monitor disease progression and treatment response without relying on subjective measures, such as clinical characteristics or patient self-assessment. In some embodiments, certain gene expression patterns for diseased patients are indicative of a response to therapy, and reversal of gene expression of this gene expression pattern in a diseased patient indicates improvement of the health of the diseased subject ("a disease gene expression signature"). Such an approach is distinct from other methods, which examines gene expression differences between patients suffering from the disease, in order to identify whether a patient has a biomarker indicative for response to therapy, as compared to other patients who do not.

[0005] In some examples, a disease gene expression signature is identified using a machine learning algorithm that identifies genes that are differentially expressed between diseased subjects, subsets of diseased subjects, and healthy subjects in a significant manner. Moreover, the present disclosure provides methods and systems that encompass an insight that certain genes within a gene expression profile of a disease subject, when compared to the gene expression profile of a healthy subject, lead to potential targets for therapy that are distinct from the differentially expressed genes in the diseased subject as compared to the healthy subject. That is, while other methods focus on differentially expressed genes in a diseased subject vs. a healthy subject, the present disclosure instead identifies targets for therapy that have significant connection (and thus impact) to these differentially expressed genes but may not be differentially expressed themselves as between diseased and healthy subjects. In some embodiments, a potential target for therapy has a significant connection to the differentially expressed genes in the diseased subject, such that modulating the target may reverse gene expression of the disease gene expression signature after treatment, thereby indicating that the subject's disease is responding to the particular therapy.

[0006] Further, the present disclosure provides methods and systems that encompass an insight that multiple targets for therapy can potentially have a significant connection to the differentially expressed genes in the diseased subject. Accordingly, it may be beneficial to provide a method for identifying which target from among the several targets yields the highest likelihood of success to reverse gene expression of the disease gene expression signature after treatment. In some embodiments, likelihood of success of target modulation to impact a disease gene expression response signature is determined using machine learning algorithms to predict response when a candidate target is modulated. In some embodiments, such a prediction is performed by assessing network proximity (which can include, for example, significance of connection) between a candidate target and each of the genes in a disease expression signature. In some embodiments, artificial intelligence software modules predict targets of highest

significance to the disease gene expression response signature, thereby providing a target of interest for therapy of a diseased subject.

[0007] In a first aspect, the present disclosure provides a method of determining or validating a target for therapy for treating a subject suffering from an autoimmune disease, disorder, or condition, the method comprising: receiving a set of response genes corresponding to a disease gene expression signature, wherein the disease gene expression signature comprises one or more genes that, when expression is reversed in whole or in part, resembles gene expression of a non-diseased subject; receiving a plurality of interactions between one or more potential therapies and a plurality of gene expressions; generating, for each response gene of the set of response genes, one or more potential therapies that alter gene expression of the response gene, based at least in part on the plurality of interactions; scoring each of the one or more potential therapies based at least in part on significance of alteration of the set of response genes, to thereby provide one or more candidate therapies; determining one or more potential targets directly modulated by the one or more candidate therapies; selecting one or more secondary targets sharing significant similarity to the one or more potential targets; compiling a set of targets comprising the one or more potential targets and the one or more secondary targets; and identifying a target from the set of targets having a significant downstream impact similarity to the set of response genes to thereby provide the target for therapy.

[0008] In some examples, the method further comprises mapping each of the one or more potential targets onto a biological network, and selecting one or more secondary targets sharing significant topological similarity to the one or more potential targets on the biological network. In some embodiments, the biological network comprises a human interactome. In some embodiments, the biological network is a human protein-protein interactome. In some embodiments, significant topological similarity of the one or more secondary targets is determined via identification of targets that are proximal to the one or more potential targets on the biological network.

[0009] In some examples, the target for therapy is directly modulated by the one or more candidate therapies. In some embodiments, the target for therapy is not associated with an approved therapy for the disease, disorder, or condition. In some embodiments, the target for therapy is associated with a second disease different from the disease, disorder, or condition. In some embodiments, the therapy comprises a member selected from Table 1. In some embodiments, the therapy comprises gene knockout or gene overexpression. In some embodiments, the therapy comprises an anti-TNF therapy. In some embodiments, the anti-TNF therapy comprises infliximab, etanercept, adalimumab, certolizumab pegol, golimumab, or a biosimilar thereof. In some embodiments, the one or more potential targets comprises JAK1,

JAK2, JAK3, IL23A, ITGA4, ITGB7, IL2RA, IL12A, IL12B, TNF, IL12RB1, IL23R, IL12RB2, or MADCAM1. In some embodiments, the significance in alteration comprises a significant change in gene expression of the set of response genes.

