

DESCRIPTION

Invention Title

NOVEL BIOMARKER FOR DIAGNOSIS OF OVARIAN CANCER

5 **Technical Field**

[0001] The present invention relates to a method of predicting the onset of ovarian cancer by measuring the expression level of a factor, which is specifically regulated in ovarian cancer, or a method of preventing or treating ovarian cancer by regulating the expression of the factor.

10 **[0002]**

Background Art

[0003] Ovarian cancer (OC) is the most lethal gynecological cancer and is responsible for 5% of all cancer-related deaths in women worldwide. Because there are no symptoms specific to ovarian cancer and due to the lack of effective screening tools, most OC patients are diagnosed at a fairly advanced stage with a 5-year survival rate of less than 30%. However, if ovarian cancer can be diagnosed at an early stage, the 5-year survival rate can be increased up to 90%.

[0004] High-grade serous ovarian cancer (HGSOC) is the most common histological type of ovarian cancer, accounting for 70% of all OC cases. HGSOC is the most aggressive type of ovarian cancer, and is often diagnosed at a fairly advanced stage due to the rapid tumor growth. In particular, the overall survival rate of HGSOC patients has barely improved for decades, and thus there is a great clinical need for diagnosing HGSOC at an early stage with high accuracy. However, currently, highly reliable biomarkers that can be used to accurately diagnose HGSOC with high sensitivity and specificity have not been discovered.

[0005] Meanwhile, blood has been the focus of extensive cancer biomarker research due to the minimally invasive nature of liquid biopsy. Previous studies have identified potential blood biomarkers for OC ranging from microRNA to exosomes, and proteins. Measurement of the cancer antigen CA-125 in the blood and human epididymis protein-4 (HE4) is currently used as the gold standard for OC detection (diagnosis) and monitoring. However, although these biomarkers are currently widely used in clinical settings, there are limitations in reducing the mortality rate of OC patients due to low sensitivity and low specificity of these biomarkers. For this reason, diagnostic method options for OC patients are significantly limited, and thus there is an urgent need to discover new additional biomarkers for diagnosing OC.

[0006] Proteome-wide analysis is a useful method for discovering and prioritizing new biomarkers that have not been previously identified. Mass spectrometry (MS) is efficiently used for comprehensive research in proteomics because it can provide unbiased profiling of the human proteome. Recently, MS analysis based on data-independent acquisition (DIA) has emerging as an analysis methodology for next-generation proteomics. DIA is based on the fragmentation of all precursor ions in a large isolation window spanning the entire range of mass-to-charge values (m/z), and has advantages over conventional data-dependent acquisition (DDA), such as improved sensitivity, improved reproducibility, and minimal fractionation requirements.

[0007] Accordingly, the present inventors have analyzed the serum proteome using DIA-based mass spectrometry to discover efficient diagnostic biomarkers for ovarian cancer, and as a result, have discovered ovarian cancer-specific biomarkers in the blood that can diagnose ovarian cancer at an early stage with high accuracy, and also evaluated the effects of the biomarkers.

[0008] Throughout the present specification, a number of publications and patent

documents are referred to and cited. The disclosure of the cited publications and patent documents is incorporated herein by reference in its entirety to more clearly describe the state of the art to which the present invention pertains and the content of the present invention.

5 **[0009]**

DISCLOSURE

Technical Problem

[0010] The present inventors have made extensive research efforts to discover efficient diagnostic biomarkers for ovarian cancer, which has the poorest prognosis among
10 gynecological cancers and has a low patient survival rate due to the lack of early diagnosis methods, thereby developing a new diagnostic method that can significantly reduce the mortality rate from ovarian cancer, especially high-grade serous ovarian cancer. As a result, the present inventors have discovered proteins or genes encoding the same that are up- or down-regulated specifically in ovarian cancer patients, and have
15 found that ovarian cancer can be diagnosed at an early stage with high accuracy by measuring the expression levels of these markers, and these markers are highly reliable markers that can also be applied to mass spectrometry-based diagnostic methods using blood, thereby completing the present invention.

[0011] Therefore, an object of the present invention is to provide a composition for
20 diagnosing ovarian cancer.

[0012] Another object of the present invention is to provide a method of providing information necessary for diagnosis of ovarian cancer.

[0013] Still another object of the present invention is to provide a composition for preventing or treating ovarian cancer.

25 **[0014]** Yet object of the present invention is to provide a method for screening a

composition for preventing or treating ovarian cancer.

[0015] Other objects and advantages of the present invention will be more apparent from the following detailed description, the appended claims and the accompanying drawings.

5 **[0016]**

Technical Solution

[0017] According to one aspect of the present invention, the present invention provides a composition for diagnosing ovarian cancer comprising, as an active ingredient, an agent for measuring the expression level of at least one gene selected from
10 the group consisting of *FGA*, *VWF*, *ARHGDIB*, and *SERPINF2*, or a protein encoded thereby.

[0018] The present inventors have made extensive research efforts to discover efficient diagnostic biomarkers for ovarian cancer, which has the poorest prognosis among gynecological cancers and has a low patient survival rate due to the lack of early
15 diagnosis methods, thereby developing a new diagnostic method that can ultimately significantly reduce the mortality rate from ovarian cancer, especially high-grade serous ovarian cancer. As a result, the present inventors have discovered proteins or genes encoding the same that are up- or down-regulated specifically in ovarian cancer patients, and have found that ovarian cancer can be diagnosed at an early stage with high
20 accuracy by measuring the expression levels of these markers, and these markers are highly reliable markers that can also be applied to mass spectrometry-based diagnostic methods using blood, thereby completing the present invention.

[0019] In the present specification, the term “*FGA*” is the name of the gene encoding fibrinogen alpha chain protein in humans. The above-described fibrinogen alpha chain
25 protein may also be referred to as FIBA, but is not limited thereto and may be referred to

by any aliases or synonyms commonly used as the name of the protein.

[0020] In the present specification, the term “VWF” is the name of the gene encoding von Willebrand factor protein in humans. The von Willebrand factor protein described above may also be referred to as VWF, but is not limited thereto and may be referred to

5 by any aliases or synonyms commonly used as the name of the protein.

[0021] In the present specification, the term “*ARHGDIB*” is the name of the gene encoding Rho GDP-dissociation inhibitor 2 protein in humans. The Rho GDP-dissociation inhibitor 2 protein described above may also be referred to as RhoGD12, but is not limited thereto and may be referred to by any aliases or synonyms commonly

10 used as the name of the protein.

[0022] In the present specification, the term “*SERPINF2*” is the name of the gene encoding alpha 2-antiplasmin protein in humans. The alpha 2-antiplasmin protein described above may be referred to as A2AP, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

15 **[0023]** All genes used in the present specification may be used independently or in combination of two or more as biomarkers for the diagnosis of ovarian cancer, and when they used in combination, the set of the genes may be a “biomarker panel”.

[0024] In the present specification, the term “biomarker panel” may also be referred to as “biomarker detection panel”, and means a set of at least two biomarkers that may be

20 used for the detection, diagnosis, determination of prognosis, staging, or monitoring of a disease or condition. The biomarker components of this biomarker set may be physically associated, such as by being packaged together, or by being reversibly or irreversibly bound to a solid support. For example, the biomarker panel of the present invention may be provided, in separate tubes that are sold or shipped together, for example as part of a
25 kit, or can be provided on a chip, membrane, strip, filter, or beads, particles, filaments,

fibers, or other supports, in or on a gel or matrix, or bound to the wells of a multi-well plate. However, the biomarker panel of the present invention is not limited to the above examples and includes all type of biomarker panel that may be used as a combination of biomarkers for diagnosis of diseases.

5 **[0025]** In the present specification, the term “diagnosis” or “diagnosing” includes determination of the susceptibility of a subject to a particular disease, determination of whether a subject currently has a particular disease, and determination of the prognosis of a subject suffering from a particular disease.

10 **[0026]** In the present specification, the term “composition for diagnosing” refers to an integrated mixture or device including a means for measuring the expression level of at least one gene selected from the group consisting of *FGA*, *VWF*, *ARHGDIB* and *SERPINF2*, or a protein encoded thereby, in order to determine whether or not a subject has developed ovarian cancer or to predict the likelihood of onset of ovarian cancer in a subject. Therefore, the term may also be expressed as a “kit for diagnosing”.

15 **[0027]** According to a specific embodiment of the present invention, the agent for measuring the expression level of the gene that is used as the marker in the present invention is a primer or probe that binds specifically to the nucleic acid molecule of the gene.

20 **[0028]** In the present specification, the term “nucleic acid molecule” is meant to encompass DNA (gDNA and cDNA) and RNA molecules. Nucleotides, which are the basic structural units in nucleic acid molecules, include not only natural nucleotides, but also analogues having modified sugar or base moieties (Scheit, *Nucleotide Analogs*, John Wiley, New York (1980); Uhlman and Peyman, *Chemical Reviews*, 90:543-584 (1990)).

25 **[0029]** As used in the present specification, the term “primer” refers to an

oligonucleotide which serves as a starting point for synthesis under conditions in which the synthesis of a primer extension product complementary to a nucleic acid chain (template) is induced, i.e., the presence of nucleotides and a polymerase such as DNA polymerase, and suitable temperature and pH. Specifically, the primer is a deoxyribonucleotide single chain. The primers that are used in the present invention may include naturally occurring dNMPs (i.e., dAMP, dGMP, dCMP and dTMP), modified nucleotides or non-natural nucleotides. In addition, the primers may also include ribonucleotides.

[0030] The primer of the present invention may be an extension primer which is annealed to a target nucleic acid and forms a sequence complementary to the target nucleic acid by a template-dependent nucleic acid polymerase. The extension primer is extended to the location where an immobilized probe is annealed and occupies the area at which the probe is annealed.

[0031] The extension primer that is used in the present invention includes a hybridizing nucleotide sequence complementary to a specific nucleotide sequence of a target nucleic acid, e.g., the gene that is used as the marker in the present invention. The term “complementary” means that a primer or a probe is complementary enough to selectively hybridize to a target nucleic acid sequence under a certain annealing or hybridization condition. The term “complementary” encompasses both “substantially complementary” and “perfectly complementary”, and specifically, it means perfectly complementary. In the present specification, the term “substantially complementary sequence” is meant to include not only a completely matching sequence, but also a sequence partially mismatching with the sequence to be compared within a range where it can anneal to a particular sequence and serve as a primer.

