

CLAIMS

1. A composition for diagnosing cancer, comprising an agent for measuring an expression level of at least one polypeptide selected from the group consisting of APOC1 (apolipoprotein C1), CHL1 (neural cell adhesion molecule L1-like), MMP9
5 (matrix metalloproteinase-9), PRDX6 (peroxiredoxin-6), PRG4 (proteoglycan 4), PPBP (platelet basic protein), FN1 (fibronectin), VWF (von Willebrand factor), and CLU (clusterin) or a fragment thereof, or a gene encoding the polypeptide or fragment thereof.

2. The composition of claim 1, wherein

10 the fragment of the APOC1 polypeptide has the amino acid sequence of SEQ ID NO: 1 (TPDVSSALDK);

the fragment of the CHL1 polypeptide has the amino acid sequence of SEQ ID NO:
2 (VIAVNEVGR);

the fragments of the MMP9 polypeptide has the amino acid sequence of SEQ ID
15 NO: 3 (AVIDDAFAR);

the fragment of the PRDX6 polypeptide has the amino acid sequence of SEQ ID
NO: 4 (LSILYPATTGR);

the fragment of the PRG4 polypeptide has the amino acid sequence of SEQ ID NO:
5 (AIGPSQTHIR);

20 the fragment of the PPBP polypeptide has the amino acid sequence of SEQ ID NO:
6 (TTSGIHPK);

the fragment of the FN1 polypeptide has the amino acid sequence of SEQ ID NO:
7 (STTPDITGYR);

the fragment of the VWF polypeptide has the amino acid sequence of SEQ ID NO:
25 8 (ILAGPAGDSNVVK); and

the fragment of the CLU polypeptide has the amino acid sequence of SEQ ID NO: 9 (TLLSNLEEAK).

3. The composition of claim 1, wherein the cancer is breast cancer.

5

4. The composition of claim 1, wherein the agent for measuring the expression level of the polypeptide comprises at least one selected from the group consisting of an antibody, an oligopeptide, a ligand, a peptide nucleic acid (PNA), and an aptamer, which bind specifically to the polypeptide or fragment thereof.

10

5. The composition of claim 4, wherein the agent for measuring the expression level of the gene encoding the polypeptide or fragment thereof comprises at least one selected from the group consisting of a primer, a probe, and an antisense oligonucleotide, which bind specifically to the gene.

15

6. A diagnostic kit comprising the composition of any one of claims 1 to 5.

7. The diagnostic kit of claim 6, wherein the kit is an RT-PCR kit, a DNA chip kit, an ELISA kit, a protein chip kit, a rapid kit or a multiple-reaction monitoring (MRM) kit.

20

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

(a) bringing a candidate substance into contact with a biological sample containing at least one polypeptide selected from the group consisting of polypeptides

25

represented by SEQ ID NO: 1 (TPDVSSALDK), SEQ ID NO: 2 (VIAVNEVGR), SEQ ID NO: 3 (AVIDDAFAR), SEQ ID NO: 4 (LSILYPATTGR), SEQ ID NO: 5 (AIGPSQTHIR), SEQ ID NO: 6 (TTSGIHPK), SEQ ID NO: 7 (STTPDITGYR), SEQ ID NO: 8 (ILAGPAGDSNVVK), and SEQ ID NO: 9 (TLLSNLEEAK), respectively, or

5 a gene encoding the polypeptide, or cells expressing the polypeptide or gene; and

(b) measuring an expression level of at least one polypeptide selected from the group consisting of polypeptides represented by SEQ ID NO: 1 (TPDVSSALDK), SEQ ID NO: 2 (VIAVNEVGR), SEQ ID NO: 3 (AVIDDAFAR), SEQ ID NO: 4 (LSILYPATTGR), SEQ ID NO: 5 (AIGPSQTHIR), SEQ ID NO: 6 (TTSGIHPK), SEQ ID NO: 7 (STTPDITGYR), SEQ ID NO: 8 (ILAGPAGDSNVVK), and SEQ ID NO: 9 (TLLSNLEEAK), respectively, or the gene encoding the polypeptide, in the biological sample,

wherein, the measured expression level of the polypeptide of SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8 or SEQ ID NO: 9 or the gene encoding the polypeptide in the biological sample isolated from the subject of interest decreased compared to that in a normal control group and;

the measured expression level of the polypeptide of SEQ ID NO: 1 or SEQ ID NO: 4 or the gene encoding the polypeptide increased compared to that in the normal control group,

20 the candidate substance is determined as the composition for preventing or treating cancer.

