

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2014-532205  
(P2014-532205A)

(43) 公表日 平成26年12月4日(2014.12.4)

(51) Int.Cl.  
G06F 19/12 (2011.01)

F I  
G O 6 F 19/12

テーマコード (参考)

審査請求 未請求 予備審査請求 未請求 (全 55 頁)

<p>(21) 出願番号 特願2014-528898 (P2014-528898)</p> <p>(86) (22) 出願日 平成24年9月7日(2012.9.7)</p> <p>(85) 翻訳文提出日 平成26年4月14日(2014.4.14)</p> <p>(86) 国際出願番号 PCT/EP2012/003760</p> <p>(87) 国際公開番号 W02013/034300</p> <p>(87) 国際公開日 平成25年3月14日(2013.3.14)</p> <p>(31) 優先権主張番号 61/532,972</p> <p>(32) 優先日 平成23年9月9日(2011.9.9)</p> <p>(33) 優先権主張国 米国 (US)</p>	<p>(71) 出願人 500586875 フィリップ モリス プロダクツ エス アー スイス国 2000 ヌーシャテル ケ ジャンルノー 3</p> <p>(74) 代理人 100078282 弁理士 山本 秀策</p> <p>(74) 代理人 100113413 弁理士 森下 夏樹</p> <p>(74) 代理人 100181674 弁理士 飯田 貴敏</p> <p>(74) 代理人 100181641 弁理士 石川 大輔</p> <p>(74) 代理人 230113332 弁護士 山本 健策</p>
---	---

最終頁に続く

(54) 【発明の名称】 ネットワークに基づく生物学的活性評価のためのシステムおよび方法

(57) 【要約】

本明細書では、生物系内の実体のサブセットからの測定された活性データに基づき1つまたは複数の攪乱に対する生物系の応答を定量化するためのシステムおよび方法が開示されている。活性データ、および測定された実体と測定されていない実体との間の関係を記述する生物系のネットワークモデルに基づき、測定されていない実体の活性が推定される。推定された活性は、処置条件への応答などの攪乱への生物系の応答を定量化するスコアを導出するために使用される。スコアは、攪乱に対するネットワークの応答の大きさおよびトポロジー分布を表すものとしてよい。

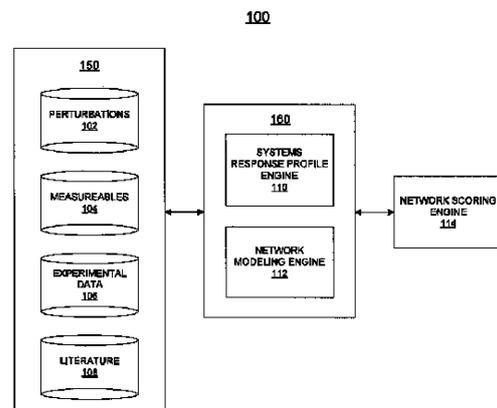


FIG. 1

## 【特許請求の範囲】

## 【請求項 1】

生物系の攪乱を定量化するためのコンピュータ化された方法であって、

第 1 のプロセッサで、第 1 の処置に対する生物学的実体の第 1 の組の応答に対応する処置データの第 1 の組を受け取るステップであって、第 1 の生物系は、生物学的実体の該第 1 の組と生物学的実体の第 2 の組とを含む生物学的実体を備え、該第 1 の生物系内のそれぞれの生物学的実体は、該第 1 の生物系内の該生物学的実体のうちの少なくとも 1 つの別のものと相互作用する、ステップと、

第 2 のプロセッサで、該第 1 の処置と異なる第 2 の処置に対する生物学的実体の該第 1 の組の応答に対応する処置データの第 2 の組を受け取るステップと、

第 3 のプロセッサで、該第 1 の生物系を表す第 1 の計算因果ネットワークモデルを提供するステップであって、該モデルは、

生物学的実体の該第 1 の組を表すノードの第 1 の組と、

生物学的実体の該第 2 の組を表すノードの第 2 の組と、

ノードを接続し、該生物学的実体の間の関係を表すエッジと、

該第 1 の処置データと該第 2 の処置データとの間の変化の予想される方向を表す、方向値とを含む、ステップと、

第 4 のプロセッサを使って、ノードの該第 1 の組における対応するノードに対する該第 1 の処置データと該第 2 の処置データとの間の差を表す活性尺度の第 1 の組を計算するステップと、

第 5 のプロセッサを使って、該第 1 の計算因果ネットワークモデルと活性尺度の該第 1 の組とに基づき、ノードの該第 2 の組における対応するノードに対する活性値の第 2 の組を生成するステップとを含む、方法。

## 【請求項 2】

第 6 のプロセッサを使って、前記第 1 の計算因果ネットワークモデルと活性値の前記第 2 の組とに基づき前記第 1 および第 2 の処置への前記第 1 の生物系の攪乱を表す該第 1 の計算因果ネットワークモデルに対するスコアを生成するステップをさらに含む、請求項 1 に記載の方法。

## 【請求項 3】

請求項 1 に記載の方法であって、活性値の前記第 2 の組を生成するステップは、ノードの前記第 2 の組におけるそれぞれの特定のノードについて、該特定のノードの該活性値と、該特定のノードが前記第 1 の計算因果ネットワークモデル内のエッジと接続されるノードの該活性値または活性尺度との間の差を表す差のステートメントを最小化する該活性値を識別するステップを含み、該差のステートメントは、ノードの該第 2 の組におけるそれぞれのノードの該活性値に依存する、方法。

## 【請求項 4】

活性値の前記第 2 の組におけるそれぞれの活性値は、活性尺度の前記第 1 の組の活性尺度の一次結合である、請求項 1 に記載の方法。

## 【請求項 5】

活性尺度の前記第 1 の組のそれぞれの活性尺度に対する変動推定値の一次結合を形成することによって活性値の前記第 2 の組のそれぞれの活性値に対する変動推定値を提供するステップをさらに含む、請求項 1 に記載の方法。

## 【請求項 6】

活性値の前記第 2 の組を第 1 の活性値ベクトルとして表すステップと、

該第 1 の活性値ベクトルを第 1 の寄与するベクトルと第 1 の寄与しないベクトルとに分解するステップであって、該第 1 の寄与するベクトルと寄与しないベクトルとの和が該第 1 の活性値ベクトルとなる、ステップとをさらに含む、請求項 2 に記載の方法。

## 【請求項 7】

前記第 1 の寄与しないベクトルは、前記第 1 の計算因果ネットワークモデルと関連付けられている符号付きラプラシアンに基づく二次関数のカーネル内にある、請求項 6 に記載

10

20

30

40

50

の方法。

【請求項 8】

前記第 1 のプロセッサで、第 3 の処置に対する生物学的実体の前記第 1 の組の応答に対応する処置データの第 3 の組を受け取るステップと、

前記第 2 のプロセッサで、第 4 の処置に対する生物学的実体の前記第 1 の組の応答に対応する処置データの第 4 の組を受け取るステップと、

前記第 4 のプロセッサを使って、ノードの前記第 1 の組に対応する活性尺度の第 3 の組を計算するステップであって、活性尺度の該第 3 の組におけるそれぞれの活性尺度はノードの該第 1 の組における対応するノードに対する処置データの該第 3 の組と処置データの該第 4 の組との間の差を表す、ステップと、

10

前記第 5 のプロセッサを使って、活性値の第 4 の組を生成するステップであって、それぞれの活性値は前記第 1 の計算因果ネットワークモデルと活性尺度の該第 3 の組とに基づきノードの前記第 2 の組における対応するノードに対する活性値を表す、ステップと、

活性値の該第 4 の組を第 2 の活性値ベクトルとして表すステップと、

該第 2 の活性値ベクトルを第 2 の寄与するベクトルと第 2 の寄与しないベクトルとに分解するステップであって、該第 2 の寄与するベクトルと該第 2 の寄与しないベクトルとの和が該第 2 の活性値ベクトルとなる、ステップと、

前記第 1 の寄与するベクトルと該第 2 の寄与するベクトルを比較するステップとをさらに含む、請求項 6 に記載の方法。

【請求項 9】

20

前記第 1 の寄与するベクトルと前記第 2 の寄与するベクトルを比較するステップは、該第 1 の寄与するベクトルと該第 2 の寄与するベクトルとの間の相関を計算して、処置データの前記第 1 の組および前記第 3 の組の比較可能性を示すステップを含む、請求項 8 に記載の方法。

【請求項 10】

前記第 1 の寄与するベクトルと前記第 2 の寄与するベクトルを比較するステップは、該第 1 の寄与するベクトルおよび該第 2 の寄与するベクトルを計算ネットワークモデルの符号付きラプラシアン像空間上に射影するステップを含む、請求項 8 に記載の方法。

【請求項 11】

請求項 1 に記載の方法であって、前記第 1 のプロセッサで、前記第 1 の処置と異なる第 3 の処置に対する生物学的実体の第 3 の組の応答に対応する処置データの第 3 の組を受け取るステップであって、第 2 の生物系は、生物学的実体の該第 3 の組と生物学的実体の第 4 の組とを含む複数の生物学的実体を備え、該第 2 の生物系内のそれぞれの生物学的実体は、該第 2 の生物系内の該生物学的実体のうちの少なくとも 1 つの別のものと相互作用する、ステップと、

30

前記第 2 のプロセッサで、該第 3 の処置と異なる第 4 の処置に対する生物学的実体の該第 3 の組の応答に対応する処置データの第 4 の組を受け取るステップと、

前記第 3 のプロセッサで、該第 2 の生物系を表す第 2 の計算因果ネットワークモデルを提供するステップであって、

生物学的実体の該第 3 の組を表すノードの第 3 の組と、

生物学的実体の該第 4 の組を表すノードの第 4 の組と、

ノードを接続し、該生物学的実体間の関係を表すエッジと、

40

該第 3 の処置データと該第 4 の処置データとの間の変化の予想される方向を表す、方向値とを含む、ステップと、

前記第 4 のプロセッサを使って、ノードの該第 3 の組に対応する活性尺度の第 3 の組を計算するステップであって、活性尺度の該第 3 の組におけるそれぞれの活性尺度はノードの該第 3 の組における対応するノードに対する処置データの該第 3 の組と処置データの該第 4 の組との間の差を表す、ステップと、

前記第 5 のプロセッサを使って、活性値の第 4 の組を生成するステップであって、それぞれの活性値は該第 2 の計算因果ネットワークモデルと活性尺度の該第 3 の組とに基づき

50

ノードの該第 4 の組における対応するノードに対する活性値を表す、ステップと、  
活性値の該第 4 の組を活性値の前記第 2 の組と比較するステップとをさらに含む、方法

【請求項 1 2】

活性値の前記第 4 の組を活性値の前記第 2 の組と比較するステップは、前記第 1 の計算因果ネットワークモデルに関連付けられている符号付きラプラシアンと前記第 2 の計算因果ネットワークモデルに関連付けられている符号付きラプラシアンとに基づき、カーネル正準相関分析を適用するステップを含む、請求項 1 1 に記載の方法。

【請求項 1 3】

前記活性尺度は、倍率変化値であり、それぞれのノードに対する該倍率変化値は、該各ノードによって表される前記生物学的実体に対する処置データの対応する組の間の差の対数を含む、前記請求項のいずれかに記載のコンピュータ化された方法。

10

【請求項 1 4】

前記第 1 の生物系および前記第 2 の生物系は、*in vitro*系、*in vivo*系、マウス系、ラット系、ヒト以外の霊長類系、およびヒト系からなる群の 2 つの異なる要素である、請求項 1 1 または請求項 1 2 に記載のコンピュータ化された方法。

【請求項 1 5】

請求項 1 に記載のコンピュータ化された方法であって、前記第 1 の処置データは、作用物質に曝露された前記第 1 の生物系に対応し、

前記第 2 の処置データは、該作用物質に曝露されない該第 1 の生物系に対応する、方法

20

【請求項 1 6】

前記生物系の前記攪乱を示す前記スコアの統計的有意性を判定するステップをさらに含む、請求項 2 に記載のコンピュータ化された方法。

【請求項 1 7】

前記スコアの統計的有意性は、複数のランダム生成検定計算因果ネットワークモデルからそれぞれ計算される複数の検定スコアに対して該スコアを比較することによって判定される、請求項 1 6 に記載のコンピュータ化された方法。

【発明の詳細な説明】

【背景技術】

30

【0001】

背景

人体は、長期間にわたって重大な健康危険要因となりうる潜在的に有害な作用物質への曝露によって常時攪乱されている。これらの作用物質への曝露で、人体内部の生物学的機構の正常な機能が損なわれる可能性がある。これらの攪乱 (*perturbation*) が人体に及ぼす作用を理解し、定量化するために、研究者らは、生物系が作用物質への曝露に応答する機構を研究している。いくつかのグループが *in vivo* 動物試験法を広範に利用してきた。しかし、動物試験法は、信頼性と目的適合性に関して疑念があるため、常に十分であるわけではない。異なる動物の生理機能には多くの相違が存在する。したがって、種が異なれば、作用物質への曝露に対する応答は異なることがある。それにより、動物試験から得られる応答がヒト生物学に外挿されうかどうかに関して疑念がある。他の方法として、ヒトの志願者での臨床研究を通じて危険性を評価することが挙げられる。しかし、これらの危険性評価は、経験に基づいて実施され、また、疾患の兆候が現れるまでに何十年もかかることもあるため、これらの評価は、有害物質を疾患にリンクする機構を解明するには十分でない場合がある。さらに他の方法として、*in vitro* 実験が挙げられる。*in vitro* の細胞および組織ベースの方法は、これに対応する動物ベースの方法に対する完全な、または部分的な代替方法として一般的な容認を受けているが、これらの方法は限られた価値を持つ。*in vitro* 法は、細胞および組織の機構の特定の態様に焦点をあわせるものであるため、生物系全体に生じる複雑な相互作用を常に考慮するわけではない。

40

50

## 【0002】

この10年間のうちに、従来の用量依存的な効力および毒性アッセイと併せた核酸、タンパク質、および代謝物レベルのハイスループット測定が、多くの生物学的過程の作用機構を解明するための手段として登場した。研究者らは、これらの異なる測定結果からの情報を科学文献からの生物学的経路に関する知識と組み合わせる意味のある生物学的モデルを構築することを試みた。この目的のために、研究者らは、可能な生物学的作用機構を識別するためにクラスタリングおよび統計的方法などの大量のデータに対するデータマイニングを実行することができる数学的および計算技術を使用し始めた。

## 【0003】

以前の研究では、生物学的過程への1つまたは複数の攪乱の結果生じる遺伝子発現の変化の特徴的サイン(signature)を明らかにすることの重要性、およびその過程の特定の活性の大きさの尺度として付加的なデータセットにおけるそのサインの存在のその後のスコア化を調査した。この点に関する大半の研究は、疾患の表現型と相関するサインを識別し、スコア化することを伴った。これらの表現型派生サインは、著しい分類能力を備えるが、単一の特定の攪乱とサインとの間の機械的關係または因果關係を欠いている。したがって、これらのサインは、多くの場合未知の機構(1つまたは複数)により、同じ疾患の表現型に至るか、またはその結果生じる複数の異なる未知の攪乱を表しうる。

10

## 【0004】

生物系におけるさまざまな個別の生物学的実体の活性が、異なる生物学的機構の活性化または抑制をどのように可能にするかを理解することには1つの難題が横たわっている。遺伝子などの、個別の実体が、複数の生物学的過程(例えば、炎症および細胞増殖)に関わることがあるため、遺伝子の活性を測定するだけでは、上記活性をトリガーする基礎をなす(underlying)生物学的過程を識別するには十分でない。

20

## 【発明の概要】

## 【課題を解決するための手段】

## 【0005】

## 要旨

本明細書では、生物系内の実体のサブセットからの測定された活性データに基づき1つまたは複数の攪乱に対する生物系の応答を定量化するためのシステムおよび方法について記載する。現行の技術はいずれも、潜在的に有害な作用物質および実験条件に応答して、マイクロスケールでの生物学的実体の活性に関与する基礎をなす機構を識別するようには適用されておらず、またこれらの実体に関わる異なる生物学的機構の活性化の定量的評価も提供しない。したがって、システム規模の生物学的データを、生物学的機構を考慮して分析し、システムが作用物質または環境の変化に応答するとき生物系の変化を定量化するための改善されたシステムおよび方法が必要である。測定された活性データおよび、測定された実体と測定されていない実体との間の關係を記述する生物系のネットワークモデルに基づき測定されていない実体の活性を推論するためのシステムおよび方法が記載される。

30

## 【0006】

一態様では、本明細書に記載されているシステムおよび方法は、生物系の攪乱を定量化する(例えば、作用物質への曝露などの処置条件に応答して、または複数の処置条件に応答して)ためのコンピュータ化された方法および1つまたは複数のコンピュータプロセッサを対象とする。コンピュータ化された方法は、第1のプロセッサで、第1の処置に対する生物学的実体の第1の組の応答に対応する処置データの第1の組を受け取るステップを含みうる。生物学的実体の第1の組および生物学的実体の第2の組は、第1の生物系に含まれる。第1の生物系内のそれぞれの生物学的実体は、第1の生物系内の生物学的実体のうちの少なくとも1つの別のものと相互作用する。コンピュータ化された方法は、第2のプロセッサで、第1の処置と異なる第2の処置に対する生物学的実体の第1の組の応答に対応する処置データの第2の組を受け取るステップも含みうる。いくつかの実装では、処置データの第1の組は、作用物質への曝露を表し、処置データの第2の組は、対照データ

40

50

である。コンピュータ化された方法は、第3のプロセッサで、第1の生物系を表す第1の計算因果ネットワークモデルを提供するステップをさらに含む。第1の計算モデルは、生物学的実体の第1の組を表すノードの第1の組と、生物学的実体の第2の組を表すノードの第2の組と、ノードを接続し、生物学的実体の間の関係を表すエッジと、第1の対照データと第1の処置データとの間の変化の予想される方向を表す、ノードまたはエッジに対する、方向値とを含む。いくつかの実装では、エッジおよび方向値は、ノード間の因果活性化関係 (causal activation relationships) を表す。

【0007】

コンピュータ化された方法は、第4のプロセッサを使って、ノードの第1の組の対応するノードに対する第1の処置データと第2の処置データとの間の差を表す活性尺度 (activity measures) の第1の組を計算するステップをさらに含む。

10

【0008】

コンピュータ化された方法は、第5のプロセッサを使って、第1の計算因果ネットワークモデルと活性尺度の第1の組とに基づき、ノードの第2の組における対応するノードに対する活性値の第2の組を生成するステップをさらに含む。いくつかの実装では、活性値の第2の組を生成するステップは、ノードの第2の組におけるそれぞれの特定のノードについて、特定のノードの活性値と、該特定のノードが第1の計算因果ネットワークモデル内のエッジと接続されるノードの活性値または活性尺度との間の差を表す差のステートメント (difference statement) を最小化する活性値を選択するステップを含み、差のステートメントは、ノードの第2の組におけるそれぞれのノードの活性値に依存する。差のステートメントは、ノードの第2の組におけるそれぞれのノードの方向値にさらに依存するものとしてよい。いくつかの実装では、活性値の第2の組におけるそれぞれの活性値は、活性尺度の第1の組の活性尺度の一次結合である。特に、この一次結合は、ノードの第1の組におけるノードと第1の計算因果ネットワークモデル内のノードの第2の組におけるノードとの間のエッジに依存するものとしてよく、また、第1の計算因果ネットワークモデル内のノードの第2の組におけるノード同士の間エッジにも依存し、第1の計算因果ネットワークモデル内のノードの第1の組におけるノード同士の間エッジには依存しない。

20

【0009】

最後に、コンピュータ化された方法は、第6のプロセッサを使って、第1の計算因果ネットワークモデルと活性値の第2の組とに基づき第1の作用物質への第1の生物系の攪乱を表す第1の計算モデルに対するスコアを生成するステップを含む。いくつかの実装では、スコアは、活性値の第2の組に対する二次従属性を有する。コンピュータ化された方法は、活性尺度の第1の組のそれぞれの活性尺度に対する変動推定値 (variation estimate) の一次結合を形成することによって活性値の第2の組のそれぞれの活性値に対する変動推定値を提供するステップもまた含む。活性値の第2の組のそれぞれの活性値に対する変動推定値は、例えば、活性尺度の第1の組のそれぞれの活性尺度に対する変動推定値の一次結合であってよい。スコアに対する変動推定値は、活性値の第2の組に対する二次従属性を有するものとしてよい。

