

What is claimed:

1. A method of determining a cancer condition of a test subject of a species, the method comprising:

at a computer system comprising at least one processor and a memory storing at least one program for execution by the at least one processor, the at least one program comprising instructions for:

A) obtaining a dataset, in electronic form, wherein the dataset comprises a corresponding methylation pattern of each respective fragment in a plurality of fragments, wherein the corresponding methylation pattern of each respective fragment (i) is determined by a methylation sequencing of one or more nucleic acid samples comprising the respective fragment in a biological sample obtained from the test subject and (ii) comprises a methylation state of each CpG site in a corresponding plurality of CpG sites in the respective fragment;

B) constructing a first patch comprising a first channel, the first patch representing a first independent set of CpG sites in a reference genome of the species, each respective CpG site in the first independent set of CpG sites corresponding to a predetermined location in the reference genome, wherein:

the first channel of the first patch comprises a plurality of instances of a first plurality of parameters, wherein each instance of the first plurality of parameters includes a parameter for a methylation status of a respective CpG site in the first independent set of CpG sites for the first patch,

the constructing B) comprises populating, for each respective fragment in the plurality of fragments that aligns to the first independent set of CpG sites, an instance of all or a portion of the first plurality of parameters based on the methylation pattern of the respective fragment; and

C) applying at least the first patch to a classifier thereby determining the cancer condition in the test subject.

2. The method of claim 1, wherein the at least one program further comprises instructions for, after the obtaining A) and prior to the constructing B):

pruning the plurality of fragments by removing from the plurality of fragments each respective fragment, whose corresponding methylation pattern across a corresponding plurality of CpG sites in the respective fragment, has a p-value that fails to satisfy a p-value threshold, wherein the p-value of the respective fragment is determined based upon a comparison of the

corresponding methylation pattern of the respective fragment to a corresponding distribution of methylation patterns of the corresponding plurality of CpG sites in a corresponding plurality of reference fragments that have the corresponding plurality of CpG sites of the respective fragment, wherein the methylation pattern of each reference fragment in the corresponding plurality of reference fragments is obtained by a methylation sequencing of nucleic acid from biological samples obtained from a cohort of healthy subjects; and optionally,

3. ~~The method of claim 2,~~ wherein:

the first patch comprises a plurality of channels including the first channel and a second channel,

the second channel comprises a corresponding instance of a second plurality of parameters for each instance of the first plurality of parameters, wherein each instance of the second plurality of parameters includes a parameter for a first characteristic, other than CpG methylation state, of a respective CpG site in the first independent set of CpG sites for the first patch, and

the constructing B) comprises populating, for each respective fragment in the plurality of fragments that aligns to the first independent set of CpG sites, an instance of all or a portion of the first plurality of parameters and an instance of all or a portion of the second plurality of parameters based on the methylation pattern of the respective fragment

; and further optionally,

(i) wherein

the constructing B), for a respective fragment in the plurality of fragments, comprises:

i) identifying, within an instance of the first plurality of parameters of the first channel, parameters, corresponding to the CpG sites in the respective fragment, that have not previously been assigned methylation states based on another fragment in the plurality of fragments;

ii) assigning for each parameter, among the identified parameters, that aligns to a respective CpG site of the respective fragment, the methylation state of the respective CpG site of the respective fragment; and

iii) assigning for each parameter, among the identified parameters, in the second plurality of parameters of the instance of the second plurality of parameters of the second channel that corresponds to the instance of the first plurality of parameters, that aligns to a

respective CpG site of the respective fragment, the first characteristic of the respective CpG site of the respective fragment;

and optionally,

(a) wherein the first characteristic of the respective CpG site is a multiplicity of the respective fragment the respective CpG site is on; or

(b) wherein the first characteristic of the respective CpG site is selected from the group consisting of:

a CpG β -value drawn from a healthy cohort,

a CpG β -value drawn from a predetermined tissue type in a healthy cohort,

a CpG β -value drawn from the test subject,

a Pearson's correlation score for methylation state of 5' and 3' neighbor CpG sites,

a Jaccard similarity, Euclidean distance, Manhattan distance, maximum value, normalized Euclidean distance, normalized maximum value, dice coefficient, or cosine coefficient of methylation state of the respective CpG site in the test subject versus a cohort of subjects,

a fragment p-value of the respective fragment,

a length of the respective fragment the respective CpG site is on,

a fragment sequence source,

a fragment mapping quality score of the respective fragment the respective CpG site is on,

a distance to a 5' adjacent CpG site in the reference genome,

a distance to a 3' adjacent CpG site in the reference genome,

a multiplicity of the respective fragment the respective CpG site is on,

a genetic element the respective CpG site is within,

a biological pathway the respective CpG site is associated with,

a gene the respective CpG site is associated with,

a value of a CpG transition impulse function for the respective CpG site,

a value of a CpG run-length encoding for the respective CpG site, and

a read strand orientation of the fragment the respective CpG site is on; or

(c) wherein more than one fragment in the plurality of fragments is assigned to a single instance of the first plurality of parameters of the first channel in the first patch provided that the more than one fragment does not have common CpG sites;

