CLAIMS

What is claimed:

1. A method of determining the presence or lack thereof cancer of a subject, the method comprising:

    (a) providing one or more sequencing reads of a ~~subject's~~ biological sample obtained from a subject;

    (b) filtering the sequencing reads with a genome database to produce a set of filtered non-human sequencing reads;

    (c) decontaminating the filtered non-human sequencing reads to remove contaminant non-human sequencing reads from the filtered non-human sequencing reads;

    ~~(c)~~(d) translating the decontaminated non-human sequencing reads to non-human proteins;

    ~~(d)~~(e) mapping the non-human proteins to a protein database, thereby producing a set of protein database associations; and

    ~~(e)~~(f) processing the set of protein database associations with a trained model thereby determining the presence ~~or lack thereof~~of cancer of the subject~~.as an output to the trained model when the trained model is provided an input of the set of protein database associations.~~

2. The method of claim 1, wherein the set of protein database associations comprises a set of functional genes, biochemical pathways, or any combination thereof.

~~3. The method of claim 1, further comprising decontaminating the filtered non-human sequencing reads prior to (c) to remove contaminant non-human sequencing reads.~~

~~4.~~3. The method of claim 1, wherein translating is completed *in silico*.

~~5.~~4. The method of claim 1, wherein the biological sample is a tissue, liquid biopsy, or any combination thereof.

~~6.~~5. The method of claim 1, wherein the subject is human or a non-human mammal.

~~7.~~6. The method of claim 1, wherein the biological sample comprises a nucleic acid composition, wherein the nucleic acid composition comprises DNA, RNA, cell-free DNA, cell-free RNA, exosomal DNA, exosomal RNA, or any combination thereof.

~~8.~~7. The method of claim 1, wherein the genome database is a human genome database.

~~9.~~8. The method of claim 1, wherein the trained model is trained with a set of functional gene and biochemical pathway abundances that are present or absent with a characteristic abundance for a cancer of interest.

~~10.~~9. The method of claim 1, wherein the non-human sequences originate from bacterial, archaeal, fungal, viral, or any combination thereof origins of life.

~~11.~~10. The method of claim 1, wherein the trained model is configured to determine a category or tissue-specific location of the cancer of the subject.

~~12.~~11. The method of claim 1, wherein the trained model is configured to determine one or more types of cancer of the subject.

~~13.~~12. The method of claim ~~12~~1, wherein the trained model is configured to determine one or more subtypes of the cancer of the subject.

~~14.~~13. The method of claim 1, wherein the trained model is configured to determine a stage of cancer of the subject, cancer prognosis of the subject, or any combination thereof.

~~15.~~14. The method of claim 1, wherein the trained model is configured to determine the presence or lack thereof cancer at a low-stage (stage I or stage II) tumor.

~~16.~~15. The method of claim 1, wherein the trained model is configured to determine an immunotherapy response of the subject when the subject is provided the immunotherapy.

~~17.~~16. The method of claim 1, further comprising ~~outputting with the~~providing as an output of the trained model a therapy for the subject to treat the subject's cancer, wherein the subject will respond with positive therapeutic efficacy when administered the therapeutic.

~~18.~~17. The method of claim 1, wherein the cancer of the subject comprises: acute myeloid leukemia, adrenocortical carcinoma, bladder urothelial carcinoma, brain lower grade glioma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, lymphoid neoplasm diffuse large B-cell lymphoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma, prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, stomach adenocarcinoma, testicular germ cell tumors, thymoma, thyroid carcinoma, uterine carcinosarcoma, uterine corpus endometrial carcinoma, uveal melanoma, or any combination thereof.

~~19.~~18. The method of claim ~~5~~4, wherein the liquid biopsy comprises: plasma, serum, whole

blood, urine, cerebral spinal fluid, saliva, sweat, tears, exhaled breath condensate, or any combination thereof.

~~20.~~19. The method of claim 1, wherein filtering comprises computationally filtering of the sequencing reads by bowtie2, Kraken, or any combination thereof programs.

