disclosure are shown and described. The present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and descriptions are to be regarded as illustrative in nature and not as restrictive.

[0005] A novel network-based system biology framework, a Predictive Response Biomarkers using Network medicine (PRoBeNet) framework, is described herein to predict treatment response biomarkers in complex I&I diseases without relying on, e.g., extensive patient datasets. The PRoBeNet framework herein can successfully determine and validate response-predicting biomarkers retrospectively for, e.g., infliximab response and prospectively for, e.g., extracellular signal-regulated kinase (ERK1/2) inhibitors. Nodes in a human interactome (HI) can be ranked by the PRoBeNet framework using a dual PageRank score, which considers proteins important to the therapy's mechanism of action and disease pathogenesis. The PRoBeNet framework's ability to determine response biomarkers can be a useful and valuable tool for, e.g., the development of companion diagnostic tests, amplifying drug efficacies, and optimizing personalized treatment strategies.

[0006] In an example, disclosed herein is a method of treating a subject suffering from a disease or disorder, the method comprising: administering to the subject a therapy, wherein the therapy has been predicted to generate a treatment response in the subject based at least in part on a trained machine learning (ML) classifier that classifies the subject as responsive or non-responsive to the therapy, wherein the trained ML classifier predicts the treatment response based at least in part on analyzing in the subject a set of cytokines or a set of transcripts. In some examples, the method further comprising: identifying one or more source nodes associated with a set of proteins targeted by the therapy; identifying one or more response nodes associated with the set of cytokines or the set of transcripts indicative of the treatment response; determining, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response; generating a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and processing a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy. In some examples, the ranked set of intermediate nodes comprises a ranked set of proteins, a ranked set of cytokines, or a ranked set of transcripts. In some examples, the ranked set of proteins, the ranked set of cytokines, or the ranked set of transcripts is used as features to obtain the trained ML classifier. In some examples, the set of biomarkers is predictive of the treatment response to the therapy. In some examples, (i) the ranked set

of cytokines is different from at least one cytokine of the set of cytokines or (ii) the ranked set of cytokines is the same as the set of cytokines. In some examples, (i) the ranked set of transcripts is different from at least one transcript of the set of transcripts or (ii) the ranked set of transcripts is the same as the set of transcripts. In some examples, the ranked set of cytokines or the ranked set of transcripts is analyzed to determine statistically significant changes after administering to the subject the therapy. In some examples, generating the ranked set of the one or more intermediate nodes comprises: determining a first rank between the one or more source nodes and the one or more intermediate nodes; determining a second rank between the one or more response nodes and the one or more intermediate nodes; and determining a rank product of the first rank and the second rank. In some examples, the method further comprising: constructing a subset of a protein-protein network associated with the disease or disorder, wherein the constructing comprises: (i) determining the one or more source nodes in the protein-protein network, based at least in part on the set of proteins targeted by the therapy; (ii) determining the one or more response nodes in the protein-protein network, based on least in part on the set of cytokines or the set of transcripts indicative of the treatment response; and (iii) determining one or more connections between the one or more source nodes and the one or more response nodes, based at least in part on one or more interactions between the one or more source nodes and the one or more response nodes. In some examples, the protein-protein network comprises a human interactome (HI). In some examples, the disease or disorder comprises an autoimmune disease, an immunology disease, or a cancer. In some examples, the immunology disease or the autoimmune disease comprises rheumatoid arthritis (RA), ulcerative colitis (UC), Crohn's disease (CD), inflammatory bowel disease, psoriatic arthritis, ankylosing spondylitis, chronic psoriasis, hidradenitis suppurativa, multiple sclerosis, juvenile idiopathic arthritis, uveitis, systemic lupus, Type 1 diabetes, Graves' Disease, Hashimoto's Thyroiditis, celiac disease, Sjögren's Syndrome, or an immune-mediated disease. In some examples, the immunology disease or the autoimmune disease comprises RA. In some examples, the immunology disease or the autoimmune disease comprises UC. In some examples, the immunology disease or the autoimmune disease comprises CD. In some examples, the therapy for the immunology disease or the autoimmune disease comprises an anti-TNF therapy. In some examples, the anti-TNF therapy comprises infliximab, adalimumab, etanercept, certolizumab pegol, or golimumab, or a biosimilar thereof. In some examples, the therapy for the immunology disease or the autoimmune disease comprises an alternative to anti-TNF therapy. In some examples, the alternative to anti-TNF therapy comprises anti-CD20, JAK, anti-IL6, rituximab, sarilumab, tofacitinib citrate, leflunomide, vedolizumab, tocilizumab,

