We claim:

1. A method for genetic screening a subject for cancer, comprising

(A) receiving a subject-specific genome-wide compendium of reads associated with a plurality of genetic markers from a biological sample of the subject, the biological sample comprising a plasma sample~~, wherein the compendium of reads each comprise reads of a single base pair length~~;

(B) filtering artefactual sites from the compendium of reads, wherein the filtering comprises (a) removing, from the compendium of reads, recurring sites generated over a cohort of reference healthy samples; and/or (b) identifying germ line mutations in the biological sample and/or identifying shared mutations between a tumor sample and peripheral blood mononuclear cells of a normal cell sample as germ line mutations, and removing said germ line mutations from the compendium of reads;

(C) filtering noise from the compendium of reads using at least one error suppression protocol to produce a filtered read set for the genome-wide compendium of reads, wherein the at least one error suppression protocol comprises (a) calculating a probability that any single nucleotide variation in the compendium is an artefactual mutation, and removing said mutation, wherein the probability is calculated as a function of features selected from the group comprising mapping-quality (MQ), variant base-quality (MBQ), position-in-read (PIR), mean read base quality (MRBQ), and combinations thereof; and/or (b) removing artefactual mutations using discordance testing between independent replicates of the same DNA fragment generated from polymerase chain reaction or sequencing processing, and/or duplication consensus wherein artefactual mutations are identified and removed when lacking concordance across a majority of a given duplication family;

(D) compiling a subject-specific signature using the filtered read set, based on comparison to specific mutational signatures associated with a pre-determined mutagenesis process;

(E) statistically quantitating a confidence estimate that the subject's biological sample, via the subject-specific signature, comprises a cancer related mutational signature based on comparison of the cancer related mutational signature exposure value to a cohort of background mutation signatures; and

(F) screening the subject's biological sample for cancer if the confidence estimate that the subject's biological sample contains the cancer related mutational signature exceeds a given threshold.

2. A method for genetic screening a subject for cancer, comprising,

(A) receiving a subject-specific genome-wide compendium of reads associated with a plurality of genetic markers from a biological sample of a subject, the biological sample comprising a plasma sample, wherein the compendium of reads each comprise a copy number variation (CNV) or structural variation (SV);

(B) dividing the compendium of reads into a plurality of windows;

(C) calculating a set of features per window, the features comprising a median depth coverage per window and a representative fragment size per window, and split reads;

(D) filtering artefactual sites from the compendium of reads, wherein the filtering comprises removing, from the compendium of reads, recurring sites generated over a cohort of reference healthy samples;

(E) normalizing the compendium of reads to produce a filtered read set for the genome-wide compendium of reads;

(F) computing an estimated tumor fraction using the filtered read set (i) by calculating a linear relationship between the set of features per window and converting the calculated relationship to estimated tumor fraction using a regression model, and/or (ii) on the basis of one or more integrative mathematical models as a function of the calculated set of features per window across the subject-specific genome-wide compendium of reads; and

(G) screening the subject's biological sample for cancer if the estimated tumor fraction exceeds an empirical threshold.

3. A system for genetic screening a subject for cancer, comprising,

an analyzing unit, the analyzing unit comprising

a pre-filter engine configured and arranged to receive a subject-specific genome-wide compendium of reads associated with a plurality of genetic markers from a biological sample of a subject, the biological sample comprising a plasma sample~~, wherein the compendium of reads each comprise reads of a single base pair length~~; and filter artefactual sites from the compendium of reads, wherein the filtering comprises removing, from the compendium of reads, recurring sites generated over a cohort of reference healthy samples; and/or identifying germ line mutations in the biological sample and/or identifying shared mutations between a tumor sample and peripheral blood mononuclear cells of the normal cell sample as germ line mutations, and removing said germ line mutations from the compendium of reads;

a correction engine configured and arranged to filter noise from the compendium of reads using at least one error suppression protocol to produce a filtered read set for the genome-wide