20. The method of claim 1, wherein the trained model comprises one or more machine learning algorithms.

~~21. The method of claim 1, wherein the protein database is the UniRef database.~~

~~22. The method of claim 1, wherein translating is accomplished by BLASTP, USEARCH, LAST, MMSeqs2, DIAMOND, or any combination thereof software packages.~~

~~23. The method of claim 2, wherein the mapping of the non-human proteins to the biochemical pathways is accomplished by mapping non-human proteins to KEGG, MetaCyc, PANTHER Pathway, PathBank or any combination thereof databases.~~

~~24. The method of claim 2, wherein the biochemical pathways are generated with the software package MinPath.~~

~~25. A method of providing a determination of the presence or lack thereof cancer of a subject, the method comprising:~~

~~(a) sequencing a nucleic acid compositions of a subject's biological sample thereby generating sequencing reads;~~

~~(b) filtering the sequencing reads with a genome database to produce a set of filtered non-human sequencing reads;~~

~~(c) translating the non-human sequencing reads to non-human proteins;~~

~~(d) mapping the non-human proteins to a protein database, thereby producing a set of protein database associations; and~~

~~(e) providing a determination of the presence or lack thereof cancer of the subject as an output of a trained model when the trained model is provided an input of the set protein database associations.~~

~~26. The method of claim 25, wherein the set of protein database associations comprises a set of functional genes, biochemical pathways, or any combination thereof.~~

~~27. The method of claim 25, further comprising decontaminating the filtered non-human sequencing reads prior to (c) to remove contaminant non-human sequencing reads.~~

~~28. The method of claim 25, wherein translating is completed *in silico*.~~

~~29. The method of claim 25, wherein the biological sample is a tissue, liquid biopsy sample, or any combination thereof.~~

30. The method of claim 25, wherein the subject is human or a non-human mammal.
31. The method of claim 25, wherein the biological sample comprises a nucleic acid composition, wherein the nucleic acid composition comprises DNA, RNA, cell-free DNA, cell-free RNA, exosomal DNA, exosomal RNA, or any combination thereof.
32. The method of claim 25, wherein the genome database is a human genome database.
33. The method of claim 25, wherein the trained model is trained with a set of functional gene and biochemical pathway abundances that are present or absent with a characteristic abundance for a cancer of interest.
34. The method of claim 25, wherein the non-human sequences originate from bacterial, archaeal, fungal, viral, or any combination thereof origins of life.
35. The method of claim 25, wherein the trained model is configured to determine a category or tissue-specific location of the cancer of the subject.
36. The method of claim 25, wherein the trained model is configured to determine one or more types of the cancer of the subject.
37. The method of claim 36, wherein the trained model is configured to determine one or more subtypes of the cancer of the subject.
38. The method of claim 25, wherein the trained model is configured to determine a stage of a cancer of the subject, cancer prognosis of the subject, or any combination thereof.
39. The method of claim 25, wherein the trained model is configured to determine the presence or lack thereof a cancer at a low-stage (stage I or stage II) tumor.
40. The method of claim 25, wherein the trained model is configured to determine an immunotherapy response of the subject when the subject is provided an immunotherapy.
41. The method of claim 25, further comprising outputting with the trained model a therapy for the subject to treat the subject's cancer, wherein the subject will respond with positive therapeutic efficacy when administered the therapy.
42. The method of claim 25, wherein the cancer of the subject comprises: acute myeloid leukemia, adrenocortical carcinoma, bladder urothelial carcinoma, brain lower grade glioma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, lymphoid neoplasm diffuse large B-cell lymphoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma,

prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, stomach adenocarcinoma, testicular germ cell tumors, thymoma, thyroid carcinoma, uterine carcinosarcoma, uterine corpus endometrial carcinoma, uveal melanoma, or any combination thereof.

43. The method of claim 29, wherein the liquid biopsy comprises: plasma, serum, whole blood, urine, cerebral spinal fluid, saliva, sweat, tears, exhaled breath condensate, or any combination thereof.

44. The method of claim 25, wherein filtering comprises computationally filtering of the sequencing reads by bowtie2, Kraken, or any combination thereof programs.