or

(ii) wherein:

the plurality of channels comprises at least three channels; and

a third channel in the first plurality of channels comprises a corresponding instance of a third plurality of parameters for each instance of the first plurality of parameters, wherein each instance of the third plurality of parameters includes a parameter for a second characteristic of a respective CpG site in the first independent set of CpG sites, wherein the second characteristic is selected from the group consisting of:

a CpG β -value drawn from a healthy cohort,

a CpG β -value drawn from a predetermined tissue type in a healthy cohort,

a CpG β -value drawn from the test subject,

a Pearson's correlation score for methylation state of 5' and 3' neighbor CpG sites,

a Jaccard similarity, Euclidean distance, Manhattan distance, maximum value, normalized Euclidean distance, normalized maximum value, dice coefficient, or cosine coefficient of methylation state of the respective CpG site in the test subject versus a cohort of subjects,

a fragment p-value of the respective fragment,

a length of the respective fragment the respective CpG site is on,

a fragment sequence source,

a fragment mapping quality score of the respective fragment the respective CpG site is on,

a distance to a 5' adjacent CpG site in the reference genome,

a distance to a 3' adjacent CpG site in the reference genome,

a multiplicity of the respective fragment the respective CpG site is on,

a genetic element the respective CpG site is within,

a biological pathway the respective CpG site is associated with,

a gene the respective CpG site is associated with,

a value of a CpG transition impulse function for the respective CpG site,

a value of a CpG run-length encoding for the respective CpG site, and

a read strand orientation of the fragment the respective CpG site is on.-

43. The method of claim 1, wherein the methylation pattern of a respective fragment does not include each CpG site in the first independent set of CpG sites of the first patch and wherein the constructing B), for a respective fragment in the plurality of fragments, comprises

populating parameters in the instance of first plurality of parameters that correspond to CpG sites present in the respective fragment; and optionally,

wherein parameters in the instance of the first plurality of parameters are zero filled.-

54. The method of claim 1, wherein the constructing B), for a respective fragment in the plurality of fragments, comprises:

i) identifying, within an instance of the first plurality of parameters of the first channel, parameters, corresponding to the CpG sites in the respective fragment, that have not previously been assigned methylation states based on another fragment in the plurality of fragments; and
 ii) assigning for each parameter, among the identified parameters, that aligns to a corresponding CpG site of the respective fragment, the methylation state of the corresponding CpG site of the respective fragment; ; and optionally,

(i) wherein a number of instances of the first plurality of parameters of the first channel are not assigned a respective fragment, and the at least one program further comprises instructions for zero filling parameters in instances of the plurality of parameters of the first channel that have not been assigned a fragment; or

(ii) wherein the identifying i) is unable to identify, within an instance of the first plurality of parameters of the first channel, parameters corresponding to the CpG sites in the respective fragment that have not previously been assigned methylation states based on another fragment in the plurality of fragments, and the at least one program further comprises instructions for discarding the respective fragment; or

(iii) wherein the identifying i) is unable to identify, within an instance of the first plurality of parameters of the first channel of the first patch, parameters corresponding to the CpG sites in the respective fragment that have not previously been assigned methylation states based on another fragment in the plurality of fragments, the at least one program further comprises instructions for creating an additional instance of the first patch and assigning the respective fragment to the additional instance of the first patch; or

(iv) wherein more than one fragment in the plurality of fragments is assigned to a single instance of the first plurality of parameters of the first channel in the first patch provided that the more than one fragment does not have common CpG sites.

6. ~~The method of claim 3, wherein the constructing B), for a respective fragment in the plurality of fragments, comprises:~~

~~i) identifying, within an instance of the first plurality of parameters of the first channel, parameters, corresponding to the CpG sites in the respective fragment, that have not previously been assigned methylation states based on another fragment in the plurality of fragments;~~

~~ii) assigning for each parameter, among the identified parameters, that aligns to a respective CpG site of the respective fragment, the methylation state of the respective CpG site of the respective fragment; and~~

~~iii) assigning for each parameter, among the identified parameters, in the second plurality of parameters of the instance of the second plurality of parameters of the second channel that corresponds to the instance of the first plurality of parameters, that aligns to a respective CpG site of the respective fragment, the first characteristic of the respective CpG site of the respective fragment.~~

~~7. The method of claim 6, wherein the first characteristic of the respective CpG site is a multiplicity of the respective fragment the respective CpG site is on.~~

~~8. The method of claim 6, wherein the first characteristic of the respective CpG site is selected from the group consisting of:~~

