gastric cancer" in J. Korean Med. Sci. 2003 Vol.18 pages 505-509 See abstract, Table 3 in page 507	1 - 13
X	JS Jang et al., "The differential proteome profile of stomach cancer : identification of the biomarker candidates" in Oncol. Res. 2004 Vol.14(10) pages 491-499 See abstract	1 -13

☐ 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	Date of mailing of the international search report
27 MARCH 2007 (27.03.2007)	**28 MARCH 2007 (28.03.2007)**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office 920 Dunsan-dong, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	CHO, Myung Sun Telephone No. 82-42-481-5594

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2006/005836

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2004121343 A1 (BIosite incorporated) 24. 06. 2004 See entire document	1 -13
A	QY He and JF Chiu "Proteomics in biomarker discovery and drug development" in J. Cell Biochem. 2003 Aug. Vol.89(5) pages 868-886 See abstract	1 - 13
A	QY He et al.,"Diverse proteomic alterations in gastric adenocarcinoma" in Proteomics 2004 Oct. Vol.4(10) pages 3276-3287 See abstract	1 -13

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/KR2006/005836

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP1394182A1	03.03.2004	AU2003250912A1	23.01.2004
		WO2004005928A2	15.01.2004
US2004121343A1	24.06.2004	AU2003302340A1	22.07.2004
		CA2511501AA	15.07.2004
		CN1751128A	22.03.2006
		EP01587955A2	26.10.2005
		JP2006526140T2	16.11.2006
		WO2004059293A2	15.07.2004