45. The method of claim 25, wherein the protein database is the UniRef database.

46. The method of claim 25, wherein translating is accomplished by BLASTP, USEARCH, LAST, MMSeqs2, DIAMOND, or any combination thereof software packages.

47. The method of claim 26, wherein the mapping of the non-human proteins to the biochemical pathways is accomplished by mapping non-human proteins to KEGG, MetaCyc, PANTHER Pathway, PathBank or any combination thereof databases.

48. The method of claim 26, wherein the biochemical pathways are generated with the software package MinPath.

49. A method of training a model configured to determine the presence or lack thereof cancer of a subject, the method comprising:

(a) providing a dataset comprising nucleic acid sequencing reads of a first set of one or more subjects' nucleic acid compositions and a corresponding one or more cancers of the first set of one or more subjects;

(b) filtering the nucleic acid sequencing reads with a build of a genome database to generate non-human sequencing reads;

(c) translating the non-human sequencing reads to non-human proteins;

(d) mapping the non-human proteins to a protein database, thereby producing a set of protein database associations; and

(e) training a model with the set of protein database associations and the corresponding one or more cancer states of the first set of one or more subjects, thereby generating a trained model configured to determine the presence or lack thereof cancer of a second set of one or more subjects.

50. The method of claim 49, wherein the set of protein database associations comprises a set of functional genes, biochemical pathways, or any combination thereof.

51. The method of claim 49, further comprising decontaminating the filtered non-human sequencing reads prior to (c) to remove contaminant non-human sequencing reads.
52. The method of claim 49, wherein translating is completed *in silico*.
53. The method of claim 49, wherein the biological sample is a tissue, liquid biopsy sample or any combination thereof.
54. The method of claim 49, wherein the first set, second set, or any combination thereof one or more subjects are human or a non-human mammal.
55. The method of claim 49, wherein the biological sample comprises a nucleic acid composition, wherein the nucleic acid composition comprises DNA, RNA, cell-free DNA, cell-free RNA, exosomal DNA, exosomal RNA, or any combination thereof.
56. The method of claim 49, wherein the genome database is a human genome database.
57. The method of claim 49, wherein the trained model is trained with a set of functional gene and biochemical pathway abundances that are present or absent with a characteristic abundance for a cancer of interest.
58. The method of claim 49, wherein the non-human sequences originate from bacterial, archaeal, fungal, viral, or any combination thereof origins of life.
59. The method of claim 49, wherein the trained model is configured to determine a category or tissue-specific location of the second set of one or more subjects' cancer.
60. The method of claim 49, wherein the trained model is configured to determine one or more types of the second set of one or more subjects' cancer.
61. The method of claim 60, wherein the trained model is configured to determine one or more subtypes of the second set of one or more subjects' cancer.
62. The method of claim 49, wherein the trained model is configured to determine a stage of the second set of one or more subjects' cancer, cancer prognosis, or any combination thereof.
63. The method of claim 49, wherein the trained is configured to determine the presence or lack thereof the second set of one or more subjects' cancer at a low-stage (stage I or stage II) tumor.
64. The method of claim 49, wherein the trained model is configured to determine an immunotherapy response of the subject when the subject is provided an immunotherapy.
65. The method of claim 49, further comprising outputting with the trained model a therapy to treat the second set of one or more subjects' cancer, wherein the second set of one or more subjects will respond with positive therapeutic efficacy when administered the therapy.
66. The method of claim 49, wherein the first and second set of one or more subjects' cancer

~~comprises: acute myeloid leukemia, adrenocortical carcinoma, bladder urothelial carcinoma, brain lower grade glioma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, lymphoid neoplasm diffuse large B-cell lymphoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma, prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, stomach adenocarcinoma, testicular germ cell tumors, thymoma, thyroid carcinoma, uterine carcinosarcoma, uterine corpus endometrial carcinoma, uveal melanoma, or any combination thereof.~~

~~67. The method of claim 53, wherein the liquid biopsy comprises: plasma, serum, whole blood, urine, cerebral spinal fluid, saliva, sweat, tears, exhaled breath condensate, or any combination thereof.~~