[0032] The primer should be long enough to prime the synthesis of an extension

product in the presence of a polymerase. The suitable length of the primer is determined depending on a number of factors, such as temperature, pH and the source of the primer, but is typically 15 to 30 nucleotides. Short primer molecules generally require lower temperatures to form a sufficiently stable hybrid complex with a template. The design of such a primer can be easily carried out by those skilled in the art with reference to a target nucleotide sequence using, for example, a program for primer design (e.g., PRIMER 3 program).

[0033] In the present specification, the term "probe" refers to a natural or modified monomer or a linear oligomer having linkages, which includes a deoxyribonucleotide and ribonucleotide that can be hybridized with a specific nucleotide sequence. Specifically, the probe is single-stranded for maximum efficiency in hybridization. More specifically, the probe is a deoxyribonucleotide. As the probe that is used in the present invention, a sequence perfectly complementary to a specific nucleotide sequence of the gene that is used as the marker in the present invention gene may be used, but a sequence substantially complementary thereto may also be used as long as specific hybridization is not interrupted. In general, the stability of a duplex formed by hybridization tends to be determined by the match of terminal sequences, and thus it is preferred to use a probe which is complementary to the 3'- end or 5'-end of the target sequence.

[0034] Conditions suitable for hybridization may be determined with reference to the contents disclosed in Joseph Sambrook, et al., *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y. (2001), and Haymes, B. D., et al., *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, D.C.(1985).

[0035] According to a specific embodiment of the present invention, the agent for measuring the expression level of the protein encoded by the gene that is used as the

marker in the present invention is an antibody or an antigen-binding fragment thereof that binds specifically to the protein, or an aptamer that binds specifically to the protein.

[0036] According to the present invention, the protein encoded by the gene that is used as the marker in the present invention may be detected according to an immunoassay method using an antigen-antibody reaction and used to analyze whether ovarian cancer has occurred. This immunoassay may be performed according to various immunoassay or immunostaining protocols which have been conventionally developed.

[0037] For example, when the method of the present invention is performed according to a radioimmunoassay method, antibodies labeled with radioisotopes (e.g., C¹⁴, I¹²⁵, P³², and S³⁵) may be used. In the present invention, the antibody that specifically recognizes the protein used as the marker is a polyclonal or monoclonal antibody, preferably a monoclonal antibody.

[0038] The antibody of the present invention may be produced by methods commonly practiced in the art, for example, the fusion method (Kohler and Milstein, *European Journal of Immunology*, 6:511-519 (1976)), the recombinant DNA method (US Patent No. 4,816,567), or the phage antibody library method (Clackson et al, *Nature*, 352:624-628 (1991), and Marks et al, *J. Mol. Biol.*, 222:58, 1-597 (1991)). General procedures for antibody production are described in detail in Harlow, E. and Lane, D., *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York, 1999; and Zola, H., *Monoclonal Antibodies: A Manual of Techniques*, CRC Press, Inc., Boca Raton, Florida, 1984.

[0039] By analyzing the intensity of the final signal by the above-described immunoassay process, it is possible to determine whether ovarian cancer has occurred or to predict the likelihood of onset of ovarian cancer. In other words, if the signal for the above-described protein in a sample from a subject is stronger than that in a normal

sample, it is determined that the subject has developed ovarian cancer or is highly likely to develop ovarian cancer in the future.

[0040] In the present specification, the term “antigen-binding fragment” refers to a portion of a polypeptide, to which an antigen can bind, among the entire structure of an immunoglobulin. Examples of the antigen-binding fragment include, but are not limited to, F(ab')₂, Fab', Fab , Fv, and scFv.

[0041] In the present specification, the term “specifically binding” has the same meaning as “specifically recognizing”, and means that an antigen and an antibody (or a fragment thereof) specifically interact with each other by an immunological reaction.

[0042] In the present invention, an aptamer that specifically binds to the above-described protein may be used instead of an antibody. In the present specification, the term “aptamer” refers to a single-stranded nucleic acid (RNA or DNA) molecule or peptide molecule that binds to a specific target substance with high affinity and specificity. General contents regarding aptamers are disclosed in detail in Hoppe-Seyler F, Butz K "Peptide aptamers: powerful new tools for molecular medicine". *J Mol Med.* 78(8):426-30(2000); Cohen BA, Colas P, Brent R. "An artificial cell-cycle inhibitor isolated from a combinatorial library". *Proc Natl Acad Sci USA.* 95(24):14272-7(1998).

[0043] In the present invention, the markers that are used for diagnosing ovarian cancer are full-length proteins which may become targets for diagnosing ovarian cancer.

In addition, fragments of the full-length proteins, which are shown in Table 2 or SEQ ID NOs. 1 to 21 of the present specification, may also become targets for diagnosing ovarian cancer. In this case, in addition to mass spectrometry, any types of protein or peptide fragment detection methods, including the protein measurement method described above, may be used.

[0044]

[0045] According to a specific embodiment of the present invention, the above-described composition for diagnosing further comprises, as an active ingredient, an agent for measuring the expression level of at least one gene selected from the group consisting of *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*,
5 *LBP*, *PPBP*, *C3*, *FCGBP*, *CD14*, *APOA1* and *APOA4*, or a protein encoded thereby.

[0046] In the present specification, the term “*S100A9*” is the name of the gene encoding calcium-binding protein A9 protein in humans. The calcium-binding protein A9 protein described above may be referred to as *S100A9*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the
10 protein.

[0047] In the present specification, the term “*VCL*” is the name of the gene encoding vinculin protein in humans. The vinculin protein described above may be referred to as *VCL*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

15 **[0048]** In the present specification, the term “*THBS1*” is the name of the gene encoding thrombospondin 1 protein in humans. The thrombospondin 1 protein described above may be referred to as *TSP1*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0049] In the present specification, the term “*MPO*” is the name of the gene encoding
20 myeloperoxidase protein in humans. The myeloperoxidase protein described above may be referred to as *MPO*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0050] In the present specification, the term “*IGFBP2*” is the name of the gene encoding insulin-like growth factor-binding protein 2 in humans. The insulin-like
25 growth factor-binding protein 2 described above may be referred to as *IBP2*, but is not

limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0051] In the present specification, the term “*SRGN*” is the name of the gene encoding serglycin protein in humans. The serglycin protein described above may be referred to as
5 *SRGN*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0052] In the present specification, the term “*PF4*” is the name of the gene encoding platelet factor 4 protein in humans. The platelet factor 4 protein described above may be referred to as *PF4*, but is not limited thereto and may be referred to by any aliases or
10 synonyms commonly used as the name of the protein.

[0053] In the present specification, the term “*GP1BA*” is the name of the gene encoding platelet glycoprotein Ib alpha chain protein in humans. The platelet glycoprotein Ib alpha chain protein described above may be referred to as *GP1BA*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as
15 the name of the protein.

[0054] In the present specification, the term “*LRG1*” is the name of the gene encoding leucine-rich alpha-2-glycoprotein 1 protein in humans. The leucine-rich alpha-2-glycoprotein 1 protein described above may be referred to as *LRG1*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name
20 of the protein.

[0055] In the present specification, the term “*PRG4*” is the name of the gene encoding the proteoglycan 4 protein in humans. The proteoglycan 4 protein described above may be referred to as *PRG4*, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0056] In the present specification, the term “*LBP*” is the name of a gene encoding

lipopolysaccharide binding protein in humans. The lipopolysaccharide binding protein described above may be referred to as LBP, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0057] In the present specification, the term “*PPBP*” is the name of the gene encoding pro-platelet basic protein in humans. The pre-platelet basic protein described above may be referred to as PPBP, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0058] In the present specification, the term “*C3*” is the name of the gene encoding complement component 3 protein in humans. The complement component 3 protein described above may be referred to as C3, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0059] In the present specification, the term “*FCGBP*” is the name of the gene encoding IgGFc-binding protein in humans. The IgGFc-binding protein described above may be referred to as FCGBP, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0060] In the present specification, the term “*CD14*” is the name of a gene encoding the cluster of differentiation 14 protein in humans. The cluster of differentiation 14 protein above described above may be referred to as CD14, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0061] In the present specification, the term “*APOA1*” is the name of the gene encoding apolipoprotein A1 protein in humans. The apolipoprotein A1 protein described above may be referred to as APOA1, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0062] In the present specification, the term “*APOA4*” is the name of the gene

encoding apolipoprotein A4 protein in humans. The apolipoprotein A4 protein described above may be referred to as APOA4, but is not limited thereto and may be referred to by any aliases or synonyms commonly used as the name of the protein.

[0063]

5 **[0064]** According to a specific embodiment of the present invention, the expression level of the *FGA*, *VWF* or *ARHGDIB* gene or the protein encoded thereby in a subject with ovarian cancer increases, and the expression level of the *SERPINF2* gene or the protein encoded thereby in the subject decreases.

[0065] In the present invention, the phrase “expression level ... increases” as used
10 while referring to “the composition for diagnosing ovarian cancer” means that the expression level of the gene or the protein encoded by the gene is significantly higher than that in a control group or a normal group. Specifically, the phrase means that the expression level increased by about 10% or more, about 20% or more, about 30% or more, about 40% or more, about 50% or more, or about 60% or more compared to that
15 in the control group or normal group, without being limited thereto.

[0066] In the present invention, the phrase “expression level ... decreases” as used while referring to “the composition for diagnosing ovarian cancer” means that the expression level of the gene or the protein encoded by the gene is significantly lower than that in a control group or a normal group. Specifically, the phrase means that the
20 expression level decreased by about 10% or more, about 20% or more, about 30% or more, about 40% or more, about 50% or more, or about 60% or more compared to that in the control group or normal group, without being limited thereto.

[0067]

[0068] According to a specific embodiment of the present invention, the expression
25 level of the *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*,

LBP, *PPBP*, *C3*, *FCGBP* or *CD14* gene or the protein encoded thereby in a subject with ovarian cancer increases, and the expression level of the *APOA1* or *APOA4* gene or the protein encoded thereby in the subject decreases.