30

40

【0010】

いくつかの実装では、活性値の第2の組は、第1の活性値ベクトルとして表され、第1の活性値ベクトルは、第1の寄与するベクトルと第1の寄与しないベクトルとに分解され、第1の寄与するベクトルと寄与しないベクトルとの和は、第1の活性値ベクトルとなる。スコアは、第1の寄与しないベクトルに依存しなくてもよく、活性値の第2の組の二次関数として計算されうる。そのような一実装では、第1の寄与しないベクトルは、二次関数のカーネル内にある。いくつかの実装では、第1の寄与しないベクトルは、計算因果ネットワークモデル (第1の計算因果ネットワークモデルなど) と関連付けられている符号付きラプラシアンに基づく二次関数のカーネル内にある。

【0011】

50

上に記載されている活性尺度および活性値は、異なる作用物質と同じ生物系に適用される処置条件との間の一致または不一致を反映する比較可能性情報を提供するために使用されうる。そうするために、コンピュータ化された方法は、第1のプロセッサで、第1の処置に対する生物学的実体の第1の組の応答に対応する処置データの第3の組を受け取るステップと、第2のプロセッサで、第2の処置に対する生物学的実体の前記第1の組の応答に対応する処置データの第4の組を受け取るステップと、第4のプロセッサを使って、ノードの第1の組に対応する活性尺度の第3の組を計算するステップであって、活性尺度の第3の組におけるそれぞれの活性尺度はノードの第1の組における対応するノードに対する処置データの第3の組と処置データの第4の組との間の差を表す、ステップとをさらに含みうる。コンピュータ化された方法は、第5のプロセッサを使って、活性値の第4の組を生成するステップであって、活性値の第4の組におけるそれぞれの活性値はノードの第2の組における対応するノードに対する活性値を表し、活性値の第4の組は計算因果ネットワークモデルと活性尺度の第3の組とに基づき、ステップと、活性値の第4の組を第2の活性値ベクトルとして表すステップとをさらに含みうる。

#### 【0012】

コンピュータ化された方法は、第2の活性値ベクトルを第2の寄与するベクトルと第2の寄与しないベクトルとに分解するステップであって、第2の寄与するベクトルと寄与しないベクトルとの和が第2の活性値ベクトルとなる、ステップと、第1と第2の寄与するベクトルを比較するステップとをさらに含みうる。いくつかの実装では、第1と第2の寄与するベクトルを比較するステップは、第1と第2の寄与するベクトルの相関を計算して、処置データの第1および第3の組の比較可能性を示すステップを含む。いくつかの実施形態では、第1と第2の寄与するベクトルを比較するステップは、第1および第2の寄与するベクトルを計算ネットワークモデルの符号付きラプラシアン像空間上に射影するステップとを含む。いくつかの実装では、処置データの第2の組は、処置データの第4の組と同じ情報を含む。

#### 【0013】

上に記載されている活性尺度および活性値は、2つの異なる生物系が同じ作用物質または処置条件による攪乱に同様に応答する程度を反映する翻訳可能性情報を提供するために使用されうる。そうするために、コンピュータ化された方法は、第1のプロセッサで、第1の処置と異なる第3の処置に対する生物学的実体の第3の組の応答に対応する処置データの第3の組を受け取るステップであって、第2の生物系は、生物学的実体の第3の組と生物学的実体の第4の組とを含む複数の生物学的実体を備え、第2の生物系内のそれぞれの生物学的実体は、第2の生物系内の生物学的実体のうちの少なくとも1つの別のものと相互作用する、ステップも含みうる。コンピュータ化された方法は、第2のプロセッサで、第3の処置と異なる第4の処置に対する生物学的実体の第3の組の応答に対応する処置データの第4の組を受け取るステップをさらに含みうる。それに加えて、コンピュータ化された方法は、第3のプロセッサで、第2の生物系を表す第2の計算因果ネットワークモデルを提供するステップを含みうる。第2の計算因果ネットワークモデルは、生物学的実体の第3の組を表すノードの第3の組と、生物学的実体の第4の組を表すノードの第4の組と、ノードを接続し、生物学的実体の間の関係を表すエッジと、第2の対照データと第2の処置データとの間の変化の予想される方向を表す、ノードに対する、方向値とを含む。

#### 【0014】

コンピュータ化された方法は、第4のプロセッサを使って、ノードの第3の組に対応する活性尺度の第3の組を計算するステップであって、活性尺度の第3の組におけるそれぞれの活性尺度はノードの第3の組の対応するノードに対する処置データの第3の組と処置データの第4の組との間の差を表す、ステップと、第5のプロセッサを使って、活性値の第4の組を生成するステップであって、ノードの第4の組における対応するノードに対する活性値の第4の組におけるそれぞれの活性値は第2の計算因果ネットワークモデルと活性尺度の第3の組とに基づき、ステップとをさらに含みうる。最後に、コンピュータ化さ

れた方法は、活性値の第4の組を活性値の第2の組と比較するステップを含むことができる。いくつかの実装では、活性値の第4の組を活性値の第2の組と比較するステップは、第1の計算因果ネットワークモデルに関連付けられている符号付きラブラシアンと第2の計算因果ネットワークモデルに関連付けられている符号付きラブラシアンとに基づきカーネル正準相関分析を適用するステップを含む。

【0015】

いくつかの実装では、第1から第6までのそれぞれのプロセッサは、単一のプロセッサまたは単一のコンピューティングデバイス内に収められている。他の実装では、第1から第6までのプロセッサの1つまたは複数が、複数のプロセッサまたはコンピューティングデバイスにわたって分散される。

10

【0016】

いくつかの実装では、計算因果ネットワークモデルは、潜在的原因を表すノードと測定された量を表すノードとの間に存在する因果関係の組を含む。このような実装では、活性尺度は、倍率変化を含みうる。倍率変化は、対照データと処置データとの間、または異なる処置条件を表すデータの2つの組の間で、初期値から最終値までノード測定値がどれだけ変化するかを記述する数であるものとしてよい。倍率変化数は、これら2つの条件の間の生物学的実体の活性の倍率変化の対数を表すものとしてよい。それぞれのノードに対する活性尺度は、各ノードによって表される生物学的実体に対する処置データと対照データとの差の対数を含みうる。いくつかの実装では、コンピュータ化された方法は、プロセッサを使って、生成されたスコアのそれぞれについて信頼区間を生成するステップを含む。

20

【0017】

いくつかの実装では、生物系のサブセットは、限定はしないが、細胞増殖機構、細胞ストレス機構、細胞炎症機構、およびDNA修復機構のうち少なくとも1つを含む。作用物質は、限定はしないが、生物系に存在も由来もしない分子または実体を含む異物を含むことができる。作用物質は、限定はしないが、毒素、治療化合物、刺激物、弛緩物質、天然物、製造物および食物を含むことができる。作用物質は、限定はしないが、タバコを加熱することによって発生したエアロゾル、タバコを燃焼させることによって発生したエアロゾル、タバコの煙、および紙巻きタバコの煙、のうち少なくとも1つを含むことができる。作用物質は、限定はしないが、カドミウム、水銀、クロム、ニコチン、タバコ特有のニトロソアミン類およびその代謝物(4-(メチルニトロソアミノ(methyl nitrosamino))-1-(3-ピリジル)-1-ブタノン(NNK)、N'-ニトロソノルニコチン(NNN)、N-ニトロソアナタピン(NAT)、N-ニトロソアナバシン(NAB)、および4-(メチルニトロソアミノ)-1-(3-ピリジル)-1-ブタノール(NNAL))を含むことができる。いくつかの実装では、作用物質は、ニコチン置換療法に使用される生成物を含む。

30

【0018】

本明細書に記載されているコンピュータ化された方法は、それぞれが1つまたは複数のプロセッサを備える1つまたは複数のコンピューティングデバイスを有するコンピュータ化されたシステムで実装されうる。一般的に、本明細書に記載されているコンピュータ化されたシステムは、コンピュータ、マイクロプロセッサ、ロジックデバイス、またはハードウェア、ファームウェア、およびソフトウェアを用いて本明細書に記載されているコンピュータ化された方法のうち1つまたは複数を実施するように構成された他のデバイスもしくはプロセッサなどの、1つまたは複数の処理デバイスを備える、1つまたは複数のエンジンを具備することができる。いくつかの実装では、上記コンピュータ化されたシステムは、システム応答プロファイルエンジン、ネットワークモデリングエンジン、およびネットワークスコア化エンジンを備える。上記エンジンは、ときどき相互接続することができ、攪乱データベース、測定可能要素データベース、実験データデータベース、および文献データベースを含む、1つまたは複数のデータベースにさらにときどき接続されうる。本明細書に記載されているコンピュータ化されたシステムは、ネットワークインターフェースを通じて通信する1つまたは複数のプロセッサおよびエンジンを有する分散型のコ

40

50

ンピュータ化されたシステムを含みうる。このような実装は、複数の通信システム上で分散コンピューティングを実行するのに適し得る。

【0019】

本開示のさらなる特徴、その特質、およびさまざまな利点は、図面全体を通して類似の参照文字が類似の部品を指す付属の図面と併せて、以下の詳細な記載を考慮に入れることで、明らかになる。

【図面の簡単な説明】

【0020】

【図1】図1は、攪乱への生物学的ネットワークの応答を定量化するための例示的なコンピュータ化されたシステムのブロック図である。

10

【0021】

【図2】図2は、ネットワーク攪乱振幅(NPA)スコアを計算することによって攪乱への生物学的ネットワークの応答を定量化するための例示的なプロセスの流れ図である。

【0022】

【図3】図3は、2種類の作用物質、2つのパラメータ、およびN個の生物学的実体に対するデータを含むシステム応答プロファイルの基礎となるデータの図形表現である。

【0023】

【図4】図4は、いくつかの生物学的実体を有する生物学的ネットワークおよびそれらの間の関係の計算モデルを示す図である。

【0024】

20

【図5】図5は、生物系の攪乱を定量化するための例示的なプロセスの流れ図である。

【0025】

【図6】図6は、ノードの組に対する活性値を生成するための例示的なプロセスの流れ図である。

【0026】

【図7】図7は、比較可能性情報を提供するための例示的なプロセスの流れ図である。

【0027】

【図8】図8は、翻訳可能性情報を提供するための例示的なプロセスの流れ図である。

【0028】

【図9】図9は、活性値およびNPAスコアに対する信頼区間を計算するための例示的なプロセスの流れ図である。

30

【0029】

【図10】図10は、バックボーンノード(backbone nodes)とサポーティングノード(supporting nodes)を有する生物学的因果ネットワークモデルを例示する図である。

【0030】

【図11】図11は、NPAスコアの統計的有意性を判定するための例示的なプロセスの流れ図である。

【0031】

【図12】図12は、NPAスコアの統計的有意性を判定するための例示的なプロセスの流れ図である。

40

【0032】

【図13】図13は、リーディングバックボーン(leading backbone)および遺伝子ノードを識別するための例示的なプロセスの流れ図である。

【0033】

【図14】図14は、生物学的攪乱の影響を定量化する例示的な分散型のコンピュータ化されたシステムのブロック図である。

【0034】

【図15】図15は、本明細書に記載されているコンピュータ化されたシステムのいずれかにおけるコンポーネントのうちのいずれかを実装するために使用されうる例示的なコンピューティングデバイスを示すブロック図である。

50

【0034】

【図16】図16は、類似している生物学的特徴 (biology) (上) および類似していない生物学的特徴 (下) による2つの実験の例示的な結果を示す図である。

【0035】

【図17】図17は、生物系の攪乱を定量化するための細胞培養実験の例示的な結果を示す図である。

【図18】図18は、生物系の攪乱を定量化するための細胞培養実験の例示的な結果を示す図である。

【発明を実施するための形態】

【0036】

詳細な説明

作用物質によって攪乱されたときに生物系内の変化の大きさを定量的に評価する計算システムおよび方法が本明細書に記載されている。いくつかの実装は、生物系の一部内の変化の大きさを表現する数値を計算するための方法を含む。この計算では、入力として、作用物質によって生物系が攪乱される制御された実験の組から得られたデータの組を使用する。次いで、データが、生物系の特徴のネットワークモデルに適用される。ネットワークモデルは、シミュレーションおよび分析のための基盤 (substrate) として使用され、生物系内の目的の特徴を使用可能にする生物学的機構および経路を表す。この機構および経路の特徴または一部は、生物系の疾病および有害作用の病理に関与しうる。通常状態下および作用物質による攪乱下を含む、さまざまな条件の下での多数の生物学的実体のステータスに関するデータによって占められるネットワークモデルを構築するために、データベースで表されている生物系の従来知識が使用される。使用されるネットワークモデルは、それが攪乱に応答するさまざまな生物学的実体のステータスの変化を表し、生物系に対する作用物質の影響の定量的および客観的評価を得ることができるという点で、動的である。これらの計算方法を運用するためのコンピュータシステムも提供される。

【0037】

本開示のコンピュータ化された方法によって生成された数値は、とりわけ、製造された産物 (安全性評価または比較のため)、栄養補強物を含む治療化合物 (効力または健康上の利益の判定のため)、および環境作用物質 (長期曝露の危険性ならびに有害作用および発症との関係の予測のため) によって引き起こされる望ましい、または有害な生物学的作用の大きさを判定するために使用されうる。

【0038】

一態様では、本明細書に記載されているシステムおよび方法は、攪乱された生物学的機構のネットワークモデルに基づき、攪乱された生物系の変化の大きさを表す計算された数値を提供する。本明細書でネットワーク攪乱振幅 (NPA) スコアと称される数値は、定義されている生物学的機構におけるさまざまな実体のステータスの変化の概略を表すために使用されうる。異なる作用物質または異なる種類の攪乱に対して得られた数値は、生物系の特徴としてそれ自体を使用可能にするか、またはそれ自体を現す生物学的機構に対する異なる作用物質または攪乱の影響を相対比較するために使用することができる。そこで、NPAスコアは、異なる攪乱に対する生物学的機構の応答を測定するために使用することができる。「スコア」という用語は、本明細書では、生物系における変化の大きさの量的尺度を与える値または値の組を一般的に指す。このようなスコアは、サンプルまたは被験体から得られた1つまたは複数のデータセットを使用して、当技術分野で公知の、本明細書で開示されている方法による、さまざまな数学的アルゴリズムおよび計算アルゴリズムのうちのいずれかを使用して計算される。

【0039】

NPAスコアは、研究者および臨床医による診断、実験計画、治療決定、およびリスクアセスメントの改善を助けることができる。例えば、NPAスコアは、毒物学的分析において候補となる生物学的機構の組をスクリーニングして、潜在的に有害な作用物質への曝露で最も影響を受けそうなものを識別するために使用することができる。攪乱へのネット

10

20

30

40

50

ワークの応答の尺度を提供することによって、これらのNPAスコアは、細胞レベル、組織レベル、器官レベル、または生物レベルで出現する表現型または生物学的転帰との分子事象の相関（実験データによって測定されている場合）を可能にすることができる。臨床医は、NPA値を使用して、作用物質によって影響される生物学的機構を患者の生理学的状態と比較し、作用物質に曝露されたときに患者がどのような健康上の危険性または利益を受ける可能性が最も高いかを判定することができる（例えば、免疫無防備状態の（immuno-compromised）患者は、強い免疫抑制応答を引き起こす作用物質に対して特に脆弱であり得る）。

#### 【0040】

同じ生物学的ネットワーク上で異なる実験同士の比較を可能にするため生物学的機構の実験データおよびネットワークモデルを定量化するためのシステムおよび方法が本明細書にさらに記載されているが、これは本明細書において「比較可能性」と称されている。いくつかの実装では、比較可能性は、実験データセットにまたがってNPAまたは他の攪乱の定量化を比較する統計的距離（statistical metric）によって定量化される。比較可能性距離（comparability metric）は、例えば、2つの刺激物（TNFおよびIL1aなど）による特定の生物学的ネットワーク（NFkBなど）の活性化に対する効果が同じ基礎をなす生物学的特徴によって支持されたかどうかを識別するのに役立つ。図16は、類似している生物学的特徴（上段）と類似していない生物学的特徴（下段）とによる2つの実験の例示的な結果を示す図である。上段の結果において、実験1では、すべての測定されたノード間で実験2と比較される実験システムの応答の約2倍となり、これは実験2が、より少ない程度においてであっても、実験1と同じ基礎をなす生物学的特徴をもたらすことを示している。下段の結果において、実験1と実験2との間のそれぞれの測定についての実験システムの応答の間に相関はなく、2つの実験によってもたらされる生物学的特徴は（両方の実験から同じ平均的実験応答が誘発されるという事実にもかかわらず）比較可能でないことを示唆している。本明細書に記載されている比較可能性尺度は、異なる曝露、または異なる用量全体にわたっての同じ曝露を比較したときにネットワーク内の類似の、または類似していない生物学的特徴を識別するために使用されうる。このような尺度は、生物学者に、NPAスコアなどの、生物学的応答の実験結果または他の定量化を適切に理解するうえでより詳細な分析を必要とするネットワークの領域を指し示しうる。

#### 【0041】

生物学的機構の実験データおよびネットワークモデルを定量化し、種、系、または機構の間の類似の生物学的ネットワーク間の比較を可能にするためのシステムおよび方法が本明細書にさらに記載されているが、これは本明細書において「翻訳可能性」と称されている。翻訳可能性尺度は、そのような種、系、または機構の間の実験的攪乱データおよびスコア（NPAスコアなど）の適用性の指標を提供する。例えば、本明細書に記載されている翻訳可能性尺度は、in vivo実験とin vitro実験の比較、マウスの実験とヒトの実験との比較、ラットの実験とヒトの実験との比較、マウスの実験とラットの実験との比較、ヒト以外の霊長類の実験とヒトの実験との比較、および異なる処置に曝される（作用物質への曝露など）他の比較可能な種、系、または機構の比較のために使用されうる。

#### 【0042】

図1は、攪乱へのネットワークモデルの応答を定量化するためのコンピュータ化されたシステム100のブロック図である。特に、システム100は、システム応答プロファイルエンジン110、ネットワークモデリングエンジン112、およびネットワークスコア化エンジン114を備える。エンジン110、112、および114は、ときどき相互接続され、攪乱データベース102、測定可能要素データベース104、実験データデータベース106、および文献データベース108を含む、1つまたは複数のデータベースにときどきさらに接続される。本明細書で使用されているように、エンジンは、コンピュータ、マイクロプロセッサ、ロジックデバイス、またはハードウェア、ファームウェア、お

10

20

30

40

50

よびソフトウェアを用いて1つまたは複数の計算オペレーションを実行するように構成された、図14を参照しつつ記載されているような他の1つまたは複数のデバイスなどの、1つまたは複数の処理デバイスを備える。

#### 【0043】

図2は、一実装による、ネットワーク攪乱振幅(NPA)スコアを計算することによって攪乱への生物学的ネットワークの応答を定量化するためのプロセス200の流れ図である。プロセス200のステップは、図1のシステム100のさまざまなコンポーネントによって実行されるように記述されるが、これらのステップはいずれも、ローカルもしくはリモートの任意の好適なハードウェアコンポーネントまたはソフトウェアコンポーネントによって実行することができ、また任意の適切な順序に配置構成されるか、または並列実行されうる。ステップ210で、システム応答プロファイル(SRP)エンジン110は、さまざまな異なるソースから生物学的データを受け取り、データそれ自体は、さまざまな異なる型のものであってよい。データは、生物系が攪乱される実験からのデータ、さらには対照データを含む。ステップ212で、SRPエンジン110は、生物系内の1つまたは複数の実体が、生物系に対する作用物質の提示に反応して変化する程度の表現である、システム応答プロファイル(SRP)を生成する。ステップ214で、ネットワークモデリングエンジン112は、その1つが作用物質または目的の特徴に関連するものとして選択される複数のネットワークモデルを含む1つまたは複数のデータベースを提供する。この選択は、系の生物学的機能の基礎をなす機構の従来知識に基づいてなされうる。いくつかの実装では、ネットワークモデリングエンジン112は、システム応答プロファイルを用いるシステム内の実体、データベース内のネットワーク、および文献にすでに記載されているネットワークの間の因果関係を抽出し、それにより、ネットワークモデルの生成、精密化、または拡張を行うことができる。ステップ216で、ネットワークスコア化エンジン114は、ネットワークモデリングエンジン112によってステップ214で識別されたネットワークおよびSRPエンジン110によってステップ212で生成されたSRPを使用してそれぞれの攪乱についてNPAスコアを生成する。NPAスコアは、生物学的実体(ネットワークによって表される)の間の基礎をなす関係の状況において攪乱または処置(SRPで表される)への生物学的応答を定量化する。開示をわかりやすくするため、また制限することなく、以下の記載を複数の節に分割する。