compendium of reads, wherein the at least one error suppression protocol comprises (a) calculating the probability that any single nucleotide variation in the compendium is an artefactual mutation, and removing said mutation, wherein the probability is calculated as a function of features selected from the group comprising mapping-quality (MQ), variant base-quality (MBQ), position-in-read (PIR), mean read base quality (MRBQ), and combinations thereof; and/or (b) removing artefactual mutations using discordance testing between independent replicates of the same DNA fragment generated from polymerase chain reaction or sequencing processing, and/or duplication consensus wherein artefactual mutations are identified and removed when lacking concordance across a majority of a given duplication family; and

a computing unit configured and arranged to compile a subject-specific signature using the filtered read set, based on comparison to specific mutational signatures associated with a pre-determined mutagenesis process; statistically quantitating a confidence estimate that the subject's biological sample, via the subject-specific signature, comprises a cancer related mutational signature based on comparison of the cancer related mutational signature exposure value to a cohort of background mutation signatures; and screen the subject for cancer if the confidence estimate that the subject's biological sample contains the cancer related mutational signature exceeds a given threshold.

4. A system for detecting residual disease in a subject in need thereof, comprising,

an analyzing unit, the analyzing unit comprising

a binning engine configured and arranged to receive a subject-specific genome-wide compendium of reads associated with a plurality of genetic markers from a biological sample of a subject, the biological sample comprising a plasma sample, wherein the compendium of reads each comprise a copy number variation (CNV); divide the compendium of reads into a plurality of windows; and calculate a set of features per window, the features comprising a median depth coverage per window and a representative fragment size per window;

a pre-filter engine configured and arranged to filter artefactual sites from the compendium of reads, wherein the filtering comprises removing, from the compendium of reads, recurring sites generated over a cohort of reference healthy samples; and

a normalization engine configured and arranged to normalize the compendium of reads to produce a filtered read set for the genome-wide compendium of reads;

and

a computing unit configured and arranged to compute an estimated tumor fraction using the filtered read set (i) by calculating a linear relationship between the set of features per window

and converting the calculated relationship to estimated tumor fraction using a regression model, and/or (ii) on the basis of one or more integrative mathematical models as a function of the calculated set of features per window across the subject-specific genome-wide compendium of reads; and screen the subject for cancer if the estimated tumor fraction exceeds an empirical threshold.

5. The method of claim 1, wherein the markers comprise single nucleotide variations (SNVs) or insertion/deletions (indels).

6. The method of claim 1, wherein filtering recurring sites generated over a cohort of reference healthy samples comprises generating a panel of normal (PON) blacklist or mask.

7. The method of claim 1, wherein the reference healthy sample comprises peripheral blood mononuclear cells (PBMC).

8. The method of claim 1, wherein step (C) comprises employing a machine learning (ML) algorithm comprising one of deep convolutional neural network (CNN), recurrent neural network (RNN), random forest (RF), support vector machine (SVM), discriminant analysis, nearest neighbor analysis (KNN), ensemble classifier, or a combination thereof to filter artefactual noise.

9. The method of claim 1, wherein in step (C)(b), correction of artefactual variation includes correction of artefactual mutations generated by PCR or sequencing using the comparison of independent replicates of the original nucleic acid fragment.

10. The method of claim 9, wherein in step (C)(b), the artefactual variations generated by paired-end 150 bp sequencing, which results in overlapping paired reads (R1 and R2), are removed by correcting back, to the corresponding reference genome, discordance between R1 and R2 pairs.

11. The method of claim 1, wherein in step (C)(b), the artefactual variations generated by duplication during sequencing and/or PCR amplification are corrected, wherein the duplication families are recognized by 5’ and 3’ similarity as well as alignment position and wherein each duplication family is used to check the consensus of a specific mutation across independent replicates, thereby correcting artefactual mutations that do not show concordance in a majority of the duplication family.

12. The method of claim 1, wherein in step (D), specific mutational signatures in a single plasma sample are identified using non-negative least square (NNLS) method.