[0069] The genes used in the present invention, the proteins encoded thereby, and the phrases “expression level... increases” and “expression level ... decreases” of the present invention have already been described in detail, and thus description thereof will be omitted to prevent excessive overlapping.

[0070]

[0071] According to a specific embodiment of the present invention, the ovarian cancer that may be diagnosed using the composition of the present invention is high-grade serous ovarian cancer (HGSOC).

[0072] In the present specification, the term “high-grade serous ovarian cancer (HGSOC)” refers to one of several subtypes of ovarian cancer, which constitutes the three major subtypes of ovarian cancer, along with the clear cell and endometrioid subtypes. HGSOC is the most malignant form of ovarian cancer and a subtype that accounts for up to 70% of all ovarian cancer cases. It is known that, because HGSOC originates in the fallopian tubes, it spreads through the abdomen early in the course of the disease, and at the time when symptoms appear, it is usually a tumor at a serious stage in many cases, indicating poor prognosis.

[0073]

[0074] According to another aspect of the present invention, the present invention provides a method of providing information necessary for diagnosing ovarian cancer, comprising a step of measuring the expression levels of *FGA*, *VWF*, *ARHGDIB* and *SERPINF2* genes or proteins encoded thereby in a biological sample isolated from a subject.

[0075] The present inventors have found for the first time that the expression levels of FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP and CD14 proteins or genes encoding these proteins in patients with ovarian cancer, especially high-grade serous ovarian cancer, are positively correlated with the likelihood of onset of ovarian cancer, and the expression levels of APOA1, APOA4 and SERPINF2 proteins or genes encoding these proteins in the patients are negatively correlated with the likelihood of onset of ovarian cancer. Accordingly, if the above-described positively correlated proteins or genes encoding them are highly expressed in a subject or a biological sample isolated from the subject, or if the above-described negatively correlated proteins or genes encoding them are under-expressed in the subject or the biological sample, the subject is determined to be a subject who has developed ovarian cancer or is likely to develop ovarian cancer in the future.

[0076] In the present specification, the term “highly expressed” means that the expression level of the gene of interest or the protein encoded by the gene is significantly higher than that in a control group or a normal group. Specifically, the term “highly expressed” means that the expression level increased by about 10% or more, about 20% or more, about 30% or more, about 40% or more, about 50% or more, or about 60% or more compared to that in the control group or the normal group, without being limited thereto.

[0077] In the present specification, the term “under-expressed” means that the expression level of the gene of interest or the protein encoded by the gene is significantly lower than that in a control group or a normal group. Specifically, the term “under-expressed” means that the expression level decreased by about 10% or more, about 20% or more, about 30% or more, about 40% or more, about 50% or more, or about 60% or

more compared to that in the control group or the normal group, without being limited thereto.

[0078] In the present specification, the term “subject” refers to a subject that provides a sample for which the expression level of the gene or the protein encoded thereby of the present invention is to be measured, and which is ultimately to be analyzed as to whether ovarian cancer has developed. Subjects include, without limitation, humans, mice, rats, guinea pigs, dogs, cats, horses, cows, pigs, monkeys, chimpanzees, baboons, or rhesus monkeys, and are specifically humans. Since the composition of the present invention provides information for not only determining whether ovarian cancer has occurred but also predicting the genetic risk of onset of ovarian cancer in the future, the subject of the present invention may be an ovarian cancer patient or a healthy subject who has not yet developed ovarian cancer.

[0079]

[0080] According to a specific embodiment of the present invention, the step of measuring the expression level is performed by mass spectrometry (MS).

[0081] In this specification, the term “mass spectrometry” may also be referred to as “mass spectrometry analysis” and refers to a method of identifying unknown compounds based on their mass spectra. Mass analysis is performed by filtering, detecting, and measuring ions using the m/z value, which is the mass-to-charge ratio. Generally, mass spectrometry includes steps of: (1) charging a compound by ionization; and (2) measuring the molecular weight of the charged compound and calculating the m/z value. The calculated m/z value is used as a reference to identify and quantify the target compound in a complex mixture. The mass spectrometry in the present specification may be performed through any type of mass spectrometry method that uses the above-described principle.

[0082]

[0083] According to a specific embodiment of the present invention, the mass spectrometry measures the expression level of at least one peptide selected from the group consisting of SEQ ID NOs: 1, 2, 12, and 21.

5 **[0084]** In the present invention, mass spectrometry for proteins used as biomarkers of ovarian cancer may be specifically performed through multiple-reaction monitoring (MRM) or parallel reaction monitoring (PRM) of the peptide fragments represented by SEQ ID NOs: 1 to 21 below, without being limited thereto:

[0085] SEQ ID NO. 1 (FGA): HPDEAAFFDTASTGK

10 **[0086]** SEQ ID NO. 2 (VWF): HIVTFDGQNFK

[0087] SEQ ID NO. 3 (S100A9): NIETIINTFHQYSVK

[0088] SEQ ID NO. 4 (VCL): LLAVAATAPPDAPNR

[0089] SEQ ID NO. 5 (THBS1): TIVTTLQDSIR

[0090] SEQ ID NO. 6 (MPO): VVLEGGIDPILR

15 **[0091]** SEQ ID NO. 7 (IGFBP2): LIQGAPTIR

[0092] SEQ ID NO. 8 (SRGN): IQDLNR

[0093] SEQ ID NO. 9 (PF4): TTSQVRPR

[0094] SEQ ID NO. 10 (GP1BA): L TSLPLGALR

[0095] SEQ ID NO. 11 (LRG1): DLLLPQPDLR

20 **[0096]** SEQ ID NO. 12 (ARHGDIB): TLLGDGPVVTDPK

[0097] SEQ ID NO. 13 (PRG4): DQYYNIDVPSR

[0098] SEQ ID NO. 14 (LBP): ITLPDFTGDLR

[0099] SEQ ID NO. 15 (PPBP): NIQSLEVIGK

[00100] SEQ ID NO. 16 (C3): TGLQEVEVK

25 **[00101]** SEQ ID NO. 17 (FCGBP): GNPAVSYYR

[00102] SEQ ID NO. 18 (CD14): FPAIQNLALR

[00103] SEQ ID NO. 19 (APOA1): VQPYLDDFQK

[00104] SEQ ID NO. 20 (APOA4): QLTPYAQR

[00105] SEQ ID NO. 21 (SERPINF2): EDFLEQSEQLFGAK.

5 **[00106]**

[00107] In the present specification, the term “multiple-reaction monitoring” refers to an analysis technique capable of monitoring a change in concentration of a specific analyte by selectively isolating, detecting and quantifying the specific analyte. MRM is a method that can quantitatively and accurately measure multiple substances such as trace
10 amounts of biomarkers present in a biological sample. In MRM, mother ions among the ion fragments generated in an ionization source are selectively transmitted to a collision tube by a first mass filter (Q1). Then, the mother ions arriving at the collision tube collide with an internal collision gas, are fragmented to generate daughter ions which are then sent to a second mass filter (Q2), where only characteristic ions are transmitted to a
15 detection unit. MRM is an analysis method with high selectivity and sensitivity that can detect only information on a component of interest as described above. MRM is used for quantitative analysis of small molecules and is used to diagnose specific genetic diseases. The MRM method has advantages in that it is easy to simultaneously measure multiple peptides, and it is possible to confirm the relative concentration difference of protein
20 diagnostic marker candidates between a normal person and a cancer patient without using an antibody. In addition, since the MRM analysis method has excellent sensitivity and selectivity, it has been introduced for the analysis of complex proteins and peptides in blood, particularly in proteomic analysis using a mass spectrometer (Anderson L. et al., Mol CellProteomics, 5: 375-88, 2006; DeSouza, L. V. et al., Anal. Chem., 81: 3462-
25 70, 2009).

[00108] In the present specification, the term “parallel reaction monitoring” refers to the parallel application of MRM. Unlike MRM, which analyzes one pair of mother ions/daughter ions at a time, parallel reaction monitoring is a method of simultaneously analyzing all daughter ions generated from one selected mother ion.

5 **[00109]**

[00110] Mass spectrometry for the proteins used as biomarkers of ovarian cancer in the present invention may be specifically performed through data-independent acquisition analysis (DIA). In the present specification, the term “data-independent acquisition” refers to a method of analyzing all ions belonging to the m/z values in a selected range
10 without the process of selecting a specific mother ion.

[00111]

[00112] According to a specific embodiment of the present invention, the above-described method of providing information necessary for diagnosing ovarian cancer further comprises a step of measuring the expression level of at least one gene selected
15 from the group consisting of *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*, *LBP*, *PPBP*, *C3*, *FCGBP*, *CD14*, *APOA1* and *APOA4*, or a protein encoded thereby.

[00113] According to a specific embodiment of the present invention, the step of measuring the expression level is performed by mass spectrometry (MS).

20 **[00114]** According to a specific embodiment of the present invention, the mass spectrometry measures the expression level of at least one peptide selected from the group consisting of SEQ ID NOs: 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19 and 20.

[00115]

25 **[00116]** According to a specific embodiment of the present invention, the biological

sample is blood.

[00117] In the present specification, the term “blood” means “whole blood”, and such blood includes plasma and serum.

[00118] In the present specification, the term “whole blood” generally refers to blood
5 composed of uncoagulated plasma and cellular components. Plasma may make up about 50 to 60% of the whole blood volume, and cellular components (e.g., red blood cells, white blood cells, or platelets) may make up about 40 to 50%.

[00119] The term “plasma” as used in the present specification refers to the liquid component of blood that functions as a transport medium in supplying nutrients to cells
10 and organs of the body.

[00120] The term “serum” as used in the present specification refers to a light yellow liquid separated from blood. Specifically, if blood is left to stand after collection, the fluidity of the blood decreases and a red clot is formed. That is, the term “serum” means a light yellow liquid remaining removal of the red clot.

[00121] According to the present invention, ovarian cancer may be diagnosed using the
15 biomarkers of the present invention, and the diagnosis may be performed through liquid biopsy using patient-derived body fluids such as blood as a biological sample. This liquid biopsy method has advantages over conventional tissue biopsy procedures in that it may minimize patient’s pain due to its non-invasiveness and provide information
20 about cancer more quickly.