#### 【0044】

本開示との関連での生物系は、機能的部分を含む、生物または生物の一部であり、該生物は本明細書では被験体と称される。上記被験体は、一般的に、ヒトを含む、哺乳類である。上記被験体は、ヒト集団における個別のヒトとすることができる。本明細書で使用されているような「哺乳類」という用語は、限定はしないが、ヒト、ヒト以外の霊長類、マウス、ラット、イヌ、ネコ、ウシ、ヒツジ、ウマ、およびブタを含む。ヒト以外の哺乳類は、有利には、ヒトの疾患のモデルを提供するために使用されうる被験体として使用されうる。ヒト以外の被験体は、非改変であるか、または遺伝子組み換え動物(例えば、トランスジェニック動物、または1つもしくは複数の遺伝子変異またはサイレンシングされた遺伝子(1つまたは複数)を持つ動物)とすることができる。被験体は、オスまたはメスとすることができる。上記操作の目的に応じて、被験体は、目的の作用物質に曝露させた被験体とすることができる。被験体は、必要に応じて研究までの時間を含む、長期間にわたって作用物質に曝露させた被験体とすることができる。被験体は、一定の期間にわたって作用物質に曝露させたか、または該作用物質ともはや接触していない被験体とすることができる。被験体は、疾患を有しているものとして診断または識別された被験体とすることができる。被験体は、疾患または有害な健康状態の処置をすでに受けたか、または今受けている最中である被験体とすることができる。被験体は、特定の健康状態または疾患に対する1つまたは複数の症状もしくは危険因子を示す被験体とすることもできる。被験体は、疾患にかかりやすい被験体とすることができ、症状を示すか、または無症候性であるかのいずれかであってよい。いくつかの実装では、目的の疾患または健康状態は、作用物質への曝露、または長期間にわたる作用物質の使用に関連する。いくつかの実装によれば

10

20

30

40

50

、上記システム100(図1)は、攪乱の種類または目的とする転帰に関連する1つまたは複数の生物系およびその機能の機構(まとめて、「生物学的ネットワーク」または「ネットワークモデル」)のコンピュータ化されたモデルを含むか、またはそれを生成する。

#### 【0045】

上記操作の環境(context)に応じて、生物系は、それが、集団における個別の生物、一般的に生物、器官、組織、細胞型、細胞小器官、細胞成分、または特定の個人の細胞(1つまたは複数)の機能に関係するとおりに異なるレベルで定義されうる。それぞれの生物系は、1つまたは複数の生物学的機構または経路を備え、上記操作はその系の機能的特徴として現れる。ヒト健康状態の定義された特徴を再現し、目的の作用物質への曝露について適している動物系は、好ましい生物系である。疾患の原因または病理に関わる細胞型および組織を反映する細胞および器官型系も、好ましい生物系である。in vivoでヒト生物学をできる限り反復する初代細胞または器官培養物を優先することも可能である。また、in vitroのヒト細胞培養物と動物モデルからin vivoで導出される最も等価の培養物とをマッチさせることも重要である。これは、基準系としてin vitroでマッチした系を使用してin vivoの動物モデルからヒト生物学への翻訳連続体(translational continuum)の創製を可能にする。したがって、本明細書に記載されているシステムおよび方法とともに使用することが企図されている生物系は、限定はしないが、機能的特徴(生物学的機能、生理学的機能、または細胞機能)、小器官、細胞型、組織種類、器官、発達段階、または上記の組み合わせによって定義されうる。生物系の例として、限定はしないが、肺系、外皮系、骨格系、筋肉系、神経系(中枢神経および末梢神経)、内分泌系、心血管系、免疫系、循環系、呼吸器系、泌尿器系、腎臓系、胃腸系、結腸直腸系、肝臓系、および生殖器系が挙げられる。生物系の他の例として、限定はしないが、上皮細胞、神経細胞、血液細胞、結合組織細胞、平滑筋細胞、骨格筋細胞、脂肪細胞、卵細胞、精子細胞、幹細胞、肺細胞、脳細胞、心臓細胞、喉頭細胞、咽頭細胞、食道細胞、胃細胞、腎細胞、肝細胞、乳腺細胞、前立腺細胞、膵臓細胞、島細胞、精巣細胞、膀胱細胞、頸部細胞、子宮細胞、結腸細胞、および直腸細胞のさまざまな細胞機能が挙げられる。これらの細胞のうちいくつかは、in vitroで培養されるか、または適切な培養条件の下で無期限にin vitroで維持される細胞系の細胞であるものとしてよい。細胞機能の例として、限定はしないが、細胞増殖(例えば、細胞分裂)、変性、再生、老化、核による細胞活性の制御、細胞間シグナル伝達、細胞分化、細胞脱分化、分泌、遊走、食作用、修復、アポトーシス、および発生プログラミングが挙げられる。生物系として考えることができる細胞成分の例として、限定はしないが、細胞質、細胞骨格、膜、リボソーム、ミトコンドリア、核、小胞体(ER)、ゴルジ体、リソソーム、DNA、RNA、タンパク質、ペプチド、および抗体が挙げられる。

#### 【0046】

生物系における攪乱は、該生物系の1つまたは複数の部分を曝露させるか、または接触させることを通じて一定期間にわたって1つまたは複数の作用物質によって引き起こされうる。作用物質は、すべての構成成分が識別や特徴付けがなされるとは限らない混合物を含む、単一の物質または物質の混合物とすることができる。作用物質またはその構成成分の化学的および物理的特性は完全に特徴付けられない場合もある。作用物質は、その構造、その構成成分、またはある条件の下で該作用物質を生成する供給源によって定義されうる。作用物質の一例は、上記生物系中に存在も由来もしない分子もしくは実体であり、該生物系と接触した後その作用物質から生成される任意の中間体または代謝産物である異物である。作用物質は、炭水化物、タンパク質、脂質、核酸、アルカロイド、ビタミン、金属、重金属、ミネラル、酸素、イオン、酵素、ホルモン、神経伝達物質、無機化合物、有機化合物、環境作用物質、微生物、粒子、環境条件、環境的影響力、または物理的力であってよい。作用物質の非限定的な例として、限定はしないが、栄養素、代謝廃棄物、毒物、麻薬、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、食物、病原体(プリオン、ウイルス、細菌、真菌、原生生物)、寸法がマイクロメートル範囲またはそれ未

10

20

30

40

50

満の粒子もしくは実体、上記のものの副産物、および上記のものの混合物が挙げられる。物理的作用物質の非限定的な例として、放射線、電磁波（太陽光を含む）、温度の上昇もしくは低下、剪断力、流体圧力、放電（1つまたは複数）またはそのシーケンス、あるいは外傷が挙げられる。

#### 【0047】

いくつかの作用物質は、閾値濃度で存在していない限り、または一定期間生物系と接触していない限り、またはその両方の組み合わせが生じていない限り生物系を攪乱しえない。攪乱を結果として引き起こす作用物質への曝露または接触は、用量に関して定量化されうる。したがって、攪乱は、作用物質への長期的曝露の結果生じうる。曝露の期間は、時間の単位で、曝露の頻度で、または上記被験体の実際のもしくは推定される寿命における時間のパーセンテージで表すことができる。攪乱は、生物系の1つまたは複数の部分に、作用物質の供給源から作用物質（上に記載されているような）を供給しないようにするか、または作用物質の供給を制限することによって引き起こされることもある。例えば、攪乱は、栄養素、水、炭水化物類、タンパク質、脂質、アルカロイド、ビタミン、ミネラル、酸素、イオン、酵素、ホルモン、神経伝達物質、抗体、サイトカイン、光の供給不足もしくは欠如によって、または生物のいくつかの部分の移動を制約することによって、または運動を抑圧もしくは要求することによって引き起こされうる。

10

#### 【0048】

作用物質は、上記生物系のどの部分（1つまたは複数）が曝露されるか、および曝露条件によって異なる攪乱を引き起こしうる。作用物質の非限定的な例は、タバコを加熱することによって発生したエアロゾル、タバコを燃焼させることによって発生したエアロゾル、タバコの煙、紙巻きタバコの煙、およびこれらのガス状構成成分または粒子状構成成分のいずれかを含みうる。作用物質のさらなる非限定的な例として、カドミウム、水銀、クロム、ニコチン、タバコ特有のニトロソアミン類およびその代謝物（4 - (メチルニトロソアミノ) - 1 - (3 - ピリジル) - 1 - ブタノン (NNK)、N' - ニトロソノルニコチン (NNN)、N - ニトロソアナタピン (NAT)、N - ニトロソアナバシン (NAB)、4 - (メチルニトロソアミノ) - 1 - (3 - ピリジル) - 1 - ブタノール (NNAL) など)、およびニコチン置換療法のために使用される生成物が挙げられる。作用物質または複合刺激物についての曝露処方計画は、毎日の設定における曝露の範囲および環境を反映すべきである。一群の標準的な曝露処方計画は、同様に定義の明確な (equally well-defined) 実験系に体系的に適用されるように設計されうる。それぞれのアッセイは、初期と後期の事象の両方を捕らえ、代表的な用量範囲が確実にカバーされるように時間および用量依存のデータを収集するように設計することが可能である。しかし、当業者であれば、本明細書に記載されているシステムおよび方法が取り扱われる適用に適しているように適合され改変されうること、また本明細書において設計されているシステムおよび方法が他の好適な適用において使用されうること、またそのような他の追加および改変が本発明の範囲から逸脱しないことを理解する。

20

30

#### 【0049】

さまざまな実装において、遺伝子の発現、タンパク質の発現もしくはタンパク質の代謝回転、マイクロRNAの発現もしくはマイクロRNAの代謝回転、翻訳後修飾、タンパク質修飾、転座、抗体産生代謝産物プロファイル、または上記のうちの2つ以上のものの組み合わせに対する系全体にわたるハイスループット測定が、各対照を含むさまざまな条件の下で生成される。これらは、一般的に、上記評価のためのアンカーとして働き、疾患の原因における明確なステップを表すことができるので、機能転帰測定は、本明細書に記載されている方法において望ましい。

40

#### 【0050】

本明細書で使用されているような「サンプル」は、被験体または実験系（例えば、細胞、組織、器官、または動物全体）から分離される任意の生体サンプルを指す。サンプルは、限定はしないが、単細胞もしくは多細胞、細胞画分、組織生検、切除組織、組織抽出物、組織、組織培養抽出物、組織培養基、吐き出されたガス、全血、血小板、血清、血漿、

50

赤血球、白血球、リンパ球、好中球、マクロファージ、B細胞もしくはそのサブセット、T細胞もしくはそのサブセット、造血細胞のサブセット、内皮細胞、滑液、リンパ液、腹水、間質液、骨髄、脳脊髄液、胸水、腫瘍浸潤物、唾液、粘液、痰、精液、汗、尿、または任意の他の体液を含むことができる。サンプルは、限定はしないが、静脈穿刺、排泄、生検、針吸引、洗浄、擦過、外科的切除、または当技術分野で公知の他の手段を含む手段によって被験体から得ることができる。

#### 【0051】

操作中に、所与の生物学的機構、転帰、攪乱、または上記の組み合わせについて、上記システム100は、処置条件に応答してネットワークにおける生物学的実体のステータスの変化の定量的尺度である、ネットワーク攪乱振幅(NPA)値を生成することができる。

10

#### 【0052】

上記システム100(図1)は、目的の健康状態、疾患、または生物学的転帰に関連する1つまたは複数のコンピュータ化されたネットワークモデル(1つまたは複数)を備える。これらのネットワークモデルのうちの一つまたは複数は、以前の生物学的知識に基づいており、外部ソースからアップロードされ、該システム100内で精選されうる。上記モデルは、測定結果に基づき上記システム100内で新たに生成することもできる。測定可能な要素は、以前の知識を用いることで生物学的ネットワークモデルへと因果的に組み込まれる。以下では、ネットワークモデルを生成もしくは精密化するために使用されうる目的の生物系における変化を表す、または攪乱への応答を表すデータの型について記載する。

20

#### 【0053】

図2を再び参照すると、ステップ210で、上記システム応答プロファイル(SRP)エンジン110は、生物学的データを受け取る。上記SRPエンジン110は、さまざまな異なるソースからこのデータを受け取ることができ、該データそれ自体は、さまざまな異なる型のものであり得る。上記SRPエンジン110によって使用される生物学的データは、文献、データベース(医薬品または医療デバイスの前臨床試験、臨床試験、および臨床後試験からのデータを含む)、ゲノムデータベース(ゲノム配列および発現データ、例えば、National Center for Biotechnology InformationによるGene Expression OmnibusまたはEuropean Bioinformatics InstituteによるArray Express(Parkinsonら、2010年、Nucl. Acids Res.、doi: 10.1093/nar/gkq1040、Pubmed ID 21071405)、市販のデータベース(例えば、Gaithersburg, MD、USAのGene Logic)、または実験研究から取り出すことができる。上記データは、特定の処置条件の効果、または特定の作用物質への曝露の効果の研究するように特に設計されている1つまたは複数の種を用いてin vitro実験、ex vivo実験、またはin vivo実験などの1つまたは複数の異なるソースからの生データを含むものとしてよい。in vitro実験系は、ヒトの疾患の重要側面を表す組織培養または器官型培養(三次元培養)を含みうる。このような実装では、これらの実験のための作用物質の用量決定および曝露の処方計画は、通常の使用もしくは活性状態の間、または特別な使用もしくは活性状態の間、ヒトについて予想されうる曝露の範囲および環境を実質的に反映することができる。実験パラメータおよび試験条件は、上記作用物質および上記曝露条件の性質、問題の生物系の分子および経路、関与する細胞型および組織、目的の転帰、および疾患の原因の側面を反映することが望まれているとおりに選択されうる。特定の動物モデル由来分子、細胞、または組織は、特定のヒト分子、細胞または組織培養物とマッチさせて、動物ベースの所見の翻訳性(translatability)を改善することができる。

30

40

#### 【0054】

ハイスループットの実験技術によって多くが生成されるSRPエンジン110によって受

50

け取られるデータは、限定はしないが、核酸に関係するもの（例えば、特定DNAもしくはRNA種の絶対的または相対的な量、DNA配列、RNA配列の変化、三次構造の変化、または、配列決定によって決定されるようなメチル化パターン、特にマイクロアレイ上の核酸に対するハイブリダイゼーション、定量的ポリメラーゼ連鎖反応、あるいは当技術分野で公知の他の技術）、タンパク質/ペプチド（例えば、絶対的または相対的な量のタンパク質、タンパク質の特定の断片、ペプチド、二次または三次構造の変化、または当技術分野で公知の方法によって決定されるような翻訳後修飾）、および機能的活性（例えば、酵素活性、タンパク質分解活性、転写調節活性、輸送活性、いくつかの結合パートナーへの結合親和力）を、いくつかの条件の下で、とりわけ含む。タンパク質またはペプチドの翻訳後修飾を含む修飾は、限定はしないが、メチル化、アセチル化、ファルネシル化、

10

#### 【0055】

上記データは、限定はしないが、細胞レベルでは細胞増殖、発生的運命、および細胞死を含むもの、生理学的レベルでは、肺気量、血圧、運動熟達度などの、測定された機能的転帰も含みうる。上記データは、限定はしないが、腫瘍転移、腫瘍寛解、機能消失、および疾患の特定の段階における平均余命などの、疾患活性または疾患重症の尺度も含みうる。疾患活性は、臨床的評価によって測定することができ、その結果は、値であるか、または定められた条件の下での1体または複数体の被験体からサンプル（またはサンプルの集団）の評価から得ることができる値の組である。臨床的評価は、被験体による面接またはアンケートに対する回答に基づくものとする。こともできる。

20

#### 【0056】

このデータは、システム応答プロファイルを決定するために使用するため明示的に生成されている場合があるか、または以前の実験でもたらされたか、または文献に公開されている場合もある。一般的に、上記データは、分子、生物学的構造、生理学的状態、遺伝形質、または表現型に関係する情報を含む。いくつかの実装では、上記データは、分子の状態、配置、量、活性、または下部構造、生物学的構造、生理学的状態、遺伝形質、または表現型の記述を含む。後に記載するように、臨床現場では、上記データは、作用物質に曝露された、ヒト被験体から得られたサンプルに対して実施されたアッセイまたはヒト被験体に関する観察結果から得られる生データまたは処理済みデータを含みうる。

30

#### 【0057】

ステップ212で、システム応答プロファイル(SRP)エンジン110は、ステップ212で受け取った生物学的データに基づいてシステム応答プロファイル(SRP)を生成する。このステップは、バックグラウンド補正、正規化、倍率変化計算、有意性決定、および差次的応答の識別（例えば、差次的に発現する遺伝子）のうちの1つまたは複数を含みうる。SRPは、生物系内の1つまたは複数の測定された実体（例えば、分子、核酸、ペプチド、タンパク質、細胞など）が生物系に適用される攪乱（例えば、作用物質への曝露）に反応して個別に変化する程度を表す表現である。一例において、SRPを生成するために、SRPエンジン110は、所与の実験系（「システム-処置」ペア）に適用されるパラメータの所与の組（例えば、処置もしくは攪乱条件）に対する測定値の組を収集する。図3は、2つのSRP、つまり、種々のパラメータ（例えば、第1の処置作用物質への曝露の用量および時間）を用いて第1の処置306を受けるN個の異なる生物学的実体についての生物活性データを含むSRP302、および第2の処置308を受けるN個の異なる生物学的実体についての生物活性データを含む類似のSRP304を示している。SRPに含まれるデータは、生の実験データ、処理済み実験データ（例えば、外れ値を除外するためにフィルタリングされている、信頼度推定でマークされている、多数の試行

40

50

にわたって平均がとられている)、計算生物学的モデルによって生成されたデータ、または科学文献から取ったデータであってよい。SRPは、絶対値、絶対変化、倍率変化、対数変化、関数、および表などの、さまざまな方法でデータを表すことができる。SRPエンジン110は、SRPをネットワークモデリングエンジン112に渡す。

#### 【0058】

前のステップで導出されたSRPはネットワーク攪乱の大きさが、それによって決定されることになる実験データを表しているが、計算および解析のための基盤であるのは生物学的ネットワークモデルである。この解析は、生物系の特徴に関連する機構および経路の詳細なネットワークモデルの開発を必要とする。このようなフレームワークは、より古典的な遺伝子発現の解析で使用されている遺伝子リストの調査を超える機構的理解の1つの層を提供する。生物系のネットワークモデルは、動的生物系を表し、生物系のさまざまな基本的特性に関する定量的情報をアセンブルすることによって構築される数学的構築体である。

10

#### 【0059】

このようなネットワークの構築は、反復プロセスである。ネットワークの境界の線引きは、目的の過程(例えば、肺における細胞増殖)に関連する機構および経路の文献調査によって導かれる。これらの経路を記述する因果関係は、ネットワークの核をなす従来の知識から抽出される。文献ベースのネットワークは、関連する表現型端点を含むハイスループットデータセットを用いて検証されうる。SRPエンジン110はデータセットを解析するために使用することができ、この結果を使用してネットワークモデルを確認し、精密化し、または生成することができる。

20

#### 【0060】

図2を再び参照すると、ステップ214で、ネットワークモデリングエンジン112は、目的の生物系の特徴の基礎をなす機構(1つまたは複数)または経路(1つまたは複数)に基づくネットワークモデルとともにSRPエンジン110からのシステム応答プロファイルを使用している。いくつかの態様では、上記ネットワークモデリングエンジン112は、SRPに基づいてすでに生成されているネットワークを識別するために使用される。上記ネットワークモデリングエンジン112は、モデルへの更新および変更を受け取るためのコンポーネントを備えることができる。上記ネットワークモデリングエンジン112は、新規データを組み込み、追加の、または精密化されたネットワークモデルを生成して、ネットワーク生成のプロセスを反復することもできる。また上記ネットワークモデリングエンジン112は、1つもしくは複数のデータセットのマージまたは1つもしくは複数のネットワークのマージを円滑にすることもできる。データベースから取り出されたネットワークの組は、追加のノード、エッジ、または全く新しいネットワークによって(例えば、特定の生物学的実体によって直接調節される追加の遺伝子の記述について文献のテキストをマイニングすることによって)手動で補うことができる。これらのネットワークは、プロセススコア化を使用可能にすることができる特徴を含む。ネットワークトポロジーが維持され、因果関係のネットワークは、ネットワークにおける任意の地点から測定可能な実体まで追跡されうる。さらに、これらのモデルは動的であり、それらのモデルを組み立てる(build)ために使用される仮定は、改変または言い換えることができ、異なる組織の環境および種に適合性を与えることができる。これは、新しい知識が利用可能になると反復試験および改善を可能にする。上記ネットワークモデリングエンジン112は、信頼度の低い、または科学文献に記載されている実験結果と食い違う対象となるノードまたはエッジを除去することができる。上記ネットワークモデリングエンジン112は、監督された学習または監督のない学習の方法(例えば、計量学習、行列補完、パターン認識)を用いて推論できる追加のノードまたはエッジを備えることもできる。

30

40

#### 【0061】

いくつかの態様において、生物系は、頂点(またはノード)と該ノード同士を接続するエッジからなる数学的なグラフとしてモデル化される。例えば、図4は、9個のノード(ノード402および404を含む)とエッジ(406および408)とを備える単純ネッ

