

CLAIMS

WHAT IS CLAIMED IS:

1. A computer-implemented method for determining a presence or absence of a genetic aberration in deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from a subject, the method comprising:
 - (a) constructing, by a computer, a multi-parametric distribution of the DNA fragments over a plurality of base positions in a genome; and
 - (b) without taking into account a base identity of each base position in a first locus, using the multi-parametric distribution to determine the presence or absence of the genetic aberration in the first locus in the subject.
2. The method of claim 1, wherein the genetic aberration comprises a sequence aberration or a copy number variation (CNV), wherein the sequence aberration is selected from the group consisting of: (i) a single nucleotide variant (SNV), (ii) an insertion or deletion (indel), and (iii) a gene fusion.
3. The method of claim 1, wherein the multi-parametric distribution comprises parameters indicative of one or more of: (i) a length of the DNA fragments that align with each of the plurality of base positions in the genome, (ii) a number of the DNA fragments that align with each of the plurality of base positions in the genome, and (iii) a number of the DNA fragments that start or end at each of the plurality of base positions in the genome.
4. The method of claim 1, further comprising using the multi-parametric distribution to determine a distribution score, wherein the distribution score is indicative of a mutation burden of the genetic aberration.
5. The method of claim 4, wherein the distribution score comprises values indicating one or more of a number of the DNA fragments with dinucleosomal protection and a number of the DNA fragments with mononucleosomal protection.
6. A computer-implemented classifier for determining genetic aberrations in a test subject using deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from the test subject, comprising:
 - (a) an input of a set of distribution scores for each of one or more populations of cell-free DNA obtained from each of a plurality of subjects, wherein each distribution score is generated based at least on one or more of: (i) a length of the DNA fragments that align with each of a plurality of base positions in a genome, (ii) a number of the DNA fragments that align with each of a plurality of base positions in a genome, and (iii) a number of the DNA fragments that start or end at each of a plurality of base positions in a genome; and
 - (b) an output of classifications of one or more genetic aberrations in the test subject.

7. A computer-implemented method for determining genetic aberrations in a test subject using deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from the test subject, the method comprising:
 - (a) providing a computer-implemented classifier configured to determine genetic aberrations in a test subject using DNA fragments from cell-free DNA obtained from the test subject, the classifier trained using a training set;
 - (b) providing as inputs into the classifier a set of distribution scores for the test subject, wherein each distribution score is indicative of one or more of: (i) a length of the DNA fragments that align with each of a plurality of base positions in a genome, (ii) a number of the DNA fragments that align with each of a plurality of base positions in a genome, and (iii) a number of the DNA fragments that start or end at each of a plurality of base positions in a genome; and
 - (c) using the classifier to generate, by a computer, a classification of genetic aberrations in the test subject.
8. A computer-implemented method for analyzing cell-free deoxyribonucleic acid (DNA) fragments derived from a subject, the method comprising:

obtaining sequence information representative of the cell-free DNA fragments; and
performing a multi-parametric analysis on a plurality of data sets using the sequence information to generate a multi-parametric model representative of the cell-free DNA fragments, wherein the multi-parametric model comprises three or more dimensions.
9. The method of claim 8, wherein the data sets are selected from the group consisting of: (a) start position of DNA fragments sequenced, (b) end position of sequenced DNA fragments, (c) number of unique sequenced DNA fragments that cover a mappable position, (d) length of sequenced DNA fragments, (e) a likelihood that a mappable base-pair position will appear at a terminus of a sequenced DNA fragment, (f) a likelihood that a mappable base-pair position will appear within a sequenced DNA fragment as a consequence of differential nucleosome occupancy, (g) a sequence motif of sequenced DNA fragments, (h) GC content, (i) sequenced DNA fragment length distribution, and (j) methylation status.
10. The method of claim 8, wherein the multi-parametric analysis comprises mapping to each of a plurality of base positions or regions of a genome, one or more distributions selected from the group consisting of:
 - (i) a distribution of the number of unique cell-free DNA fragments containing a sequence that covers the mappable position in the genome,
 - (ii) a distribution of the fragment lengths for each of at least some of the cell-free DNA fragments such that the DNA fragment contains a sequence that covers the mappable