[00122]

[00123] According to a specific embodiment of the present invention, the ovarian cancer is high-grade serous ovarian cancer (HGSOC).

[00124] The meaning of ovarian cancer or high-grade serous ovarian cancer in the
25 present invention has already been described in detail, description thereof will be

omitted below to avoid excessive overlapping.

[00125]

[00126] According to another aspect of the present invention, the present invention provides a composition for preventing or treating ovarian cancer, comprising, as an
5 active ingredient, either an inhibitor of at least one selected from the group consisting of FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP, and CD14, or an activator of at least one selected from the group consisting of APOA1, APOA4, and SERPINF2.

[00127] The genes and the proteins encoded thereby used in the present invention have
10 already been described in detail, description thereof will be omitted to avoid excessive overlapping.

[00128] According to the present invention, the present inventors have experimentally demonstrated that if the FGA, VWF and ARHGDIB proteins or the genes encoding them in ovarian cancer patients are inhibited, or if the SERPINF2 protein or the genes
15 encoding the same in ovarian cancer patients is overexpressed or activated, the development of ovarian cancer can be suppressed.

[00129] In the present specification, the term “inhibitors” refers to substances that reduce the activity or expression of target genes, specifically the *FGA*, *VWF* and *ARHGDIB* genes, so that the activity or expression of the target genes becomes
20 undetectable or is at an insignificant level or the biological function of the target genes may be significantly reduced.

[00130] Examples of the inhibitors of the target genes include, but are not limited to, shRNA, siRNA, miRNA, ribozyme, peptide nucleic acid (PNA), and an antisense oligonucleotide, which inhibit the expression of the genes whose sequences are already
25 known in the art at the gene level; a CRISPR system comprising a guide RNA that

recognizes the target gene; antibodies or aptamers that inhibit the expression at the protein level; and compounds, peptides and natural products that inhibit the activity of the target genes. In addition, any means for gene and protein level inhibition known in the art may be used.

5 **[00131]** In the present specification, the term “small hairpin RNA (shRNA)” refers to a single-stranded RNA sequence consisting of 50 to 70 nucleotides, which forms a stem-loop structure *in vivo* and creates a tight hairpin structure for inhibiting target gene expression by RNA interference. Usually, a double-stranded stem is formed by complementary base pairing of a long RNA consisting of 19 to 29 nucleotides on both
10 sides of the loop consisting of 5 to 10 nucleotides, and for constitutive expression, the double-stranded stem is transduced into cells via a vector including U6 promoter and is usually delivered to daughter cells so that inhibition of expression of the target gene is inherited.

[00132] In the present specification, the term “siRNA” refers to a short double-stranded
15 RNA capable of inducing RNA interference (RNAi) through cleavage of a specific mRNA. The siRNA is composed of a sense RNA strand having a sequence homologous to the mRNA of the target gene and an antisense RNA strand having a sequence complementary thereto. The total length thereof may be 10 to 100 bases, preferably 15 to 80 bases, most preferably 20 to 70 bases, and the ends thereof may be blunt or cohesive
20 as long as the siRNA is capable of inhibiting the expression of the target gene by the RNAi effect. The cohesive ends may have a 3' end overhang and a 5' end overhang.

[00133] In the present specification, the term “microRNA (miRNA)” refers to an oligonucleotide that is not expressed in cells, and means a single-stranded RNA molecule that inhibits target gene expression by complementary binding to the mRNA of
25 the target gene while having a short stem-loop structure.

[00134] In the present specification, the term “ribozyme” refers to a kind of RNA which is an RNA molecule having the same function as an enzyme that recognizes and cleaves a specific RNA nucleotide sequence by itself. The ribozyme is a nucleotide sequence complementary to a target mRNA strand and is composed of a region that binds specifically to the target mRNA and a region that cleaves the target RNA.

[00135] In the present specification, the term “peptide nucleic acid (PNA)” refers to a molecule capable of complementary binding to DNA or RNA while having both nucleic acid and protein properties. PNA is not found in nature, is artificially synthesized by a chemical method, and regulates target gene expression by forming a double strand through hybridization with a natural nucleic acid having a nucleotide sequence complementary thereto.

[00136] In the present specification, the term “antisense oligonucleotide” refers to a nucleotide sequence complementary to a specific mRNA sequence. Specifically, it refers to a nucleic acid molecule which binds to a complementary sequence in a target mRNA and inhibits the activities of the target mRNA, which are essential for translation into proteins, translocation into cytoplasm, maturation, or all other overall biological functions. The antisense oligonucleotide may be modified at one or more bases, sugars or backbone positions to enhance efficacy thereof (De Mesmaeker et al., *Curr Opin Struct Biol.*, 5(3):343-55, 1995). The oligonucleotide backbone may be modified with phosphorothioate, phosphotriester, methylphosphonate, short-chain alkyl, cycloalkyl, short-chain heteroatomic, heterocyclic sugar sulfonate, or the like.

[00137] In the present specification, the term “guide RNA (gRNA)” refers to an RNA molecule used in a gene editing system that recognizes a target gene and induces a nuclease to specifically cleave the recognized site. Typical examples of this gene editing system include a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)

system.

[00138] According to the present invention, the expression inhibitor of the present invention may be a specific antibody that inhibits the activity of the protein encoded by the gene. An antibody that specifically recognizes the target protein is a polyclonal or
5 monoclonal antibody, preferably a monoclonal antibody.

[00139] Since the antibodies or aptamers that may be used in the present invention have already been described in detail, description thereof will be omitted to avoid excessive overlapping.

[00140]

10 **[00141]** In the present specification, the term “activator” refers to an active ingredient that enhances the expression level or activity of a target gene, specifically SERPINF2 protein. Examples of the activator include, but are not limited to, nucleic acid molecules, peptides, proteins, compounds, and natural products, which enhance the expression of the SERPINF2 protein whose sequence and structure are already known in the art at the
15 gene or protein level or enhance the inherent biological activity of the SERPINF2 protein. Accordingly, the term “activator of SERPINF2 protein” is used in the same sense as the term “SERPINF2 protein agonist”.

[00142] In the present specification, the term "prevention" or “preventing” means inhibiting the occurrence of a disorder or disease in a subject who has never been
20 diagnosed as having the disorder or disease, but is likely to suffer from such disorder or disease.

[00143] In the present specification, the term "treatment" or “treating” means (a) inhibiting the progress of a disorder, disease or symptom; (b) alleviating the disorder, disease or symptom; or (c) eliminating the disorder, disease or symptom. When the
25 composition of the present invention is administered to a subject, it functions to inhibit

the generation of ovarian cancer cells while inhibiting the expression of the FGA, VWF and ARHGDIB proteins or the genes encoding them or activating the expression of the SERPINF2 protein or the gene encoding the same, thereby inhibiting, eliminating, or alleviating the progress of symptoms resulting from ovarian cancer. Thus, the composition of the present invention may serve as a therapeutic composition for the disease by itself, or may be administered in combination with other pharmacological ingredients and applied as a therapeutic aid for the disease. Accordingly, in the present specification, the term "treatment" or "therapeutic agent" is meant to encompass "treatment aid" or "therapeutic aid agent".

10 **[00144]**

[00145] According to a specific embodiment of the present invention, the ovarian cancer that can be prevented or treated using the composition of the present invention is high-grade serous ovarian cancer (HGSOC).

[00146] Since the meaning of ovarian cancer or high-grade serous ovarian cancer in the present invention has already been described in detail, description thereof will be omitted to avoid excessive overlapping.

[00147]

[00148] According to another aspect of the present invention, the present invention provides a method for screening a composition for preventing or treating ovarian cancer, comprising steps of:

[00149] (a) bringing a test substance into contact with a biological sample containing at least one protein selected from the group consisting of FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP, CD14, APOA1, APOA4, and SERPINF2, a gene encoding the protein, or cells expressing the protein or gene; and

[00150] (b) measuring the expression level of the protein or the gene in the biological sample,

[00151] wherein, the activity or expression level of the FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP or CD14 protein or the gene encoding the same in the biological sample decreased, or the activity or expression level of the APOA1, APOA4 or SERPINF2 protein or the gene encoding the same in the biological sample increased, the test substance is determined as the composition for preventing or treating ovarian cancer.

[00152] Since the ovarian cancer-related genes or proteins encoded thereby that are used as the markers in the present invention, and the ovarian cancer that can be prevented or treated by regulating the expression of the genes or proteins have already been described in detail, description thereof will be omitted to avoid excessive overlapping.

[00153] In the present invention, the term “biological sample” refers to any sample containing cells expressing the above-described gene or the protein encoded by the gene, which is obtained from mammals, including humans. Examples of the biological sample include, but are not limited to, tissues, organs, cells, or cell cultures. More specifically, the biological sample may be cancer tissue, cancer cells, a culture thereof, or blood.

[00154] The term “test substance” as used while referring to the screening method of the present invention refers to an unknown substance which is used in screening to examine whether it affects the activity or expression level of the gene of the present invention by being added to a sample containing cells expressing the gene. Examples of the test substance include, but are not limited to, compounds, nucleotides, peptides, and natural extracts. The step of measuring the expression level or activity of the gene in the biological sample treated with the test substance may be performed by various

expression level and activity measurement methods known in the art.

[00155]

[00156] According to a specific embodiment of the present invention, the ovarian cancer is high-grade serous ovarian cancer (HGSOC).

5 **[00157]** Since the meaning of ovarian cancer or high-grade serous ovarian cancer in the present invention has already been described in detail, description thereof will be omitted to avoid excessive overlapping.

[00158]

[00159] According to another aspect of the present invention, the present invention
10 provides a method for diagnosing ovarian cancer, comprising a step of administering to a subject a composition for diagnosing ovarian cancer comprising, as an active ingredient, an agent for measuring the expression level of at least one gene selected from the group consisting of *FGA*, *VWF*, *ARHGDIB* and *SERPINF2*, or a protein encoded thereby.

[00160] According to a specific embodiment of the present invention, the method for
15 diagnosing ovarian cancer according to the present invention further comprises a step of administering to the subject the composition further comprising, as an active ingredient, an agent for measuring the expression level of at least one gene selected from the group consisting of *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*, *LBP*, *PPBP*, *C3*, *FCGBP*, *CD14*, *APOA1* and *APOA4*, or a protein encoded thereby.