- ~~a CpG β value drawn from a healthy cohort,~~
- ~~a CpG β value drawn from a predetermined tissue type in a healthy cohort,~~
- ~~a CpG β value drawn from the test subject,~~
- ~~a Pearson's correlation score for methylation state of 5' and 3' neighbor CpG sites,~~
- ~~a Jaccard similarity, Euclidean distance, Manhattan distance, maximum value, normalized Euclidean distance, normalized maximum value, dice coefficient, or cosine coefficient of methylation state of the respective CpG site in the test subject versus a cohort of subjects,~~
- ~~a fragment p value of the respective fragment,~~
- ~~a length of the respective fragment the respective CpG site is on,~~
- ~~a fragment sequence source,~~
- ~~a fragment mapping quality score of the respective fragment the respective CpG site is on,~~
- ~~a distance to a 5' adjacent CpG site in the reference genome,~~
- ~~a distance to a 3' adjacent CpG site in the reference genome,~~
- ~~a multiplicity of the respective fragment the respective CpG site is on,~~
- ~~a genetic element the respective CpG site is within,~~

~~_____ a biological pathway the respective CpG site is associated with,~~
~~_____ a gene the respective CpG site is associated with,~~
~~_____ a value of a CpG transition impulse function for the respective CpG site,~~
~~_____ a value of a CpG run-length encoding for the respective CpG site, and~~
~~_____ a read strand orientation of the fragment the respective CpG site is on.~~

9. ~~The method of claim 5 or 6, wherein more than one fragment in the plurality of fragments is assigned to a single instance of the first plurality of parameters of the first channel in the first patch provided that the more than one fragment does not have common CpG sites.~~

10. ~~The method of claim 4, wherein parameters in the instance of the first plurality of parameters are zero filled.~~

115. The method of claim 1, wherein the first independent set of CpG sites are in a CpG index of the reference genome; and optionally,

12. ~~The method of claim 11, _____~~ wherein the CpG index of the reference genome includes a first CpG site, not present in the first independent set of CpG sites, that is located in the reference genome between a second CpG site and a third CpG site that are present in the first independent set of CpG sites.

136. The method of claim 1,

(i) wherein:

the first independent set of CpG sites includes a first CpG site and a second CpG site that are adjacent to each other in a CpG index of the reference genome,

a first fragment in the plurality of fragments includes the first CpG site but not the second CpG site, and

a second fragment in the plurality of fragments includes the second CpG site but not the first CpG site; or

14. ~~The method of claim 1, _____~~ (ii) wherein a parameter in an instance of the first plurality of parameters, for a respective fragment in the plurality of fragments, is:

methyated when the corresponding CpG site in the respective fragment is determined by the methylation sequencing to be methylated,

unmethylated when the corresponding CpG site in the respective fragment is determined by the methylation sequencing to not be methylated,

other when the corresponding CpG site in the respective fragment is determined by the methylation sequencing to be other than methylated or unmethylated; or

~~15. The method of claim 5, wherein a number of instances of the first plurality of parameters of the first channel are not assigned a respective fragment, and the at least one program further comprises instructions for zero filling parameters in instances of the plurality of parameters of the first channel that have not been assigned a fragment.~~

~~16. The method of claim 5, wherein the identifying i) is unable to identify, within an instance of the first plurality of parameters of the first channel, parameters corresponding to the CpG sites in the respective fragment that have not previously been assigned methylation states based on another fragment in the plurality of fragments, and the at least one program further comprises instructions for discarding the respective fragment.~~

~~17. The method of claim 5, wherein the identifying i) is unable to identify, within an instance of the first plurality of parameters of the first channel of the first patch, parameters corresponding to the CpG sites in the respective fragment that have not previously been assigned methylation states based on another fragment in the plurality of fragments, the at least one program further comprises instructions for creating an additional instance of the first patch and assigning the respective fragment to the additional instance of the first patch.~~

~~18. The method of claim 3, wherein:~~

~~—— the plurality of channels comprises at least three channels; and~~

~~—— a third channel in the first plurality of channels comprises a corresponding instance of a third plurality of parameters for each instance of the first plurality of parameters, wherein each instance of the third plurality of parameters includes a parameter for a second characteristic of a respective CpG site in the first independent set of CpG sites, wherein the second characteristic is selected from the group consisting of:~~

~~—— a CpG β value drawn from a healthy cohort,~~

~~—— a CpG β value drawn from a predetermined tissue type in a healthy cohort,~~

~~—— a CpG β value drawn from the test subject,~~

~~—— a Pearson's correlation score for methylation state of 5' and 3' neighbor CpG sites,~~

~~—— a Jaccard similarity, Euclidean distance, Manhattan distance, maximum value, normalized Euclidean distance, normalized maximum value, dice coefficient, or cosine coefficient of methylation state of the respective CpG site in the test subject versus a cohort of subjects,~~

~~—— a fragment p-value of the respective fragment,~~

~~—— a length of the respective fragment the respective CpG site is on,~~

~~—— a fragment sequence source,~~

~~—— a fragment mapping quality score of the respective fragment the respective CpG site is on,~~

~~—— a distance to a 5' adjacent CpG site in the reference genome,~~

~~—— a distance to a 3' adjacent CpG site in the reference genome,~~

~~—— a multiplicity of the respective fragment the respective CpG site is on,~~

~~—— a genetic element the respective CpG site is within,~~

~~—— a biological pathway the respective CpG site is associated with,~~

~~—— a gene the respective CpG site is associated with,~~

~~—— a value of a CpG transition impulse function for the respective CpG site,~~

~~—— a value of a CpG run-length encoding for the respective CpG site, and~~

~~—— a read strand orientation of the fragment the respective CpG site is on.~~

19. ~~The method of claim 1,~~ (iii) wherein the first independent set of CpG sites is drawn from across the entire reference genome.