[0010] In some examples, the disease, disorder, or condition comprises an autoimmune disease, disorder, or condition. In some embodiments, 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. In some embodiments, the disease, disorder, or condition comprises ulcerative colitis. In some embodiments, the disease, disorder, or condition comprises rheumatoid arthritis. In some embodiments, the disease, disorder, or condition comprises Alzheimer's disease. In some embodiments, the disease, disorder, or condition comprises multiple sclerosis.

[0011] In another example, the present disclosure provides a method of treating a subject suffering from a disease, disorder, or condition, wherein the subject exhibits a disease gene expression signature associated with the disease, disorder, or condition, the method comprising administering to the subject a therapy that has been determined to revert the disease gene expression signature toward a non-diseased gene expression signature, wherein the therapy has been determined at least in part by: receiving a set of response genes corresponding to the disease gene expression signature, wherein the disease gene expression signature comprises one or more genes that, when expression is reversed in whole or in part, resembles gene expression of a non-diseased subject; receiving a plurality of interactions between one or more potential therapies and a plurality of gene expressions; generating, for each response gene of the set of response genes, one or more potential therapies that alter gene expression of the response gene, based at least in part on the plurality of interactions; scoring each of the one or more potential therapies based at least in part on significance of alteration of the set of response genes, to thereby provide one or more candidate therapies; determining one or more potential targets directly modulated by the one or more candidate therapies; selecting one or more secondary targets sharing significant similarity to the one or more potential targets; compiling a set of targets comprising the one or more potential targets and the one or more secondary targets; selecting a target from the list of targets for the therapy having a significant downstream impact similarity to the set of response genes; and determining that the therapy directly modulates the target.

[0012] In some examples, the therapy has been determined at least in part by further mapping each of the one or more potential targets onto a biological network, and selecting one

or more secondary targets sharing significant topological similarity to the one or more potential targets on the biological network. In some embodiments, the biological network comprises a human interactome. In some embodiments, the biological network is a human protein-protein interactome. In some embodiments, significant topological similarity of the one or more secondary targets is determined via identification of targets that are proximal to the one or more potential targets.

[0013] In some examples, the disease gene expression signature is determined at least in part by: analyzing 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 of prior subjects based at least in part on the gene expression data; and selecting one or more genes having significant differences in gene expression between the two or more groups of prior subjects and a group of non-diseased subjects ("disease candidate genes"), to thereby provide the disease gene expression signature. In some embodiments, stratifying the cohort of subjects into two or more groups of prior subjects is based on whether the prior subjects do or do not respond to a particular therapy.

[0014] In some examples, the target for the therapy is directly modulated by the one or more candidate therapies. In some embodiments, target for therapy is not associated with an approved therapy for the disease, disorder, or condition. In some embodiments, the therapy comprises an anti-TNF therapy. In some embodiments, the anti-TNF therapy comprises infliximab, etanercept, adalimumab, certolizumab pegol, golimumab, or a biosimilar thereof. In some embodiments, the therapy comprises gene knockout or gene overexpression. In some embodiments, the therapy comprises a member selected from Table 1. In some embodiments, the one or more potential targets comprises JAK1, JAK2, JAK3, IL23A, ITGA4, ITGB7, IL2RA, IL12A, IL12B, TNF, IL12RB1, IL23R, IL12RB2, or MADCAM1. In some embodiments, the significance in alteration comprises a significant change in gene expression of the set of response genes.

[0015] In some examples, the disease, disorder, or condition comprises an autoimmune disease, disorder, or condition. In some embodiments, 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. In some embodiments, the disease, disorder, or condition comprises ulcerative colitis. In some embodiments, the disease, disorder, or condition comprises rheumatoid arthritis. In some embodiments, the

disease, disorder, or condition comprises Alzheimer's disease. In some embodiments, the disease, disorder, or condition comprises multiple sclerosis.

[0016] In some examples, scoring of each of the one or more potential therapies comprises: determining a difference in expression level of the set of response genes after treatment with the one or more potential therapies relative to the set of response genes before treatment with the one or more potential therapies; and calculating a p-value for each of the one or more potential therapies.

