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, 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; 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, 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.

4. The method of claim 3, 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.

5. 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.

6. 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.

7. A computer-implemented method for detecting an epigenetic aberration in cell-free deoxyribonucleic acid (DNA) fragments derived from a subject, the method comprising:

obtaining sequence information representative of the cell-free DNA fragments;

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, and parameters indicative of two 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, 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; and

using the multi-parametric model to detect the epigenetic aberration in the cell-free DNA fragments.

8. The method of claim 7, 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.

9. The method of claim 7, 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.

10. The method of claim 9, 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 following genes:

AKT1, ALK, APC, AR, ARAF, ARID1A, ATM, BRAF, BRCA1, BRCA2, CCND1, CCND2, CCNE1, CDH1, CDK4, CDK6, CDKN2A, CDKN2B, CTNNB1, EGFR, EGFR (exons 19 and 20), ERBB2, ERBB2 (exons 19 and 20), ESR1, EZH2, FBXW7, FGFR1, FGFR2, FGFR3, GATA3, GNA11, GNAQ, GNAS, HNF1A, HRAS, IDH1, IDH2, JAK2, JAK3, KIT, KRAS, MAP2K1, MAP2K2, MET, MET (exon 14 skipping), MLH1, MPL, MYC, NF1, NFE2L2, NOTCH1, NPM1, NRAS, NTRK1, PDGFRA, PIK3CA, PTEN, PTPN11, RAF1, RB1, RET, RHEB, RHOA, RIT1, ROS1, SMAD4, SMO, SRC, STK11, TERT, TP53, TSC1, VHL.

11. The method of claim 9, 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.

12. The method of claim 11, 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.

13. The method of claim 7, wherein the multi-parametric analysis comprises applying, by a computer, one or more mathematical transforms to generate the multi-parametric model.

14. The method of claim 7, further comprising identifying in the multi-parametric model, one or more peaks, each peak having a peak distribution width and a peak coverage.

15. The method of claim 14, 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.

16. The method of claim 15, 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.

17. The method of claim 7, 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.

18. The method of claim 7, 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.

19. The method of claim 7, 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.

20. The method of claim 7, further comprising estimating a mutation burden of the subject.