20 **[00161]**

[00162] According to another aspect of the present invention, the present invention provides a method for preventing or treating ovarian cancer, comprising a step of administering to a subject a composition for preventing or treating ovarian cancer comprising, as an active ingredient, either an inhibitor of at least one selected from the
25 group consisting of *FGA*, *VWF*, *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*,

GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP and CD14, or an activator of at least one selected from the group consisting of APOA1, APOA4 and SERPINF2.

[00163] Since the genes or the proteins encoded by the genes that may be used for the diagnosis, prevention or treatment of ovarian cancer in the present invention, and the inhibitors or activators of expression of the genes and proteins have already been described in detail, description thereof will be omitted to avoid excessive overlapping.

[00164]

Advantageous Effects

[00165] The features and advantages of the present invention are summarized as follows:

[00166] (a) The present invention provides a method of diagnosing ovarian cancer by measuring the expression levels of genes or proteins involved in the development of ovarian cancer, and a method of preventing or treating ovarian cancer by regulating the expression of these genes or proteins.

[00167] (b) The present invention provides effective biomarkers for ovarian cancer, especially high-grade serous ovarian cancer, and thus provides a method capable of predicting the development of ovarian cancer at an early stage with high reliability. Thus, the present invention may be useful for efficient prevention or treatment of ovarian cancer by inhibiting or activating the expression of these factors.

[00168]

Brief Description of Drawings

[00169] FIG. 1a schematically shows comprehensive serum proteome profiling experiments followed by the experimental procedure for targeted MS analysis.

[00170] FIG. 1b is a graph showing the protein concentration dynamic range of serum proteins identified by comprehensive serum proteome profiling for spectral library

generation. Here, protein concentration estimates were obtained from the Human Protein Atlas (blood protein).

[00171] FIG. 1c shows the annotated MS/MS spectrum of the peptide (QCYLQQVK) of RNF213.

5 **[00172]** FIG. 2a depicts graphs showing the number of peptides (top) and proteins (bottom) identified by LC-DIA-MS/MS in HGSOC and HC, and indicates the total and average numbers of the identified peptides and proteins.

[00173] FIG. 2b shows the results of PCA analysis of the HGSOC and HC serum proteomes. Here, the variation explained by each principal component is indicated.

10 **[00174]** FIG. 2c is a heatmap of the expression of DEProteins in HGSOC and HC, with the number of up- and down-regulated proteins indicated.

[00175] FIG. 2d shows the pathway enriched by DEProteins. Here, the dot plot shows the significance of the pathway enriched by DEProteins in log₁₀ (p-value).

[00176] FIG. 2e is a diagram illustrating a network model showing the interactions
15 between DEProteins involved in the enriched pathway of FIG. 2d.

[00177] FIG. 3a is a box plot showing normalized peak areas (light/heavy) of 21 potential HGSOC diagnostic biomarkers in a validation cohort using targeted MS experiments.

[00178] FIG. 3b is a box plot showing normalized protein abundance of 21 biomarker
20 candidates in a discovery cohort using LC-DIA-MS/MS.

[00179] FIG. 4a depicts graphs showing the results of measuring the viability values of SK-OV-3 cells using proliferation assays after treatment with control siRNA (siControl) or the indicated siRNAs for FGA, VWF and ARHGDIB, or overexpression of SERPINF2 or treatment with DMSO (n =3 per condition).

25 **[00180]** FIG. 4b shows representative images (left) and quantification (right) of wound

areas measured at 0 hr or 12 hr after treating SK-OV-3 cells with the indicated siRNAs or overexpressing SERPINF2 (n =3 per condition).

[00181] FIG. 4c shows representative images (left) and result graphs (right) of colony formation of SK-OV-3 cells upon treatment with the indicated siRNAs or overexpression of SERPINF2 (n =3 per condition).

[00182] FIG. 4d shows the results of measuring expression levels by Western blotting after treatment with control siRNA (siControl) or the indicated siRNAs for FGA, VWF and ARHGDIB, or overexpression of SERPINF2 or treatment with DMSO (n =3 per condition).

[00183] FIG. 4e depicts graphs showing the results of measuring the viability values of SK-OV-3 cells using proliferation assays after treatment with control siRNA (siControl) or the indicated siRNAs for FGA, VWF and ARHGDIB (n =3 per condition).

[00184] FIG. 4f shows representative images (left) and quantification results (right) of wound areas measured at 0 hr or 12 hr after treating SK-OV-3 cells with the indicated siRNAs (n =3 per condition).

[00185] FIG. 4g shows representative images (left) and result graph (right) of colony formation of SK-OV-3 cells upon treatment with the indicated siRNAs (n =3 per condition).

[00186]

Mode for Invention

[00187] Hereinafter, the present invention will be described in more detail by way of examples. These examples are only for illustrating the present invention in more detail, and it will be apparent to those skilled in the art that the scope of the present invention according to the subject matter of the present invention is not limited by these examples.

[00188]

[00189] Examples

[00190] Experimental Methods and Analysis Methods

[00191] Study Population

[00192] In this study, serum samples from patients with high-grade serous ovarian cancer (HGSOC) and controls (HCs) were used. For study population groups, patients were identified who met the following criteria: (1) patients diagnosed with HGSOC; those who received primary surgical treatment between August 2012 and December 2020; and (3) people who donated blood samples taken the day before surgery for scientific purposes after providing written consent. Meanwhile, patients with malignant tumors other than HGSOC were excluded.

[00193] The present inventors collected the patients' baseline clinicopathologic characteristics, such as the age at diagnosis, International Federation of Obstetrics and Gynecology (FIGO) stage, and initial serum CA-125 levels. In terms of survival outcomes, progression-free survival (PFS) was defined as the time interval from the date of diagnosis to the date of disease progression. The serum samples from the test subjects were frozen, kept fresh, and collected from the Seoul National University Hospital Human Biobank. The control group was composed of healthy people who received regular health checkups at the Seoul National University Hospital Healthcare System Gangnam Center.

[00194]

[00195] Reagents

[00196] The following were used: multiple affinity removal column (Agilent Technology, Palo Alto, CA, USA); S-Trap (Protifi, Huntington, NY, USA); triethylammonium bicarbonate buffer (TEAB; Sigma); acetonitrile (ACN; Macron, Avantor Performance Materials, Center Valley, PA); DL-dithiothreitol (DTT; Sigma);

iodoacetamide (IAA; Sigma); trifluoroacetic acid (TFA; Pierce, Rockford, IL); formic acid (FA; Sigma); dimethyl sulfoxide (DMSO; Sigma).

[00197]

[00198] *Serum Sample Preparation*

5 **[00199]** Using a multiple affinity removal column (MARS14), the 14 most abundant proteins in plasma (albumin, IgA, IgG, IgM, α 1-antitrypsin, α 1-acid glycoprotein, apolipoprotein A1, apolipoprotein A2, Complement C3, transferrin, α 2-macroglobulin, transthyretin, haptoglobin, and fibrinogen) were removed. 40 μ L portion of serum diluted 4-fold with proprietary “Buffer A” was loaded onto a MARS14 depletion column
10 on a binary HPLC system (20A Prominence, Shimadzu, Tokyo, Japan). The free fraction was collected in a collection tube containing 100 μ L of 5% SDS in 100 mM TEAB solution and then completely dried using a speed-vac concentrator (Thermo Fisher Scientific, Waltham, MA, USA). The dried sample was resuspended in 100 μ L of 50 mM TEAB solution and sonicated for 10 min. Then, reduction (DTT, 10 mM,
15 56°C, 30 min) and alkylation (IAA, 20 mM, room temperature in the dark, 30 min) was done for 100 μ g proteins. Thereafter, sample preparation was done by suspension-trapping (S-Trap)-based tryptic digestion according to slight modifications of the manufacturer’s instructions. Briefly, after loading, the sample was washed with 90:10% methanol/50 mM ammonium bicarbonate. The sample was then digested with
20 50 mM ammonium bicarbonate, and trypsin (1:25 trypsin/protein) was added to this fraction and incubation was done overnight at 37°C. The digested peptides were eluted by centrifugation at 1,000g for 60 seconds. Two additional elutions were performed using 0.2% formic acid and 0.2% formic acid contained in 50% acetonitrile. The three elutions were combined, dried by vacuum centrifugation, and stored at -80°C until use.

25 **[00200]**

[00201] *Comprehensive Serum Proteome Profiling Experiments for Serum Spectral Library Generation*

[00202] For serum proteome profiling experiments, a total of 157 serum peptide samples were combined to make 6 pooled sample groups (Benign, Clear Cell, Endometriod, High-Grade Serous, Mucinous, and Healthy).

[00203]

[00204] A previously developed dual online noncontiguous fractionating and concatenating reverse phase/reverse phase liquid chromatography (DO-NCFC-RP/RPLC) system was modified to generate 24 NCFC fractions. Briefly, online NCFC device consists of mid-pH RP column (150 μ m x 50 cm) and two NCFC valves (25-port, 1-channel, C5M-66024D, VICI) interconnected with 24 NCFC loops (200 μ m). For the dual online liquid chromatography system, two SPE columns (150 μ m $\times$ 3 cm) and two analytical columns (75 μ m $\times$ 150 cm) were used. All the columns were manufactured in-house packed with Jupiter C18 materials (3 μ m, 300 Å, Phenomenex) and were heated at 60°C.

[00205]

[00206] The serum peptides (25 μ g, 40 μ g or 50 μ g) from each pooled sample group were injected and separated on the mid-pH column for the online 1st dimensional separation. A mid-pH solvent A (10 mM TEAB in water solvent, pH 7.5) and a mid-pH solvent B (10 mM TEAB in 99% ACN solvent, pH 7.5) were used to generate a gradient (1-50 % Solvent B for 120 min at a flow rate of 1 μ L/min). During the mRP gradient elution, two NCFC valves were synchronously rotated with a 1-minute interval for 5 cycles so that each of 24 fraction loops contained five 1-minute eluates that are noncontiguously concatenated. For 2-dimensional separation, each of 24 online NCFC fractions was transferred to a SPE column while acidified and diluted 10 times with

dilution buffer (0.2 % TFA in water). A low-pH solvent A (0.1 % formic acid in water) and a low-pH solvent B (0.1 % formic acid in ACN) were used to generate a gradient (10-37.5% Sol B for 160 min, 37.5-80% Sol B for 5 min, 80% Sol B for 13 min, 10% Sol B for 2 min) at a flow rate of 0.3 μ L/min.

