CLAIMS

What is claimed:

1. A method of determining a disease of a subject, comprising:

(a) providing a biological sample of a subject;

(b) enriching one or more nucleic acid molecules of the biological sample by affinity targeting of an epigenetic feature common to the one or more 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 the disease of the subject as an output of a predictive model when the predictive model is provided the enriched one or more nucleic acid molecules as an input.

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

(a) providing a biological sample of one or more subjects with a disease;

(b) enriching the biological sample of the one or more subjects by affinity targeting of an epigenetic feature common to one or more nucleic acid molecules of the biological sample;

(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 disease of the one or more subjects.

3. A computer system to determine a disease of a subject, 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 biological sample, wherein the one or

more nucleic acid molecules are enriched by affinity targeting of an epigenetic feature common to the one or more nucleic acid molecules; and

(ii) determine a disease of the subject as an output of a predictive model when the predictive model is provided the one or more nucleic acid molecule sequencing.

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.

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.

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.

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

8. The method of claim 1 or 2 or the computer system of claim 3, wherein the biological sample comprises a tissue, liquid biopsy sample or a combination thereof.

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

10. 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.

11. 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.

12. 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.

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

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

15. The method or computer system of claim 14, 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.

16. The method or computer system of claim 14, wherein the epigenetic writers and erasers are catalytically inactive.

17. The method or computer system of claim 14, 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.

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

19. The method of claim 1 or 2 or the computer system of claim 3, wherein the predictive model is trained with one or more mammalian, one or more non-mammalian, or a combination thereof features determined from one or more subjects' one or more nucleic acid molecules of a biological sample and a corresponding disease of the one or more subjects or the one or more features comprise one or more mammalian features, one or more non-mammalian features, or a combination thereof features; and wherein the one or more mammalian feature comprises mammalian genomic coordinates or annotated genomic loci and a number of sequencing reads associated therewith; or the one or more mammalian feature comprises mammalian functional gene and biochemical pathway abundances; or the one or more non-mammalian feature comprises microbial taxonomic assignments and a number of sequencing reads associated therewith; or the one or more non-mammalian features comprises microbial functional gene and biochemical pathway abundances.

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