CLAIMS

WHAT IS CLAIMED IS:

1. Use of a therapy in the manufacture of a medicament for determining a disease gene expression signature,

wherein a test subject is to be administered a therapeutically effective amount of (i) the therapy, based at least in part on a trained machine learning classifier analyzing a disease gene expression signature to predict responsiveness of the test subject to the therapy, or (ii) a second therapy different from the therapy, based at least in part on the trained machine learning classifier analyzing the disease gene expression signature to predict non-responsiveness of the test subject to the therapy,

wherein the disease gene expression signature is determined at least in part by:

receiving gene expression data from a cohort of subjects suffering from the disease, disorder, or condition;

stratifying the cohort of subjects into two or more groups based at least in part on the gene expression data;

calculating differences in gene expression between the two or more groups of subjects and a group of non-diseased subjects;

selecting one or more genes having significant differences in gene expression between the two or more groups of subjects and the group of non-diseased subjects ("disease candidate genes");

compiling a set of disease genes comprising the disease candidate genes; and

selecting at least a subset of the set of disease genes to thereby determine the disease gene expression signature.

2. The use of claim 1, wherein the disease gene expression signature is determined at least in part by further mapping the disease candidate genes onto a biological network, and selecting adjacent genes on the biological network having significant connection to each other or to the disease candidate genes, wherein the set of disease genes comprises the disease candidate genes and the adjacent genes.

3. The use of claim 2, wherein the biological network comprises a human interactome.

4. The use of claim 2 or 3, wherein the adjacent genes form a significant sub-network with each other or to the disease candidate genes.

5. The use of claim 2, wherein the adjacent genes are identified via a machine-learning algorithm.

6. The use of claim 5, wherein the machine-learning algorithm comprises a random walk.

7. The use of any one of claims 1-6, wherein the disease, disorder, or condition comprises ulcerative colitis, Crohn's disease, rheumatoid arthritis, juvenile arthritis, psoriatic arthritis, plaque psoriasis, ankylosing spondylitis, Guillain-Barre syndrome, Sjogren's syndrome, scleroderma, vitiligo, bipolar disorder, Graves' disease, schizophrenia, Alzheimer's disease, multiple sclerosis, Parkinson's disease, or a combination thereof.

8. The use of claim 7, wherein the disease, disorder, or condition comprises ulcerative colitis.

9. The use of claim 7, wherein the disease, disorder, or condition comprises rheumatoid arthritis.

10. The use of claim 7, wherein the disease, disorder, or condition comprises Alzheimer's disease.

11. The use of claim 7, wherein the disease, disorder, or condition comprises multiple sclerosis.

12. The use of any one of claims 1-11, wherein stratifying the cohort of subjects into two or more groups is random or based at least in part on whether the prior subjects do or do not respond to the therapy.

13. The use of any one of claims 1-12, wherein the therapy comprises a member selected from Table 1.

14. The use of any one of claims 1-12, wherein the therapy comprises an anti-TNF therapy.

15. The use of any one of claims 1-14, wherein the stratifying further comprises grouping subjects from the same cohort having similar gene expression.

16. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with an accuracy of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

17. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a sensitivity of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

18. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a specificity of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

19. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a positive predictive value of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

20. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a negative predictive value of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

21. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a true positive rate of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

22. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a true negative rate of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

23. The use of claim 1, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with an area under curve (AUC) of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

24. A method of determining a disease gene expression signature for quantifying responsiveness to a therapy for subjects suffering from a disease, disorder, or condition, the method comprising:

receiving gene expression data from a cohort of subjects suffering from the same disease, disorder, or condition;

stratifying the cohort of subjects into two or more groups based at least in part on the gene expression data;

calculating differences in gene expression between the two or more groups of subjects and a group of non-diseased subjects;

selecting one or more genes having significant differences in gene expression between the two or more groups of subjects and the group of non-diseased subjects ("disease candidate genes");

compiling a set of disease genes comprising the disease candidate genes; and

selecting at least a subset of the set of disease genes to thereby determine the disease gene expression signature.

25. The method of claim 24, further comprising mapping the disease candidate genes onto a biological network, and selecting adjacent genes on the biological network having significant connection to each other or to the disease candidate genes, wherein the set of disease genes comprises the disease candidate genes and the adjacent genes.

26. The method of claim 25, wherein the biological network comprises a human interactome.

27. The method of claim 25 or 26, wherein the adjacent genes form a significant sub-network with each other or to the disease candidate genes.

28. The method of claim 25, wherein the adjacent genes are identified via a machine-learning algorithm.

29. The method of claim 28, wherein the machine-learning algorithm comprises a random walk.

30. The method of any one of claims 24-29, wherein the disease, disorder, or condition comprises ulcerative colitis, Crohn's disease, rheumatoid arthritis, juvenile arthritis, psoriatic

arthritis, plaque psoriasis, ankylosing spondylitis, Guillain-Barre syndrome, Sjogren's syndrome, scleroderma, vitiligo, bipolar disorder, Graves' disease, schizophrenia, Alzheimer's disease, multiple sclerosis, Parkinson's disease, or a combination thereof.

31. The method of claim 30, wherein the disease, disorder, or condition comprises ulcerative colitis.

32. The method of claim 30, wherein the disease, disorder, or condition comprises rheumatoid arthritis.

33. The method of claim 30, wherein the disease, disorder, or condition comprises Alzheimer's disease.

34. The method of claim 30, wherein the disease, disorder, or condition comprises multiple sclerosis.

35. The method of any one of claims 24-34, wherein stratifying the cohort of subjects into two or more groups is random or based at least in part on whether the prior subjects do or do not respond to the therapy.