5 **[00207]**

[00208] DO-NCFC-RP/RPLC was online coupled to an Orbitrap ExplorisTM 480 mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) with high field asymmetric waveform ion mobility spectrometry (FAIMS ProTM, Thermo Fisher Scientific, Bremen, Germany). The peptides were ionized by an in-house nanoelectrospray source at 3 kV of spray voltage and the temperature of the desolvation capillary was set at 250°C. Three compensation voltages (CVs) of -35 V, -50 V and -60 V were used with 1 second cycle time. The temperatures of both inner and outer electrodes were set at 100°C. For each CV, full MS scans (300-1,400 Th) were obtained at a resolution of 60,000 and full MS AGC target was 500 with an IT of 20 ms. The isolation window of 1.2 Th and
10 normalized collision energy (NCE) of 28 were used for higher-energy collisional dissociation (HCD). The MS/MS spectra were obtained at a resolution of 15,000 and AGC target of 1,000 and IT of 32 ms. The MS/MS data were subjected to Post-Experiment Monoisotopic Mass Refinement (mPE-MMR) and the resultant MS/MS data (e.g., mgf files) were searched against a protein database consisting of UniProt human
15 reference database (released on August 2022, including 97,093 entries) and common contaminants (179 entries) using the MS-GF+ search engine (v9949, <http://proteomics.ucsd.edu/software-tools/ms-gf/>). The search parameters were semi-tryptic cleavage and a precursor mass tolerance of 10 ppm. Carbamidomethylation on cysteine was used for static modification while oxidation on methionine and
20 deamidation on asparagine and glutamine were used as variable modifications. The
25

search results were filtered by the target-decoy analysis at a false discovery rate (FDR) of 1% at peptide spectrum match (PSM) level. A bipartite graph analysis was used for protein inference. The protein with the highest number of sibling peptides was selected as the representative protein from each protein group. When more than one protein had the same number of sibling peptides, the protein with the highest sequence coverage was chosen as the representative protein. In addition, a protein that has unique peptides was considered as a separate representative protein.

[00209]

[00210] The representative MS/MS spectra of the final peptides were used to construct a spectral library database. Briefly, the MS2 spectrum with the highest search score (i.e., $-\log(\text{SpecE-value})$) from each peptide ion was selected as the representative MS/MS spectrum. For each peptide, fragment peaks from the representative MS/MS spectra were matched with theoretical fragments (b- and y-type) within 0.01 Da. Resulting annotated fragments were obtained while other fragments of no annotation were discarded. Finally, the representative MS/MS spectra having more than five annotated fragments were included in the spectral library with experimental mass information and retention time information. The experimental retention time was converted to the normalized elution time using the iRT peptide (Ki-3002-2, Biognosys) which was spiked in each NCFC fraction.

[00211]

[00212] *Data-Independent Acquisition Analysis and Data Processing*

[00213] Digested peptides were separated using a Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific). The tryptic peptides were reconstituted in 0.1% TFA and separated on an Acclaim™ Pepmap RSLC C18 column (150 mm x 150 μm i.d., 2 μm, 100 Å) equipped with a C18 Pepmap trap column (20 mm × 100 μm i.d., 5 μm, 100

Å, Thermo Scientific) over 60 min (1 µL/min) using a 5-40% acetonitrile gradient in 0.1% formic acid and 5% DMSO at 50°C. The LC was coupled to an Orbitrap Exploris™ 480 mass spectrometer with EASY-SPRAY™ source (Thermo Fisher Scientific). Mass spectrometry runs were operated in DIA mode. For DIA experiments, full MS resolutions were set to 60,000 and full MS AGC target was 300% with an IT of 25 ms. The m/z range was set to 300-1,400. AGC target value for MS2 spectra was set to 1,000%. Forty-four windows of 24 Da were used with an overlap of 1 Da. Resolution was set to 15,000 and IT to 22 ms. NCE was set at 27. Default settings were used for FAIMS with dual CV of -35 V and -45 V, except the gas flow was not applied and spray voltage was set to 2.5 kV. The LC-DIA-MS/MS data were processed using DIA-NN (version 1.7.10) with the in-house serum spectral library. Spectra were searched with default settings, except m/z ranges were 300-1,300 for precursor and 200-1,300 for fragment ions. Identification results were filtered at FDR of 1% at precursor level. Peptide and protein quantities were obtained using the MaxLFQ algorithm, and the quantity data were subjected to quantile normalization. All LC-MS/MS data were deposited in the ProteomeXchange consortium to the PRIDE partner repository under dataset identifier PXD033169.

[00214]

[00215] *Differential Expression Analysis*

[00216] An integrative statistical method was applied to define differentially expressed peptides and proteins (DEPeptides and DEProteins, respectively). In brief, the present inventors calculated test statistics for each peptide or protein using Students' t-test, Wilcoxon-Ranksum test, and log2-median-ratio in each comparison. Then, empirical distributions of the test statistics and log2 median ratios for the null hypothesis were estimated by randomly permutating all samples 10,000 times. Using the estimated

empirical distributions for each peptide or protein, the present inventors computed adjusted p-values for the observed test statistics and log2 median ratio, and then calculated the overall p-value by combining these p-values using Stouffer's method. Finally, the present inventor identified DEPeptides as those with overall p-values < 0.05 and absolute log2 median ratios greater than the 10th and 90th percentile means of the empirical distribution in each comparison. The present inventors defined DEPeptides as those with t-test p-values <0.05, absolute log2-median-ratios greater than the 10th and 90th percentile means of the empirical distribution for log2-median-ratios, and the number of consistent DEPeptides ≥ 1 . Only peptides and proteins expressed in more than half of the total samples of the two test groups in each comparison were used for hypothesis testing.

[00217]

[00218] *Pathway Analysis*

[00219] Pathway over-representation analysis was performed using ConsensusPathDB.

Pathways represented by DEProteins were identified as those with q-values of less than 0.05 and the number of molecules involved in the pathways being 3 or more. The entire set of genes identified in the current study was used as the background gene list. The identified pathways were classified according to their functional relevance into "immune system", "IGF signaling", "TGF- β signaling", "ECM organization" or "lipoprotein assembly". A network model was constructed for DEProteins related to over-represented pathways. The network was constructed by setting nodes as proteins and adding protein-protein interactions and activation/inhibition relationships. Nodes were aligned based on localization and activation/repression relationships obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and literature search. The network was drawn using Cytoscape.

[00220]

[00221] *Multiple and Parallel Reaction Monitoring Analysis*

[00222] Multiple reaction monitoring-mass spectrometry (MRM-MS) analysis was applied with Qtrap5500 plus (Sciex). Tryptic peptides were separated on a ZORBAX 300SB-C18 reversed-phase column (0.5 mm × 150 µm i.d., 3.5 µm, 100 Å; Agilent) using a 5-40% acetonitrile gradient in 0.1% formic acid for 15 min (20 µL/min). Collision energy (CE) values for each ionized peptide were determined using SKYLINE, which provides values. Parallel reaction monitoring-mass Spectrometry (PRM-MS) analysis was applied with the Orbitrap Exploris™ 480 mass spectrometer. Tryptic peptides were separated on an Acclaim™ Pepmap RSLC C18 column (150 mm × 150 µm i.d., 2 µm, 100 Å) equipped with a C18 Pepmap trap column (20 mm × 100 µm i.d., 5 µm, 100 Å, Thermo Scientific) using a 5-40% acetonitrile gradient in 0.1% formic acid and 5% DMSO at 50°C over 60 min (1.5 µL/min). Mass spectrometry was performed in targeted mass mode. Precursor target masses, m/z values, and charge states were listed through data-dependent acquisition analysis using synthetic peptides. For the PRM experiment, full MS resolution was set to 60,000, full MS AGC target was set to 300 with an IT of 25 ms, and the m/z range was set to 280-1,200. MS/MS resolution was set to 15,000, IT was set to 22ms, and NCE was set to 27. The spray voltage was set at 2.5 kV, and the PRM-MS spectral peaks were selected and calculated using SKYLINE.

[00223]

[00224] *In-Vitro Functional Analysis*

[00225] *Cell Culture*

[00226] Human HGSOc cancer cell line SK-OV-3 was obtained from Seoul National University College of Medicine (Korea). SK-OV-3 cells were cultured in medium containing RPMI 1640 (Corning; 10-041-CV), 10% fetal bovine serum (FBS;

Invitrogen), and 1% penicillin/streptomycin (WelGENE Inc.) for 37°C under 5% CO₂. Cells were maintained with fresh medium for 3 days.

[00227]

[00228] *Plasmid and siRNA Transfection*

5 **[00229]** Myc-DDK tagged plasmid was purchased from OriGene Technologies, Inc. (Rockville, MD). The Myc-DDK tagged SERPINF2-human serpin peptidase inhibitor (RC206435) plasmid was used for transfection. The plasmid was transfected into SK-OV-3 cells using Lipofectamine 3000 reagent according to the manufacturer's protocol from Invitrogen (#L3000008). For knockdown experiments, the following siRNAs were
10 purchased from Genolution, Inc. (Korea):

[00230] Negative Control 5'-CCUCGUGCCGUUCCAUCAGGUAGUU-3' (SEQ ID NO: 22),

[00231] siFGA-1 5'-CUCUGUAUCUGGUAGUACUUU-3' (SEQ ID NO: 23),

[00232] siFGA-2 5'-GGAUUUAGGCACAUUGUCUUU-3' (SEQ ID NO: 24),

15 **[00233]** siVWF-1 5'-CCUUCUGAGCCCACAAUAAUU-3' (SEQ ID NO: 25),

[00234] siVWF-2 5'-GCUGUAAGUCUGAAGUAGAUU-3' (SEQ ID NO: 26),

[00235] siARHGDIB-1 5'-GGAAAUGGACAAAGAUGAUUU-3' (SEQ ID NO: 27),

and

[00236] siARHGDIB-2 5'-AAACCAUUGUGUUAAAGGAUU-3' (SEQ ID NO: 28).