207. The method of claim 1, the at least one program further comprising instructions for:

constructing a second patch comprising a corresponding first channel, the second patch representing a second independent set of CpG sites in the reference genome of the species, each respective CpG site in the second independent set of CpG sites corresponding to a predetermined location in the reference genome, wherein the corresponding first channel of the second patch comprises a corresponding plurality of instances of a first plurality of parameters, wherein each instance of the corresponding first plurality of parameters of the second channel includes a parameter for a methylation status of a respective CpG site in the second independent set of CpG sites for the second patch; and

populating, for each respective fragment in the plurality of fragments that aligns to the second independent set of CpG sites, an instance of all or a portion of the first plurality of

parameters of the second patch based on the methylation pattern of the respective fragment thereby constructing the second patch; and wherein

the applying C) further comprises applying the first and second patches to the classifier thereby determining the cancer condition in the test subject; and optionally,

~~21. The method of claim 20, (i) wherein:~~

the second patch comprises a corresponding plurality of channels including the corresponding first channel;

a corresponding second channel in the corresponding plurality of channels of the second patch comprises a corresponding instance of a second plurality of parameters for each instance of the first plurality of parameters, wherein each instance of the second plurality of parameters of the second patch includes a parameter for a first characteristic, other than CpG methylation state, of a respective CpG site in the second independent set of CpG sites for the second patch; and

the instructions for populating, for each respective fragment in the plurality of fragments that aligns to the second independent set of CpG sites, further populates an instance of all or a portion of the instance of the second plurality of parameters of the second patch based on the methylation pattern of the respective fragment; or

~~22. The method of claim 20, (ii) wherein the first independent set of CpG sites does not overlap with the second independent set of CpG sites.~~

~~23. The method of claim 20, (iii) wherein the first independent set of CpG sites overlaps with the second independent set of CpG sites; or~~

~~24. The method of claim 20, (iv) wherein the first patch represents either: an equally sized, but different, portion of the reference genome than the second patch; or~~

~~25. The method of claim 20, wherein the first patch represents a first portion of the reference genome and the second patch represents a second portion of the reference genome, wherein a size of the first portion is different than a size of the second portion; and optionally,~~

~~26. The method of claim 24 or 25, (a) wherein:~~

the first independent set of CpG sites consists of a first number of CpG sites,

the second independent set of CpG sites consists of a second number of CpG sites, and the first number of CpG sites is the same as the second number of CpG sites; or

~~27. The method of claim 24 or 25, _____~~ (b) wherein:

the first independent set of CpG sites consists of a first number of CpG sites, the second independent set of CpG sites consists of a second number of CpG sites, and the first number of CpG sites is different than the second number of CpG sites.

288. The method of claim 1, wherein the methylation sequencing of one or more nucleic acid samples is i) whole genome methylation sequencing or ii) targeted DNA methylation sequencing using a plurality of nucleic acid probes; and optionally.

~~29. The method of claim 28, _____~~ wherein the methylation sequencing of one or more nucleic acid samples uses a plurality of nucleic acid probes and the plurality of nucleic acid probes comprises one hundred or more probes.

~~309.~~ The method of claim 1,

(i) wherein the methylation sequencing of one or more nucleic acid samples detects one or more 5-methylcytosine (5mC) and/or 5-hydroxymethylcytosine (5hmC) in the respective fragment; or

~~31. The method of claim 1, _____~~ (ii) wherein the methylation sequencing of one or more nucleic acid samples comprises conversion of one or more unmethylated cytosines or one or more methylated cytosines, in the respective fragment, to a corresponding one or more uracils; and optionally.

~~32. The method of claim 31, _____~~ (a) wherein the one or more uracils are detected during the methylation sequencing as one or more corresponding thymines; or

~~33. The method of claim 31, _____~~ (b) wherein the conversion of one or more unmethylated cytosines or one or more methylated cytosines comprises a chemical conversion, an enzymatic conversion, or combinations thereof; or

~~34. The method of claim 1, _____~~ (iii) wherein:

the at least one program further comprises instructions for constructing a plurality of patches including the first patch, each respective patch being for a different independent set of CpG sites in the reference genome;

the constructing B) constructs a plurality of patches including the first patch;

the classifier comprises a plurality of trained first stage models and a second stage model;

the applying the at least the first patch to a classifier comprises:

obtaining a feature vector comprising a plurality of feature elements, wherein each feature element in the plurality of feature elements is an output of a corresponding trained first stage model in the plurality of trained first stage models upon application of a respective patch in the plurality of patches to the corresponding trained first stage model; and

applying the feature vector to the second stage model thereby determining the cancer condition in the test subject; and optionally,

~~35. The method of claim 34,~~ (a) wherein:

each respective trained first stage model in the plurality of trained first stage models is a corresponding trained convolutional neural network and the second stage model is a logistic regression model; and

the first channel of the first patch is two dimensional with each respective instance of the plurality of instances of the first plurality of parameters of the first patch forming a first dimension and the first plurality of parameters of the first patch forming the second dimension; or

~~36. The method of claim 34,~~ (b) wherein the plurality of patches is between 10 patches and 10000 patches; or

~~37. The method of claim 34,~~ (c) wherein the plurality of patches is between 100 patches and 3000 patches.

