CLAIMS

What is claimed:

1. A method of identifying microbial features for determining a disease of the subject, the method comprising:

(a) exposing a biological sample of the subject to one or more probes, wherein the one or more probes bind non-specifically to one or more nucleic acid molecules of the biological sample;

(b) obtaining a first set of sequencing reads of the one or more nucleic acid molecules bound to the one or more probes;

(c) identifying a second set of sequencing reads within the first set of sequencing reads, wherein the second set of sequencing reads comprise non-human sequencing reads obtained through non-specific hybridizations; and

(d) identifying one or more microbial features for determining the disease of the subject from the second set of sequencing reads.

2. The method of claim 1, further comprising generating taxonomic assignments and abundances for the second set of sequencing reads.

3. The method of claim 1, wherein identifying of step (c) comprises comparing the second set of sequencing reads with a genome database.

4. The method of claim 1, wherein the one or more probes comprise multiplexed oligonucleotide probes that couple non-specifically to one or more microbial nucleic acid molecules.

5. A method of validating microbial features, comprising:

(a) receiving a first set of one or more microbial features of a first biological sample from a first subject with a disease determined by non-specific interactions of one or more probes with one or more nucleic acid molecules of the first biological sample;

(b) training a predictive model with the first set of one or more microbial features of the first biological sample and the disease of the first subject, thereby producing a trained predictive model;

(c) receiving a second set of one or more microbial features of a second biological sample of a second subject with a disease; and

(d) validating the first set of one or more microbial features by comparing a predicted disease provided by the trained predictive model and the disease of the second subject, wherein the predicted disease provided by the trained predictive model is generated when the second set of one or more microbial features are provided as an input to the trained predictive model.

6. The method of claim 5, wherein the first set of one or more microbial features comprises taxonomic assignment and abundances of a first set of microbial sequencing reads, and wherein the second set of one or more microbial features comprises taxonomic assignment and abundance of a second set of microbial sequencing reads.

7. The method of claim 5, further comprising removing one or more contaminant microbial features from the first set of one or more microbial features, the second set of one or more microbial features, or a combination thereof.

8. The method of claim 5, wherein the one or more probes or the second set of one or more probes comprise multiplexed oligonucleotide probes target mammalian genomic regions.

9. The method of claim 5, wherein the first set of one or more microbial features or the second set of one or more microbial features are determined by:

(a) sequencing one or more nucleic acid molecules bound to the first set of one or more probes or a second set of one or more probes, thereby generating one or more sequencing reads;

(b) mapping the one or more sequencing reads to a human genome database to identify one or more non-human sequencing reads; and

(c) determining a first set of one or more microbial features or a second set of one or more microbial features from the one or more non-human sequencing reads.

10. The method of claim 5, wherein the first set of one or more probes or the second set of one or more probes comprise multiplexed oligonucleotide probes that couple non-specifically to one or more microbial nucleic acid molecules.

11. The method of claim 5, wherein the one or more microbial features of the second biological sample are determined by sequencing enriched or non-enriched microbial nucleic acid molecules of the second biological sample.

12. A method, comprising:

(a) exposing a biological sample of a first subject with a first disease to one or more probes, wherein the one or more probes bind non-specifically to one or more nucleic acid molecules of the biological sample;

(b) sequencing the one or more nucleic acid molecules bound to the one or more probes, thereby generating one or more sequencing reads;

(c) mapping the one or more sequencing reads to a genome database, thereby identifying one or more non-human sequencing reads; and

(d) generating a predictive model for predicting a second disease of a second subject, wherein the predictive model is trained with one or more microbial features of the one or more non-human sequencing reads and the first disease of the first subject.

13. The method of claim 12, wherein the one or more microbial features comprise taxonomic assignments and abundances of the one or more non-human sequencing reads.

14. The method of claim 12**,** further comprising removing one or more contaminant microbial features from the one or more microbial features prior to training the predictive model.

15. The method of claim 12, wherein the predictive model is configured to predict a subject's response to chemotherapy, immunotherapy, neoadjuvant therapy, or any combination thereof therapy administered to treat a disease.

16. The method of claim 12, wherein the predictive model is configured to identify and remove one or more contaminate microbial features, while selectively retaining one or more non-contaminant microbial features.

17. The method of claim 12**,** wherein the predictive model comprises a machine learning model.

18. The method of claim 12, wherein the one or more probes comprise multiplexed oligonucleotide probes that couple non-specifically to one or more microbial nucleic acid molecules.

19. A method, comprising:

(a) exposing a biological sample of a subject with a disease to one or more probes, wherein the one or more probes bind non-specifically to one or more nucleic acid molecules of the biological sample;

(b) identifying one or more sequencing reads of the one or more nucleic acid molecule bound to the one or more probes;

(c) mapping the one or more sequencing reads to a genome database, thereby identifying one or more non-human sequencing reads of the one or more sequencing reads; and

(d) identifying one or more microbial features of the one or more non-human sequencing reads to classify the subject’s disease.

20. The method of claim 19, wherein the one or more microbial features comprise taxonomic assignments and abundances of the non-human sequencing reads.

21. The method of claim 19, wherein the one or more probes comprise multiplexed oligonucleotide probes that couple non-specifically to one or more microbial nucleic acid molecules.

22. The method of claim 19, wherein the one or more probes comprise multiplexed oligonucleotide probes that target mammalian nucleic acid molecules.

23. A system, comprising:

(a) one or more processors; and

(b) a non-transient computer readable storage medium comprising software, wherein the software comprises executable instructions that, as a result of execution, cause the one or more processors of a computer system to:

(i) receive one or more nucleic acid molecule sequencing reads of subject's biological sample, wherein the subject has a disease, and wherein the one or more nucleic acid molecule sequencing reads are obtained from one or more nucleic acid molecules enriched by one or more probes exposed to the subject's biological sample;

(ii) map the one or more nucleic acid molecule sequencing reads to a genome database, thereby identifying one or more non-human sequencing reads of the one or more nucleic acid molecule sequencing reads; and

(iii) identify one or more microbial features of the one or more non-human sequencing reads to classify the subject's disease.