50

トワーク400を示している。上記ノードは、限定はしないが、化合物、DNA、RNA、タンパク質、ペプチド、抗体、細胞、組織、および器官などの、生物系における生物学的実体を表すものとするができる。上記エッジは、上記ノード間の関係を表しうる。上記グラフ内のエッジは、上記ノード間の関係を表しうる。例えば、エッジは、「に結合する」関係、「で表される」関係、「発現プロファイリングに基づいて共調節される」関係、「阻害する」関係、「原稿中に共出現する」関係、または「構造要素を共有する」関係を表しうる。一般的に、これらの種類の関係は、一对のノードの間の関係を記述する。上記グラフにおけるノードは、ノード間の関係も表しうる。したがって、上記グラフで表される、関係の間の関係 (relationships between relationships)、または1つの関係と別の種類の生物学的実体との間の関係を表すことが可能である。例えば、化学物質を表す2つのノードの間の関係は、反応を表すものとすることができる。この反応は、反応と反応を阻害する化学物質との間の関係のノードとすることができる。

10

20

30

40

50

#### 【0062】

グラフは、無向グラフであってもよく、それぞれのエッジに関連付けられている2つの頂点を区別しないことを意味する。代替的に、グラフのエッジは、一方の頂点から別の頂点へ有向であってもよい。例えば、生物学的な環境において、転写調節ネットワークおよび代謝ネットワークは、有向グラフとしてモデル化されうる。転写調節ネットワークのグラフモデルでは、ノードは遺伝子を表し、エッジはそれらの遺伝子の間の転写関係を示す。別の例として、タンパク質間相互作用ネットワークは、生物のプロテオームにおけるタンパク質間の直接的な物理的相互作用を記述し、そのようなネットワークにおける相互作用に関連付けられている方向はないことが多い。そのため、これらのネットワークは、無向グラフとしてモデル化することができる。いくつかのネットワークは、有向と無向の両方のエッジを有することができる。グラフを構成する実体および関係（つまり、上記ノードおよびエッジ）は、システム100内のデータベースにおける相互に関連付けられているノードのウェブとして記憶されうる。

#### 【0063】

上記データベース内で表される知識は、さまざまな異なるソースから引き出される、さまざまな異なる種類の知識であってもよい。例えば、特定のデータは、遺伝子に関する情報、および遺伝子間の関係を含む、ゲノムデータベースを表しうる。このような一例では、ノードは、がん遺伝子を表し、そのがん遺伝子ノードに接続されている別のノードは、該がん遺伝子を阻害する遺伝子を表すことができる。上記データは、タンパク質、およびタンパク質間の関係、疾患およびそれらの相互関係、ならびにさまざまな疾患状態を表すことができる。図形表現で組み合わせることができる多くの異なる型のデータが存在する。計算モデルは、例えば、DNAデータセット、RNAデータセット、タンパク質データセット、抗体データセット、細胞データセット、組織データセット、器官データセット、医療データセット、疫学データセット、化学データセット、毒物学データセット、患者データセット、および集団データセットにおける知識を表すノード間の関係のウェブを表すものとしてよい。本明細書で使用される場合、データセットは、定められた条件の下でサンプル（またはサンプルの群）の評価の結果得られる数値の集合である。データセットは、例えば、サンプルの定量化可能な実体を実験的に測定することによって、または代替的に、または研究所、臨床研究組織などのサービスプロバイダーから、または公開もしくは専用データベースから得ることができる。データセットは、データ、およびノードによって表される生物学的実体を含むことができ、該データセットのそれぞれにおけるノードは、同じデータセットにおける、または他のデータセットにおける他のノードと関係していてもよい。さらに、上記ネットワークモデリングエンジン112は、例えば、DNA、RNA、タンパク質、または抗体のデータセットにおける遺伝情報から、医療データセットにおける医療情報、患者データセットにおける、また集団全体では、疫学データセットにおける個別の患者に関する情報までを表す計算モデルを生成することができる。上に記載されているさまざまなデータセットに加えて、他の多くのデータセット、または計算モデル

を生成するときを含めることができる生物学的情報の種類がありうる。例えば、データベースはさらに、医療記録データ、構造/活性関係データ、伝染病理に関する情報、臨床試験に関する情報、曝露パターンデータ、生成物の使用履歴に関するデータ、および他の任意の種類生命科学関係の情報を含むことも可能である。

#### 【0064】

上記ネットワークモデリングエンジン112は、例えば、遺伝子間の調節相互作用、タンパク質間の相互作用、または細胞もしくは組織における複雑な生化学的相互作用を表す1つまたは複数のネットワークモデルを生成することができる。上記ネットワークモデリングエンジン112によって生成されたネットワークは、静的モデルおよび動的モデルを含むことができる。上記ネットワークモデリングエンジン112は、任意の適用可能な数学的スキームを使用して、ハイパーグラフおよび重みづけ二部構成グラフ (weighted bipartite graph) などの、システムを表すことができ、ここでは、ノードの2つの種類が反応および化合物を表すために使用される。上記ネットワークモデリングエンジン112は、発現量に差のある遺伝子内の機能関係遺伝子の過剰表現に基づく解析、ベイジアンネットワーク解析、グラフィカルガウスモデル技術、または遺伝子関連性ネットワーク技術などの他の推論技術を用いてネットワークモデルを生成して、実験データの組(例えば、遺伝子発現、代謝産物濃度、細胞応答など)に基づき関連する生物学的ネットワークを識別することもできる。

10

#### 【0065】

上に記載されているように、上記ネットワークモデルは、生物系の機能的特徴の基礎をなす機構および経路に基づく。上記ネットワークモデリングエンジン112は、作用物質の長期的な健康上のリスクまたは健康上の利益の研究に関連する生物系の特徴に関する結果を表すモデルを生成するか、または含むことができる。したがって、上記ネットワークモデリングエンジン112は、細胞機能、特に、限定はしないが、細胞増殖、細胞性ストレス、細胞再生、アポトーシス、DNA損傷/修復、または炎症応答を含む、生物系内の目的の特徴に関するか、または寄与する機能の、さまざまな機構についてのネットワークモデルを生成するか、または含むことができる。他の実施形態では、上記ネットワークモデリングエンジン112は、急性全身毒性、発がん性、皮膚透過、心血管疾患、肺疾患、生態毒性、目の洗浄/腐食、遺伝毒性、免疫毒性、神経毒性、薬物動態、薬物代謝、器官毒性、生殖および発達毒性、皮膚刺激/腐食、または皮膚感作性に関連する、計算モデルを含むか、または生成することができる。一般的に、上記ネットワークモデリングエンジン112は、核酸(DNA、RNA、SNP、siRNA、miRNA、RNAi)、タンパク質、ペプチド、抗体、細胞、組織、器官、および任意の他の生物学的実体のステータス、ならびにそれらの各相互作用についての計算モデルを含むか、または生成することができる。一例において、計算ネットワークモデルは、免疫応答または炎症反応の間の免疫系のステータスおよびさまざまな種類の白血球の機能を表すために使用されうる。他の例において、計算ネットワークモデルは、心血管系の性能ならびに内皮細胞の機能および代謝を表すために使用するとともに可能である。

20

30

#### 【0066】

本開示のいくつかの実装では、上記ネットワークは、生物学的因果関係知識のデータベースから引き出される。このデータベースは、異なる生物学的機構の実験研究を実施して、そのいくつかは因果関係であってもよい機構間の関係(例えば、活性化または阻害関係)を抽出することによって生成され、Cambridge、Massachusetts、USAのSelventa Inc.によって精選された、Genstruct Technology PlatformまたはSelventa Knowledgebaseなどの、市販のデータベースと組み合わせることができる。生物学的因果関係知識のデータベースを使用することで、上記ネットワークモデリングエンジン112は、攪乱102および測定可能要素104(measurable 104)をリンクするネットワークを識別することができる。いくつかの実装では、上記ネットワークモデリングエンジン112は、SRPエンジン110からのシステム応答プロファイルと文献においてすでに

40

50

生成されているネットワークとを使用して生物学的実体間の因果関係を抽出する。他の処理ステップのうちで、上記データベースをさらに処理して、論理的矛盾を取り除き、生物学的実体の異なる組の間に相同的推論を適用することによって新しい生物学に関する知識を生み出すことができる。

【0067】

いくつかの実装では、上記データベースから抽出されたネットワークモデルは、逆因果的推論(RCR)に基づくが、これは因果関係のネットワークを処理して機構仮説を立て、次いで、示差測定結果のデータセットに対してそれらの機構仮説を評価する自動化推論技術である。それぞれの機構仮説は、生物学的実体を、その実体が影響を及ぼしうる測定可能な量にリンクさせる。例えば、測定可能な量として、とりわけ、生物学的実体の濃度、個数、または相対存在量の増減、生物学的実体の活性化もしくは阻害、または生物学的実体の構造、機能、または論理の変化が挙げられ得る。RCRでは、計算のための基盤として生物学的実体間の実験的に観察される因果的相互作用の有向ネットワークを使用する。上記有向ネットワークは、生物学的実体間の相互関係を記録するための構文である、Biological Expression Language(商標)(BEL(商標))で表すことができる。上記RCRの計算では、限定はしないが、経路長(上流ノードと下流ノードとを接続するエッジの最大数)などのネットワークモデル生成、および上流ノードを下流ノードに接続する可能な因果的経路に対するいくつかの制約条件を指定する。RCRの出力は、関連性および精度を評価する統計量によって順位化された、実験測定結果の差異の上流制御機構(upstream controller)を表す機構仮説の組である。上記機構仮説出力をアSEMBルして、因果連鎖およびより大きなネットワークを形成し、相互接続されている機構および経路のより高いレベルで上記データセットを解釈することができる。

10

20

【0068】

機構仮説の一種は、潜在的原因を表すノード(上流ノードまたは制御機構)と測定された量を表すノード(下流ノード)との間に存在する因果関係の組を含む。この種類の機構仮説は、上流ノードによって表されている実体の存在量が増えた場合に、因果増大関係によってリンクされる下流ノードが増大すると推論され、因果減少関係によってリンクされる下流ノードが減少すると推論されるなどの、予測を行うために使用することができる。

30

【0069】

機構仮説は、測定されたデータ、例えば、遺伝子発現データの組と、それらの遺伝子の公知の制御機構である生物学的実体との間の関係を表す。それに加えて、これらの関係は、上流実体と下流実体(例えば、下流遺伝子)の差示的発現との間の影響の符号(正または負)を含む。機構仮説の下流実体は、文献で精選されている生物学的因果関係知識のデータベースから引き出されうる。いくつかの実装では、計算可能な因果関係ネットワークモデルの形態の、上流実体を下流実体にリンクする機構仮説の因果関係は、上記NPAスコア化法によるネットワーク変化の計算のための基盤である。

【0070】

いくつかの実施形態では、生物学的実体の複雑な因果関係ネットワークモデルは、該モデルにおける生物系のさまざまな特徴を表す個別の機構仮説を収集し、すべての上記下流実体(例えば、下流遺伝子)と単一の上流実体または過程との接続を再編成することによって単一の因果関係ネットワークモデルに変換され、これにより、複雑な因果関係ネットワークモデル全体を表すことができ、これは本質的に基礎をなすグラフ構造の平坦化である。したがって、ネットワークモデルで表されているような生物系の特徴および実体の変化は、個別の機構仮説を組み合わせることによって評価することができる。いくつかの実装では、因果ネットワークモデル内のノードのサブセット(本明細書では「バックボーンノード」と称される)は、測定されない、または簡便に、もしくは経済的に測定されえない実体に対応する生物学的実体の第1の組、例えば、生物系内の重要な主体(key actor)の生物学的機構または活動を表し、ノードの別のサブセット(本明細書では「サポーターノード」と称される)は、測定することができ、また計算のため値が実験

40

50

的に決定され、データセットで提示される生物系内の生物学的実体の第2の組、例えば、生物系内の複数の遺伝子の発現レベルを表す。図10は、4つのバックボーンノード1002、1004、1006、および1008、ならびにバックボーンノードの間の、またバックボーンノードからサポート遺伝子発現ノード1010、1012、および1014の群へのエッジを含む例示的なネットワークを示している。図10内のそれぞれのエッジは、有向であり（つまり、因果関係の方向を表す）、符号付き（つまり、正または負の調節を表す）である。このタイプのネットワークは、いくつかの生物学的実体または機構（例えば、特定の酵素の存在量または活性化の増大と同じくらい固有である量から増殖因子シグナル伝達経路のステータスを反映するものと同じくらい複雑である量までの範囲の）と、正または負に調節される他の下流の実体（例えば、遺伝子発現レベル）との間に存在する因果関係の組を表しうる。

10

#### 【0071】

いくつかの実装では、上記システム100は、上記細胞が紙巻きタバコの煙に曝露されているときの細胞増殖の機構についてのコンピュータ化されたモデルを含むか、または生成することができる。このような一例では、上記システム100は、限定はしないが、がん、肺疾患、および心血管疾患を含む、紙巻きタバコの煙の曝露に関連するさまざまな健康状態を表す1つまたは複数のネットワークモデルを含むか、または生成することもできる。いくつかの態様において、これらのネットワークモデルは、適用される攪乱（例えば、作用物質への曝露）、さまざまな条件の下での応答、目的の測定可能な量、調査されている転帰（例えば、細胞増殖、細胞性ストレス、炎症、DNA修復）、実験データ、臨床データ、疫学データ、および文献のうちの少なくとも1つに基づく。

20

#### 【0072】

図示されている一例として、上記ネットワークモデリングエンジン112は、細胞性ストレスのネットワークモデルを生成するように構成されうる。上記ネットワークモデリングエンジン112は、文献データベースから公知のストレス応答に関わる関連する機構を記述するネットワークを受け取ることができる。上記ネットワークモデリングエンジン112は、肺および心血管の環境でのストレスに応答して動作することが公知の生物学的機構に基づいて1つまたは複数のネットワークを選択することができる。いくつかの実装では、上記ネットワークモデリングエンジン112は、生物系内の1つまたは複数の機能単位を識別し、より小さなネットワークをそれらの機能性に基づいて組み合わせることによってより大きなネットワークモデルを組み立てる。特に、細胞性ストレスモデルについては、上記ネットワークモデリングエンジン112は、酸化的ストレス、遺伝毒性ストレス、低酸素ストレス、浸透ストレス、生体異物ストレス、および剪断応力への応答に関連する機能単位を考慮することができる。したがって、細胞性ストレスモデルに対するネットワーク成分（network component）は、生体異物代謝応答、遺伝毒性ストレス、内皮剪断応力、低酸素応答、浸透ストレス、および酸化的ストレスを含みうる。上記ネットワークモデリングエンジン112は、特定の細胞群において実施されたストレス関連実験からの公に入手可能なトランスクリプトームデータの計算解析からの内容を受け取ることができる。

30

#### 【0073】

生物学的機構のネットワークモデルを生成するときに、ネットワークモデリングエンジン112は、1つまたは複数のルールを含むことができる。このようなルールは、ネットワーク内容、ノードの種類などを選択するためのルールを含んでよい。上記ネットワークモデリングエンジン112は、*in vitro*および*in vivo*の実験結果の組み合わせを含む、実験データのデータベース106から1つまたは複数のデータセットを選択することができる。上記ネットワークモデリングエンジン112は、実験データを利用して、文献において識別されているノードおよびエッジを検証することができる。細胞性ストレスのモデリングの例において、上記ネットワークモデリングエンジン112は、疾患のない肺または心血管組織において実験が生理学的に関連するストレスをどれほどうまく表しているかに基づき実験についてのデータセットを選択することができる。データセ

40

50

ットの選択は、例えば、表現型ストレスのエンドポイントデータの利用可能性、遺伝子発現プロファイリング実験の統計的厳密さ、および通常の疾患のない肺または心血管の生物学との実験の環境との関連性に基づくものとする事ができる。

【0074】

関連するネットワークの集合を識別した後、上記ネットワークモデリングエンジン112はさらに、これらのネットワークを処理し、精密化することができる。例えば、いくつかの実装では、複数の生物学的実体およびそれらの接続は、グループ化され、新しい1つまたは複数のノードによって表されうる（例えば、クラスタリングまたは他の技術を使用して）。

【0075】

上記ネットワークモデリングエンジン112はさらに、識別されたネットワークにおけるノードおよびエッジに関する記述的情報を含むものとしてよい。上に記載されているように、ノードは、その関連する生物学的実体、該関連する生物学的実体が測定可能な量であるか否かの指示、または該生物学的実体の任意の他の記述子によって記述され、その一方、エッジは、例えば、エッジが表す関係の種類（例えば、アップレギュレーションまたはダウンレギュレーション、相関、条件付き依存性、または非依存性などの因果関係）、その関係の強さ、またはその関係における統計的信頼度によって記述されうる。いくつかの実装では、それぞれの処置について、測定可能な実体を表すそれぞれのノードは、上記処置に応答する活性の変化の予測される方向（つまり、増加または減少）に関連付けられている。例えば、気管支上皮細胞が、腫瘍壊死因子（TNF）などの作用物質に曝露される場合、特定の遺伝子の活性が増大しうる。この増大は、文献から公知である（またネットワークモデリングエンジン112によって識別されたネットワークのうちの1つで表される）直接的調節関係があるため、またはネットワークモデリングエンジン112によって識別されたネットワークのうちの1つまたは複数のエッジを通じて多数の調節関係（例えば、自己分泌シグナリング）を追跡することによって生じうる。いくつかの場合において、上記ネットワークモデリングエンジン112は、上記測定可能な実体のそれぞれについて、特定の攪乱に反応して、変化の予測される方向を識別することができる。上記ネットワークにおける異なる経路が特定の实体についての変化の相反する予測される方向を示す場合、それら2つの経路は、変化の正味の方向を決定するためにさらに詳しく調査されうるか、またはその特定の实体の測定結果が破棄されうる。

【0076】

本明細書において提示されている計算方法およびシステムは、実験データおよび計算ネットワークモデルに基づきNPAスコアを計算する。計算ネットワークモデルは、システム100によって生成されるか、システム100内にインポートされるか、またはシステム100内で（例えば、生物学的知識のデータベースから）識別されうる。ネットワークモデル内の攪乱の下流の効果として識別される実験測定値は、ネットワーク特有の応答スコアの生成において組み合わされる。したがって、ステップ216で、ネットワークスコア化エンジン114は、ネットワークモデリングエンジン112によってステップ214で識別されたネットワークおよびSRPエンジン110によってステップ212で生成されたSRPを使用してそれぞれの攪乱についてNPAスコアを生成する。NPAスコアは、生物学的実体（識別されたネットワークによって表される）の間の基礎をなす関係の状況において処置（SRPで表される）への生物学的応答を定量化する。ネットワークスコア化エンジン114は、ネットワークモデリングエンジン112内に含まれるか、またはネットワークモデリングエンジン112によって識別されたネットワークのそれぞれについてNPAスコアを生成するためのハードウェア構成要素およびソフトウェア構成要素を備えることができる。

【0077】

ネットワークスコア化エンジン114は、攪乱に対するネットワークの応答の大きさおよびトポロジー分布を示すスカラー値またはベクトル値のスコアを生成する技術を含む、一群のスコア化技術のうちのいずれかを実装するように構成されうる。

10

20

30

40

50

## 【0078】

追加のスコア化技術は、いくつかの用途において有利に応用することができ、また、同じ生物学的ネットワーク上の異なる実験同士の比較を可能にする（本明細書では「比較可能性」と称される）、または種、系、または機構の間の類似の生物学的ネットワーク同士の比較を可能にする（本明細書では「翻訳可能性」と称される）ように拡張されうる。次に、一群のスコア化技術について、比較可能性および翻訳可能性を評価するための技術と併せて記載する。

## 【0079】

図5は、作用物質に応答する生物系の攪乱を定量化するための例示的なプロセス500の流れ図である。プロセス500は、例えば、ネットワークスコア化エンジン114、またはシステム100の任意の他の適切に構成された1つまたは複数のコンポーネントによって実装されうる。特に、生物学的実体の第1の組が測定されうる（つまり、生物学的実体の第1の組について処置データと対照データとが測定される）が、生物学的実体の第2の組は、測定されえない（つまり、生物学的実体の第2の組について処置データも対照データも測定されない）。データは、いくつもの理由から生物学的実体の第2の組について容易に入手可能であるとは限らない（または限られた量が入手可能であり得る）。例えば、生物学的実体の第2の組に対応するデータは、取得が特に困難であるか、または生物学的実体の第2の組は、生物学的実体の別の容易に測定可能な組に関係付けられ、これにより、データは測定可能な組から容易に推論することができる。