20 **[00237]**

[00238] siRNA was transfected into SK-OV-3 cells using Lipofectamine RNAiMAX transfection reagent according to the manufacturer's protocol from Invitrogen (#13778075). SK-OV-3 cells were exposed to siRNA reagent for 72 hours and harvested with PBS (WelGENE Inc.). Additionally, DNA transfections were incubated in Opti-
25 MEM with diluted lipofectamine for 28 hours and harvested for Western blotting. For

proliferation assay and cell migration, cells were transfected with siRNA reagent at 48 hours and with DNA transfection reagent at 24 hours.

[00239]

[00240] *Western Blotting*

5 **[00241]** Cells were lysed in RIPA lysis buffer (Cell Biolabs, Inc.) containing phosphatase inhibitor (100X cocktail GenDEPOT). Lysates were boiled in SDS sample buffer (Alfa Aesar Thermo) at 95°C for 5 min. Proteins were separated on 4-20% precast SDS-PAGE gels (BIORAD) and transferred to FL-PVDF membranes (MERCK) at 15 minutes of the ATTO EZ Fast protocol. Membrane blocking was performed by
10 incubation with primary antibody and 5% albumin bovine serum (BSA; EcoCell) in 1X TBST overnight at 4°C. The membranes were first reduced, and then washed three times with 1X TBST and incubated with secondary antibody for 30 minutes at room temperature. Then, for the detected proteins, Western blotting reagent solution (Roche) and ECL (ElectroChemiLuminescence) of ODYSSEY Fc System (Li-Cor Bioscience)
15 were used. The following antibodies were used for Western blotting: FGA (Mα 1:500; ab92572, Abcam), VWF (Rα 1:500; ab189500, Abcam), ARHGDIB (D4GDI 1:1000; ab181252, Abcam), α-Tubulin DM1A (Mα 1:1000; Cell Signaling Technology Inc.), FLAG (GT231 Mα 1:1000; GTX629631, GeneTex Inc), anti-mouse IgG antibody HRP (1:3000; 170-6516, BIORAD), and anti-rabbit IgG antibody HRP (1:3000; 170-6515,
20 BIORAD).

[00242]

[00243] *Cell Proliferation Assay*

[00244] The WST-1 assay kit was used for cell proliferation assay according to the manufacturer's protocol from TaKaRa (MK400; Premix WST-1 Cell Proliferation Assay
25 System). Transfected cells were plated on 96-well plates at a density of 1,000 cells per

well in 100 μ L medium and incubated for 5 days. Then, the cells were treated with 10 μ L WST-1 reagent, and after 4 hours, cell proliferation was analyzed by measuring the absorbance of formazan dye at 450 nm.

[00245]

5 **[00246]** *Wound Healing Assay*

[00247] Cell migration was assessed by scratching SK-OV-3 cells cultured at a density of 500 cells per well in 12-well plates. Wound healing images at 0 and 12 hours were acquired using a LEICA DMI1 microscope system with a 10 $\times$ /0.22 PH1 lens and MC170 HD camera. The experiments were performed in triplicate using three wells. The
10 average width of wound closure at four different positions in each experiment was imaged using ImageJ software.

[00248]

[00249] *Colony Formation Assay*

[00250] Transfected cells were plated on 6-well plates at a density of 500 cells per well.
15 After 10 days, the colonies were washed once with PBS at room temperature, stained with 0.5% crystal violet (V5265, Sigma) for 1 hour, and washed three times to remove residual dye. Then, the colonies were dried overnight and analyzed using ImageJ software.

[00251]

20 **[00252]** *Statistical Analysis of In Vitro Functional Analysis*

[00253] A two-tailed t-test was used to estimate the statistical significance of all functional test results. A p-value of less than 0.05 was considered significant, and each experiment was performed with three biological replicates.

[00254]

25 **[00255]** **Experimental Results**

[00256] *Characterization of Samples*

[00257] For the discovery of potential diagnostic biomarkers for HGSOC, 54 HGSOC patient cases and 48 healthy control (HC) subjects were prepared: (1) a discovery set for proteomics profiling consisting of 26 HGSOC and 24 HC, and (2) a validation set for targeted MS experiments consisting of 28 HGSOC and 24 HC (Table 1). The average age of HGSOC patients was 59.5 and 56 years in the discovery and validation sets, respectively, and the average age of HC was 46 and 48.5 years in the discovery and validation sets, respectively. The numbers of FIGO stage I, II, and III HGSOC patients were 3, 3, and 20, respectively, in the discovery set, and 3, 10, and 15, respectively, in the validation set. During the median observation period of 25 months, 42.3% of HGSOC patients in the discovery set and 57.1% of HGSOC patients in the validation set experienced recurrence after primary treatment surgery.

[00258] [Table 1] Characteristics of samples

Sample set	Discovery set		Validation set	
Histology	HC ¹	HGSOC ²	HC	HGSOC
Number of cases	24	26	25	28
Age (median)	25-65 (46)	42-78 (59.5)	33-70 (48.5)	33-76 (56)
FIGO ³				
Stage I	-	3	-	3
Stage II	-	3	-	10
Stage III	-	20	-	15
Recurrence				
Yes	-	11	-	16
No	-	15	-	12

¹ Healthy control (HC), ² high-grade serous ovarian cancer (HGSOC), ³ Federation of International of Gynecologists and Obstetricians (FIGO)

[00259]

[00260] *Comprehensive OC Serum Spectral Library*

5 **[00261]** Comprehensive global proteome profiling experiments on serum samples were performed using dual online non-contiguous fractionating and concatenating reverse-phase/reverse-phase liquid chromatography coupled to a FAIMS-MS/MS instrument (DO-NCFC-RP/RPLC-FAIMS-MS/MS). A total of 9 sets of DO-NCFC-RP/RPLC-FAIMS-MS/MS experiments were performed on 6 groups of depleted serum samples,
10 each group containing 24 online NCFC fractions and 3 CVs (FIG. 1a). A total of 216 LC-FAIMS-MS/MS global proteome datasets were obtained, and a group of 137,086 peptides and a group of 12,066 proteins were identified through MS-GF+ search. The identified serum proteome was estimated to cover approximately 8 orders of dynamic ranges (440 µg of ceruloplasmin to 3.5 pg of ring finger protein 213) (FIGS. 1b and 1c).
15 The serum proteome of the present invention provides high coverage of the serum proteome and can be used to construct an OC serum spectral library. Among the tandem spectra of the identified serum peptides, the tandem spectrum with the highest search score for the peptide and five or more annotated fragment ion peaks were used to construct a spectral library. The spectral library contained 127,444 spectra of 94,212
20 non-redundant peptides covering a group of 8,458 proteins.

[00262]

[00263] *Discovery of Potential Diagnostic Biomarkers for HGSOC*

[00264] LC-DIA-MS/MS analysis was performed on 24 HC and 26 HGSOC in the discovery cohort to discover HGSOC diagnostic biomarkers. A DIA-NN search using
25 the OC serum spectral library identified 29,721 peptides and 2,439 proteins across all

samples (FIG. 2a), which correspond to the largest quantification of the serum proteome of OC to date. The number of peptides identified across all samples varied from 13,818 to 15,987 with an average of 14,797, and the number of proteins varied from 1,012 to 1,240 with an average of 1,104. Through principal component analysis (PCA), it could be confirmed that the identified serum proteome clearly separated HGSOC from HC (FIG. 2b). For downstream analysis, 14,755 peptides and 951 proteins quantified in more than half of the total samples in both the HGSOC and HC groups were used. Differential expression analysis was then performed to identify potential diagnostic biomarkers for HGSOC. As a result, the present inventors could identify 100 up- and 54 down-regulated proteins in the serum samples from HGSOC patients compared to HC (FIG. 2c).

[00265] Pathway analysis showed that the up-regulated proteins were involved in complement and coagulation cascades, platelet activation and aggregation, toll-like receptor 4 (TLR4) signaling, and neutrophil extracellular trap (NET) formation, insulin-like growth factor (IGF) formation and transforming growth factor β (TGF- β) signaling, and immune processes involving extracellular matrix (ECM) organization. On the other hand, proteins involved in Fc gamma receptor (FC γ R) activation and plasma lipoprotein assembly were significantly down-regulated (FIG. 2d).

[00266] Next, a network model was constructed, which describe the interactions between proteins involved in the over-represented processes in HGSOC serum (FIG. 2e).

A network model involving showed that the coagulation and complement cascades showed that (1) the coagulation cascade, including up-regulation of procoagulants (F11, VWF, and FGA/B) and down-regulation of anticoagulant factors (SERPINA5/F2 and A2M), and (2) the complement cascade (MBL2, C3/9/C4A/B/BPA/BPB and CFHR2/5) that cross-talks with the coagulation cascade to promote coagulation or carry out the reverse reaction were commonly up-regulated.

[00267] A network involving neutrophils and macrophages showed co-regulation of (1) NET formation (LBP, ACTB and MPO), (2) TLR4 signaling, including up-regulation of TLR4 signaling activation proteins (LBP, FGA/B, CD14, S100A8) and down-regulation of TLR4 signaling inhibition proteins (APOA1/4/C3), and (3) down-regulation of FC γ R-mediated phagocytosis (IGLC2/3/HG2/HG3/KC/KV1-5) (KV1D-33/KV3-20/KV3D-20/KV4-1). Additionally, it was confirmed that up-regulation of FCGBP along with activation of TLR4 signaling could cause macrophage recruitment to the tumor microenvironment (TME). In addition, a network involving platelets showed coordinated up-regulation of platelet activation and aggregation (e.g., VCL, GP1BA, PF4, PPBP, SRGN, VWF, and FGA), which is promoted by NET formation and can induce the coagulation cascade.

[00268] Finally, a network model involving IGF, TGF- β signaling, and ECM organization showed that (1) up-regulation of ECM organization (EFEMP1, SERPINE1, PCOLCE, SPARC, TIMP1, FN1, NID1, ENO1, LCP1, ARHGDIB and FGA/B), (2) IGF signaling (IGFBP2 and APP), and (3) TGF- β signaling (LRG1, LTBP1, THBS1, PRG4 and SERPINE1) collectively contributed to cancer cell proliferation, migration, and invasion in HGSOC.