36. The method of any one of claims 24-35, wherein the therapy comprises a member selected from Table 1.

37. The method of any one of claims 24-35, wherein the therapy comprises an anti-TNF therapy.

38. The method of any one of claims 24-37, wherein the cohort of subjects is suffering from the same disease, disorder, or condition as the subjects being assessed for therapy responsiveness.

39. The method of any one of claims 24-38, wherein the stratifying further comprises grouping subjects from the same cohort having similar gene expression.

40. The method of any one of claims 24-39, further comprising using the disease gene expression signature to train a machine learning classifier, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of a test subject suffering from the disease, disorder, or condition to the therapy, based at least in part on analyzing gene expression data of the test subject.

41. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with an accuracy of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

42. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a sensitivity of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

43. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a specificity of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

44. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a positive predictive value of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

45. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a negative predictive value of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

46. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a true positive rate of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

47. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with a true negative rate of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

48. The method of claim 40, wherein the trained machine learning classifier is configured to predict responsiveness or non-responsiveness of the test subject with an area under curve (AUC) of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

49. The method of claim 40, wherein the test subject is to be administered a therapeutically effective amount of the therapy, when the trained machine learning classifier predicts responsiveness of the test subject to the therapy.

50. The method of claim 40, wherein the test subject is to be administered a therapeutically effective amount of a second therapy that is different from the therapy, when the trained machine learning classifier predicts non-responsiveness of the test subject to the therapy.

51. A method of validating response to a therapy for a subject suffering from a disease, disorder, or condition, the method comprising:

analyzing changes in a disease gene expression signature in the subject after administration of the therapy, wherein the disease gene expression signature is determined to quantify responsiveness to the therapy.

52. The method of claim 51, wherein the disease gene expression signature is determined at least in part by:

receiving gene expression data from a cohort of subjects suffering from the disease, disorder, or condition;

stratifying the cohort of subjects into two or more groups based at least in part on the gene expression data;

calculating differences in gene expression between the two or more groups of subjects and a group of non-diseased subjects;

selecting one or more genes having significant differences in gene expression between the two or more groups of subjects and the group of non-diseased subjects ("disease candidate genes");

compiling a set of disease genes comprising the disease candidate genes; and

selecting at least a subset of the set of disease genes to thereby determine the disease gene expression signature.

53. A method of monitoring therapeutic efficacy in a subject suffering from a disease, disorder, or condition, the method comprising:

monitoring changes in a disease gene expression signature after administration of a therapy, wherein the disease gene expression signature has been determined at least in part by:

analyzing gene expression data from a cohort of subjects suffering from the same disease, disorder, or condition as the subject;

stratifying the cohort of subjects into two or more groups based on the gene expression data;

determining differences in gene expression between the two or more groups of subjects and a group of non-diseased subjects;

selecting one or more genes having significant differences in gene expression between the two or more groups of subjects and the group of non-diseased subjects ("disease candidate genes");

compiling a set of disease genes comprising the disease candidate genes; and

selecting at least a subset of the set of disease genes to thereby determine the disease gene expression signature.

54. The method of claim 53, wherein the disease gene expression signature is determined at least in part by further mapping the disease candidate genes onto a biological network, and selecting adjacent genes on the biological network having significant connection to each other or to the disease candidate genes, wherein the set of disease genes comprises the disease candidate genes and the adjacent genes.

55. The method of claim 54, wherein the biological network comprises a human interactome.

56. The method of claim 53 or 54, wherein the adjacent genes form a significant sub-network with each other or to the disease candidate genes.

57. The method of claim 53, wherein the adjacent genes are selected by a machine-learning process.

58. The method of claim 57, wherein the machine-learning process comprises a random walk.

59. The method of any one of claims 53-58, wherein the disease, disorder, or condition comprises ulcerative colitis, Crohn's disease, rheumatoid arthritis, juvenile arthritis, psoriatic arthritis, plaque psoriasis, ankylosing spondylitis, Guillain-Barre syndrome, Sjogren's syndrome, scleroderma, vitiligo, bipolar disorder, Graves' disease, schizophrenia, Alzheimer's disease, multiple sclerosis, Parkinson's disease, or a combination thereof.

60. The method of any one of claims 53-59, wherein stratifying the cohort of subjects into two or more groups is random or based at least in part on whether the prior subjects do or do not respond to the therapy.

61. The method of any one of claims 53-60, wherein the therapy comprises a member selected from Table 1.

62. The method of any one of claims 53-61, wherein the therapy comprises an anti-TNF therapy.

63. The method of any one of claims 53-62, wherein the stratifying further comprises grouping subjects from the same cohort having similar gene expression.

64. The method of any one of claims 53-63, further comprising selecting a test subject for a clinical trial, based at least in part on whether the disease gene expression signature of the test subject exhibits a quantifiable change toward a disease gene expression signature of a non-diseased subject.

65. A method of identifying and selecting subjects for a clinical trial, the method comprising:

receiving gene expression data of a cohort of subjects;

analyzing the gene expression data to detect the presence of a disease gene expression signature;

identifying changes in the disease gene expression signature relative to gene expression of a non-diseased subject in the cohort of subjects after the cohort of subjects have been administered at least one dose of a therapy; and

selecting subjects for the clinical trial who exhibit a quantifiable change in the disease gene expression signature towards gene expression of a healthy subject, wherein the disease gene expression signature is determined by the method of any one of claims 1-64.

66. A system comprising:

a processor of a computing device; and

a memory having instructions stored thereon, wherein the instructions, when executed by the processor cause the processor to perform the method of any one of claims 1-65.