anakinra, abatacept, TL1A, co-stimulation blockade, or a biosimilar thereof. In some examples, the cancer comprises bladder cancer, breast cancer, colon cancer, rectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, melanoma, Non-Hodgkin's lymphoma, pancreatic cancer, prostate cancer, or thyroid cancer. In some examples, the therapy for the cancer comprises a targeted therapy, chemotherapy, hormone therapy, immunotherapy, photodynamic therapy, radiation therapy, stem cell transplant therapy, or any combination thereof. In some examples, the trained ML classifier predicts the treatment response to the therapy using a non-linear relationship between (i) measurements of the ranked set of intermediate nodes and (ii) the treatment response indicated by statistically significant changes in the ranked set of intermediate nodes after administering to the subject the therapy. In some examples, the trained ML classifier is trained using measurements of the one or more intermediate nodes in (i) a first set of subjects with the disease or disorder determined to be responsive to the therapy and (ii) a second set of subjects with the disease or disorder determined to be non-responsive to the therapy. In some examples, the trained ML classifier comprises a neural network, deep learning, perceptron, random forest, Bayes, Markov, Gaussian process, clustering algorithm, support vector machine, generative model, or kernel. In some examples, the set of cytokines or the ranked set of cytokines indicative of the treatment response to an immunology disease or an autoimmune disease comprises: CCL2, CSF2, CXCL8, IFNγ, IL10, IL12A, IL-12A/B, IL12B, IL13, IL15, IL16, IL17A, IL18, IL1B, IL2, IL21, IL22, IL23A, IL27, IL4, IL5, IL6, IL7, IL8, MIF, TGFB1, TGFB2, TGFB3, or TNF. In some examples, the set of cytokines or the ranked set of cytokines indicative of the treatment response to ulcerative colitis (UC) comprises: CCL2, CSF2, CXCL8, IFNγ, IL10, IL12A, IL13, IL17A, IL1B, IL21, IL22, IL23A, IL27, IL4, IL5, IL6, IL7, TGFB1, or TNF. In some examples, the set of cytokines or the ranked set of cytokines indicative of the treatment response to rheumatoid arthritis (RA) comprises: CSF2, CXCL8, IFNγ, IL10, IL12A, IL12B, IL15, IL16, IL17A, IL18, IL1B, IL4, IL6, MIF, TGFB1, TGFB2, TGFB3, TNF. In some examples, the set of cytokines or the ranked set of cytokines indicative of the treatment response to Crohn's disease (CD) comprises: IL10, IFNγ, IL13, IL17A, IL1B, IL2, IL4, IL5, IL6, IL7, IL8, CCL2, TNF, or IL-12A/B. In some examples, the set of cytokines or the ranked set of cytokines indicative of the treatment response to ulcerative colitis (UC) comprises: IL10, IFNγ, IL13, IL17A, IL1B, IL2, IL4, IL5, IL6, IL7, IL8, CCL2, TNF, or IL-12A/B. In some examples, the trained ML classifier predicts the treatment response with an area under receiver operating characteristic curve (AUC) of at least about 0.8. In some examples, the set of cytokines or the ranked set of cytokines is measured using an immunoassay. In some examples, the set of transcripts or the ranked set of transcripts is measured using reverse transcription quantitative real-

time polymerase chain reaction (RT-qPCR), microarrays, or RNA-sequencing (RNA-seq). In some examples, the therapy comprises an experimental therapy in a clinical trial. In some embodiments, the experimental therapy comprises KO-947.

[0007] In another examples, disclosed herein is a method for determining biomarkers predictive of treatment response to a therapy for a subject suffering from a disease or disorder, the method comprising: identifying one or more source nodes associated with a set of proteins targeted by the therapy; determining, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response; generating a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and processing a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy. Thus, in one aspect, there is provided a method for determining biomarkers predictive of treatment response to a therapy for a subject suffering from rheumatoid arthritis (RA), the method comprising: identifying one or more source nodes associated with a set of proteins targeted by the therapy; determining, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response; generating a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and processing a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy. In some embodiments, the ranked set of intermediate nodes comprises a ranked set of proteins, a ranked set of cytokines, or a ranked set of transcripts. In some embodiments, the ranked set of proteins, the ranked set of cytokines, or the ranked set of transcripts is used as features to obtain a trained ML classifier. In some embodiments, the set of biomarkers is predictive of the treatment response to the therapy. In some embodiments, (i) the ranked set of cytokines is different from at least one cytokine of the set of cytokines or (ii) the ranked set of cytokines is the same as the set of cytokines. In some embodiments, (i) the ranked set of transcripts is different from at least one transcript of the set of transcripts or (ii) the ranked set of transcripts is the same as the set of transcripts. In some embodiments, the ranked set of cytokines or the ranked set of transcripts is analyzed to determine statistically significant changes after administering to the subject the therapy. In some embodiments, generating the ranked set of the one or more intermediate nodes comprises: determining a first rank between the one or more source nodes and