## 【0080】

作用物質に応答する生物系の攪乱を定量化するために、ネットワークスコア化エンジン114は、攪乱への生物学的機構の応答を表す数値である、NPAスコアを計算することができる。NPAスコアを計算するための一方法では、直接的に測定されるデータ（つまり、上記の例における生物学的実体の第1の組に対応する）のみを使用する。しかし、このアプローチは、生物学的機構に対する攪乱の影響を判定するために潜在的に使用されるデータのサブセットに制限される。特に、直接測定されない生物学的実体の別の組（つまり、上記の例における生物学的実体の第2の組に対応する）があってもよいが、NPAスコアに対する情報を提供することができる。この場合、生物学的実体の未測定の組は、測定された組に関係していてもよく、それによりネットワークスコア化エンジン114は、この測定可能な組から未測定の組に関係するデータを推論することができる。そこで、NPAスコアは、測定されたデータ、推論されたデータ、または両方の組み合わせに基づくものとしてよい。図5のプロセス500は、推論されたデータに基づきNPAスコアを計算するための方法を記述する。

## 【0081】

ステップ502で、ネットワークスコア化エンジン114は、生物系内の生物学的実体の第1の組に対する処置データおよび対照データを受け取る。処置データは、作用物質への生物学的実体の第1の組の応答に対応するが、対照データは、作用物質の非存在への生物学的実体の第1の組の応答に対応する。生物系は、生物学的実体の第1の組（処置データおよび対照データをステップ502で受け取る）、さらには生物学的実体の第2の組（処置データおよび対照データを受け取ることはできない）を含む。生物系内のそれぞれの生物学的実体は、生物系内の生物学的実体の他の少なくとも1つと相互作用し、特に、第1の組の少なくとも1つの生物学的実体は第2の組内の少なくとも1つの生物学的実体と相互作用する。生物系内の生物学的実体の間の関係は、生物学的実体の第1の組を表すノードの第1の組、生物学的実体の第2の組を表すノードの第2の組、およびノードを接続し生物学的実体の間の関係を表すエッジを含む計算ネットワークモデルによって表すことができる。計算ネットワークモデルは、ノードに対する方向値も含むことができ、これは対照データと処置データとの間の変化（例えば、活性化または抑制）の予想される方向を表す。このようなネットワークモデルの例は、上に詳しく記載されている。

## 【0082】

ステップ504で、ネットワークスコア化エンジン114は、生物学的実体の第1の組

における生物学的実体に対する活性尺度を計算する。活性尺度の第1の組におけるそれぞれの活性尺度は、第1の組内の特定の生物学的実体に対する処置データと対照データとの間の差を表す。計算ネットワークモデルにおける生物学的実体の第1の組とノードの第1の組との間に対応関係があるため、ステップ504は、計算ネットワークモデルにおけるノードの第1の組に対する活性尺度も計算する。いくつかの実装では、活性尺度は、倍率変化を含みうる。倍率変化は、対照データと処置データとの間、または異なる処置条件を表すデータの2つの組の間で、初期値から最終値までノード測定値がどれだけ変化するかを記述する数であるものとしてよい。倍率変化数は、これら2つの条件の間の生物学的実体の活性の倍率変化の対数を表すものとしてよい。それぞれのノードに対する活性尺度は、各ノードによって表される生物学的実体に対する処置データと対照データとの間の差の対数を含みうる。いくつかの実装では、コンピュータ化された方法は、プロセッサを使って、生成されたスコアのそれぞれについて信頼区間を生成するステップを含む。

10

## 【0083】

ステップ506で、ネットワークスコア化エンジン114は、生物学的実体の第2の組における生物学的実体に対する活性値を生成する。第2の組における生物学的実体について処置データも対照データも受け取らなかったため、ステップ506で生成された活性値は、推論された活性値を表し、活性尺度の第1の組および計算ネットワークモデルに基づく。生物学的実体の第2の組（計算ネットワークモデルにおけるノードの第2の組に対応する）について推論された活性値は、一群の推論技術のうちいずれかに従って生成することができ、いくつかの実装が、図6に関して以下に記載されている。ステップ506で測定されていない実体について生成される活性値は、ネットワークモデルによって提供される実体間の関係を使用して、直接測定されていない生物学的実体の挙動を明らかにする。

20

## 【0084】

ステップ508で、ネットワークスコア化エンジン114は、ステップ506で生成された活性値に基づきNPAスコアを計算する。NPAスコアは、作用物質への生物系の攪乱を表し（対照データと処置データとの差に反映されるような）、ステップ506で生成された活性値および計算ネットワークモデルに基づく。いくつかの実装では、ステップ508で計算されたNPAスコアは、

30

## 【化1】

$$NPA(G, \beta) = \frac{1}{|\{(x \rightarrow y) \text{ s.t. } x, y \in V_0\}|} \sum_{\text{s.t. } x, y \in V_0}^{x \rightarrow y} (f(x) + \text{sign}(x \rightarrow y)f(y))^2, \quad (1)$$

に従って計算することができ、ただし、 $V_0$ は、生物学的実体の第1の組を表し（つまり、処置データおよび対照データをステップ502で受け取ったもの）、 $f(x)$ は、生物学的実体 $x$ についてステップ508で生成された活性値を表し、 $\text{sign}(x \rightarrow y)$ は、生物学的実体 $x$ を表すノードを生物学的実体 $y$ を表すノードに接続する計算ネットワークモデルにおけるエッジの方向値を表す。生物学的実体の第2の組と関連付けられている活性値のベクトルが、 $f_2$ で表されるとすると、ネットワークスコア化エンジン114は、二次形式

40

## 【化2】

$$NPA = f_2^T Q f_2, \quad (2)$$

を介してNPAスコアを計算するように構成することができ、ただし、

## 【化3】

$$Q = \frac{1}{|\{(x \rightarrow y) \text{ s.t. } x, y \in V_0\}|} \left[ \left( \text{diag}(\text{out}|_{l^2(V \setminus V_0)}) + \text{diag}(\text{in}|_{l^2(V \setminus V_0)}) - (-A - A^T) \right) |_{l^2(V \setminus V_0)} \right] \in l^2(V \setminus V_0) \quad (3)$$

であり、 $\text{diag}(\text{out})$ は、ノードの第2の組におけるそれぞれのノードの出次数を

50

持つ対角行列を表し、 $\text{diag}(i, n)$  は、ノードの第 2 の組におけるそれぞれのノードの入次数を持つ対角行列を表し、 $A$  は、第 2 の組のノードのみに限定され、

【化 4 - 1】

$$A_{xy} = \begin{cases} \text{sign}(x \rightarrow y) & x \rightarrow y \text{ の場合} \\ 0 & \text{そうでない場合} \end{cases} \quad (4)$$

に従って定義されている計算ネットワークモデルの隣接行列を表す。 $A$  が重み付け隣接行列である場合、 $A$  の要素  $(x, y)$  に、重み因子  $w(x, y)$  を掛けることができる。

【0085】

ステップ 508 は、NPA スコアに対する信頼区間を計算するステップも含みうる。いくつかの実装では、活性値  $f_2$  は、多変量正規分布  $N(\mu, \Sigma)$  に従うと仮定され、次いで、式 2 により計算された NPA スコアは、以下の式に従って計算されうる関連する分散を有することになる。

【化 4 - 2】

$$\text{var}(f^T Q f) = 2\text{tr}(Q \Sigma Q \Sigma) + 4\mu^T Q \Sigma Q \mu. \quad (5)$$

いくつかの実装では、式 5 に従って動作するものなどにおいて、NPA スコアは、活性値に対する二次従属性を有する。ネットワークスコア化エンジン 114 は、式 5 に従って計算された分散を使用して、とりわけ、チェビシェフの不等式を適用するか、または中心極限定理に頼ることによって控えめな信頼区間を生成するようにさらに構成されうる。

【0086】

図 6 は、ノードの組に対する活性値を生成するための例示的なプロセス 600 の流れ図である。例えば、プロセス 600 は、図 5 のプロセス 500 のステップ 506 で実行することができる。説明しやすくするためネットワークスコア化エンジン 114 によって実行されるものとして記述される。ステップ 602 で、ネットワークスコア化エンジン 114 は、差のステートメントを識別する。差のステートメントは、特定の生物学的実体の活性尺度または値と特定の生物学的実体が接続される生物学的実体の活性尺度または値との間の差を表す式または他の実行可能なステートメントであるものとしてよい。目的の生物系を表す計算ネットワークモデルの専門用語では、差のステートメントは、ネットワークモデルにおける特定のノードの活性尺度または値と特定のノードがエッジを介して接続されるノードの活性尺度または値との間の差を表す。差のステートメントは、計算ネットワークモデルにおけるノードの 1 つまたは複数のいずれかに依存するものとしてよい。いくつかの実施形態では、差のステートメントは、図 5 のステップ 506 に関して上に記載されているノードの第 2 の組におけるそれぞれのノード（つまり、処置データまたは対照データが利用可能でなく、活性値が他のノードおよび計算ネットワークモデルに関連付けられている処置データまたは対照データから推論されるノード）の活性値に依存する。

【0087】

いくつかの実装では、ネットワークスコア化エンジン 114 は、ステップ 602 で以下の差のステートメント

【化 5 - 1】

$$\sum_{x \rightarrow y} (f(x) - \text{sign}(x \rightarrow y) f(y))^2 w(x \rightarrow y), \quad (6)$$

を識別するが、ただし、 $f(x)$  は、活性値（ノードの第 2 の組におけるノード  $x$  に対する）または尺度（ノードの第 1 の組におけるノード  $x$  に対する）を表し、 $\text{sign}(x, y)$  は、生物学的実体  $x$  を表すノードを、生物学的実体  $y$  を表すノードに接続する計算ネットワークモデルにおけるエッジの方向値を表し、 $w(x, y)$  は、実体  $x$  および  $y$  を表すノードを接続するエッジに関連付けられている重みを表す。説明を簡単にするため、残りの説明では、 $w(x, y)$  は 1 に等しいと仮定するが、当業者であれば、式 6 の差のステートメントの説明を通して（つまり、式 4 を参照しつつ上に記載されているような重み

10

20

30

40

50

付け隣接行列を使用することによって)、1以外の重みも容易にたどれる。

【0088】

ネットワークスコア化エンジン114は、以下の等価なステートメントのいずれかを含む、多くの異なる方法で式6の差のステートメントを実装することができる。

【化5-2】

$$\begin{aligned} & \sum_{x \rightarrow y} (f(x) - \text{sign}(x \rightarrow y)f(y))^2 \\ &= \sum_x \sum_{y: x \rightarrow y} f(x)^2 + f(y)^2 - 2\text{sign}(x \rightarrow y)f(x)f(y) \\ &= \sum_x f(x)^2 \cdot \text{out}(x) + \sum_y f(y)^2 \cdot \text{in}(y) - 2\sum_{x \rightarrow y} \text{sign}(x \rightarrow y)f(x)f(y) \\ &= f^T(\text{diag}(\text{out}) + \text{diag}(\text{in}))f - f^T(A + A^T)f. \end{aligned} \quad (7)$$

10

【0089】

ステップ604で、ネットワークスコア化エンジン114は、差の目標(difference objective)を識別する。差の目標は、ネットワークスコア化エンジン114が生物学的実体の第2の組に対する活性値を選択する際に向かう差のステートメントの値に対する最適化ゴールを表す。差の目標は、差のステートメントが最大化されるか、最小化されるか、または可能な限りターゲット値に近づけられることを指定することができる。差の目標は、活性値が選択される際の生物学的実体を指定することができ、それぞれの実体について許容される活性値の範囲に対して制約を課することができる。いくつかの実装では、差の目標は、図5のステップ506を参照しつつ上に記載されているノードの第2の組におけるすべての生物学的実体にわたって式6の差のステートメントを最小化することであり、ただしその際に、生物学的実体の第1の組の活性(処置データおよび対照データが利用可能であるもの)が図5のステップ504で計算された活性尺度に等しいという制約が課される。この差の目標は、以下の計算最適化問題

20

【化6】

$$\text{argmin}_{f \in \mathcal{L}^2(V)} \sum_{x \rightarrow y} (f(x) - \text{sign}(x \rightarrow y)f(y))^2 \cdot w(x \rightarrow y) \quad \text{ここで} \quad f|_{V_0} = \beta, \quad (8)$$

として書くことができる。ただし、 $\beta$ は、第1の組における実体のそれぞれに対する図5のステップ504で計算された活性尺度を表す。

30

【0090】

ステップ604で識別された差の目標を扱うために、ネットワークスコア化エンジン114は、ステップ606に進んで、差の目標に基づきネットワークモデルを計算により特徴付けるように構成される。生物系を表す計算ネットワークモデルは、いくつかの方法で特徴付けることができる(例えば、上に記載されているような重み付けまたは非重み付け隣接行列Aを介して)。異なる特徴付けが、異なる差の目標に適している場合もあり、NPAスコアを計算する際のネットワークスコア化エンジン114の性能を改善する。例えば、差の目標が、上記の式8に従って定式化されると、ネットワークスコア化エンジン114は、

【化7-1】

$$L = \text{diag}(\text{out}) + \text{diag}(\text{in}) - (A + A^T). \quad (9)$$

40

に従って定義された符号付きラプラシアン行列を使用して計算ネットワークモデルを特徴付けるように構成される。この特徴付けが与えられた場合、式8の差の目標は、

【化7-2】

$$\text{argmin}_{f \in \mathcal{L}^2(V)} f^T L f \quad \text{ここで} \quad f|_{V_0} = \beta. \quad (10)$$

として表すことができる。

【0091】

50

ネットワークスコア化エンジン 114 は、ネットワークモデルを、ノードの第 1 の組内の接続、ノードの第 1 の組からノードの第 2 の組への接続、ノードの第 2 の組からノードの第 1 の組への接続、およびノードの第 2 の組内の接続の 4 つの構成要素に分割することによって第 2 のレベルで計算ネットワークモデルを特徴付けるように構成されうる。計算に関して、ネットワークスコア化エンジン 114 は、ラプラシアン行列を 4 つの部分行列に分割し（これらの構成要素のそれぞれについて 1 つ）、活性のベクトル  $f$  を 2 つの部分ベクトルに分割する（ノードの第 1 の組  $f_1$  の活性に 1 つ、ノードの第 2 の組  $f_2$  の活性に 1 つ）ことによってこの追加の特徴付けを実装することができる。式 10 の差のステートメントのこの再特徴付けは、

【化 8】

$$f^T \begin{pmatrix} L_1 & L_2 \\ L_2^T & L_3 \end{pmatrix} f = (f_1^T \ f_2^T) \begin{pmatrix} L_1 & L_2 \\ L_2^T & L_3 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = f_1^T L_1 f_1 + f_1^T L_2 f_2 + f_2^T L_2^T f_1 + f_2^T L_3 f_2. \quad (11)$$

として書くことができる。

【0092】

ステップ 606 で、ネットワークスコア化エンジン 114 は、差の目標を達成するか、または近似するために活性値を選択する。当技術分野では多くの異なる計算最適化ルーチンが公知であり、ステップ 604 で識別された任意の差の目標に適用されうる。式 10 の差の目標がステップ 604 で識別される実装では、ネットワークスコア化エンジン 114 は、 $f_2$  に関して式 11 の（数値的または解析的）微分をとり、その導関数をゼロと置き、並べ替えて  $f_2$  の式をくくり出すことによって式 11 の式を最小化する  $f_2$  の値を選択するように構成されうる。

【化 9】

$$\frac{\partial}{\partial f_2} (f^T L f) = 2L_2^T f_1 + 2L_3 f_2, \quad (12)$$

なので、ネットワークスコア化エンジン 114 は、

【化 10】

$$f_2 = -L_3^{-1} L_2^T f_1 \equiv K f_1. \quad (13)$$

に従って  $f_2$  を計算するように構成されうる。

【0093】

$f_1$  は、生物学的実体の第 1 の組に対する計算された活性尺度のベクトルなので（これに対して処置データおよび対照データが利用可能である）、生物学的実体の第 2 の組に対する活性値は、式 13 による計算された活性尺度の一次結合として表すことができる。式 13 のように、活性値は、ノードの第 1 の組におけるノードと第 1 の計算ネットワークモデル（つまり、 $L_2$ ）内のノードの第 2 の組におけるノードとの間のエッジに依存し、また計算因果ネットワークモデル（つまり、 $L_3$ ）内のノードの第 2 の組におけるノード間のエッジにも依存しうる。いくつかの実装では（式 13 に従って動作するものなど）、活性値は、計算ネットワークモデル内のノードの第 1 の組におけるノード間のエッジに依存しない。

【0094】

ステップ 608 で、ネットワークスコア化エンジン 114 は、ステップ 606 で生成された活性値を提供する。いくつかの実装では、活性値は、ユーザーに対して表示される。いくつかの実装では、活性値は、上に記載されているように NPA スコアを計算するために図 5 のステップ 508 で使用される。いくつかの実装では、この活性値に対する分散および信頼情報も、ステップ 608 で生成されうる。例えば、活性値および活性尺度は、おおそ多変量正規分布  $N(\mu, \quad)$  に従うものと仮定されうる場合、 $A f$  も、

10

20

30

40

【化 1 1 - 1】

$$\text{var}(Af) = A\Sigma A^T \quad (14)$$

となる多変量正規分布に従う。この場合、推論される活性値に対する信頼区間は、

【化 1 1 - 2】

$$A = -L_3^{-1}L_2^T \text{ および } \Sigma = \text{diag}(\text{var}(\beta))$$

として標準的な統計的手法を使用して計算できる。

【0 0 9 5】

図 5 のステップ 5 0 4 で計算された活性尺度および図 5 のステップ 5 0 6 で生成された活性値（例えば、図 6 のプロセス 6 0 0 による）は、異なる作用物質と同じ生物系に適用される処置条件との間の一致または不一致を反映する比較可能性情報を提供するために使用されうる。図 7 は、比較可能性情報を提供するための例示的なプロセス 7 0 0 の流れ図である。プロセス 7 0 0 は、図 5 のステップ 5 0 6 でノードの第 2 の組について活性値を生成した後に、例えば、ネットワークスコア化エンジン 1 1 4、またはシステム 1 0 0 の任意の他の適切に構成された 1 つまたは複数のコンポーネントによって実行されうる。

10

【0 0 9 6】

ステップ 7 0 2 で、ネットワークスコア化エンジン 1 1 4 は、活性値の第 1 の組を第 1 の活性値ベクトルとして表す。この種の表現は、式 1 1 を参照しつつ上に記載されており、活性値の組は、ベクトル  $f_2$  として表された。ステップ 7 0 4 で、ネットワークスコア化エンジン 1 1 4 は、第 1 の活性値ベクトルを第 1 の寄与するベクトルと第 1 の寄与しないベクトルとに分解する。第 1 の寄与するベクトルおよび第 1 の寄与しないベクトルは、活性値ベクトルと N P A スコアとの間の関係に依存する。N P A スコアが

20

$$N P A = g(h(v_1)) \quad (15)$$

となるように、第 1 の活性値ベクトル  $v_1$  の変換  $g$  として表される場合、 $v_1$  は、ステップ 7 0 4 で

$$v_1 = v_{1c} + v_{1nc} \quad (16)$$

および

$$g(v_{1nc}) = 0 \quad (17)$$

となるように 2 つのベクトル  $v_{1c}$  と  $v_{1nc}$  との和に分解されうる。数学的に、寄与しないベクトル  $v_{1nc}$  は、 $g$  が厳密に正の符号に定まっている場合に変換  $h$  のカーネル内にあると言われ、寄与するベクトル  $v_{1c}$  は、変換  $h$  の像空間内にあると言われる。標準的な計算手法を適用して、さまざまな種類の変換のカーネルおよび像空間を決定することができる。ネットワークスコア化エンジン 1 1 4 が、式 5 および 1 3 に従って活性値ベクトル  $v_1$  から N P A スコアを計算する場合、その N P A スコア変換のカーネルは、行列積

30

【化 1 2】

$$(L_3^{-1}L_2^T)$$

のカーネルであり、その N P A スコア変換の像空間は、行列積

【化 1 3】

$$(L_3^{-1}L_2^T)$$

の像空間である。したがって、活性値ベクトルは、標準的な計算射影法を使用して、行列積

40

【化 1 4】

$$(L_3^{-1}L_2^T)$$

の像空間内の寄与する成分  $v_{1c}$  と行列積

【化 15】

$(L_3^{-1}L_2^T)$

のカーネル内の寄与しない成分  $v_1 n c$  に分解することができ、NPA は、寄与しない成分  $v_1 n c$  に依存しえない。

【0097】

NPA スコアは二次形式として計算できるので（上で示されているように）、ネットワークスコア化エンジン 114 は、入力データがモデルにおける機構の実際の攪乱を反映しないとしても有意な（生物学的変動に関して）スコアを生成することができる。ネットワークが実際に攪乱されるか（つまり、モデルに記述されている生物学がデータに反映されていること）を評価するために、随伴する統計量を使用して、抽出された信号がネットワーク構造に特有であるか、または収集されたデータに内在するかどうかを判定するのを補助することができる。いくつかの種類の間で並べ替え検定は、観察された信号が生物学的因果ネットワークモデルによって与えられるデータまたは構造に内在する特性をより適切に表現しているかどうかを評価する際に特に有益でありうる。