[0017] In some examples, the potential targets are identified via a machine-learning algorithm. In some embodiments, the machine-learning algorithm comprises a random walk.

[0018] In a second aspect, the present disclosure provides a method for determining a personalized therapy for a subject, the method comprising: receiving or generating a disease gene expression signature comprising a set of response genes; receiving or generating one or more potential therapies that alter expression of the set of response genes; ranking each of the one or more potential therapies based at least in part on significance of alteration of the set of response genes, to thereby provide one or more candidate therapies; determining one or more potential targets directly modulated by the one or more candidate therapies; ranking one or more secondary targets based at least in part on significance of similarity to the one or more potential targets; compiling a set of targets comprising the one or more potential targets and the one or more secondary targets; selecting a target from the set of targets for the personalized therapy having a significant downstream impact similarity to the set of response genes; and determining that the personalized therapy directly modulates the target.

[0019] In some examples, the method further comprises mapping each of the one or more potential targets onto a biological network, and ranking one or more secondary targets based at least in part on significance of topological similarity to the one or more potential targets on the biological network. In some embodiments, the biological network comprises a human interactome.

[0020] In some examples, the disease gene expression signature is determined at least in part by: analyzing 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 of prior subjects based at least in part on the gene expression data; and selecting one or more genes having significant differences in gene expression between the two or more groups of prior subjects and a group of non-diseased subjects ("disease candidate genes"), to thereby provide the disease gene expression signature.

[0021] In a third aspect, the present disclosure provides 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 any of the methods according to the first and second aspects.

[0022] In another example, the present disclosure provides a method of determining or validating a target for therapy for treating a subject suffering from a disease, disorder, or condition, the method comprising: receiving a set of response genes corresponding to a disease gene expression signature, wherein the disease gene expression signature is or comprises one or more genes that, when expression is reversed in whole or in part, resembles gene expression of a healthy subject; receiving a plurality of interactions between one or more potential therapies and a plurality of gene expressions; generating for each gene of the set of response genes, one or more potential therapies that alter gene expression of the set of response genes; scoring each of the one or more potential therapies based on significance of alteration (e.g., the change in gene expression) of the set of response genes, to thereby provide one or more candidate therapies; determining one or more potential targets directly modulated by the one or more candidate therapies; mapping each of the one or more potential targets onto a biological network; selecting one or more secondary targets sharing significant topological similarity to the one or more potential targets on the biological network; compiling a list of targets comprising the one or more potential targets and the one or more secondary targets; identifying a target having a significant downstream impact similarity to the set of response genes from the list of targets to thereby provide the target for therapy.

[0023] In some examples, the target for therapy is directly modulated by the one or more candidate therapies.

[0024] In some examples, significant topological similarity of the one or more secondary targets is determined via identification of targets that are proximal to the one or more potential targets.

[0025] In some examples, the target for therapy is not associated (e.g., is not approved for use) with a therapy.

[0026] In some examples, the target for therapy is associated (e.g., is approved for use) with a disease distinct from the disease afflicting the subject (e.g., is a “novel target”).

[0027] In some examples, the therapy comprises a member selected from Table 1.

[0028] In some examples, the therapy comprises gene knockout or gene overexpression.

[0029] In some examples, the therapy comprises an anti-TNF therapy.

[0030] In some examples, the one or more potential targets is selected from JAK1, JAK2, JAK3, IL23A, ITGA4, ITGB7, IL2RA, IL12A, IL12B, TNF, IL12RB1, IL23R, IL12RB2, and MADCAM1.

[0031] In another example, the present disclosure provides a method of treating a subject that exhibits a disease gene expression signature, the method comprising administering a therapy determined to revert the disease gene expression signature toward a healthy gene expression signature, wherein the therapy has been determined by: selecting a set of response genes from the disease gene expression signature; identifying one or more potential therapies that alter gene expression of the set of response genes; scoring each of the one or more potential therapies based on significance of alteration of the set of response genes to provide one or more candidate therapies; determining one or more potential targets directly modulated by the one or more candidate therapies; mapping each of the one or more potential targets onto a biological network; selecting one or more secondary targets sharing significant topological similarity to the one or more potential targets on the biological network; compiling a list of targets comprising the one or more potential targets and the one or more secondary targets; selecting a target for treatment from the list of targets by identifying a target having a significant downstream impact to the set of response genes; and identifying the therapy that directly modulates the target for treatment.

