WHAT IS CLAIMED IS:

1. A computer implemented method comprising:

generating sequence reads from a tumor nucleic acid sample, a noncancerous nucleic acid sample, and a non-tissue nucleic acid sample collected from the same patient, wherein the sequence reads are generated using whole genome sequencing (WGS);

generating a tumor variant call file, a noncancerous variant call file, and a non-tissue variant call file by analyzing the sequence reads corresponding respectively to the tumor nucleic acid sample, the noncancerous nucleic acid sample, and the non-tissue sample;

comparing the tumor variant call file to the noncancerous variant call file to generate a list of somatic variants;

comparing the list of somatic variants to the non-tissue variant call file to generate a list of candidate somatic variants;

generating, by a classification machine learning model, scores for each of the candidate somatic variants in the list of candidate somatic variants, wherein the scores are generated based on a plurality of classifications generated by the classification machine learning model;

determining, based on the scores, a ctDNA status for the patient, wherein the ctDNA status is either positive or negative; and

generating a report that provides the ctDNA status for the patient.

2. The computer implemented method of claim 1, wherein the tumor nucleic acid sample is cancer positive tissue, wherein the noncancerous nucleic acid sample is white blood cells, and wherein the non-tissue nucleic acid sample is plasma.

3. The computer implemented method of claim 2, wherein the noncancerous nucleic acid sample and the non-tissue nucleic acid sample are collected from the same whole blood sample.

4. The computer implemented method of claim 1, wherein the patient is diagnosed with cancer, received surgery to remove one or more tumors, and received a therapeutic treatment post-surgery.

5. The computer implemented method of claim 4, wherein the patient is diagnosed with colorectal cancer, head and neck cancer, lung cancer, breast cancer, or melanoma.

6. The computer implemented method of claim 1, wherein the tumor nucleic acid sample, the noncancerous samples, and the non-tissue samples are collected (i) pre-surgery, (ii) during surgery, (iii) about 3 days to about 65 days post-surgery and before receiving a therapeutic treatment, (iv) about every 6 months up to 3 years post-surgery and after receiving the therapeutic treatment, or (v) any combination thereof.

7. The computer implemented method of claim 1, wherein the tumor variant call file and the noncancerous variant call file are filtered using a set of filtering criteria, and wherein the set of filtering criteria include removing: (i) variants annotated as low confidence, (ii) variants annotated as indels, (iii) variants observed in genomic databases, (iv) variants overlapping simple tandem repeat tracks, (v) variants at genomic positions with less than 10x coverage, (vi) variants at genomic positions with an alternate allele count less than 4 in the tumor nucleic acid sample or greater than 1 in the noncancerous nucleic acid sample, (vii) variants with a variant allele frequency less than 0.05, or (viii) any combination thereof.

8. The computer implemented method of claim 1, wherein the list of candidate somatic variants comprises substitutions, small indels, chromosomal rearrangements, copy number variation, microsatellite instabilities, or any combination thereof.

9. The computer implemented method of claim 1, wherein prior to generating the scores, the classification model filters, using a set of noncancerous donor samples, the list of candidate somatic variants to generate a filtered list of candidate somatic variants.

10. The computer implemented method of claim 1, wherein the classification machine learning model is a random forest classifier comprising an ensemble of trees having at least 500 decision trees, wherein:

each of the trees generates a score for an input candidate somatic variant,

the random forest classifier averages the scores generated by each of the trees to determine a final score,

the final score is compared to a predetermined threshold to determine whether a ctDNA status of the non-tissue nucleic acid sample is positive or negative,

the ensemble of trees considers at least 50 features associated with the candidate somatic variants, and

each tree considers a different subset of features from the at least 50 features to make a prediction for the class.

11. The computer implemented method of claim 10, wherein the final score is greater than or equal to the predetermined threshold and the ctDNA status is positive, and wherein the final score is less than the predetermined threshold and the ctDNA status is negative.

12. The computer implemented method of claim 1, wherein the ctDNA status is correlated with clinicopathological risk factors to predict survival rate, wherein the clinicopathological risk factors predict recurrence risk, and wherein the clinicopathological risk factors include depth of tumor invasion and spread of tumor to neighboring lymph nodes.

13. The computer implemented method of claim 12, wherein the correlation between the ctDNA status and the clinicopathological risk factors is included in the report, and wherein the report further describes a recurrence risk and a predicted survival rate of the patient, based on the ctDNA status and clinicopathological risk factors of the patient.

14. A system comprising:

one or more processors; and

one or more computer-readable media storing instructions which, when executed by the one or more processors, cause the system to perform operations comprising:

generating sequence reads from a tumor nucleic acid sample, a noncancerous nucleic acid sample, and a non-tissue nucleic acid sample collected from the same patient, wherein the sequence reads are generated using whole genome sequencing (WGS);

generating a tumor variant call file, a noncancerous variant call file, and a non-tissue variant call file by analyzing the sequence reads corresponding respectively to the tumor nucleic acid sample, the noncancerous nucleic acid sample, and the non-tissue sample;

comparing the tumor variant call file to the noncancerous variant call file to generate a list of somatic variants;

comparing the list of somatic variants to the non-tissue variant call file to generate a list of candidate somatic variants;

generating, by a classification machine learning model, scores for each of the candidate somatic variants in the list of candidate somatic variants, wherein the scores are generated based on a plurality of classifications generated by the classification machine learning model;

determining, based on the scores, a ctDNA status for the patient, wherein the ctDNA status is either positive or negative; and

generating a report that provides the ctDNA status for the patient.

15. One or more non-transitory computer-readable media storing instructions which, when executed by one or more processors, cause a system to perform operations comprising:

generating sequence reads from a tumor nucleic acid sample, a noncancerous nucleic acid sample, and a non-tissue nucleic acid sample collected from the same patient, wherein the sequence reads are generated using whole genome sequencing (WGS);

generating a tumor variant call file, a noncancerous variant call file, and a non-tissue variant call file by analyzing the sequence reads corresponding respectively to the tumor nucleic acid sample, the noncancerous nucleic acid sample, and the non-tissue sample;

comparing the tumor variant call file to the noncancerous variant call file to generate a list of somatic variants;

comparing the list of somatic variants to the non-tissue variant call file to generate a list of candidate somatic variants;

generating, by a classification machine learning model, scores for each of the candidate somatic variants in the list of candidate somatic variants, wherein the scores are generated based on a plurality of classifications generated by the classification machine learning model;

determining, based on the scores, a ctDNA status for the patient, wherein the ctDNA status is either positive or negative; and

generating a report that provides the ctDNA status for the patient.