13. The method of claim 1, wherein in step (E), specific mutational signatures are further validated for confidence using a comparison of the cancer-specific mutation signature exposure values to the exposure values inferred for a plurality of random background signatures.

14. The method of claim 13, wherein in step (F), the subject is identified as having cancer if the confidence estimate that the subject's biological sample contains the cancer related mutational signature exceeds a given threshold for z-score>2std.

15. The method of claim 1, wherein in step (D), additionally or alternatively comprises employing a machine learning (ML) algorithm to distinguish between cancer altered sequencing reads and reads altered by sequencing errors.

16. The method of claim 15, wherein the ML is trained on a plurality of true mutated reads and error using a large collection of tumor and normal WGS data and the trained ML is capable of distinguishing between a read that contains a true variant versus a read that contain a sequencing artifact.

17. The method of claim 1, further comprising orthogonal integration of a secondary feature comprising fragment size shift.

18. The method of claim 17, wherein intra-patient fragment size shifts in list of tumor-specific markers and random markers are analyzed using statistical methods that include tests for significance or Gaussian mixture model (GMM).

19. The method of claim 2, wherein the markers comprise copy number variations (CNVs).

20. The method of claim 2, wherein in step (B), each window is at least ≥ 150 bp.

21. The method of claim 2, wherein step (C) comprises extraction of depth coverage (Log2) and fragment size (COM) relationship (slope, R^2) from the genome-wide feature vectors.

22. The method of claim 2, wherein step (D) comprises filtering recurring sites generated over a cohort of reference healthy plasma samples by generating a panel of normal (PON) blacklist or mask; and/or filtering windows of low mappability or coverage.

23. The method of claim 2, wherein the normalization step includes normalizing depth coverage values to correct for GC-content and mappability biases by performing two LOESS regression curve-fitting on the bin-wise GC-fraction and mappability score.

24. The method of claim 2, wherein the normalization step includes batch-effect correction using a robust-zscore normalization, which is applied to each sample separately.

25. The method of claim 24, wherein the zscore normalization includes calculation of median and median-absolute-deviation (MAD) based on the neutral regions of each sample and normalizing all CNV bins are normalized by subtracting the median value and dividing the differential by MAD.

26. The method of claim 2, wherein step (E) includes calculating depth coverage skew and/or fragment size center-of-mass (COM) skew in the plasma sample in comparison to a panel of normal (PON) healthy plasma samples.

27. The method of claim 2, wherein step (F) includes copy-number-variation (CNV) calling and calculation of tumor fraction of the filtered read set using a hidden Markov model or a self-organizing neural network that includes a neural network based on Adaptive Resonance Theory (ART) or selforganizing map (SOM).

28. The method of claim 2, further comprising orthogonal integration of a secondary feature comprising fragment size shift.

29. The method of claim 28, wherein intra-patient fragment size shifts in list of tumor-specific markers and random markers are analyzed using statistical methods that include tests for significance or Gaussian mixture model (GMM).

30. The method of claim 1, wherein in step (E), statistically quantitating the confidence estimate comprises:

determining a cancer related mutational signature exposure value compiled from the subject-specific signature using the filtered read set and comparing the exposure value to the cohort of background mutation signatures.

31. The method of claim 1, wherein in step (E), statistically quantitating the confidence estimate comprises:

assigning the confidence estimate to each individual mutation detected in a sample from tumor patients;

integrating the confidence estimates across the entire genome; and

employing a rigid analysis of specific cosmic mutational signatures.

32. The system of claim 3, wherein statistically quantitating the confidence estimate comprises:

determining a cancer related mutational signature exposure value compiled from the subject-specific signature using the filtered read set and comparing the exposure value to the cohort of background mutation signatures.

33. The system of claim 3, wherein statistically quantitating the confidence estimate comprises:

assigning the confidence estimate to each individual mutation detected in a sample from tumor patients;

integrating the confidence estimates across the entire genome; and

employing a rigid analysis of specific cosmic mutational signatures.