- position in the genome, and
- (iii) a distribution of the likelihoods that a mappable base-pair position will appear at a terminus of a sequenced DNA fragment.
11. The method of claim 10, wherein the plurality of base positions or regions of a genome include at least one base position or region associated with one or more of the genes listed in **Table 1**.
 12. The method of claim 10, wherein the mapping comprises mapping a plurality of values from each of a plurality of the data sets, to each of a plurality of base positions or regions of a genome.
 13. The method of claim 12, wherein at least one of the plurality of values is a data set selected from the group consisting of (a) start position of DNA fragments sequenced, (b) end position of sequenced DNA fragments, (c) number of unique sequenced DNA fragments that cover a mappable position, (d) length of sequenced DNA fragments, (e) a likelihood that a mappable base-pair position will appear at a terminus of a sequenced DNA fragment, (f) a likelihood that a mappable base-pair position will appear within a sequenced DNA fragment as a consequence of differential nucleosome occupancy, or (g) a sequence motif of sequenced DNA fragments.
 14. The method of claim 8, wherein the multi-parametric analysis comprises applying, by a computer, one or more mathematical transforms to generate the multi-parametric model.
 15. The method of claim 8, wherein the multi-parametric model is a joint distribution model of a plurality of variables selected from the group consisting of: (a) start position of DNA fragments sequenced, (b) end position of sequenced DNA fragments, (c) number of unique sequenced DNA fragments that cover a mappable position, (d) length of sequenced DNA fragments, (e) a likelihood that a mappable base-pair position will appear at a terminus of a sequenced DNA fragment, (f) a likelihood that a mappable base-pair position will appear within a sequenced DNA fragment as a consequence of differential nucleosome occupancy, and (g) a sequence motif of sequenced DNA fragments.
 16. The method of claim 8, further comprising identifying in the multi-parametric model, one or more peaks, each peak having a peak distribution width and a peak coverage.
 17. The method of claim 16, further comprising detecting one or more deviations between the multi-parametric model representative of the cell-free DNA fragments and a reference multi-parametric model.
 18. The method of claim 17, wherein the deviation is selected from the group consisting of:
 - (i) an increase in the number of reads outside a nucleosome region,
 - (ii) an increase in the number of reads within a nucleosome region,

- (iii) a broader peak distribution relative to a mappable genomic location,
 - (iv) a shift in location of a peak,
 - (v) identification of a new peak,
 - (vi) a change in depth of coverage of a peak,
 - (vii) a change in start position around a peak, and
 - (viii) a change in fragment sizes associated with a peak.
19. The method of claim 8, further comprising determining a contribution of the multi-parametric model attributed to (i) apoptotic processes in cells from which the cell-free DNA originated or (ii) necrotic processes in cells from which the cell-free DNA originated.
20. The method of claim 8, further comprising performing a multi-parametric analysis to (i) measure RNA expression of the cell-free DNA fragments, (ii) measure methylation of the cell-free DNA fragments, (iii) measure a nucleosomal mapping of the cell-free DNA fragments, or (iv) identify the presence of one or more somatic single nucleotide polymorphisms in the cell-free DNA fragments or one or more germline single nucleotide polymorphisms in the cell-free DNA fragments.
21. The method of claim 8, further comprising generating a distribution score comprising values indicating a number of the DNA fragments with dinucleosomal protection or a number of the DNA fragments with mononucleosomal protection.
22. The method of claim 8, further comprising estimating a mutation burden of the subject.
23. A computer-implemented method for analyzing cell-free deoxyribonucleic acid (DNA) fragments derived from a subject, the method comprising:
obtaining a multi-parametric model representative of the cell-free DNA fragments; and
performing, with the computer, statistical analysis to classify the multi-parametric model as being associated with one or more nucleosomal occupancy profiles representing distinct cohorts.
24. A computer-implemented method for creating a trained classifier, comprising:
(a) providing a plurality of different classes, wherein each class represents a set of subjects with a shared characteristic;
(b) for each of a plurality of populations of cell-free DNA obtained from each of the classes, providing a multi-parametric model representative of cell-free deoxyribonucleic acid (DNA) fragments from the populations of cell-free DNA, thereby providing a training data set; and
(c) training, by a computer, a learning algorithm on the training data set to create one or more trained classifiers, wherein each trained classifier is configured to classify a test