10

【0098】

図 11 および 12 は、因果ネットワークモデルおよび特定のデータセットを与えられた場合に提案されている NPA スコアの統計的有意性を判定するためにネットワークスコア化エンジン 114 によって使用できるプロセス 1100 および 1200 を例示している。提案されている NPA スコアの統計的有意性を判定するステップは、ネットワークによってモデル化されている生物系が攪乱されたかどうかを示すために役立つ。提案されている NPA スコアの統計的有意性を判定するために、ネットワークスコア化エンジン 114 は、以下に記載されているようにデータに一方または両方の検定を適用することができる。

20

【0099】

両方の検定（本明細書ではそれぞれ並べ替え検定と称されている）は、因果ネットワークモデルの 1 つまたは複数の態様のランダム置換を生成するステップと、その結果得られる検定モデルを使用して提案されている NPA スコアを生成したのと同じデータセットおよびアルゴリズムに基づき検定 NPA スコアを計算するステップと、検定 NPA スコアを、提案されている NPA スコアと比較するか、または順位付けして、提案されている NPA スコアの統計的有意性を判定するステップとに基づく。検定モデルを生成するためにランダムに類別される因果ネットワークモデルの態様は、サポートリングノードのラベル、バックボーンノードをサポートリングノードに接続するエッジ、またはバックボーンノード同士を接続するエッジを含む。

30

【0100】

一実装では、本明細書で「O 統計」検定と称されている並べ替え検定は、因果ネットワークモデル内のサポートリングノードの位置の重要度を評価する。プロセス 1100 は、計算された NPA スコアの統計的有意性を評価するための方法を含む。特に、ステップ 1102 で、第 1 の提案されている NPA スコアは、無修正ネットワーク（`unmodified network`）とも称される、生物系内の実体の因果関係に関する知識に基づくネットワークに基づき計算される。ステップ 1106 で、遺伝子標識およびその結果として、それぞれのサポートリングノードの対応する値は、ネットワークモデル内のサポートリングノードの間にランダムに再割り当てされる。ランダムな再割り当ては、何回も、例えば、C 回繰り返され、ステップ 1112 で、検定 NPA スコアは、ランダムな再割り当てに基づき計算され、その結果、C 個の検定 NPA スコアが分布する。ネットワークスコア化エンジン 114 は、ネットワークに基づき NPA スコアを計算するために上に記載されている方法のいずれかに従って提案された NPA スコアおよび検定 NPA スコアを計算することができる。ステップ 1114 で、提案された NPA スコアは、検定 NPA スコアの分布と比較されるか、またはその分布と突き合わせて順位付けされ、これにより、提案された NPA スコアの統計的有意性を判定する。

40

50

## 【0101】

特定の植え込みにおいて、生物系の攪乱を定量化する方法は、因果ネットワークモデルに基づき提案されているNPAスコアを計算するステップと、スコアの統計的有意性を判定するステップとを含む。この有意性は、因果ネットワークモデルのサポータリングノードのラベルをランダムに再割り当てして検定モデルを作成するステップと、検定モデルに基づき検定NPAスコアを計算するステップと、提案されたNPAスコアと検定NPAスコアとを比較して生物系が攪乱されているかどうかを判定するステップとを含む方法によって計算されうる。サポータリングノードのラベルは、活性尺度に関連付けられる。

## 【0102】

整数Cは、ネットワークスコア化エンジンによって決定される任意の数であり、ユーザー入力に基づくものとしてよい。整数Cは、ランダムな再割り当てに基づくNPAスコアの結果得られる分布がほぼ滑らかであるように十分に大きいものとしてよい。整数Cは、再割り当てが所定の回数だけ実行されるように固定されてもよい。あるいは、整数Cは、結果として得られるNPAスコアに応じて変化してもよい。例えば、整数Cは、繰り返し大きくすることができ、追加の再割り当ては、結果として得られるNPA分布が滑らかでない場合に実行されうる。それに加えて、分布がガウス分布または任意の他の好適な分布などの特定の形態に似てくるまでCを大きくするなどこの分布に対する任意の他の追加の要件を使用することもできる。いくつかの実装では、整数Cは、約500から約1000までの範囲内である。

## 【0103】

ステップ1110で、ネットワークスコア化エンジン114は、ステップ1106で生成されたランダムな再割り当てに基づきC個のNPAスコアを計算する。特に、NPAスコアは、ステップ1106で生成されるそれぞれの再割り当てについて計算される。いくつかの実装では、C個すべての再割り当ては、最初に、ステップ1106で生成され、次いで、対応するNPAスコアが、ステップ1110でC個の再割り当てに基づき計算される。他の実装では、対応するNPAスコアは、再割り当てのそれぞれの組が生成された後に計算され、このプロセスは、C回繰り返される。後者のシナリオではメモリに関するコストを節約することができ、Cに対する値がすでに計算されているN個の値に依存する場合には望ましい場合がある。ステップ1112で、ネットワークスコア化エンジン114は、ステップ1106で生成されたランダムな再割り当てに対応して、結果として得られるC個のNPAスコアを集計して、NPA値の分布を形成または生成する。分布は、NPA値のヒストグラムまたはこのヒストグラムの正規化バージョンに対応しうる。

## 【0104】

ステップ1114で、ネットワークスコア化エンジン114は、第1のNPAスコアをステップ1112で生成されたNPAスコアの分布と比較する。例えば、この比較は、提案されたNPAスコアと分布との間の関係を表す「p値」を決定するステップを含みうる。特に、p値は、提案されたNPAスコア値より高い、または低い分布のパーセンテージに対応しうる。小さい、例えば、0.5%未満、1%未満、5%未満、または任意の他の割合のp値は、提案されたNPAスコアが統計的に有意であることを示す。例えば、ステップ1114で計算された低いp値（例えば、 $< 0.05$ または5%未満）を有する提案されたNPAスコアは、提案されたNPAスコアはランダムな遺伝子標識再割り当ての結果得られる有意な数の検定NPAスコアに関して高いことを示している。

## 【0105】

特定の实装において、本明細書で「K統計」検定と称されている別の並べ替え検定は、因果ネットワークモデル内のバックボーンノードの構造の重要度を評価する。プロセス1200は、提案されたNPAスコアの統計的有意性を評価するための方法を含む。プロセス1200は、因果ネットワークモデルの一態様がランダムに類別され、これにより、複数の検定モデルを作成し、その後、複数の検定NPAスコアが計算されるという点でプロセス1100に類似している。生物系内の実体の因果関係の知識に基づき構築された因果ネットワークモデルは、無修正ネットワークとも称される。このようなモデルでは、エッ

10

20

30

40

50

ジは、符号付きであるものとしてよく、したがって、エッジは、2つのバックボーンノード間の正または負の関係を表すことができる。したがって、因果ネットワークモデルは、結果として正の影響をもたらすバックボーンノードを接続する  $n$  個のエッジと、結果として負の影響をもたらすバックボーンノードを接続する  $m$  個のエッジとを備える。

#### 【0106】

ステップ1202で、提案されたNPAスコアは、生物系における実体の因果関係の知識に基づき構築されたネットワークに基づき計算される。次いで、ステップ1204で、個数  $n$  の負のエッジと個数  $m$  の正のエッジが決定される。ステップ1206で、バックボーンノードの対は、それぞれ  $n$  個の負のエッジのうち1つまたは  $m$  個の正のエッジのうち1つとランダムに接続される。 $n + m$  個のエッジを持つランダムな接続を生成するこのプロセスは、 $C$  回繰り返される。すでに説明されているように、繰り返しの数  $C$  は、ユーザー入力によって、または検定NPAスコアの分布の平滑度によって決定される。ステップ1212で、複数の検定NPAスコアは、他のバックボーンノードにランダムに接続されているバックボーンノードを含む複数の検定モデルに基づき計算される。ネットワークスコア化エンジン114は、ネットワークに基づきNPAスコアを計算するために上に記載されている方法のいずれかに従って提案されたNPAスコアおよび検定NPAスコアを計算することができる。ステップ1214で、提案されたNPAスコアは、検定NPAスコアの分布と比較されるか、またはその分布と突き合わせて順位付けされ、これにより、提案されたNPAスコアの統計的有意性を判定する。

10

#### 【0107】

ステップ1210で、ネットワークスコア化エンジン114は、ステップ1206で形成されたランダムな再接続に基づき  $C$  個のNPAスコアを計算する。ステップ1212で、ネットワークスコア化エンジン114は、ステップ1106で生成されたランダムな再接続の結果得られる検定モデルに基づき、検定NPA値の分布を生成するために結果として得られる  $C$  個のNPAスコアを集計する。分布は、NPA値のヒストグラムまたはこのヒストグラムの正規化バージョンに対応しうる。

20

#### 【0108】

ステップ1214で、ネットワークスコア化エンジン114は、提案されたNPAスコアをステップ1212で生成されたNPAスコアの分布と比較する。例えば、この比較は、提案されたNPAスコアと分布との間の関係を表す「 $p$  値」を決定するステップを含みうる。特に、 $p$  値は、提案されたNPAスコア値より高い、または低い分布のパーセンテージに対応しうる。小さい、例えば、0.1%未満、0.5%未満、1%未満、5%未満、または中間の割合の  $p$  値は、提案されたNPAスコアが統計的に有意であることを示す。例えば、ステップ1214で計算された低い  $p$  値（例えば、 $< 0.05$  または5%未満）を有する提案されたNPAスコアは、提案されたNPAスコアはバックボーンノードのランダムな再接続の結果得られる有意な数の検定NPAスコアに関して高いことを示している。

30

#### 【0109】

いくつかの実装では、両方の  $p$  値（図11および12で計算された）は、提案されたNPAスコアが統計的に有意であるとみなされるためには低い値である必要がありうる。他の実装では、ネットワークスコア化エンジン114は、提案されたNPAスコアが有意であるとわかるために1つまたは複数の  $p$  値が低いことを要求しうる。

40

#### 【0110】

図13は、リーディングバックボーンおよび遺伝子ノードを識別するための例示的なプロセス1300の流れ図である。ステップ1302で、ネットワークスコア化エンジン114は、識別されたネットワークモデルに基づきバックボーン演算子（backbone operator）を生成する。バックボーン演算子は、サポーターノードの活性尺度のベクトルに作用し、バックボーン演算子は、バックボーンノードに対する活性値のベクトルを出力する。いくつかの実装における好適なバックボーン演算子は、上の式13で定義されている演算子  $K$  である。

50

## 【0111】

ステップ1304で、ネットワークスコア化エンジン114は、ステップ1302で生成されたバックボーン演算子を使用してリーディングバックボーンノードのリストを生成する。リーディングバックボーンノードは、処置および対照データならびに生物学的因果ネットワークモデルの解析時に識別される最も有意なバックボーンノードを表すものとしてよい。このリスト生成するために、ネットワークスコア化エンジン114は、バックボーン演算子を使用して、その後バックボーンノードに対する活性値のベクトルとそれ自体との内積で使用されうるカーネルを形成することができる。いくつかの実装では、ネットワークスコア化エンジン114は、リーディングバックボーンノードのリストを生成するのに、そのような内積の結果である和の中の項を降順に並べ替えるステップと、その和への最大の寄与因子に対応するノードの固定された数、または総和の指定されたパーセンテージ（例えば、60%）を達成するために必要な最も有意に寄与するノードの数のいずれかを選択するステップとを実行する。それと同等のことであるが、ネットワークスコア化エンジン114は、リーディングバックボーンノードのリストを生成するのに、式1の順序付けられた項の累積合計を計算することによってNPAスコアの80%を占めるバックボーンノードを含めるステップを実行することができる。上に記載されているように、この累積合計は、以下の内積（バックボーン演算子Kを使用する）

10

## 【化16】

$$f_1^T K^T K f_1. \quad (18)$$

20

の項の累積合計として計算されうる。したがって、リーディングノードの識別は、活性尺度およびネットワークポロジの両方に依存する。

## 【0112】

ステップ1306で、ネットワークスコア化エンジン114は、ステップ1302で生成されたバックボーン演算子を使用してリーディング遺伝子ノードのリストを生成する。式2で示されているように、NPAスコアは、倍率変化の二次形式として表されうる。そこで、いくつかの実装では、リーディング遺伝子リストは、以下のスカラー積

$$(f_1 | L_2 (L_3^{-1})^T L_3^{-1} L_2^T f_1). \quad (19)$$

30

の順序付けられた和の項を識別することによって生成される。リーディング遺伝子リストの両端は、NPAスコアに負方向に寄与する遺伝子も生物学的有意性を有するので重要な場合がある。

## 【0113】

いくつかの実装では、ネットワークスコア化エンジン114は、ステップ1306でそれぞれの遺伝子について構造的重要度値も生成する。構造的重要度値は、実験データとは無関係であり、いくつかの遺伝子がモデルにおける遺伝子の位置によりバックボーンノードの値を推論するうえで他の遺伝子より重要である可能性があるという事実を表す。構造的な重要度は、遺伝子jについて

## 【化18】

$$l_j = \sum_{i=1}^N |(L_3^{-1} L_2^T)_{ij}| \quad (20)$$

40

によって定義されうる。

## 【0114】

リーディングバックボーンノードリスト内の生物学的実体およびリーディング遺伝子ノードリスト内の遺伝子は、（対照条件に対する）処置条件による基礎をなすネットワークの活性化のバイオマーカーに対する候補である。これら2つのリストは、将来の研究に対するターゲットを識別するために別々に、もしくは一緒に使用されうるか、または以下に記載されているように、他のバイオマーカー識別プロセスで使用されうる。

## 【0115】

50

次に図7を参照すると、いくつかの実装では、ネットワークスコア化エンジン114は、ステップ704での第1の活性ベクトルを、以下のラプラシアン行列

【化19】

$$L_{l^2(V \setminus V_0)} = (\text{diag}(\text{out}_{l^2(V \setminus V_0)}) + \text{diag}(\text{in}_{l^2(V \setminus V_0)}) - (A + A^T))_{l^2(V \setminus V_0)} \in l^2(V \setminus V_0) \quad (21)$$

のカーネルおよび像空間に基づきそれぞれ寄与しない成分と寄与する成分とに分解するが、そこでは、計算ネットワークモデルは、図5のステップ506を参照しつつ上に記載されているように生物学的実体の第2の組における生物学的実体に対応するノードに制限されている。ネットワークスコア化エンジン114は、以下に記載されているように、さらに分析するため「符号付き」拡散カーネルを式21のラプラシアンの行列指数として計算し、第1の活性値ベクトルをスペクトル成分上に射影して少なくとも1つの寄与する成分を生成するようにさらに構成されうる。

10

【0116】

ステップ706で、ネットワークスコア化エンジン114は、第1の寄与するベクトル（ステップ704で決定される）を異なる実験からの活性値の第2の組から決定された第2の寄与するベクトルと比較する。この第2の寄与するベクトルを決定するために、ノードの第1の組に対して異なる処置および対照データを使用してステップ702および704を繰り返すことができる（図5により）。いくつかの実施形態では、第2の寄与するベクトルを決定するために、同じ処置および/または対照データが使用されうる。第2の寄与するベクトルは、異なる実験に対するNPAスコアに寄与する異なる処置（および適宜異なる対照データ）を伴う異なる実験から導出される活性値の成分を表す。両方の実験において目的の生物系は、同じであるので、基礎をなす計算ネットワークモデルは同じであり、したがって、第2の寄与しないベクトルおよび寄与するベクトルは、行列積

20

【化20】

$$(L_3^{-1}L_2^T)$$

のカーネルおよび行列積

【化21】

$$(L_3^{-1}L_2^T)$$

30

の像空間にそれぞれ依存する。

【0117】

ステップ708で、ネットワークスコア化エンジン114は、ステップ706の比較結果に基づく比較可能性情報を提供する。いくつかの実装では、比較可能性情報は、第1の寄与するベクトルと第2の寄与するベクトルとの間の相関である。いくつかの実装では、比較可能性情報は、第1の寄与するベクトルと第2の寄与するベクトルとの間の距離である。ステップ708で比較可能性情報を提供するために、ベクトルを比較するための多くの技術のうちどれでも使用することができる。

【0118】

図5のステップ504で計算された活性尺度および図5のステップ506で生成された活性値（例えば、図6のプロセス600による）は、2つの異なる生物系が同じ作用物質または処置条件による攪乱に類似の応答を示す程度を反映する翻訳可能性情報を提供するために使用されうる。一例では、2つの異なる生物系は、in vitro系、in vivo系、マウス系、ラット系、ヒト以外の霊長類系、およびヒト系の任意の組み合わせであってよい。図8は、翻訳可能性情報を提供するための例示的なプロセス800の流れ図である。プロセス800は、図5のステップ506でノードの第2の組について活性値を生成した後に、例えば、ネットワークスコア化エンジン114、またはシステム100の任意の他の適切に構成された1つまたは複数のコンポーネントによって実行されうる。ステップ802で、ネットワークスコア化エンジン114は、第1の生物系における実体について活性値の第1の組を決定し、ステップ804で、ネットワークスコア化エンジン

40

50

114は、第2の生物系における実体について活性値の第2の組を決定する。第1の生物系および第2の生物系のそれぞれは、対応する第1の計算ネットワークモデルおよび第2の計算ネットワークモデルによって表される。活性値は、例えば、図5のステップ506または図6のプロセス600により決定されうる。

【0119】

ステップ806で、ネットワークスコア化エンジン114は、ステップ802で決定された活性値の第1の組をステップ804で決定された活性値の第2の組と比較する。いくつかの実装では、ネットワークスコア化エンジン114は、第1の生物系 ( $V^{(1)}$ ) に対する第1の活性値と第2の生物系 ( $V^{(2)}$ ) に対する第2の活性値との間の以下の関係

10

【化22】

$$\begin{array}{ccc} l^2(V_0^{(1)}) & \xrightarrow{h_1} & l^2(V_0^{(2)}) \\ (L_3^{(1)})^{-1}(L_2^{(1)})^T \downarrow & \circlearrowleft & \downarrow (L_3^{(2)})^{-1}(L_2^{(2)})^T \\ l^2(V^{(1)} \setminus V_0^{(1)}) & \xrightarrow{h_2} & l^2(V^{(2)} \setminus V_0^{(2)}), \end{array} \quad (22)$$

を解析するように構成されるが、ただし、 $h_1$  および  $h_2$  は、それぞれ、活性尺度レベルにおける第1の生物系と第2の生物系との間のマッピング（例えば、第1の生物系における実験に対する処置および対照データから第2の生物系における実験に対する処置および対照データへのマッピング）、および推論された活性値レベルにおける第1の生物系と第2の生物系との間のマッピング（例えば、第1の生物系に対する推論された活性値から第2の生物系に対する推論された活性値へのマッピング）を表す。これらのマッピングは未知である可能性があるが、ネットワークスコア化エンジン114は、活性尺度レベルおよび推論された活性値レベルで比較を実行することによってこれらのマッピングに関する情報を決定するように構成されうる。例えば、いくつかの実装では、ネットワークスコア化エンジン114は、各行列積

20

【化23】

$$(L_3^{(i)})^{-1} (L_2^{(i)})^T$$

30

の像空間内に射影されるか、または関連する行列（式21を参照しつつ上に記載されているラプラシアン行列など）のスペクトル成分上に射影される活性値の間の相関を計算するように構成される。いくつかの実装では、ネットワークスコア化エンジン114は、カーネル正準相関分析（KCCA）法を適用することによって活性値の第1の組と第2の組とを比較することができ、その多くは当技術分野で周知である。

【0120】

ステップ808で、ネットワークスコア化エンジン114は、ステップ806における比較結果に基づく翻訳可能性情報を提供する。図7のステップ708で提供される比較可能性情報を参照しつつ上に記載されているように、ベクトルを比較するための多くの技術のうちのいずれかを使用して、ステップ808で比較可能性情報を提供することができる。例えば、いくつかの実装では、ネットワークスコア化エンジン114は、各行列積

40

【化24】

$$(L_3^{(i)})^{-1} (L_2^{(i)})^T$$

の像空間内に射影されるか、または関連する行列（式21を参照しつつ上に記載されているラプラシアン行列など）のスペクトル成分上に射影される活性値の間の相関を計算するように構成される。いくつかの実装では、ネットワークスコア化エンジン114は、カーネル正準相関分析（KCCA）法を適用することによって活性値の第1の組と第2の組とを比較し、翻訳可能性情報を提供することができ、その多くは当技術分野で周知である。