~~38~~10. The method of claim 1, wherein:

the classifier comprises a plurality of first stage models and a dynamic neural network;

the at least one program further comprises instructions for constructing a plurality of patches including the first patch, each respective patch being for a different set of CpG sites in the reference genome;

the constructing B) constructs a respective patch including the first patch;

the applying the at least the first patch to a classifier C) comprises:

C1) applying each respective patch in the plurality of patches to a corresponding first stage model in the plurality of first stage models, wherein the corresponding first stage model comprises:

i) a respective input layer for receiving the respective patch, wherein the respective patch comprises a first number of dimensions;

ii) a respective fully connected embedding layer that comprises a corresponding set of weights, wherein the respective fully connected embedding layer directly or indirectly receives output of the respective input layer, and wherein a respective output of the respective embedding layer is a second number of dimensions that is less than the first number of dimensions; and

iii) a respective output layer that directly or indirectly receives output from the respective fully connected embedding layer; and

C2) inputting an aggregate of the respective output from each respective fully connected embedding layer of each trained first stage model in the plurality of first stage models into the dynamic neural network thereby determining the cancer condition in the test subject; and optionally

~~39. The method of claim 38, _____ wherein the respective output of the respective embedding layer of each respective first stage model in the plurality of first stage models is a set of between 32 and 1048 values; and further optionally,~~

~~40. The method of claim 39, _____ wherein~~

~~(i) the at least one program further comprises instructions for training the plurality of first stage models and the dynamic neural network using a cohort of subjects, wherein the cohort of subjects comprises a first subset of subjects that have a first label for the cancer condition and a second subset of subjects that have a second label for the cancer condition; and optionally,~~

~~(a)~~

~~(i) wherein the cancer condition is a stage of a specified cancer and each subject in the cohort of subjects is labeled with a stage of a specified cancer; and optionally~~

(ii) wherein the cohort includes subjects that have a stage of an anorectal cancer, a stage of bladder cancer, a stage of breast cancer, a stage of cervical cancer, a stage of colorectal cancer, a stage of head and neck cancer, a stage of hepatobiliary cancer, a stage of endometrial cancer, a stage of kidney cancer, a stage of leukemia, a stage of liver cancer, a stage of lung cancer, a stage of lymphoid neoplasm, a stage of melanoma, a stage of multiple myeloma, a stage of myeloid neoplasm, a stage of ovary cancer, a stage of non-Hodgkin lymphoma, a stage of pancreatic cancer, a stage of prostate cancer, a stage of renal cancer, a stage of thyroid cancer, a stage of upper gastrointestinal tract cancer, a stage of urothelial carcinoma, or a stage of uterine cancer; or optionally

~~41. The method of claim 40,~~ (b) wherein the instructions for training comprise:

ai) stratifying, on a random basis, the cohort of subjects into a plurality of groups based on any combination of cancer condition, age, smoking status, or sex;

bi) using a first group in the plurality of groups as a training group and the remainder of the plurality of groups as test groups to train the plurality of models and the dynamic neural network against the training group;

cii) repeating the using (i)(b)(ii) for each group in the plurality of groups so that each group in the plurality of groups is used as the training group in an iteration of the using (i)(b)(ii); and

dii) repeating the stratifying (i)(b)(ia), using (i)(b)(iib) and repeating (i)(b)(iie) until a classifier performance criterion is satisfied;

and optionally,

wherein the cancer condition is whether or not a subject has cancer and the stratifying (i)(b)(i) ensures that each group in the plurality groups has equal numbers of subject that have cancer and that do not have cancer; and/or optionally

~~42. The method of claim 40 or 41,~~ (c) wherein the cancer condition is tissue of origin and each subject in the cohort of subjects is labeled with a tissue of origin; and optionally,

~~43. The method of claim 42,~~ (i) wherein the cohort includes subjects that an anorectal cancer, a bladder cancer, a breast cancer, a cervical cancer, a colorectal cancer, a head and neck cancer, a hepatobiliary cancer, an endometrial cancer, a kidney cancer, a leukemia, a liver cancer, a lung cancer, a lymphoid neoplasm, a melanoma, a multiple

myeloma, a myeloid neoplasm, an ovary cancer, a non-Hodgkin lymphoma, a pancreatic cancer, a prostate cancer, a renal cancer, a thyroid cancer, an upper gastrointestinal tract cancer, a urothelial carcinoma, or a uterine cancer; or

~~44. The method of claim 40,~~ (d) wherein the cancer condition is a stage of a specified cancer and each subject in the cohort of subjects is labeled with a stage of a specified cancer; and optionally