[0032] In some examples, the disease gene expression signature is determined 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 of prior subjects based on the gene expression data; and selecting one or more genes having significant differences in gene expression between the two or more groups of prior subjects and a group of healthy subjects ("disease candidate genes"), to thereby provide the disease gene expression signature.

[0033] In some examples, the target for treatment is directly modulated by the one or more candidate therapies.

[0034] In some examples, significant topological similarity of the one or more secondary targets is determined via identification of targets that are proximal to the one or more potential targets

[0035] In some examples, target for therapy is not associated with a therapy.

[0036] In some examples, the therapy comprises an anti-TNF therapy.

[0037] In some examples, the anti-TNF therapy is selected from infliximab, etanercept, adalimumab, certolizumab pegol, golimumab, and biosimilars thereof.

[0038] In some examples, the therapy comprises a member selected from Table 1.

[0039] In some examples, the one or more potential targets are selected from JAK1, JAK2, JAK3, IL23A, ITGA4, ITGB7, IL2RA, IL12A, IL12B, TNF, IL12RB1, IL23R, IL12RB2, and MADCAM1

[0040] In some examples, 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.

[0041] In some examples, scoring of each of the one or more potential therapies comprises: determining a difference in expression level of the set of response genes after treatment with the one or more potential therapies relative to the set of response genes before treatment with the one or more potential therapies; and calculating a p-value for each of the one or more potential therapies.

[0042] In some examples, the potential targets are identified by a machine-learning algorithm.

[0043] In some examples, the machine-learning algorithm comprises a random walk.

[0044] In some examples, stratifying the cohort of subjects into two or more groups of prior subjects is based at least in part on whether the prior subjects do or do not respond to a particular therapy.

[0045] In another example, the present disclosure provides a method for engineering a personalized therapy for a subject, the method comprising: receiving or generating a disease gene expression signature comprising a set of response genes; receiving or generating a set of one or more potential therapies that alter expression of the one or more response genes; ranking each of the set of the one or more potential therapies according to significance of alteration of the one or more response genes, to provide a set of one or more candidate therapies; determining one or more potential targets directly modulated by the set of one or more candidate therapies, optionally by mapping the one or more potential targets onto a biological network; and ranking significance of topological similarity between each of the one or more potential targets and the set of response genes; mapping each of the one or more potential targets onto a biological network; identifying one or more secondary targets sharing significant downstream impact to the one or more potential targets; compiling a list of targets comprising the one or more potential targets and the one or more secondary targets; selecting a target for treatment from the list of targets; and selecting the personalized therapy that modulates the target for treatment.

[0046] In some examples, the disease gene expression signature is determined by: receiving or generating 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 of prior subjects based on the gene expression data; and selecting one or more genes

having significant differences in gene expression between the two or more groups of prior subjects and

a group of healthy subjects ("disease candidate genes"), to thereby provide the disease gene expression signature.

[0047] In another example, the present disclosure provides a system for determining or validating a target for therapy for treating a subject suffering from a disease, the 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 one or more operations of any method described herein.

[0048] Another example of the present disclosure provides a non-transitory computer readable medium comprising machine executable code that, upon execution by one or more computer processors, implements any of the methods above or elsewhere herein.

[0049] Another example of the present disclosure provides a system comprising one or more computer processors and computer memory coupled thereto. The computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements any of the methods above or elsewhere herein.

[0050] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, 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 description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0051] 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 or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0052] FIG. 1 depicts an example workflow for identifying a disease expression signature.

[0053] FIG. 2A depicts a plot illustrating network similarity analysis that shows that: TNF has significantly closer network impact similarity to experimentally derived treatment module than to randomly selected treatment module.