【0121】

50

図9は、活性値およびNPAスコアに対する信頼区間を計算するための例示的なプロセス900の流れ図である。ステップ902で、ネットワークスコア化エンジン114は、図5のステップ504を参照しつつ上に記載されているように活性化尺度（本明細書では表す）を計算する。いくつかの実装では、活性尺度は、Limma R統計分析パッケージによって、または別の標準的な統計的手法によって決定される倍率変化値または重み付け倍率変化値（例えば、関連する偽不発見率を使用して重み付けされる）であってよい。ステップ904で、ネットワークスコア化エンジン114は、ステップ902で計算された活性尺度（または重み付け活性尺度）に関連付けられている分散を計算する。いくつかの実装では、行列は、ステップ904で、 $\Sigma = \text{diag}(\text{var}(\cdot))$ として定義される。ステップ906で、ラプラシアン行列を生成するために、関連するネットワークの構造が使用される（例えば、式9を参照しつつ以下に記載されているように）。ネットワークは、重み付け、符号付き、および有向、またはこれらの組み合わせであるものとしてよい。ステップ908で、ネットワークスコア化エンジン114は、式12のラプラシアンの表現を、左辺がゼロに等しいと置いて解き、 $f_2$ （活性値のベクトル）を生成する。ステップ910で、ネットワークスコア化エンジン114は、活性値のベクトルの分散を計算する。いくつかの実装では、このベクトルは、

【化25】

$$\text{var}(f_2) = L_3 L_2^T \Sigma L_2 L_3^T, \quad (23)$$

により計算され、 $L_2$  および  $L_3$  は、式11において定義されているとおりである。ステップ912で、ネットワークスコア化エンジン114は、

【化26】

$$f_2(x) \pm z(1 - \frac{\alpha}{2}) \sqrt{\text{var}(f_2(x))}, \quad (24)$$

により  $f_2$  のそれぞれの入力の信頼区間を計算するが、ただし、

【化27】

$$z(1 - \frac{\alpha}{2})$$

は、関連する  $N(0, 1)$  分位数である（例えば、 $\alpha = 0.05$  ならば  $1.96$ ）。ステップ914で、ネットワークスコア化エンジン114は、ステップ916で使用される二次形式行列を計算してステップ916でNPAスコアを計算する。いくつかの実装では、二次形式行列は、上記の式3により計算される。ステップ916で、ネットワークスコア化エンジン114は、式2による二次形式行列  $Q$  を使用してNPAスコアを計算する。ステップ918で、ネットワークスコア化エンジン114は、ステップ916で計算されたNPAスコアの分散を計算する。いくつかの実装では、この分散は、

【化28】

$$\text{var}(NPA) = \text{var}(f_2^T Q f_2) = 2\text{tr}(Q \Sigma^2 Q \Sigma^2) + 4f_2^T Q \Sigma^2 Q f_2, \quad (25)$$

により計算されるが、ただし、 $\Sigma^2 = \text{var}(f_2)$  である。ステップ920で、ネットワークスコア化エンジン114は、ステップ916で計算されたNPAスコアの信頼区間を計算する。いくつかの実装では、信頼区間は、

【化29】

$$NPA \pm \sqrt{(\frac{1}{1-\alpha})} \sqrt{\text{var}(NPA)}. \quad (26)$$

または

【化30】

$$NPA \pm z(1 - \frac{\alpha}{2}) \sqrt{\text{var}(NPA)}. \quad (27)$$

により計算される。

【0122】

10

20

30

40

50

図14は、生物学的攪乱の影響を定量化する分散型のコンピュータ化されたシステム1400のブロック図である。システム1400のコンポーネントは、図1のシステム100内のものと似ているが、システム100の配置構成は、それぞれのコンポーネントがネットワークインターフェース1410を通じて通信するような構成をとる。そのような実装は、「クラウドコンピューティング」パラダイムなどの共通ネットワークリソースへのアクセスを共有することができるワイヤレス通信システムを含む複数の通信システム上での分散コンピューティングに適している可能性がある。

#### 【0123】

図15は、本明細書において説明されているプロセスを実行するための図1のシステム100または図11のシステム1100のコンポーネントのうちのいずれかなどのコンピューティングデバイスのブロック図である。システム応答プロファイルエンジン110、ネットワークモデリングエンジン112、ネットワークスコア化エンジン114、集約エンジン116、ならびに転帰データベース、攪乱データベース、および文献データベースを含むデータベースのうちの一つまたは複数を用意する、システム100のコンポーネントのそれぞれは、一つまたは複数のコンピューティングデバイス1500上に実装される。いくつかの態様において、複数の上記のコンポーネントおよびデータベースは、一つのコンピューティングデバイス1500内に収めることができる。いくつかの実装では、複数のコンピューティングデバイス1500にまたがって一つのコンポーネントおよび一つのデータベースを実装することができる。

10

#### 【0124】

上記コンピューティングデバイス1500は、少なくとも一つの通信インターフェースユニット、入力/出力コントローラ1510、システムメモリー、および一つまたは複数のデータ記憶デバイスを備える。上記システムメモリーは、少なくとも一つのランダムアクセスメモリー(RAM1502)および少なくとも一つのリードオンリーメモリー(ROM1504)を備える。これらの要素はすべて、中央処理装置(CPU1506)と通信して、該コンピューティングデバイス1500の動作を円滑に行わせる。上記コンピューティングデバイス1500は、多くの異なる方法で構成される。例えば、上記コンピューティングデバイス1500は、従来のスタンドアロン型コンピュータであってもよいが、代替的に、コンピューティングデバイス1500の機能を複数のコンピュータシステムおよびアーキテクチャにまたがって分散させることもできる。上記コンピューティングデバイス1500は、モデリング、スコア化、および集約演算の一部または全部を実行するように構成される。図15では、上記コンピューティングデバイス1500は、ネットワークまたはローカルネットワークを介して、他のサーバもしくはシステムにリンクされる。

20

30

#### 【0125】

上記コンピューティングデバイス1500は、分散型アーキテクチャで構成することができ、データベースおよびプロセッサは、別のユニットまたは場所に収納される。いくつかのこのようなユニットは、一次処理機能を実行し、最低限、汎用コントローラまたはプロセッサおよびシステムメモリーを含む。このような一態様では、これらのユニットのそれぞれは、通信インターフェースユニット1508を介して、他のサーバ、クライアントもしくはユーザコンピュータおよび他の関係するデバイスとの一次通信リンクとして働く通信ハブまたはポート(図示せず)に接続する。上記通信ハブまたはポートは、通信ルーターとしてもっぱら使用される、最小処理機能をそれ自体有することができる。さまざまな通信プロトコルが、システムの一部であってもよく、これは、限定はしないがイーサネット(登録商標)(Ethernet(登録商標))、SAP、SAS(商標)、ATP、BLUETOOTH(登録商標)、GSM(登録商標)、およびTCP/IPを含む。

40

#### 【0126】

上記CPU1506は、一つまたは複数の従来のマイクロプロセッサなどのプロセッサ、および該CPU1506の操作負荷をオフロードする数値演算コプロセッサ(math co-processor)などの一つまたは複数の補助コプロセッサを備える。上記

50

C P U 1 5 0 6 は、上記通信インターフェースユニット 1 5 0 8 および上記入力/出力コントローラ 1 5 1 0 と通信し、これを通じて該 C P U 1 5 0 6 は他のサーバ、ユーザ端末、またはデバイスなどの他のデバイスと通信する。上記通信インターフェースユニット 1 5 0 8 および上記入力/出力コントローラ 1 5 1 0 は、例えば、他のプロセッサ、サーバ、またはクライアント端末と同時通信するための複数の通信チャネルを備えることができる。互いに通信するデバイスであっても、互いにひっきりなしに送信している必要はない。それと反対に、そのようなデバイスは、必要に応じて互いに送信するだけでもよく、実際には大半の時間においてデータの交換を差し控えることができ、該デバイス間の通信リンクを確立するために実行するのにいくつかのステップを必要とするものとしてよい。

【 0 1 2 7 】

上記 C P U 1 5 0 6 は、上記データ記憶デバイスとも通信する。上記データ記憶デバイスとして、磁気メモリー、光メモリー、または半導体メモリーの適切な組み合わせを含み得、例えば、R A M 1 5 0 2、R O M 1 5 0 4、フラッシュドライブ、コンパクトディスクなどの光ディスク、またはハードディスクもしくはドライブが挙げられる。上記 C P U 1 5 0 6 および上記データ記憶デバイスはそれぞれ、例えば、単一のコンピュータまたは他のコンピューティングデバイス内に丸ごと配置されるか、または U S B ポート、シリアルポートケーブル、同軸ケーブル、イーサネット（登録商標）型ケーブル、電話回線、無線周波トランシーバ、または他の類似のワイヤレスもしくは有線媒体または上記のもの組み合わせなどの、通信媒体によって互いに接続されうる。例えば、上記 C P U 1 5 0 6 は、上記通信インターフェースユニット 1 5 0 8 を介して上記データ記憶デバイスに接続されうる。上記 C P U 1 5 0 6 は、1 つまたは複数の特定の処理機能を実行するように構成されうる。

【 0 1 2 8 】

上記データ記憶デバイスは、例えば、( i ) 上記コンピューティングデバイス 1 5 0 0 用のオペレーティングシステム 1 5 1 2、( i i ) 本明細書に記載されているシステムおよび方法により、また特に上記 C P U 1 5 0 6 に関して詳しく記載されているプロセスにより、該 C P U 1 5 0 6 に指示するように適合された 1 つまたは複数のアプリケーション 1 5 1 4 (例えば、コンピュータプログラムコードまたはコンピュータプログラム製品)、または ( i i i ) 上記プログラムが必要とする情報を記憶するために利用されうる情報を記憶するように適合されたデータベース ( 1 つまたは複数 ) 1 5 1 6 を記憶することができる。いくつかの態様では、上記データベース ( 1 つまたは複数 ) として、実験データを記憶するデータベース、および公開文献モデルが挙げられる。

【 0 1 2 9 】

上記オペレーティングシステム 1 5 1 2 およびアプリケーション 1 5 1 4 は、例えば、圧縮形式、非コンパイル形式、および暗号化形式で記憶され、コンピュータプログラムコードを含むことができる。上記プログラムの命令は、上記 R O M 1 5 0 4 または上記 R A M 1 5 0 2 などの、データ記憶デバイス以外のコンピュータ可読媒体から上記プロセッサのメインメモリーへと読み込むことができる。上記プログラムにおける命令のシーケンスの実行により上記 C P U 1 5 0 6 が本明細書に記載されているプロセスステップを実行するが、ハード配線回路を、本開示のプロセスの実装のためのソフトウェア命令の代わりに、または該ソフトウェア命令と組み合わせて使用することができる。したがって、記載されているシステムおよび方法は、ハードウェアとソフトウェアとの特定の組み合わせに限定されない。

【 0 1 3 0 】

本明細書に記載されているようなモデリング、スコア化、および集約に関して 1 つまたは複数の機能を実行するのに適したコンピュータプログラムコードが提供されうる。上記プログラムは、オペレーティングシステム 1 5 1 2、データベース管理システム、および上記プロセッサが上記入力/出力コントローラ 1 5 1 0 を介してコンピュータ周辺デバイス (例えば、ビデオディスプレイ、キーボード、コンピュータマウスなど) とインターフェースすることを可能にする「デバイスドライバ」などのプログラム要素を含むことがで

10

20

30

40

50

きる。

【0131】

本明細書で使用されているような「コンピュータ可読媒体」という用語は、実行のため命令を上記コンピューティングデバイス1500（または本明細書に記載されているデバイスの任意の他のプロセッサ）のプロセッサに与えるか、または与えることに関わる任意の非一時的媒体を指す。このような媒体は、限定はしないが、不揮発性媒体および揮発性媒体を含む、多くの形態をとりうる。不揮発性媒体としては、例えば、光ディスク、磁気ディスク、もしくは光磁気ディスク、またはフラッシュメモリなどの集積回路メモリが挙げられる。揮発性媒体としては、典型的にはメインメモリを構成するダイナミックランダムアクセスメモリ（DRAM）を含む。コンピュータ可読媒体の一般的な形態としては、例えば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、CD-ROM、DVD、任意の他の光媒体、パンチカード、紙テープ、穴の形状を有する任意の他の物理的媒体、RAM、PROM、EPROMまたはEEPROM（電子的に消去可能なプログラム可能リードオンリーメモリ）、FLASH-EEPROM、任意の他のメモリチップもしくはカートリッジ、またはコンピュータが読み取ることができる任意の他の非一時的媒体が挙げられる。

10

【0132】

さまざまな形態のコンピュータ可読媒体が、実行のため1つまたは複数の命令の1つまたは複数のシーケンスを上記CPU1506（または本明細書に記載されているデバイスの任意の他のプロセッサ）に伝えることに関与しうる。例えば、上記命令は、最初に、リモートコンピュータ（図示せず）の磁気ディスクで伝えることができる。上記リモートコンピュータは、命令をそのリモートコンピュータのダイナミックメモリにロードし、モデムを使用してイーサネット（登録商標）接続、ケーブル線、さらには電話回線を介して該命令を送ることができる。コンピューティングデバイス1500（例えば、サーバ）に対してローカルの通信デバイスは、各通信回線上でデータを受け取り、該データを上記プロセッサのシステムバスに出すことができる。上記システムバスは、データをメインメモリに伝え、上記プロセッサはそのメインメモリから命令を取り出して実行する。メインメモリに入った命令は、必要に応じて、上記プロセッサによる実行前または実行後にメモリに記憶することができる。それに加えて、命令は、通信ポートを介して、電気信号、電磁気信号、または光信号として受け取ることができ、これらはさまざまな種類の情報を伝えるワイヤレス通信またはデータストリームの形態の例である。

20

30

【0133】

本開示の実装は、特定の例を参照しつつ特に示され記載されているが、当業者であれば、添付の特許請求の範囲に定められているような本開示の範囲から逸脱することなく本開示において形態および細部にさまざまな変更を加えることができることを理解するべきである。ゆえに、本開示の範囲は、添付の特許請求の範囲によって示され、したがって、該特許請求の範囲の等価性の意味および範囲内にあるすべての変更は、包含されることが意図されている。

【0134】

本明細書に記載されているシステムおよび方法は、よく理解されている細胞培養実験を使用して試験されている。正常ヒト気管支上皮（NHBE）細胞は、G1期で細胞を停止させるCDK4/6阻害剤（CDKI）である、PD-0332991に曝して処理された。処理された細胞は、次いで、培地からCDKIを取り除き、洗浄することによって細胞周期への再入が可能となった。細胞周期の再入は、CDKIが取り除かれ、細胞が洗浄されてから2、4、6、および8時間後のS期で細胞を蛍光標識することによって実験的に確認された。CDKIを取り除いてから2、4、6、および8時間後に試料採取された細胞の遺伝子転写プロファイルを得た。培地中でCDKIに連続的に曝露された細胞のプロファイルも得られた。CDKIが取り除かれたときに差次的に活性化された生物学的過程および機構を識別するために、さまざまな時点において得られた洗浄細胞の遺伝子転写プロファイルを使用してネットワーク攪乱振幅スコアが計算された。CDKIの取り除き

40

50

に関連する攪乱に対するNPAスコアの計算のため、127個のノードと240個のエッジを備える細胞周期サブネットワークが使用された。これは、Schlagelらの論文(参照によりその全体が本明細書に組み込まれている、2011年、「A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue」BMC Syst Biol. 10月19日、5巻、168頁)で公開されている細胞増殖ネットワークモデルのサブネットワークである。

【0135】

NPAスコア(図18)は、S期の細胞の数の対応する増加を示す蛍光活性化セルソーティング(FACS)分析(図17)の結果と一致する2時間の時点から8時間の時点までの時点の範囲にわたって増加することが判明した。NPAスコアは、P値<0.05において上に記載されているように2つの並べかえ検定を受けさせ、統計量(「O」および「K」統計量)は両方とも、実験のNHBE細胞におけるこの特定の生物系、つまり、細胞周期が、実際に攪乱されることを示した。この分析では、S期への進行に関わることが知られている重要な機構に正確に対応する細胞周期ネットワークモデルにおけるリーディングノードも識別した。E2Fタンパク質は、p53およびCHEK1の(間接的)制御の下でCdkによって順にリン酸化されるRbPとの複合体を形成する。また、Cdkと併せて、G1/S-サイクリンは、予想通り、リーディングノードの過程(leading nodes processes)の一部である。この方法によって識別されるリーディングノードは、taof(TFDP1)、taof(E2F2)、CHEK1、TFDP1、kaof(CHEK1)、taof(E2F3)、taof(E2F1)、taof(RB1)、有糸分裂細胞周期のG1/S移行、CDC2、E2F2、CCNA2、CCNE1、THAP1、CDKN1A、TP53 P@S20、E2F3、kaof(CDK2)である。Taofは、「transcriptional activity of」(の転写活性)の略語であり、kaofは、「kinase activity of」(のキナーゼ活性)の略語である。TP53 P@S20は、TP53の位置20におけるセリンがリン酸化されていることに対する略語である。その結果は、遺伝子発現データと、因果ネットワークモデルに具体化される生物系の知識を活用する機構駆動アプローチ(mechanism-driven approach)との組み合わせを使用して生物系の攪乱を定量することができることを示す。

【0136】

本発明は、以下の番号を振られている段落においてさらに定義される。

【0137】

生物系の攪乱を定量化するためのコンピュータ化された方法であって、

第1のプロセッサで、第1の処置に対する生物学的実体の第1の組の応答に対応する処置データの第1の組を受け取るステップであって、第1の生物系は、生物学的実体の第1の組と生物学的実体の第2の組とを含む生物学的実体を備え、第1の生物系内のそれぞれの生物学的実体は、第1の生物系内の生物学的実体のうちの少なくとも1つの別のものと相互作用する、ステップと、

第2のプロセッサで、第1の処置と異なる第2の処置に対する生物学的実体の第1の組の応答に対応する処置データの第2の組を受け取るステップと、

第3のプロセッサで、第1の生物系を表す第1の計算因果ネットワークモデルを提供するステップであって、該モデルは、

生物学的実体の第1の組を表すノードの第1の組と、

生物学的実体の第2の組を表すノードの第2の組と、

ノードを接続し、生物学的実体の間の関係を表すエッジと、

第1の処置データと第2の処置データとの間の変化の予想される方向を表す、方向値とを含む、ステップと、

第4のプロセッサを使って、ノードの第1の組における対応するノードに対する第1の処置データと第2の処置データとの間の差を表す活性尺度の第1の組を計算するステップ

と、

第5のプロセッサを使って、第1の計算因果ネットワークモデルと活性尺度の第1の組とに基づき、ノードの第2の組における対応するノードに対する活性値の第2の組を生成するステップとを含む、コンピュータ化された方法。

【0138】

第6のプロセッサを使って、第1の計算因果ネットワークモデルと活性値の第2の組とに基づき、第1および第2の処置への第1の生物系の攪乱を表す第1の計算因果ネットワークモデルに対するスコアを生成するステップをさらに含む、段落137に記載の方法。

【0139】

段落137に記載の方法であって、活性値の第2の組を生成するステップは、ノードの第2の組におけるそれぞれの特定のノードについて、特定のノードの活性値と、該特定のノードが第1の計算因果ネットワークモデル内のエッジと接続されるノードの活性値または活性尺度との間の差を表す差のステートメント ( *d i f f e r e n c e s t a t e m e n t* ) を最小化する活性値を識別するステップを含み、差のステートメントは、ノードの第2の組におけるそれぞれのノードの活性値に依存する、方法。

10

【0140】

差のステートメントは、ノードの第2の組におけるそれぞれのノードの方向値にさらに依存する、段落139に記載の方法。

【0141】

活性値の第2の組におけるそれぞれの活性値は、活性尺度の第1の組の活性尺度の一次結合である、段落137に記載の方法。

20

【0142】

一次結合は、ノードの第1の組におけるノードと第1の計算因果ネットワークモデル内のノードの第2の組におけるノードとの間のエッジに依存し、かつ、第1の計算因果ネットワークモデル内のノードの第2の組におけるノード同士の間のエッジにも依存する、段落141に記載の方法。

【0143】

一次結合は、第1の計算因果ネットワークモデル内のノードの第1の組におけるノード同士の間のエッジには依存しない、段落141に記載の方法。

【0144】

スコアは、活性値の第2の組に対する二次従属性を有する、138に記載の方法。

30

【0145】

活性尺度の第1の組のそれぞれの活性尺度に対する変動推定値の一次結合を形成することによって活性値の第2の組のそれぞれの活性値に対する変動推定値を提供するステップをさらに含む、段落137に記載の方法。

【0146】

スコアに対する変動推定値は、活性値の第2の組に対する二次従属性を有する、段落138に記載の方法。

【0147】

活性値の第2の組を第1の活性値ベクトルとして表すステップと、第1の活性値ベクトルを第1の寄与するベクトルと第1の寄与しないベクトルとに分解するステップであって、第1の寄与するベクトルと第1の寄与しないベクトルとの和は、第1の活性値ベクトルとなる、ステップとをさらに含む、段落138に記載の方法。

