CLAIMS

What is claimed is:

1. A method of classifying a biological sample with respect to a plurality of states associated with colorectal cancer (CRC) or advanced adenoma (AA), the method comprising:

receiving peptide structure data corresponding to a set of proteins in the biological sample, wherein the peptide structure data comprises at least one peptide structure from Table 10;

inputting quantification data identified from the peptide structure data into a machine-learning model trained to identify a disease indicator based on the quantification data;

identifying, by the machine-learning model, the disease indicator; and

classifying the biological sample with respect to the plurality of states associated with the CRC or the AA based upon the identified disease indicator.

2. The method of claim 1, wherein the set of proteins comprises one or more glycoproteins, and

wherein the at least one peptide structure comprises a glycopeptide.

3. The method of claim 1 or 2, wherein the machine-learning model comprises a least absolute shrinkage and selection operator (LASSO) regression model; and

wherein the quantification data is generated using a liquid chromatography-mass spectrometry (LC-MS) system or reaction monitoring mass spectrometry (MRM-MS).

4. The method of claim 1- 3, wherein the quantification data for a peptide structure of the set of peptide structures comprises at least one of an abundance, a relative abundance, a normalized abundance, a differential abundance, a relative quantity, an adjusted quantity, a normalized quantity, a relative concentration, an adjusted concentration, or a normalized concentration.

5. The method of claim 1 - 4, further comprising:

receiving biological samples from a plurality of subjects; and

performing a differential expression analysis using the quantification data for the plurality of subjects.

6. A method of diagnosing an individual with colorectal cancer (CRC) or advanced adenoma (AA), comprising:

detecting a presence or amount of at least one peptide structure from a plurality of peptide structures from Table 10;

inputting a quantification of the detected at least one peptide structure into a machine-learning model trained to generate a class label;

determining if the class label is above or below a threshold for a classification;

identifying a diagnostic classification for the individual based on whether the class label is above or below the threshold for the classification; and

diagnosing the individual as having the CRC or the AA based on the diagnostic classification.

7. The method of claim 6, wherein the quantification is generated using a liquid chromatography-mass spectrometry (LC-MS) system, and the at least one peptide structure is generated using multiple reaction monitoring mass spectrometry (MRM-MS).

8. The method of claim 6-7, wherein the CRC is one of stage I CRC, stage II CRC, stage III CRC, or stage IV CRC; and

wherein the individual is determined have a healthy state, in response to an absence of the CRC or the AA.

9. The method of claim 6-7, further comprising, training the machine-learning model based on:

receiving quantification data for a panel of peptide structures for a plurality of subjects diagnosed with a plurality of states associated with the CRC or the AA, wherein the plurality of states comprises at least one of a CRC state, an AA state, or a healthy state; and

training the machine-learning model to determine a state of the plurality of states based on a biological sample from the subject based on the quantification data.

10. The method of claim 9, wherein the training of the machine-learning model to determine the state of the plurality of states comprises training the machine-learning model to generate a class label for the state of the plurality of states.

11. The method of claim 7-10, wherein the at least one peptide structure comprises a peptide sequence and a glycan structure, and wherein the glycan structure of the peptide sequence corresponds to a glycan structure GL number in accordance with the Table 10, wherein the glycan structure comprises a symbol structure in accordance with the glycan structure GL number according to the Table 10, Table 11A, and Table 11B.

12. The method of claim 11,

wherein a rightmost N-acetylgalactosamine of the glycan structure in Table 11A is attached to a linking site position in the peptide sequence in accordance with Table 10, and

wherein a bottommost N-acetylglucosamine of the glycan structure in Table 11B is attached to a linking site position in the peptide sequence in accordance with Table 10.

13. A composition comprising one or more peptide structures from Table 10.

14. A composition comprising one or more peptides comprising a sequence set forth in SEQ ID NOs: 1-38.

15. A method of classifying a biological sample obtained from a subject with respect to a plurality of states associated with colorectal cancer (CRC) or advanced adenoma (AA), the method comprising:

receiving mass spectrometry (MS) quantification data obtained from the biological sample,

wherein the quantification data comprises a quantification level associated with each of one or more peptides derived from one or more proteins of Table 9;

inputting the MS quantification data into a machine-learning model,

wherein the machine-learning model is trained on one or more training MS quantification data sets comprising quantification data from training samples characterized as having the CRC, having the AA, or not having the CRC or the AA,

wherein, for each training sample, the associated training MS quantification data comprises a quantification level associated with each of one or more peptides derived from the one or more proteins of the Table 9; and

classifying the biological sample with respect to the plurality of states associated with the CRC or the AA.

16. The method of claim 15, wherein the biological sample is classified as having the CRC, or the AA, or not having the CRC and the AA.

17. The method of claim 15-16, wherein the MS quantification data comprises one or more of peptide sequence information, post-translational modification information comprising glycan information, the quantification level associated with one or more peptides derived from each protein of Model 1 or Model 2 of Table 9, the quantification level associated with one or more peptides of Table 10, and a quantification level associated with at least one peptide derived from each protein of Model 1 or Model 2 of Table 9.

18. The method of claim 16, wherein the training MS quantification data comprises one or more of the quantification level associated with one or more peptides derived from each protein of Model 1 or Model 2 of Table 9 and the quantification level associated with one or more peptides of Table 10.

19. A method of determining a glycopeptide profile of a biological sample obtained from a subject,

wherein the glycopeptide profile is based on a quantification level associated with one or more peptides derived from one or more proteins of Table 9;

the method comprising:

subjecting the biological sample, or a derivative thereof, to a mass spectrometry (MS) technique configured to assess the one or more peptides derived from one or more proteins of Table 9 to obtain MS information;

determining the quantification level associated with the one or more peptides derived from one or more proteins of Table 9 based on the MS information; and

determining the glycopeptide profile based on the quantification level associated with the one or more peptides derived from one or more proteins of Table 9.

20. The method of claim 19, wherein the one or more peptides comprise a sequence set forth in SEQ ID NO:5 and/or SEQ ID NO:6.