- population of cell-free DNA from a test subject into one or more of the plurality of different classes.
25. A method of classifying a test sample from a subject, comprising:
- (a) providing a multi-parametric model representative of cell-free deoxyribonucleic acid (DNA) fragments from a test population of cell-free DNA from the subject; and
 - (b) classifying the test population of cell-free DNA using a trained classifier.
26. A computer-implemented method comprising:
- (a) generating, by a computer, sequence information from cell-free DNA fragments from a subject;
 - (b) mapping, by a computer, the cell-free DNA fragments to a reference genome based on the sequence information; and
 - (c) analyzing, by a computer, the mapped cell-free DNA fragments to determine, at each of a plurality of base positions in the reference genome, a plurality of measures selected from the group consisting of:
 - (i) number of cell-free DNA fragments mapping to the base position,
 - (ii) length of each cell-free DNA fragment mapping to the base position,
 - (iii) number of cell-free DNA fragments mapping to the base position as a function of length of the cell-free DNA fragment;
 - (iv) number of cell-free DNA fragments starting at the base position;
 - (v) number of cell-free DNA fragments ending at the base position;
 - (vi) number of cell-free DNA fragments starting at the base position as a function of length, and
 - (vii) number of cell-free DNA fragments ending at the base position as a function of length.
27. A computer-implemented method of analyzing cell-free DNA fragments derived from a subject, the method comprising:
- (a) receiving, by a computer, sequence information representative of the cell-free DNA fragments, and
 - (b) performing an analysis per mappable base position or genome position, comprising a plurality of:
 - (i) the number of sequence fragments that start or end at the base position or genome position,
 - (ii) sequence or fragment lengths at the base position or genome position,
 - (iii) fragment or sequence coverage at the base position or genome position, and
 - (iv) sequence motif distribution at the base position or genome position.

28. A method of generating a classifier for determining a likelihood that a subject belongs to one or more classes of clinical significance, the method comprising:
- a) providing a training set comprising, for each of the one or more classes of clinical significance, populations of cell-free DNA from each of a plurality of subjects of a species belonging to the class of clinical significance and from each of a plurality of subjects of the species not belonging to the class of clinical significance;
 - b) sequencing cell-free DNA fragments from the populations of cell-free DNA to produce a plurality of DNA sequences;
 - c) for each population of cell-free DNA, mapping the plurality of DNA sequences to each of one or more genomic regions in a reference genome of the species, each genomic region comprising a plurality of genetic loci;
 - d) preparing, for each population of cell-free DNA, a dataset comprising, for each of a plurality of the genetic loci, values indicating a quantitative measure of at least one characteristic selected from:
 - (i) DNA sequences mapping to the genetic locus, (ii) DNA sequences starting at the locus, and (iii) DNA sequences ending at the genetic locus, to yield a training set; and
 - e) training a computer-based machine learning system on the training set, thereby generating a classifier for determining a likelihood that the subject belongs to one or more classes of clinical significance.
29. A method of determining an abnormal biological state in a subject, the method comprising:
- a) sequencing cell-free DNA fragments from cell-free DNA from the subject to produce DNA sequences;
 - b) mapping the DNA sequences to each of one or more genomic regions in a reference genome of a species of the subject, each genomic region comprising a plurality of genetic loci;
 - c) preparing a dataset comprising, for each of a plurality of the genetic loci, values indicating a quantitative measure of at least one feature selected from:
 - (i) DNA sequences mapping to the genetic locus, (ii) DNA sequences starting at the locus, and (iii) DNA sequences ending at the genetic locus; and
 - d) based on the dataset, determining a likelihood of the abnormal biological state.
30. A computer-implemented method for generating an output indicative of a presence or absence of a genetic aberration in deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from a subject, the method comprising:
- (a) constructing, by a computer, a distribution of the DNA fragments from the cell-free DNA over a plurality of base positions in a genome; and