~~45. The method of claim 44,~~ (i) wherein the cohort includes subjects that have a stage of an anorectal cancer, a stage of bladder cancer, a stage of breast cancer, a stage of cervical cancer, a stage of colorectal cancer, a stage of head and neck cancer, a stage of hepatobiliary cancer, a stage of endometrial cancer, a stage of kidney cancer, a stage of leukemia, a stage of liver cancer, a stage of lung cancer, a stage of lymphoid neoplasm, a stage of melanoma, a stage of multiple myeloma, a stage of myeloid neoplasm, a stage of ovary cancer, a stage of non-Hodgkin lymphoma, a stage of pancreatic cancer, a stage of prostate cancer, a stage of renal cancer, a stage of thyroid cancer, a stage of upper gastrointestinal tract cancer, a stage of urothelial carcinoma, or a stage of uterine cancer; and optionally

~~46. The method of claim 41, wherein the cancer condition is whether or not a subject has cancer and the stratifying a) ensures that each group in the plurality groups has equal numbers of subject that have cancer and that do not have cancer.~~

~~47. The method of any one of claims 40-46,~~ wherein the training eliminates one or more patches in the plurality of patches using L1 or L2 regularization based upon values provided by the respective output layer of each respective patch in the plurality of patches during the training.

~~48~~11. The method of claim 1, wherein the plurality of instances of the first plurality of parameters is between 24 and 2048.

~~49~~12. The method of claim 1, wherein a number of instances in the plurality of instances of the first plurality of parameters is determined based on expected read depth of the plurality of fragments plus one standard deviation across the plurality of fragments.

~~50~~13. The method of claim 1, wherein the constructing B) further comprises sorting respective fragments assigned to the first patch based on their respective p-values or their starting position in the reference genome.

~~51~~14. The method of any one of the preceding ~~claims 1-50~~, wherein the at least one program further comprises instructions for selecting the first independent set of CpG sites of the first patch through evaluation of a plurality of CpG methylation patterns determined by a methylation sequencing of a plurality of clinical fragments obtained from a plurality of clinical nucleic acid samples of a plurality of clinical biological samples obtained from a clinical cohort comprising a plurality of clinical subjects, wherein the plurality of clinical subjects includes a first set of clinical subjects that have a first indication for the cancer condition and a second set of clinical subjects that have a second indication for the cancer condition.

~~52~~15. The method of claim ~~51~~14, wherein

(i) the instructions for selecting comprise:

determining a first ranking of a plurality of CpG sites in the reference genome based upon a respective first mutual information score for a methylation status of each CpG site in the plurality of CpG sites between the first set of clinical subjects and the second set of clinical subjects; and

selecting a first threshold number of CpG sites for the corresponding independent set of CpG sites for the first patch using the ranking; or

~~53. The method of claim 51,~~ (ii) wherein:

the plurality of clinical subjects includes a third set of clinical subjects that have a third indication for the cancer condition and a fourth set of clinical subjects that have a fourth indication for the cancer condition and the instructions for selecting further comprise:

determining a second ranking of the plurality of CpG sites in the reference genome based upon a respective second mutual information score for a methylation status of each CpG site in the plurality of CpG sites between the third set of clinical subjects and the fourth set of clinical subjects; and

selecting a second threshold number of CpG sites for the first independent set of CpG sites of the first patch using the second ranking; and optionally,

54. ~~The method of claim 52,~~ _____ (a) wherein the constructing B) further comprises sorting respective fragments assigned to the first patch based on their respective first mutual information score; or

55. ~~The method of claim 51,~~ _____ (iii) wherein the first indication for the cancer condition is a first cancer type and the second indication for the cancer condition is a second cancer type; and optionally,

56. ~~The method of claim 52,~~ _____ (a) wherein each respective CpG site in the first threshold number of CpG sites for the first independent set of CpG sites of the first patch is padded in the reference genome from all other CpG sites in the first threshold number of CpG sites by a threshold number of residues; or

57. ~~The method of claim 51,~~ _____ (iv) wherein the instructions for selecting further comprise:

determining a first ranking of a plurality of fixed length regions in the reference genome based upon a respective first mutual information score for a methylation status of a CpG site methylation pattern of each fixed length region in the plurality of fixed length regions between the first set of clinical subjects and the second set of clinical subjects; and

selecting a first threshold number of CpG sites for the first independent set of CpG sites of the first patch from those fixed length regions in the plurality of fixed length regions using the first ranking; and optionally,

58. ~~The method of claim 57,~~ _____ (a) wherein:

the plurality of clinical subjects includes a third set of clinical subjects that have a third indication for the cancer condition and a fourth set of clinical subjects that have a fourth indication for the cancer condition and the instructions for selecting further comprise:

determining a second ranking of the plurality of fixed length regions in the reference genome based upon a respective second mutual information score for a methylation status of a CpG site methylation pattern of each fixed length region in the plurality of fixed length regions between the third set of clinical subjects and the fourth set of clinical subjects; and

selecting a second threshold number of CpG sites for the first independent set of CpG sites of the first patch using the second ranking; or

~~59. The method of claim 57,~~ (b) wherein the constructing B) further comprises sorting respective fragments assigned to the first patch based on their respective first mutual information score; or

~~60. The method of claim 57,~~ (c) wherein the first indication for the cancer condition is a first cancer type and the second indication for the cancer condition is a second cancer type; or

~~61. The method of claim 57,~~ (d) wherein each respective CpG site in the first threshold number of CpG sites for the first independent set of CpG sites of the first patch is padded in the reference genome from all other CpG sites in the first threshold number of CpG sites by a threshold number of residues.