the one or more intermediate nodes; determining a second rank between the one or more response nodes and the one or more intermediate nodes; and determining a rank product of the first rank and the second rank. In some embodiments, the method further comprising: constructing a subset of a protein-protein network associated with the disease or disorder, wherein the constructing comprises: (i) determining the one or more source nodes in the protein-protein network, based at least in part on the set of proteins targeted by the therapy; (ii) determining the one or more response nodes in the protein-protein network, based on least in part on the set of cytokines or the set of transcripts indicative of the treatment response; and (iii) determining one or more connections between the one or more source nodes and the one or more response nodes, based at least in part on one or more interactions

between the one or more source nodes and the one or more response nodes. In some embodiments, the protein-protein network comprises a human interactome (HI). In some embodiments, the disease or disorder comprises an autoimmune disease, an immunology disease, or a cancer. In some embodiments, the immunology disease or the autoimmune disease comprises rheumatoid arthritis (RA), ulcerative colitis (UC), Crohn's disease (CD), inflammatory bowel disease, psoriatic arthritis, ankylosing spondylitis, chronic psoriasis, hidradenitis suppurativa, multiple sclerosis, juvenile idiopathic arthritis, uveitis, systemic lupus, Type 1 diabetes, Graves' Disease, Hashimoto's Thyroiditis, celiac disease, Sjögren's Syndrome, or an immune-mediated disease. In some embodiments, the immunology disease or the autoimmune disease comprises RA. In some embodiments, the immunology disease or the autoimmune disease comprises UC. In some embodiments, the immunology disease or the autoimmune disease comprises CD. In some embodiments, the therapy for the immunology disease or the autoimmune disease comprises an anti-TNF therapy. In some embodiments, the anti-TNF therapy comprises infliximab, adalimumab, etanercept, certolizumab pegol, or golimumab, or a biosimilar thereof. In some embodiments, the therapy for the immunology disease or the autoimmune disease comprises an alternative to anti-TNF therapy. In some embodiments, the alternative to anti-TNF therapy comprises anti-CD20, JAK, anti-IL6, rituximab, sarilumab, tofacitinib citrate, leflunomide, vedolizumab, tocilizumab, anakinra, abatacept, TL1A, co-stimulation blockade, or a biosimilar thereof. In some embodiments, the cancer comprises bladder cancer, breast cancer, colon cancer, rectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, melanoma, Non-Hodgkin's lymphoma, pancreatic cancer, prostate cancer, or thyroid cancer. In some embodiments, the therapy for the cancer comprises a targeted therapy, chemotherapy, hormone therapy, immunotherapy, photodynamic therapy, radiation therapy, stem cell transplant therapy, or any combination thereof. In some embodiments, the trained ML classifier predicts the treatment response to the therapy using a non-linear relationship between (i) measurements of the ranked set of intermediate nodes and (ii) the treatment response indicated by statistically significant changes in the ranked set of intermediate nodes after administering to the subject the therapy. In some embodiments, the trained ML classifier is trained using measurements of the one or more intermediate nodes in (i) a first set of subjects with the disease or disorder determined to be responsive to the therapy and (ii) a second set of subjects with the disease or disorder determined to be non-responsive to the therapy. In some embodiments, the trained ML classifier comprises a neural network, deep learning, perceptron, random forest, Bayes, Markov, Gaussian process, clustering algorithm, support vector machine, generative model, or kernel. In some embodiments, the set of cytokines or the ranked set of cytokines indicative of the treatment response to an immunology disease or an autoimmune disease