- (b) for each of one or more genetic loci, calculating, by a computer, a quantitative measure indicative of a ratio of (1) a number of the DNA fragments with dinucleosomal protection associated with a genetic locus from the one or more genetic loci, and (2) a number of the DNA fragments with mononucleosomal protection associated with the genetic locus, or vice versa; and
- (c) determining, using the quantitative measure for each of the one or more genetic loci, said output indicative of a presence or absence of the genetic aberration in the one or more genetic loci in the subject.
31. A computer-implemented method for generating an output indicative of a presence or absence of a genetic aberration in deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from a subject, the method comprising:
- (a) constructing, by a computer, a distribution of the DNA fragments from the cell-free DNA over a plurality of base positions in a genome; and
- (b) using the distribution to determine said output indicative of a presence or absence of the genetic aberration in the subject, wherein the presence or absence is determined (i) without comparing the distribution of the DNA fragments to a reference distribution from a source external to a genome of the subject, (ii) without comparing parameters derived from the distribution of the DNA fragments to reference parameters, and (iii) without comparing the distribution of the DNA fragments to a reference distribution from a control of the subject.
32. The method of claim 31, wherein the genetic aberration comprises a copy number variation (CNV) or a single nucleotide variant (SNV).
33. A computer-implemented method for deconvolving a distribution of deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from a subject, the method comprising:
- (a) constructing, by a computer, a distribution of a coverage of the DNA fragments from the cell-free DNA over a plurality of base positions in a genome; and
- (b) for each of one or more genetic loci, deconvolving, by a computer, the distribution of the coverage, thereby generating fractional contributions associated with one or more members selected from the group consisting of a copy number (CN) component, a cell clearance component, and a gene expression component.
34. The method of claim 33, further comprising generating an output indicative of a presence or absence of a genetic aberration based at least on a portion of the fractional contributions.
35. A computer-implemented method for generating an output indicative of a presence or absence of a genetic aberration in deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from a subject, the method comprising:

- (a) constructing, by a computer, a distribution of the DNA fragments from the cell-free DNA over a plurality of base positions in a genome;
 - (b) identifying, by a computer, one or more peaks at one or more base positions of the plurality of base positions in the distribution of the DNA fragments, wherein each peak comprises a peak value and a peak distribution width; and
 - (c) determining, by a computer, based at least on (i) the one or more base positions, (ii) the peak value, and (iii) the peak distribution width, the presence or absence of the genetic aberration in the subject.
36. The method of claim 35, wherein the one or more peaks comprises a dinucleosomal peak or a mononucleosomal peak.
37. The method of claim 36, wherein said output indicative of a presence or absence of the genetic aberration is determined based at least on a quantitative measure indicative of a ratio of a first peak value associated with the dinucleosomal peak and a second peak value associated with the mononucleosomal peak, or vice versa.
38. A computer-implemented method for generating an output indicative of a presence or absence of a genetic aberration in deoxyribonucleic acid (DNA) fragments from cell-free DNA obtained from a subject, the method comprising:
- (a) constructing, by a computer, a distribution of the DNA fragments from the cell-free DNA over a plurality of base positions in a genome;
 - (b) analyzing, by a computer, the distribution of the DNA fragments at one or more genetic loci, which analyzing comprises detecting deviations between the distribution of the DNA fragments and a plurality of reference distributions selected from: (i) one or more healthy reference distributions associated with one or more cohorts of healthy controls, and (ii) one or more diseased reference distributions associated with one or more cohorts of diseased subjects; and
 - (c) determining, by a computer, based at least on the deviations detected in (b), said output indicative of a presence or absence of the genetic aberration in the subject.
39. The method of claim 38, wherein analyzing comprises calculating one or more delta signals, each delta signal comprising a difference between the distribution of the DNA fragments and a reference distribution of the plurality of reference distributions.
40. A method for processing a biological sample of a subject, comprising:
- (a) obtaining said biological sample of said subject, wherein said biological sample comprises deoxyribonucleic acid (DNA) fragments;
 - (b) assaying said biological sample to generate a signal(s) indicative of a presence or absence of DNA fragments with (i) dinucleosomal protection associated with a genetic

- locus from one or more genetic loci, and (ii) mononucleosomal protection associated with the genetic locus; and
- (c) using said signal(s) to generate an output indicative of said presence or absence of DNA fragments with (i) dinucleosomal protection associated with a genetic locus from one or more genetic loci, and (ii) mononucleosomal protection associated with the genetic locus.
41. The method of claim 40, wherein assaying comprises (i) enriching said biological sample for DNA fragments for a set of one or more genetic loci or (ii) sequencing said DNA fragments of said biological sample.
42. A method for analyzing a biological sample comprising cell-free DNA fragments derived from a subject, the method comprising detecting DNA fragments from the same genetic locus which correspond to each of mononucleosomal protection and dinucleosomal protection.
43. A method for analyzing a biological sample comprising cell-free DNA fragments derived from a subject, the method comprising detecting DNA fragments with dinucleosomal protection associated with a genetic locus.
44. The method of claim 43, wherein the genetic locus comprises ERBB2, TP53, or NF1.