CLAIMS

What is claimed:

Claim 1. A method of determining a disease of a subject wherein the disease comprises cancer, comprising:

(a) providing a tissue biopsy or a liquid biopsy sample of a subject;

(b) enriching one or more nucleic acid molecules of the liquid biopsy sample by affinity targeting of an epigenetic feature common to the one or more nucleic acid molecules, wherein the nucleic acid molecules are cell-free DNA and comprise mammalian nucleic acid molecules and microbial nucleic acid molecules;

(c) sequencing the enriched one or more nucleic acid molecules to generate one or more nucleic acid molecule sequencing reads; and (d) determining detection of cancer of the subject as an output of a machine-learning predictive model when the predictive model is provided input based on the enriched one or more nucleic acid molecules,

wherein the epigenetic feature common to the one or more nucleic acid molecules comprises any one of 5-methylcytosine (5mC), 5- hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxy cytosine (5caC), N4- acetylcytosine (N4AcC), 6-methyladenosine (6mA), or 4-methylcytosine (4mC).

Claim 2. A method of training a predictive model, comprising:

(a) providing a tissue biopsy or a liquid biopsy sample of one or more subjects with a cancer;

(b) enriching the liquid biopsy sample of the one or more subjects by affinity targeting of an epigenetic feature common to one or more nucleic acid molecules of the liquid biopsy the sample, wherein the nucleic acid molecules are cell-free DNA and comprise mammalian nucleic acid molecules and microbial nucleic acid molecules;

(c) sequencing the enriched one or more nucleic acid molecules to generate one or more nucleic acid molecule sequencing reads; and (d) training the predictive model with one or more features of the one or more nucleic acid molecule sequencing reads and the cancer of the one or more subjects,

wherein the epigenetic feature common to the one or more nucleic acid molecules comprises any one of 5-methylcytosine (5mC), 5- hydroxymethylcytosine (5hmC), 5-

formylcytosine (5fC), 5-carboxy cytosine (5caC), N4- acetylcytosine (N4AcC), 6-methyladenosine (6mA), or 4-methylcytosine (4mC).

Claim 3. A computer system to determine a disease of a subject wherein the disease comprises cancer, comprising:

(a) one or more processors; and

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

(i) receive a subject's one or more nucleic acid molecule sequencing reads of one or more nucleic acid molecules of a tissue biopsy or a liquid biopsy sample, wherein the one or more nucleic acid molecules are cell-free DNA and comprise mammalian nucleic acid molecules and microbial nucleic acid molecules, and are enriched by affinity targeting of an epigenetic feature common to the one or more nucleic acid molecules; and

(ii) determine detection of cancer of the subject as an output of a machine-learning predictive model when the predictive model is provided the one or more nucleic acid molecule sequencing reads,

wherein the epigenetic feature common to the one or more nucleic acid molecules comprises any one of 5-methylcytosine (5mC), 5- hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxy cytosine (5caC), N4- acetylcytosine (N4AcC), 6-methyladenosine (6mA), or 4-methylcytosine (4mC).

Claim 4. The method of claim 1 or 2 or the computer system of claim 3, further comprising or wherein the executable instruction further comprises filtering the one or more nucleic acid molecules sequencing reads to identify one or more non-mammalian sequencing reads and one or more mammalian sequencing reads.

Claim 5. The method of claim 1 or 2 or the computer system of claim 3, wherein the epigenetic feature comprises a nucleic acid epigenetic feature.

Claim 6. The method of claim 1 or 2 or the computer system of claim 3, wherein the epigenetic feature comprises a mammalian nucleic acid epigenetic feature or a non-mammalian nucleic acid epigenetic feature.

Claim 7. The method or computer system of claim 6, wherein the non-mammalian nucleic acid epigenetic feature comprises phosphorothioate-linked nucleotides.

Claim 8. The method of claim 1 or 2 or the computer system of claim 3, wherein the subject is human or a non-human mammal.

Claim 9. The method or computer system of claim 5, wherein the affinity targeting of the nucleic acid epigenetic feature comprises concentrating the nucleic acid epigenetic feature.

Claim 10. The method or computer system of claim 5, wherein the nucleic acid epigenetic feature comprises methylated CpG dinucleotides pairs, unmethylated CpG dinucleotide pairs, or a combination thereof; or nucleobases 5-methylcytosine, 5-hydroxymethylcytosine, N4-acetylcytosine, N6-methyladenine, or any combination thereof.

Claim 11. The method of claim 1 or 2 or the computer system of claim 3, wherein the affinity targeting utilizes specific affinity reagents to bind to the epigenetic feature.

Claim 12. The method or computer system of claim 11, wherein the specific affinity reagents comprise streptavidin, NeutrAvidin, polyclonal, monoclonal, recombinant antibodies, aptamers, recombinant epigenetic proteins, or any combination thereof.

Claim 13. The method or computer system of claim 12, wherein the recombinant epigenetic proteins comprise epigenetic readers, writers, erasers, or any combination thereof.

Claim 14. The method or computer system of claim 13, wherein the epigenetic readers comprise recombinant methyl-CpG binding proteins Mecp2, Mbd1-6, SETDB1, SETDB2, TIP5/BAZ2A, Zbtb38, Kaiso, Zbtb4, Np95, Np97 or the recombinant methyl-binding domains derived therefrom; recombinant zinc finger CXXC domain-containing proteins KDM2A, KDM2A, KDM2B, FBXL19, CFP1, DNMT1, MLL1, MLL2, MDB1, TET1, TET3, IDAX,

CXXC5, CGBP, or the recombinant CXXC domains derived therefrom; or the microbial proteins Dam, CcrM, ModA13, SpnD39III, Dcm, JHP1050, M2.Hpy.AII, or the recombinant methyl-binding domains derived therefrom.

Claim 15. The method or computer system of claim 13, wherein the epigenetic writers and erasers are catalytically inactive.

Claim 16. The method or computer system of claim 13, wherein the epigenetic readers, writers, and erasers comprise an epitope tag, wherein the epitope tag comprises a N- or C-terminal 6x-histidine tag, green fluorescent protein (MA), myc, hemagglutinin (HA), Fc fusion, molecular recognition motif, or any combination thereof.

Claim 17. The method or computer system of claim 16, wherein the molecular recognition motif comprises a birA or sortase motif.

Claim 18. The method of claim 1 or 2 or the computer system of claim 3, wherein

(a) the predictive model is trained with

(i) one or more mammalian features, one or more non-mammalian features, or a combination of one or more mammalian and non-mammalian features determined from one or more subjects' one or more nucleic acid molecules of a biological sample and

(ii) a corresponding cancer of the one or more subjects; or

(b) the one or more features comprise one or more mammalian features, one or more non-mammalian features, or a combination of one or more mammalian and non-mammalian features;

and wherein

(i) the one or more mammalian features comprises mammalian genomic coordinates or annotated genomic loci and a number of sequencing reads thereof; or

(ii) the one or more mammalian features comprises mammalian functional gene and biochemical pathway abundances; or

(iii) the one or more non-mammalian features comprises microbial taxonomic assignments and a number of sequencing reads thereof; or

(iv) the one or more non-mammalian features comprise microbial functional gene and biochemical pathway abundances.

Claim 19. The method or computer system of claim 18, wherein the predictive model removes contaminate non-mammalian features while selectively retaining other non-contaminate non-mammalian features.