~~62~~16. The method of any one of the preceding ~~claims 1-61~~, wherein the one or more nucleic acid samples are cell-free nucleic acid samples.

~~63~~17. A computer system for determining a cancer condition of a test subject of a species, the computer system comprising:

at least one processor; and

a memory storing at least one program for execution by the at least one processor, the at least one program comprising instructions for:

A) obtaining a dataset, in electronic form, wherein the dataset comprises a corresponding methylation pattern of each respective fragment in a plurality of fragments, wherein the corresponding methylation pattern of each respective fragment (i) is determined by a methylation sequencing of one or more nucleic acid samples comprising the respective fragment in a biological sample obtained from the test subject and (ii) comprises a methylation state of each CpG site in a corresponding plurality of CpG sites in the respective fragment;

B) constructing a first patch comprising a first channel, the first patch representing a first independent set of CpG sites in a reference genome of the species, each respective CpG site in the first independent set of CpG sites corresponding to a predetermined location in the reference genome, wherein:

the first channel of the first patch comprises a plurality of instances of a first plurality of parameters, wherein each instance of the first plurality of parameters includes a

parameter for a methylation status of a respective CpG site in the first independent set of CpG sites for the first patch,

the constructing B) comprises populating, for each respective fragment in the plurality of fragments that aligns to the first independent set of CpG sites, an instance of all or a portion of the first plurality of parameters based on the methylation pattern of the respective fragment; and

C) applying at least the first patch to a classifier thereby determining the cancer condition in the test subject.

6418. A non-transitory computer-readable storage medium having stored thereon program code instructions that, when executed by a processor, cause the processor to perform a method of determining a cancer condition of a test subject of a species, the method comprising:

A) obtaining a dataset, in electronic form, wherein the dataset comprises a corresponding methylation pattern of each respective fragment in a plurality of fragments, wherein the corresponding methylation pattern of each respective fragment (i) is determined by a methylation sequencing of one or more nucleic acid samples comprising the respective fragment in a biological sample obtained from the test subject and (ii) comprises a methylation state of each CpG site in a corresponding plurality of CpG sites in the respective fragment;

B) constructing a first patch comprising a first channel, the first patch representing a first independent set of CpG sites in a reference genome of the species, each respective CpG site in the first independent set of CpG sites corresponding to a predetermined location in the reference genome, wherein:

the first channel of the first patch comprises a plurality of instances of a first plurality of parameters, wherein each instance of the first plurality of parameters includes a parameter for a methylation status of a respective CpG site in the first independent set of CpG sites for the first patch,

the constructing B) comprises populating, for each respective fragment in the plurality of fragments that aligns to the first independent set of CpG sites, an instance of all or a portion of the first plurality of parameters based on the methylation pattern of the respective fragment; and

C) applying at least the first patch to a classifier thereby determining the cancer condition in the test subject.

6519. A method of determining a cancer condition of a test subject of a species, the method comprising:

A) obtaining, via one or more processors, a training dataset from one or more training subjects, wherein the training dataset comprises one or more training methylation patterns of a plurality of fragments in one or more biological samples obtained from the one or more training subjects and one or more predetermined cancer conditions associated with the one or more training methylation patterns;

B) constructing, via the one or more processors, one or more patches based on the training dataset, each patch of the one or more patches comprising one or more channels and representing one or more CpG sites in a reference genome of the species, each CpG site of the one or more CpG sites corresponding to a predetermined location in the reference genome;

C) training, via the one or more processors, a computational model based on the one or more patches and the training dataset;

D) obtaining, via the one or more processors, a test dataset from the test subject, wherein the test dataset comprises one or more testing methylation patterns of a plurality of fragments in one or more biological samples obtained from the test subject; and

E) determining, via the one or more processors, the cancer condition of the test subject based on the test dataset and the computational model.