~~68. The method of claim 49, wherein filtering comprises computationally filtering of the sequencing reads by bowtie2, Kraken, or any combination thereof programs.~~

~~69. The method of claim 49, wherein the protein database is the UniRef database.~~

~~70. The method of claim 49, wherein translating is accomplished by BLASTP, USEARCH, LAST, MMSeqs2, DIAMOND, or any combination thereof software packages.~~

~~71. The method of claim 50, wherein the mapping of the non-human proteins to the biochemical pathways is accomplished by mapping non-human proteins to KEGG, MetaCyc, PANTHER Pathway, PathBank or any combination thereof databases.~~

~~72. The method of claim 50, wherein the biochemical pathways are generated with the software package MinPath.~~

~~73. The method of claim 51, wherein the dataset further comprises a corresponding previous or current treatment administered to the first set of one or more subjects.~~

~~74. The method of claim 73, wherein the dataset further comprises a treatment efficacy of the first set of one or more subjects' previous or current treatment administration.~~

~~75. A computer-implemented method for utilizing a trained predictive model to provide a therapeutic treatment prediction for one or more subjects, the method comprising:~~

~~(a) receiving a first set of one or more subjects' nucleic acid sequencing reads of a biological sample and corresponding cancer classification;~~

(b) filtering the nucleic acid sequencing reads with a build of a genome database to generate non-human sequencing reads;

(c) translating the non-human sequencing reads to non-human proteins;

(d) mapping the non-human proteins to a protein database, thereby producing a set of protein database associations; and

(e) utilizing a trained predictive model to provide a treatment prediction for the first set of one or more subjects when the set of protein database associations are provided as an input to the trained predictive model.

76. The method of claim 75, wherein the trained predictive model is trained on a second set of one or more subjects' nucleic acid sequencing reads of a biological sample, corresponding cancer classification, corresponding treatment administered, corresponding treatment response, or any combination thereof.

77. The method of claim 76, wherein the second set of one or more subjects are different than the first set of one or more subjects.

78. The method of claim 75, wherein the set of protein database associations comprises a set of functional genes, biochemical pathways, or any combination thereof.

79. The method of claim 75, further comprising decontaminating the filtered non-human sequencing reads prior to (c) to remove contaminant non-human sequencing reads.

80. The method of claim 75, wherein translating is completed *in silico*.

81. The method of claim 75, wherein the biological sample is a tissue, liquid biopsy sample or any combination thereof.

82. The method of claim 75, wherein the first set of one or more subjects are human or a non-human mammal.

83. The method of claim 75, wherein the biological sample nucleic acid composition comprises DNA, RNA, cell-free DNA, cell-free RNA, exosomal DNA, exosomal RNA, or any combination thereof.

84. The method of claim 75, wherein the genome database is a human genome database.

85. The method of claim 75, wherein the non-human sequences originate from bacterial, archaeal, fungal, viral, or any combination thereof origins of life.

86. The method of claim 75, wherein the treatment prediction comprises an immunotherapy response of the first set of one or more subjects when the first set of one or more subjects are administered an immunotherapy.

87. The method of claim 75, wherein the treatment prediction comprises a therapeutic efficacy that the first set of one or more subjects will respond with positive efficacy.

88. The method of claim 75, wherein the cancer classification comprises: acute myeloid leukemia, adrenocortical carcinoma, bladder urothelial carcinoma, brain lower grade glioma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, lymphoid neoplasm diffuse large B-cell lymphoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma, prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, stomach adenocarcinoma, testicular germ cell tumors, thymoma, thyroid carcinoma, uterine carcinosarcoma, uterine corpus endometrial carcinoma, uveal melanoma, or any combination thereof.

89. The method of claim 79, wherein the liquid biopsy comprises: plasma, serum, whole blood, urine, cerebral spinal fluid, saliva, sweat, tears, exhaled breath condensate, or any combination thereof.

90. The method of claim 75, wherein filtering comprises computationally filtering of the sequencing reads by bowtie2, Kraken, or any combination thereof programs.