[00269] In the network model, potential HGSOC-related diagnostic biomarkers were selected for targeted MS experiments according to the following criteria: fold change, consistency of expression across samples, and association with previously reported OC pathophysiology.

[00270] The present inventors selected an initial set of 33 biomarker candidates related to the aforementioned biological processes, including coagulation and complement cascades (VWF, SERPINF2, MBL2, C3 and C4B), NET formation (MPO), TLR4 signaling (LBP, CD14 and S100A8/9), lipoprotein assembly (APOA1/4), macrophage

recruitment to TME (FCGBP), platelet activation and aggregation (VCL, PFN1, FLNA, GP1BA, VWF, FN1, FGA, PF4, PPBP, SRGN and CALU), ECM organization (SPARC, TIMP1, FN1, ENO1, LCP1 and ARHGDIB); IGF signaling (IGFBP2); and TGF- β signaling (LRG1, LTBP1, THBS1 and PRG4) (shown in red in FIG. 2e).

5 **[00271]**

[00272] *Validation of Potential Diagnostic Biomarkers for HGSOc*

[00273] Targeted MS analysis (MRM or PRM) was performed in an independent cohort of 25 HCs and 28 HGSOc patients using stable isotope-labeled (SIL) peptides of 33 biomarker candidates to validate their potential diagnostic value. Seven candidates
10 (FLNA, FN1, LTBP1, PFN1, S100A8, TIMP1, and CALU) did not show detectable transition signals in MRM or PRM experiments and were therefore excluded from the final set of biomarker candidates. The expression patterns of the final set of 26 diagnostic biomarker candidates were compared with each other using the results of targeted MS analysis (FIG. 3a) and LC-DIA-MS/MS analysis (FIG. 3b). Targeted MS
15 analysis indicated that the expression of 21 of 26 candidates showed statistically significant ($P < 0.05$) changes between HGSOc and HC (Table 2), which were consistent with LC-DIA-MS/MS analysis.

[00274]

[00275] [Table 2] MRM quantification results

Gene name	Protein name	UniProt ID	LC-DIA-MS/MS	Targeted MS			
			DEProtein ¹	Method	Peptide	log ₂ -fold-change ²	P-value ³
FGA	FIBA	P02671	UP	PRM	HPDEAAFFDTASTGK	2.61	<0.005
VWF	VWF	P04275	UP	PRM	HIVTFDQGQNFK	2.48	<0.001
S100A9	S100A9	P06702	UP	PRM	NIETIINTFHQYSVK	1.75	<0.01
VCL	VCL	P18206	UP	PRM	LLAVAATAPPDAPNR	1.70	<0.05
THBS1	TSP1	P07996	UP	MRM	TIVTTLQDSIR	1.50	<0.001
MPO	MPO	P05164	UP	PRM	VVLEGGIDPILR	1.19	<0.005
IGFBP2	IBP2	P18065	UP	MRM	LIQGAPTIR	1.13	<0.001
SRGN	SRGN	P10124	UP	MRM	IQDLNR	1.03	<0.001
PF4	PF4	P02776	UP	MRM	TTSQVRPR	0.91	<0.001
GP1BA	GP1BA	P07359	UP	MRM	LTSLPLGALR	0.87	<0.001
LRG1	LRG1	P02750	UP	MRM	DLLLPQPDLR	0.85	<0.001
ARHGDIB	RhoGDI2	P52566	UP	PRM	TLLGDGPVVTDPK	0.85	<0.005
PRG4	PRG4	Q92954	UP	MRM	DQYYNIDVPSR	0.83	<0.001
LBP	LBP	P18428	UP	MRM	ITLPDFTGDLR	0.80	<0.001
PPBP	PPBP	P02775	UP	MRM	NIQSLEVIGK	0.54	<0.001
C3	C3	P01024	UP	MRM	TGLQEVEVK	0.43	<0.005
FCGBP	FCGBP	Q9Y6R7	UP	MRM	GNPAVSYYR	0.42	<0.005
CD14	CD14	P08571	UP	MRM	FPAIQNLALR	0.41	<0.05
APOA1	APOA1	P02647	DOWN	MRM	VQPYLDDFQK	-0.30	<0.01
APOA4	APOA4	P06727	DOWN	MRM	QLTPYAQR	-0.74	<0.001
SERPINF2	A2AP	P08697	DOWN	PRM	EDFLEQSEQLFGAK	-1.08	<0.001

¹UP, Up-regulated, DOWN, Down-regulated, ²HGSOC vs. HC, ³Student's t-test p-value with multiple testing correction using the Benjamini-Hochberg procedure

[00276]

5 [00277] *In-Vitro Functional Experiments of Serum Biomarkers for HGSOC*

[00278] Under the assumption that the HGSOC serum biomarkers may play a functional role rather than simply being a result of biological processes regulated in HGSOC, functional analysis was performed on three strongly up-regulated biomarkers (FGA, VWF, and ARHGDIB) and one strongly down-regulated biomarker (SERPINF2).

5 First, the present inventors compared the cell viability, migration, and colony formation abilities of SK-OV-3 human HGSOC cells expressing control siRNA and siRNA for each of the three up-regulated biomarkers (FIG. 4d). It was confirmed that cell viability, migration, and colony formation ability were significantly reduced by knockdown of FGA, VWF and ARHGDIB in SK-OV-3 cells ($P < 0.05$) (FIGS. 4a to 4c and 4e to 4g).

10 Meanwhile, it was confirmed that overexpression of SERPINF2 in SK-OV-3 cells (FIG. 4d) significantly ($P < 0.05$) reduced the viability, migration and colony formation ability of the cells (FIGS. 4a to 4c and 4e to 4g). The above data suggest that HGSOC serum biomarkers such as FGA, VWF, ARHGDIB, and SERPINF2 may play an important regulatory role in cancer cell proliferation and migration in HGSOC.

15 **[00279]**

[00280] Although the present invention has been described in detail with reference to the specific features, it will be apparent to those skilled in the art that this description is only of a preferred embodiment thereof, and does not limit the scope of the present invention. Thus, the substantial scope of the present invention will be defined by the
20 appended claims and equivalents thereto.

CLAIMS

1. A composition for diagnosing ovarian cancer comprising, as an active ingredient, an agent for measuring an expression level of at least one gene selected from the group consisting of *FGA*, *VWF*, *ARHGDIB*, and *SERPINF2*, or a protein encoded thereby.

2. The composition of claim 1, further comprising, as an active ingredient, an agent for measuring an expression level of at least one gene selected from the group consisting of *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*, *LBP*, *PPBP*, *C3*, *FCGBP*, *CD14*, *APOA1*, and *APOA4*, or a protein encoded thereby.

3. The composition of claim 1, wherein the expression level of the *FGA*, *VWF* or *ARHGDIB* gene or the protein encoded thereby in a subject with ovarian cancer increases, and the expression level of *SERPINF2* gene or the protein encoded thereby in the subject decreases.

4. The composition of claim 2, wherein the expression level of the *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*, *LBP*, *PPBP*, *C3*, *FCGBP* or *CD14* gene or the protein encoded thereby in a subject with ovarian cancer increases, and the expression level of the *APOA1* or *APOA4* gene or the protein encoded thereby in the subject decreases.

5. The composition of claim 1, wherein the ovarian cancer is high-grade serous ovarian cancer (HGSOC).

6. A method of providing information necessary for diagnosing ovarian cancer, comprising a step of measuring expression levels of *FGA*, *VWF*, *ARHGDIB* and *SERPINF2* genes or proteins encoded thereby in a biological sample isolated from a subject.

5

7. The method of claim 6, wherein the step of measuring the expression levels is performed by mass spectrometry (MS).

8. The method of claim 7, wherein the mass spectrometry measures an expression level of at least one peptide selected from the group consisting of SEQ ID NOs: 1, 2, 12 and 21.

9. The method of claim 6, further comprising a step of measuring an expression level of at least one gene selected from the group consisting of *S100A9*, *VCL*, *THBS1*, *MPO*, *IGFBP2*, *SRGN*, *PF4*, *GP1BA*, *LRG1*, *PRG4*, *LBP*, *PPBP*, *C3*, *FCGBP*, *CD14*, *APOA1* and *APOA4*, or a protein encoded thereby.

15

10. The method of claim 9, wherein the step of measuring the expression level is performed by mass spectrometry (MS).

20

11. The method of claim 10, wherein the mass spectrometry measures an expression level of at least one peptide selected from the group consisting of SEQ ID NOs: 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19 and 20.

25

12. The method of claim 6, wherein the biological sample is blood.

13. The method of claim 6, wherein the ovarian cancer is high-grade serous ovarian cancer (HGSOC).

5 14. A composition for preventing or treating ovarian cancer comprising, as an active ingredient, either an inhibitor of at least one selected from the group consisting of FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP, and CD14, or an activator of at least one selected from the group consisting of APOA1, APOA4, and SERPINF2.

10 15. The composition of claim 14, wherein the ovarian cancer is high-grade serous ovarian cancer (HGSOC).

16. A method for screening a composition for preventing or treating ovarian cancer, comprising steps of:

15 (a) bringing a test substance into contact with a biological sample containing at least one protein selected from the group consisting of FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP, CD14, APOA1, APOA4, and SERPINF2, a gene encoding the protein, or
20 cells expressing the protein or gene; and

(b) measuring an expression level of the protein or the gene in the biological sample,

wherein, the activity or expression level of the FGA, VWF, S100A9, VCL, THBS1, MPO, IGFBP2, SRGN, PF4, GP1BA, LRG1, ARHGDIB, PRG4, LBP, PPBP, C3, FCGBP or CD14 protein or the gene encoding the protein in the biological sample
25

decreased, or the activity or expression level of the APOA1, APOA4 or SERPINF2 protein or the gene encoding the protein in the biological sample increased, the test substance is determined as the composition for preventing or treating ovarian cancer.

- 5 17. The method of claim 16, wherein the ovarian cancer is high-grade serous ovarian cancer (HGSOC).