6620. The method of claim 6519, wherein:

(i) the one or more training methylation patterns (~~ia~~) are determined by at least one methylation sequencing of one or more nucleic acid samples comprising the plurality of fragments in the one or more biological samples obtained from the one or more training subjects; and (~~ib~~) comprise at least one methylation state of each CpG site in the plurality of fragments in the one or more biological samples obtained from the one or more training subjects; or

~~67. The method of claim 65,~~ (ii) wherein the one or more testing methylation patterns (i) are determined by a methylation sequencing of one or more nucleic acid samples comprising the plurality of fragments in the one or more biological samples obtained from the test subject and (ii) comprise at least one methylation state of each CpG site in the plurality of fragments in the one or more biological samples obtained from the test subject; or

68. ~~The method of claim 65,~~ (iii) wherein the computational model comprises a convolutional neural network and a second stage model-; and optionally:

(a) wherein the convolutional neural network is a pre-trained convolutional neural network; and further optionally:

(b) wherein the pre-trained convolutional neural network comprises a customized VGG-11 convolutional neural network, wherein the customized VGG-11 convolutional neural network comprises customized filter size and activation function;
or

69. ~~The method of claim 65,~~(iv) further comprising, prior to step B), pruning the plurality of fragments by removing from the plurality of fragments each respective fragment, whose corresponding methylation pattern across a corresponding plurality of CpG sites in the respective fragment, has a p-value that fails to satisfy a p-value threshold-; and optionally,

70. ~~The method of claim 69,~~ (a) wherein the p-value of the respective fragment is determined based upon methylation patterns associated with a plurality of reference fragments obtained by a methylation sequencing of nucleic acid from one or more biological samples obtained from a cohort of healthy subjects-; or

71. ~~The method of claim 65,~~ (v) wherein the one or more channels comprise a first channel, the first channel comprising a plurality of instances of a first plurality of parameters, wherein each instance of the first plurality of parameters includes a parameter for a methylation status of a respective CpG site in a first independent set of CpG sites for a patch of the one or more patches-; and optionally,

72. ~~The method of claim 71,~~ (a) wherein the constructing B) comprises populating, for each respective fragment in the plurality of fragments in one or more biological samples obtained from the one or more training subjects that aligns to the first independent set of CpG sites, an instance of all or a portion of the first plurality of parameters based on the training methylation pattern of the respective fragment-; or

73. ~~The method of claim 71,~~ (b) wherein the constructing B), for a respective fragment in the plurality of fragments, comprises:

i) identifying, within an instance of the first plurality of parameters of the first channel, parameters, corresponding to the CpG sites in the respective fragment, that have not previously been assigned methylation states based on another fragment in the plurality of fragments; and

_____ii) assigning for each parameter, among the identified parameters, that aligns to a corresponding CpG site of the respective fragment, the methylation state of the corresponding CpG site of the respective fragment; or

74. ~~The method of claim 71,~~ _____(c) wherein the one or more channels comprise a second channel, the second channel comprising a corresponding instance of a second plurality of parameters for each instance of the first plurality of parameters, wherein each instance of the second plurality of parameters includes a parameter for a first characteristic, other than CpG methylation state, of a respective CpG site in the first independent set of CpG sites for the first patch; and optionally,

75. ~~The method of claim 74,~~ _____(i) wherein the one or more channels comprise a third channel, the third channel comprising a corresponding instance of a third plurality of parameters for each instance of the first plurality of parameters, wherein each instance of the third plurality of parameters includes a parameter for a second characteristic of a respective CpG site in the first independent set of CpG sites; or

76. ~~The method of claim 74,~~ _____(ii) wherein the first characteristic of the respective CpG site is a multiplicity of the respective fragment the respective CpG site is on; or

77. ~~The method of claim 74,~~ _____(iii) wherein the first characteristic of the respective CpG site comprises at least one of:

- a CpG β -value drawn from a healthy cohort,
- a CpG β -value drawn from a predetermined tissue type in a healthy cohort,
- a CpG β -value drawn from the test subject,
- a Pearson's correlation score for methylation state of 5' and 3' neighbor CpG sites,
- a Jaccard similarity, Euclidean distance, Manhattan distance, maximum value, normalized Euclidean distance, normalized maximum value, dice coefficient, or cosine

coefficient of methylation state of the respective CpG site in the test subject versus a cohort of subjects,

- a fragment p-value of the respective fragment,
- a length of the respective fragment the respective CpG site is on,
- a fragment sequence source,
- a fragment mapping quality score of the respective fragment the respective CpG site is on,
- a distance to a 5' adjacent CpG site in the reference genome,
- a distance to a 3' adjacent CpG site in the reference genome,
- a multiplicity of the respective fragment the respective CpG site is on,
- a genetic element the respective CpG site is within,
- a biological pathway the respective CpG site is associated with,
- a gene the respective CpG site is associated with,
- a value of a CpG transition impulse function for the respective CpG site,
- a value of a CpG run-length encoding for the respective CpG site, and
- a read strand orientation of the fragment the respective CpG site is on; or

~~78. The method of claim 65, (vi) further comprising transmitting, via the one or more processors, the cancer condition to an electronic record associated with a user device of the test subject; or~~

~~79. The method of claim 65, (vii) wherein the cancer condition comprises level of cancer, tissues of origin, and metastatic disease statuses; or~~

~~80. The method of claim 68, wherein the convolutional neural network is a pre-trained convolutional neural network.~~

~~81. The method of claim 80, wherein the pre-trained convolutional neural network comprises a customized VGG-11 convolutional neural network, wherein the customized VGG-11 convolutional neural network comprises customized filter size and activation function.~~

~~82. The method of claim 65, (viii) further comprising updating the computational model using one or more biological priors.~~