91. The method of claim 75, wherein the protein database is the UniRef database.

92. The method of claim 75, wherein translating is accomplished by BLASTP, USEARCH, LAST, MMSeqs2, DIAMOND, or any combination thereof software packages.

93. The method of claim 76, wherein the mapping of the non-human proteins to the biochemical pathways is accomplished by mapping non-human proteins to KEGG, MetaCyc, PANTHER Pathway, PathBank or any combination thereof databases.

94. The method of claim 76, wherein the biochemical pathways are generated with the software package MinPath.

95. A method of changing a subject's cancer treatment with a trained predictive model, the method comprising:

(a) providing one or more sequencing reads of a subject's biological sample with cancer, cancer type, and treatment administered to treat the cancer;

(b) filtering the sequencing reads with a genome database to produce a set of filtered non-human sequencing reads;

(c) translating the non-human sequencing reads to non-human proteins;

(d) mapping the non-human proteins to a protein database, thereby producing a set of protein database associations; and

(e) changing the subject's cancer treatment when the treatment administered differs from a treatment recommendation outputted by a trained predictive model when inputted with the set of protein database associations.

96. The method of claim 95, wherein the trained predictive model is trained on a second set of one or more subjects' nucleic acid sequencing reads of a biological sample, corresponding cancer classification, corresponding treatment administered, corresponding treatment response, or any combination thereof.

97. The method of claim 96, wherein the second set of one or more subjects are different than the first set of one or more subjects.

98. The method of claim 95, wherein the set of protein database associations comprises a set of functional genes, biochemical pathways, or any combination thereof.

99. The method of claim 95, further comprising decontaminating the filtered non-human sequencing reads prior to (c) to remove contaminant non-human sequencing reads.

100. The method of claim 95, wherein translating is completed *in silico*.

101. The method of claim 95, wherein the biological sample is a tissue, liquid biopsy sample or any combination thereof.

102. The method of claim 95, wherein the subject is human or a non-human mammal.

103. The method of claim 95, wherein the biological sample nucleic acid composition comprises DNA, RNA, cell-free DNA, cell-free RNA, exosomal DNA, exosomal RNA, or any combination thereof.

104. The method of claim 95, wherein the genome database is a human genome database.

105. The method of claim 95, wherein the non-human sequences originate from bacterial, archaeal, fungal, viral, or any combination thereof origins of life.

106. The method of claim 95, wherein the treatment recommendation comprises an immunotherapy response of the subject when the subject is administered an immunotherapy.

107. The method of claim 95, wherein the treatment recommendation comprises a therapeutic that the subject will respond with positive efficacy.

108. The method of claim 95, wherein the subject's cancer comprises: acute myeloid leukemia, adrenocortical carcinoma, bladder urothelial carcinoma, brain lower grade glioma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma,

~~glioblastoma multiforme, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, lymphoid neoplasm diffuse large B cell lymphoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma, prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, stomach adenocarcinoma, testicular germ cell tumors, thymoma, thyroid carcinoma, uterine carcinosarcoma, uterine corpus endometrial carcinoma, uveal melanoma, or any combination thereof.~~

~~109. The method of claim 101, wherein the liquid biopsy comprises: plasma, serum, whole blood, urine, cerebral spinal fluid, saliva, sweat, tears, exhaled breath condensate, or any combination thereof.~~

~~110. The method of claim 95, wherein filtering comprises computationally filtering of the sequencing reads by bowtie2, Kraken, or any combination thereof programs.~~

~~111. The method of claim 95, wherein the protein database is the UniRef database.~~

~~112. The method of claim 95, wherein translating is accomplished by BLASTP, USEARCH, LAST, MMSeqs2, DIAMOND, or any combination thereof software packages.~~

~~113. The method of claim 96, wherein the mapping of the non-human proteins to the biochemical pathways is accomplished by mapping non-human proteins to KEGG, MetaCyc, PANTHER Pathway, PathBank or any combination thereof databases.~~

~~114. The method of claim 96, wherein the biochemical pathways are generated with the software package MinPath.~~