comprises: CCL2, CSF2, CXCL8, IFNγ, IL10, IL12A, IL-12A/B, IL12B, IL13, IL15, IL16, IL17A, IL18, IL1B, IL2, IL21, IL22, IL23A, IL27, IL4, IL5, IL6, IL7, IL8, MIF, TGFB1, TGFB2, TGFB3, or TNF. In some embodiments, the set of cytokines or the ranked set of cytokines indicative of the treatment response to ulcerative colitis (UC) comprises: CCL2, CSF2, CXCL8, IFNγ, IL10, IL12A, IL13, IL17A, IL1B, IL21, IL22, IL23A, IL27, IL4, IL5, IL6, IL7, TGFB1, or TNF. In some embodiments, the set of cytokines or the ranked set of cytokines indicative of the treatment response to rheumatoid arthritis (RA) comprises: CSF2, CXCL8, IFNγ, IL10, IL12A, IL12B, IL15, IL16, IL17A, IL18, IL1B, IL4, IL6, MIF, TGFB1, TGFB2, TGFB3, TNF. In some embodiments, the set of cytokines or the ranked set of cytokines indicative of the treatment response to Crohn's disease (CD) comprises: IL10, IFNγ, IL13, IL17A, IL1B, IL2, IL4, IL5, IL6, IL7, IL8, CCL2, TNF, or IL-12A/B. In some embodiments, the set of cytokines or the ranked set of cytokines indicative of the treatment response to ulcerative colitis (UC) comprises: IL10, IFNγ, IL13, IL17A, IL1B, IL2, IL4, IL5, IL6, IL7, IL8, CCL2, TNF, or IL-12A/B. In some embodiments, the trained ML classifier predicts the treatment response with an area under receiver operating characteristic curve (AUC) of at least about 0.8. In some embodiments, the set of cytokines or the ranked set of cytokines is measured using an immunoassay. In some embodiments, the set of transcripts or the ranked set of transcripts is measured using reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR), microarrays, or RNA-sequencing (RNA-seq). In some embodiments, the therapy comprises an experimental therapy in a clinical trial.

[0008] In another examples, disclosed herein is a method for administering an experimental therapy to a subject in a clinical trial, comprising: administering to the subject the experimental therapy, wherein the experimental therapy has been predicted to generate a treatment response in the subject based at least in part on a trained machine learning (ML) classifier that classifies the subject as responsive or non-responsive to the experimental therapy, wherein the trained ML classifier predicts the treatment response based at least in part on analyzing in the subject a set of cytokines or a set of transcripts.

[0009] In another examples, disclosed herein is a method for selecting a subject to receive an experimental therapy in a clinical trial, comprising: selecting to the subject to receive the experimental therapy, wherein the experimental therapy has been predicted to generate a treatment response in the subject based at least in part on a trained machine learning (ML) classifier that classifies the subject as responsive or non-responsive to the therapy, wherein the trained ML classifier predicts the treatment response based at least in part on analyzing in the subject a set of cytokines or a set of transcripts.

[0010] In another example, disclosed herein is a computer program product for determining biomarkers predictive of treatment response to a therapy for a subject suffering from a disease or disorder, the computer program product comprising at least one non-transitory computer-readable medium having computer-readable program code portions embodied therein, the computer-readable program code portions comprising: an executable portion configured to identify one or more source nodes associated with a set of proteins targeted by the therapy; an executable portion configured to determine, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response; an executable portion configured to generate a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and an executable portion configured to process a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy. Hence, in one aspect, there is provided A computer program product for determining biomarkers predictive of treatment response to a therapy for a subject suffering from rheumatoid arthritis (RA), the computer program product comprising at least one non-transitory computer-readable medium having computer-readable program code portions embodied therein, the computer-readable program code portions comprising: an executable portion configured to identify one or more source nodes associated with a set of proteins targeted by the therapy; an executable portion configured to determine, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response; an executable portion configured to generate a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and an executable portion configured to process a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy.

INCORPORATION BY REFERENCE

[0011] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the

disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

FIGs. 1A-1H depict an overview, implementation, and validation of the biomarker-discovery framework. **FIG. 1A** depicts a schematic representation of a human interactome (HI) after running the personalized PageRank (PPR) algorithm herein starting from the drug target as a source node. **FIG. 1B** depicts a schematic representation of the HI after running the personalized PageRank algorithm starting from the diseases-signature cytokines as readout nodes. **FIG. 1C** depicts a schematic representation of the HI reflecting a final node ranking based on two rankings in **FIGs. 1A-1B**. Schematic representations of framework implementation: **FIG. 1D** depicts rheumatoid (RA) cytokines and tumor necrosis factor (TNF) (e.g., infliximab target) mapped on

CLAIMS

WHAT IS CLAIMED IS:

1. A method for determining biomarkers predictive of treatment response to a therapy for a subject suffering from rheumatoid arthritis (RA), the method comprising:

identifying one or more source nodes associated with a set of proteins targeted by the therapy;

determining, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response;

generating a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and

processing a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy.