CLAIMS

WHAT IS CLAIMED IS:

1. A method of determining or validating a target for therapy for treating a subject suffering from an autoimmune disease, disorder, or condition, the method comprising:

    receiving a set of response genes corresponding to a disease gene expression signature, wherein the disease gene expression signature comprises one or more genes that, when expression is reversed in whole or in part, resembles gene expression of a non-diseased subject;

    receiving a plurality of interactions between one or more potential therapies and a plurality of gene expressions;

    generating, for each response gene of the set of response genes, one or more potential therapies that alter gene expression of the response gene, based at least in part on the plurality of interactions;

    scoring each of the one or more potential therapies based at least in part on a significance of alteration of the set of response genes, to thereby provide one or more candidate therapies;

    determining one or more potential targets directly modulated by the one or more candidate therapies;

    selecting one or more secondary targets sharing significant similarity to the one or more potential targets;

    compiling a set of targets comprising the one or more potential targets and the one or more secondary targets; and

    identifying a target from the set of targets having a significant downstream impact similarity to the set of response genes to thereby provide the target for therapy.

2. The method of claim 1, further comprising mapping each of the one or more potential targets onto a biological network, and selecting one or more secondary targets sharing significant topological similarity to the one or more potential targets on the biological network.
3. The method of claim 2, wherein the biological network comprises a human interactome.
4. The method of claim 2, wherein the biological network is a human protein-protein interactome.
5. The method of claim 2, wherein significant topological similarity of the one or more secondary targets is determined via identification of targets that are proximal to the one or more potential targets on the biological network.

6. The method of claim 1, wherein the target for therapy is directly modulated by the one or more candidate therapies.
7. The method of claim 1, wherein the target for therapy is not associated with an approved therapy for the autoimmune disease, disorder, or condition.
8. The method of claim 1, wherein the target for therapy is associated with a second autoimmune disease different from the autoimmune disease, disorder, or condition.
9. The method of claim 1, wherein the therapy comprises an anti-TNF therapy.
10. The method of claim 9, wherein the anti-TNF therapy comprises infliximab, etanercept, adalimumab, certolizumab pegol, golimumab, or a biosimilar thereof.
11. The method of claim 1, wherein the therapy comprises gene knockout or gene overexpression.
12. The method of claim 1, wherein the therapy comprises a member selected from Table 1.
13. The method of claim 1, wherein the one or more potential targets comprises JAK1, JAK2, JAK3, IL23A, ITGA4, ITGB7, IL2RA, IL12A, IL12B, TNF, IL12RB1, IL23R, IL12RB2, or MADCAM1.
14. The method of claim 1, wherein the significance of alteration comprises a significant change in gene expression of the set of response genes.
15. The method of claim 1, wherein the autoimmune disease, disorder, or condition comprises rheumatoid arthritis, ulcerative colitis, Crohn's disease, juvenile arthritis, psoriatic arthritis, plaque psoriasis, ankylosing spondylitis, Guillain-Barre syndrome, Sjogren's syndrome, scleroderma, vitiligo, Graves' disease, multiple sclerosis, or a combination thereof.
16. The method of claim 1, wherein the autoimmune disease, disorder, or condition comprises rheumatoid arthritis.
17. The method of claim 1, wherein the autoimmune disease, disorder, or condition comprises ulcerative colitis.
18. The method of claim 1, wherein the autoimmune disease, disorder, or condition comprises multiple sclerosis.
19. A method for determining a personalized therapy for a subject, the method comprising:

    receiving or generating a disease gene expression signature comprising a set of response genes;

    receiving or generating one or more potential therapies that alter expression of the set of response genes;

ranking each of the one or more potential therapies based at least in part on significance of alteration of the set of response genes, to thereby provide one or more candidate therapies;

determining one or more potential targets directly modulated by the one or more candidate therapies;

ranking one or more secondary targets based at least in part on significance of similarity to the one or more potential targets;

compiling a set of targets comprising the one or more potential targets and the one or more secondary targets;

selecting a target from the set of targets for the personalized therapy having a significant downstream impact similarity to the set of response genes; and

determining that the personalized therapy directly modulates the target.

20. The method of claim 19, further comprising mapping each of the one or more potential targets onto a biological network, and ranking one or more secondary targets based at least in part on significance of topological similarity to the one or more potential targets on the biological network.

21. The method of claim 20, wherein the biological network comprises a human interactome.

22. The method of claim 21,wherein the biological network is a human protein-protein interactome.

23. The method of claim 19, wherein the disease gene expression signature is determined at least in part by:

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

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

selecting one or more genes having significant differences in gene expression between the two or more groups of prior subjects and a group of non-diseased subjects (“disease candidate genes”), to thereby provide the disease gene expression signature.

24. 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 1-23.