9. A system for diagnosing breast cancer, comprising:

an input unit configured to receive an input value;

25 a reading unit comprising a machine learning model pre-trained to read whether

breast cancer has occurred; and

an output unit configured to output whether breast cancer has occurred,

wherein the input value is a measured value for an expression level of at least one polypeptide selected from the group consisting of SEQ ID NO: 1 (TPDVSSALDK), SEQ
5 ID NO: 2 (VIAVNEVGR), SEQ ID NO: 3 (AVIDDAFAR), SEQ ID NO: 4 (LSILYPATTGR), SEQ ID NO: 5 (AIGPSQTHIR), SEQ ID NO: 6 (TTSGIHPK), SEQ ID NO: 7 (STTPDITGYR), SEQ ID NO: 8 (ILAGPAGDSNVVK), and SEQ ID NO: 9 (TLLSNLEEK), in a biological sample.

10 10. The system of claim 9, wherein the machine learning model is a deep learning model.

11. The system of claim 9, wherein the biological sample is whole blood, leukocytes, peripheral blood mononuclear cells, buffy coat, plasma, serum, sputum, tears,
15 mucus, nasal washes, nasal aspirate, breath, urine, semen, saliva, peritoneal washings, ascites, cystic fluid, meningeal fluid, amniotic fluid, glandular fluid, pancreatic fluid, lymph fluid, pleural fluid, nipple aspirate, bronchial aspirate, synovial fluid, joint aspirate, organ secretions, cells, cell extract, or cerebrospinal fluid..

20 12. The system of claim 9, wherein the measured value for the expression level of the polypeptide is a quantitative value obtained by mass spectrometry.

13. The system of claim 9, wherein the measuring of the expression level of the polypeptide is performed by protein chip assay, immunoassay, ligand binding assay,
25 MALDI-TOF (Matrix Assisted Laser Desorption/Ionization Time of Flight Mass

Spectrometry) assay, SELDI-TOF (Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry) assay, radioimmunoassay, radioimmunodiffusion, Ouchterlony immunodiffusion, rocket immunoelectrophoresis, immunohistochemical staining, complement fixation assay, two-dimensional electrophoresis assay, liquid chromatography-mass spectrometry (LC-MS), LC-MS/MS (liquid chromatography-mass spectrometry/mass spectrometry), Western blotting, or ELISA (enzyme-linked immunosorbent assay).

14. The system of claim 9, wherein the measuring of the expression level of the polypeptide is performed by a multiple-reaction monitoring (MRM) method.

15. The system of claim 9, wherein

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 1 is 1032.5280 or in the range of 1032.5280 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 2 is 966.5524 or in the range of 966.5524 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 3 is 977.5051 or in the range of 977.5051 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 4 is 1191.6732 or in the range of 1191.6732 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 5 is 1180.6433 or in the range of 1180.6433 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 6 is 840.4574 or in the range of 840.4574 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 7 is

1110.5426 or in the range of 1110.5426 ± 1 when the z value is 1;

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 8 is 1240.6896 or in the range of 1240.6896 ± 1 when the z value is 1; and

a mass-to-charge ratio (m/z) of the polypeptide represented by SEQ ID NO: 9 is
5 1117.6099 or in the range of 1117.6099 ± 1 when the z value is 1.

16. The system of claim 14, wherein the multiple-reaction monitoring method uses a synthetic peptide obtained by substituting certain amino acids of each of the polypeptides with an isotope or E. coli beta-galactosidase as an internal standard
10 substance.

17. The system of claim 16, wherein the synthetic peptide has the same sequence as the sequence represented by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, or 9 and comprise a stable isotope.
15

18. The system of claim 9, wherein the measuring of the expression level of the gene encoding the polypeptide is performed by reverse transcription-polymerase chain reaction (RT-PCR), competitive RT-PCR, real-time RT-PCR, RNase protection assay (RPA), Northern blotting, or DNA chip assay.
20

19. The system of claim 9,
wherein the measured expression level of the CHL1 (neural cell adhesion molecule L1-like), MMP9 (matrix metalloproteinase-9), PRG4 (proteoglycan 4), PPBP (platelet basic protein), FN1 (fibronectin), VWF (von Willebrand factor) or CLU
25 (clusterin) polypeptide or the gene encoding the polypeptide in the biological sample

isolated from the subject of interest is higher than that in a normal control group and;

if the measured expression level of the APOC1 (apolipoprotein C1) or PRDX6 (peroxiredoxin-6) polypeptide or the gene encoding the polypeptide in the biological sample is lower than that in the normal control group,

5 a likelihood of onset of the cancer is highly predicted.

20. The system of claim 9, which predicts responsiveness of the subject of interest to an anticancer drug.