2. The method of claim 1, wherein the ranked set of intermediate nodes comprises a ranked set of proteins, a ranked set of cytokines, or a ranked set of transcripts.
3. The method of claim 2, wherein the ranked set of proteins, the ranked set of cytokines, or the ranked set of transcripts is used as features to obtain a trained ML classifier.
4. The method of claim 2, wherein (i) the ranked set of cytokines is different from at least one cytokine of the set of cytokines or (ii) the ranked set of cytokines is the same as the set of cytokines.
5. The method of claim 2, wherein (i) the ranked set of transcripts is different from at least one transcript of the set of transcripts or (ii) the ranked set of transcripts is the same as the set of transcripts.
6. The method of claim 2, wherein the ranked set of cytokines or the ranked set of transcripts is analyzed to determine statistically significant changes after administering to the subject the therapy.
7. The method of claim 2, wherein generating the ranked set of the one or more intermediate nodes comprises:

determining a first rank between the one or more source nodes and the one or more intermediate nodes;

determining a second rank between the one or more response nodes and the one or more intermediate nodes; and

determining a rank product of the first rank and the second rank.

8. The method of claim 1, further comprising:

constructing a subset of a protein-protein network associated with the disease or disorder, wherein the constructing comprises:

(i) determining the one or more source nodes in the protein-protein network, based at least in part on the set of proteins targeted by the therapy;

(ii) determining the one or more response nodes in the protein-protein network, based on least in part on the set of cytokines or the set of transcripts indicative of the treatment response; and

(iii) determining one or more connections between the one or more source nodes and the one or more response nodes, based at least in part on one or more interactions between the one or more source nodes and the one or more response nodes.

9. The method of claim 1, wherein the therapy for rheumatoid arthritis (RA)comprises an anti-TNF therapy.

10. The method of claim 9, wherein the anti-TNF therapy comprises infliximab, adalimumab, etanercept, certolizumab pegol, or golimumab, or a biosimilar thereof.

11. The method of claim 3, wherein the trained ML classifier predicts the treatment response to the therapy using a non-linear relationship between (i) measurements of the ranked set of intermediate nodes and (ii) the treatment response indicated by statistically significant changes in the ranked set of intermediate nodes after administering to the subject the therapy.

12. The method of claim 3, wherein the trained ML classifier is trained using measurements of the one or more intermediate nodes in (i) a first set of subjects with rheumatoid arthritis (RA)determined to be responsive to the therapy and (ii) a second set of subjects with rheumatoid arthritis (RA)determined to be non-responsive to the therapy.

13. The method of claim 2, wherein the set of cytokines or the ranked set of cytokines indicative of the treatment response to rheumatoid arthritis (RA) comprises: CSF2, CXCL8, IFNγ, IL10, IL12A, IL12B, IL15, IL16, IL17A, IL18, IL1B, IL4, IL6, MIF, TGFB1, TGFB2, TGFB3, TNF.

14. The method of claim 3, wherein the trained ML classifier predicts the treatment response with an area under receiver operating characteristic curve (AUC) of at least about 0.8.

15. A computer program product for determining biomarkers predictive of treatment response to a therapy for a subject suffering from rheumatoid arthritis (RA), the computer program product comprising at least one non-transitory computer-readable medium having computer-

readable program code portions embodied therein, the computer-readable program code portions comprising:

an executable portion configured to identify one or more source nodes associated with a set of proteins targeted by the therapy;

an executable portion configured to determine, by a network-based diffusion process, one or more intermediate nodes associated with the one or more source nodes or the one or more response nodes, wherein the one or more intermediate nodes are indicative of the treatment response;

an executable portion configured to generate a ranked set of the one or more intermediate nodes, wherein the ranked set is based at least in part on one or more interactions of the intermediate nodes with the one or more source nodes or the one or more response nodes; and

an executable portion configured to process a subset of the ranked set of intermediate nodes to determine a set of biomarkers that are predictive of the treatment response to the therapy.