40

【0148】

スコアは、第1の寄与しないベクトルに依存しない、段落147に記載の方法。

【0149】

スコアは、活性値の第2の組の二次関数として計算され、第1の寄与しないベクトルは、二次関数のカーネル内にある、段落148に記載の方法。

【0150】

第1の寄与しないベクトルは、第1の計算因果ネットワークモデルと関連付けられてい

50

る符号付きラプラシアンに基づく二次関数のカーネル内にある、段落 1 4 7 に記載の方法。

【 0 1 5 1 】

第 1 のプロセッサで、第 3 の処置に対する生物学的実体の第 1 の組の応答に対応する処置データの第 3 の組を受け取るステップと、

第 2 のプロセッサで、第 4 の処置に対する生物学的実体の第 1 の組の応答に対応する処置データの第 4 の組を受け取るステップと、

第 4 のプロセッサを使って、ノードの第 1 の組に対応する活性尺度の第 3 の組を計算するステップであって、活性尺度の第 3 の組におけるそれぞれの活性尺度はノードの第 1 の組における対応するノードに対する処置データの第 3 の組と処置データの第 4 の組との間の差を表す、ステップと、

第 5 のプロセッサを使って、活性値の第 4 の組を生成するステップであって、それぞれの活性値は第 1 の計算因果ネットワークモデルと活性尺度の第 3 の組とに基づきノードの第 2 の組における対応するノードに対する活性値を表す、ステップと、

活性値の第 4 の組を第 2 の活性値ベクトルとして表すステップと、

第 2 の活性値ベクトルを第 2 の寄与するベクトルと第 2 の寄与しないベクトルとに分解するステップであって、第 2 の寄与するベクトルと第 2 の寄与しないベクトルとの和が第 2 の活性値ベクトルとなる、ステップと、

第 1 の寄与するベクトルと第 2 の寄与するベクトルを比較するステップとをさらに含む、段落 1 4 7 に記載の方法。

【 0 1 5 2 】

第 1 と第 2 の寄与するベクトルを比較するステップは、第 1 と第 2 の寄与するベクトルの間の相関を計算して、処置データの第 1 および第 3 の組の比較可能性を示すステップを含む、段落 1 5 1 に記載の方法。

【 0 1 5 3 】

第 1 と第 2 の寄与するベクトルを比較するステップは、第 1 および第 2 の寄与するベクトルを計算ネットワークモデルの符号付きラプラシアンの像空間上に射影するステップを含む、段落 1 5 1 に記載の方法。

【 0 1 5 4 】

処置データの第 2 の組は、処置データの第 4 の組と同じ情報を含む、段落 1 5 1 に記載の方法。

【 0 1 5 5 】

前記第 1 のプロセッサで、第 1 の処置と異なる第 3 の処置に対する生物学的実体の第 3 の組の応答に対応する処置データの第 3 の組を受け取るステップであって、第 2 の生物系は、生物学的実体の第 3 の組と生物学的実体の第 4 の組とを含む複数の生物学的実体を備え、第 2 の生物系内のそれぞれの生物学的実体は、第 2 の生物系内の生物学的実体のうちの少なくとも 1 つの別のものと相互作用する、ステップと、

第 2 のプロセッサで、第 3 の処置と異なる第 4 の処置に対する生物学的実体の第 3 の組の応答に対応する処置データの第 4 の組を受け取るステップと、

第 3 のプロセッサで、第 2 の生物系を表す第 2 の計算因果ネットワークモデルを提供するステップであって、このモデルは、

生物学的実体の第 3 の組を表すノードの第 3 の組と、

生物学的実体の第 4 の組を表すノードの第 4 の組と、

ノードを接続し、生物学的実体の間の関係を表すエッジと、

第 3 の処置データと第 4 の処置データとの間の変化の予想される方向を表す、方向値とを含む、ステップと、

第 4 のプロセッサを使って、ノードの第 3 の組に対応する活性尺度の第 3 の組を計算するステップであって、活性尺度の第 3 の組におけるそれぞれの活性尺度は、ノードの第 3 の組における対応するノードに対する処置データの第 3 の組と処置データの第 4 の組との間の差を表す、ステップと、

10

20

30

40

50

第5のプロセッサを使って、活性値の第4の組を生成するステップであって、それぞれの活性値は、第2の計算因果ネットワークモデルと活性尺度の第3の組とに基づきノードの第4の組における対応するノードに対する活性値を表す、ステップと、

活性値の第4の組を活性値の第2の組と比較するステップとをさらに含む、段落137に記載の方法。

【0156】

活性値の第4の組を活性値の第2の組と比較するステップは、第1の計算因果ネットワークモデルに関連付けられている符号付きラプラシアンと第2の計算因果ネットワークモデルに関連付けられている符号付きラプラシアンとに基づきカーネル正準相関分析を適用するステップを含む、段落155に記載の方法。

10

【0157】

活性尺度は、倍率変化値であり、それぞれのノードに対する倍率変化値は、各ノードによって表される生物学的実体に対する処置データの対応する組の間の差の対数を含む、上記段落137～156のいずれかに記載のコンピュータ化された方法。

【0158】

生物系は、細胞増殖機構、細胞ストレス機構、細胞炎症機構、およびDNA修復機構のうち少なくとも1つを含む、上記段落137～157のいずれかに記載のコンピュータ化された方法。

【0159】

第1の処置は、タバコを加熱することによって発生したエアロゾルへの曝露、タバコを燃焼させることによって発生したエアロゾルへの曝露、タバコの煙への曝露、および紙巻きタバコの煙への曝露、のうちの少なくとも1つを含む、上記段落137～158のいずれかに記載のコンピュータ化された方法。

20

【0160】

第1の処置は、生物系に存在も由来もしない分子または実体を含む、異物への曝露を含む、上記段落137～159のいずれかに記載のコンピュータ化された方法。

【0161】

第1の処置は、毒素、治療化合物、刺激物、弛緩物質、天然物、製造物および食物への曝露を含む、上記段落137～160のいずれかに記載のコンピュータ化された方法。

【0162】

第1の生物系および第2の生物系は、*in vitro*系、*in vivo*系、マウス系、ラット系、ヒト以外の霊長類系、およびヒト系からなる群の2つの異なる要素である、段落155および段落156のいずれかに記載のコンピュータ化された方法。

30

【0163】

第1の処置データは、作用物質に曝露された第1の生物系に対応し、第2の処置データは、作用物質に曝露されない第1の生物系に対応する、段落137に記載のコンピュータ化された方法。

【0164】

生物系の攪乱を示すスコアの統計的有意性を判定するステップをさらに含む、段落138に記載のコンピュータ化された方法。

40

【0165】

スコアの統計的有意性は、それぞれ複数のランダムに生成した試験の計算因果ネットワークモデルから計算される複数の検定スコアに対してスコアを比較することによって判定される、段落164に記載のコンピュータ化された方法。

【0166】

ランダム生成検定計算因果ネットワークモデルは、第1の計算因果ネットワークモデルの1つまたは複数の態様をランダムに類別することによって生成される、段落165に記載のコンピュータ化された方法。

【0167】

第1の計算因果ネットワークモデルの1つまたは複数の態様は、ノードの第1の組のラ

50

ベル、ノードの第2の組をノードの第1の組に接続するエッジ、またはノードの第2の組同士を接続するエッジを備える、段落166に記載のコンピュータ化された方法。

【 図 1 】

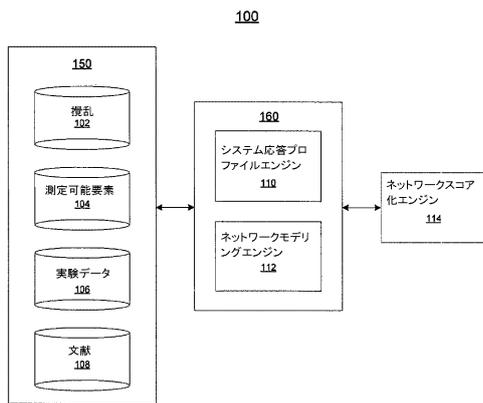


FIG. 1

【 図 2 】

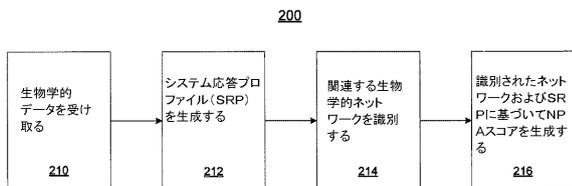


FIG. 2

【 図 3 】

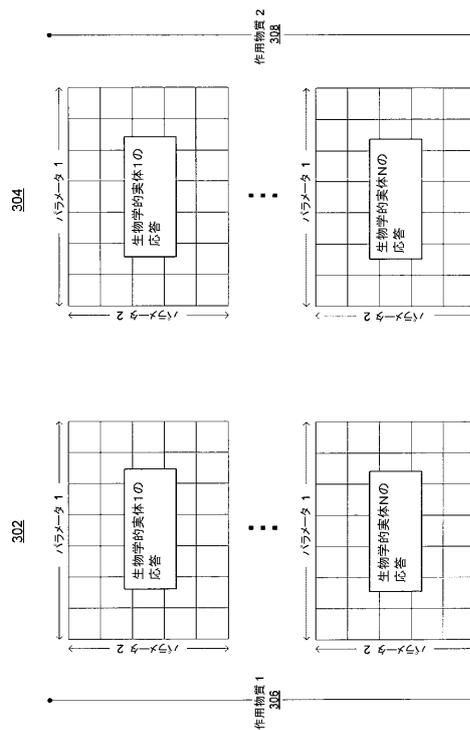


FIG. 3

【 図 4 】

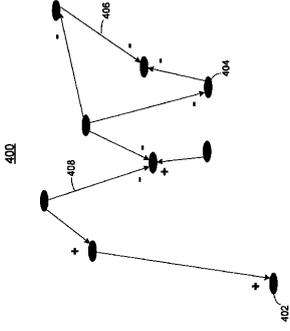


FIG. 4

【 図 6 】

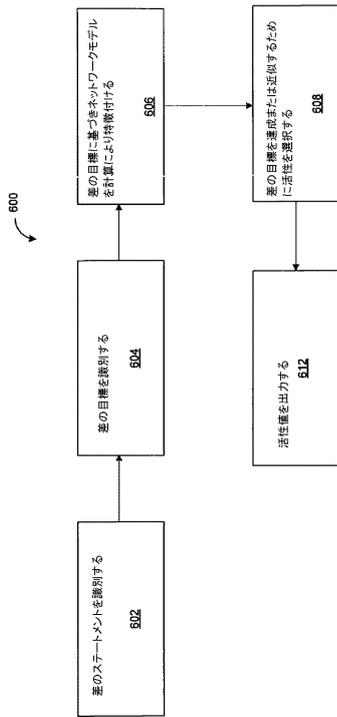


FIG. 6

【 図 5 】

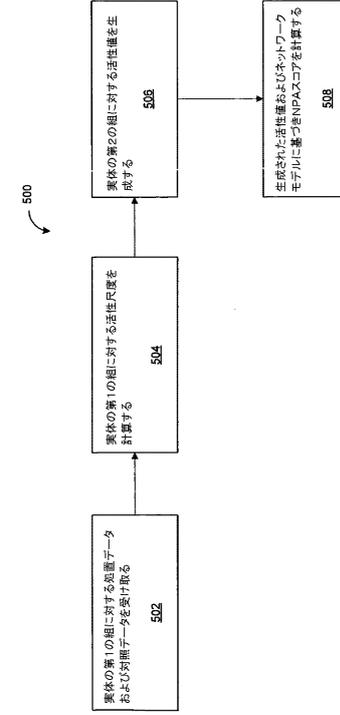


FIG. 5

【 図 7 】

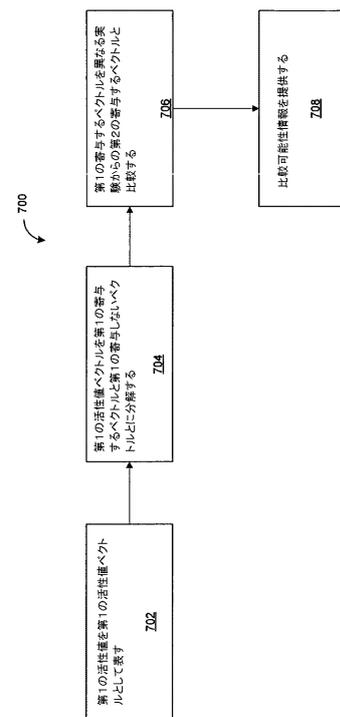


FIG. 7



【 図 1 2 】

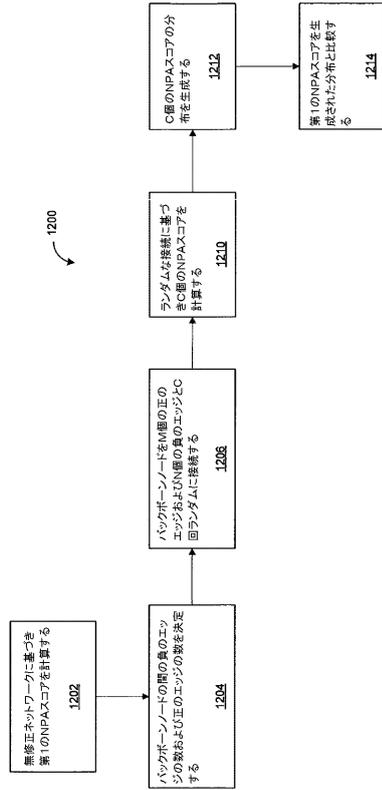


FIG. 12

【 図 1 3 】

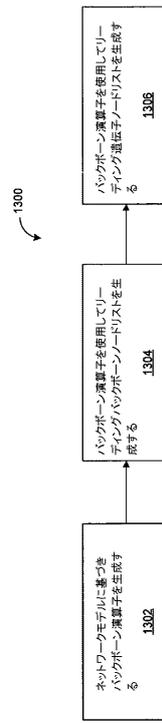


FIG. 13

【 図 1 4 】

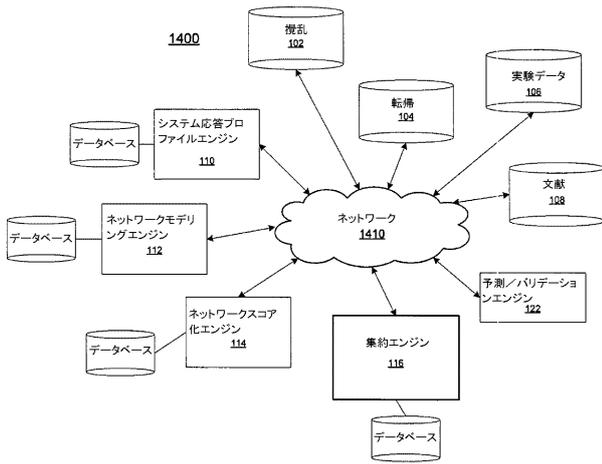


FIG. 14

【 図 1 5 】

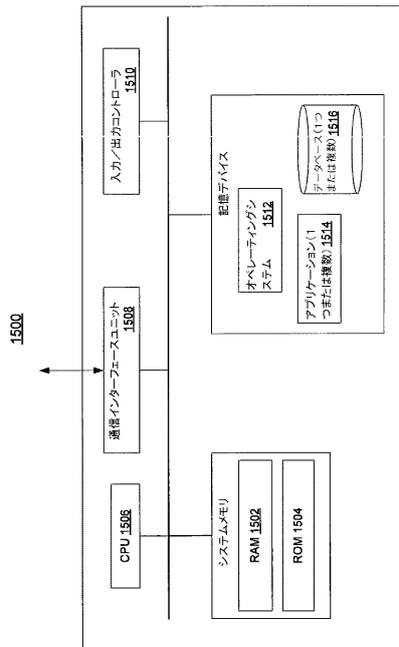
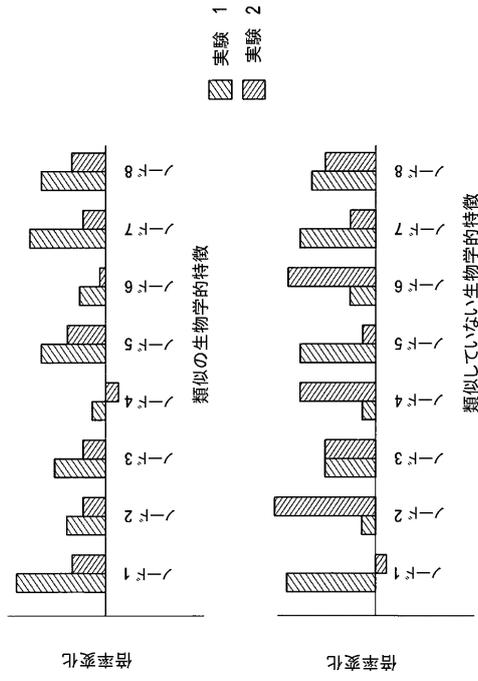


FIG. 15

【 図 1 6 】



【 図 1 7 】

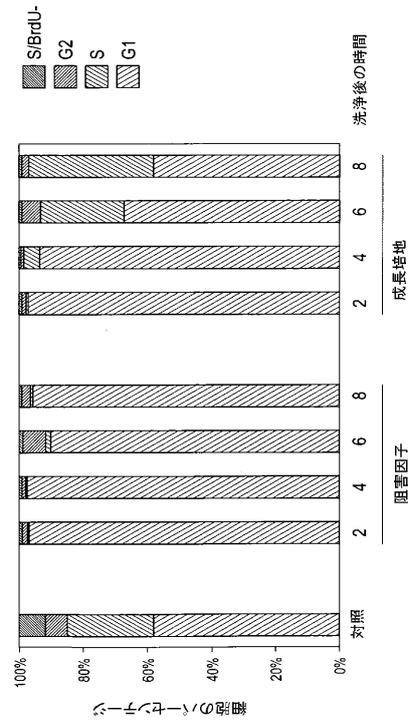


FIG. 16

FIG. 17

【 図 1 8 】

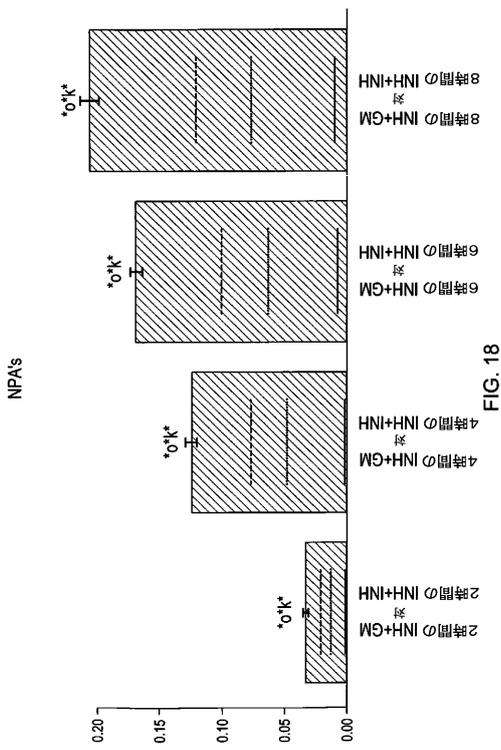


FIG. 18

## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/EP2012/003760**Box No. II Observations where certain claims were found unsearchable (Continuation of Item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2.  Claims Nos.: **1-17(partially)**  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
**see FURTHER INFORMATION sheet PCT/ISA/210**
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of Item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2012/003760

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>	
INV. G06F19/00 G06F19/12 ADD.	
According to International Patent Classification (IPC) or to both national classification and IPC	
<b>B. FIELDS SEARCHED</b>	
Minimum documentation searched (classification system followed by classification symbols) G06F	
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched	
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data	
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>	
Category*	Citation of document, with indication, where appropriate, of the relevant passages
	Relevant to claim No.
X	ANONYMOUS: "Reverse Causal Reasoning Methods - White Paper", INTERNET CITATION, 4 February 2011 (2011-02-04), pages 1-26, XP002681944, Retrieved from the Internet: URL:http://www.selventa.com/attachments/white_papers/reverse-causal-reasoning.pdf [retrieved on 2012-08-17] p. 3, 19-13 p. 3, paragraph 3 p. 4, RCR Overview p.6, paragraph 2 p. 8, paragraph 3 p.8-9 HYP Evaluation Statistics; figure 6 ----- -/--
	1-17
<input checked="" type="checkbox"/>	Further documents are listed in the continuation of Box C.
<input checked="" type="checkbox"/>	See patent family annex.
* Special categories of cited documents :	
*A* document defining the general state of the art which is not considered to be of particular relevance	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*E* earlier application or patent but published on or after the international filing date	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*O* document referring to an oral disclosure, use, exhibition or other means	*Z* document member of the same patent family
*P* document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search	Date of mailing of the international search report
18 July 2013	25/07/2013
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Schmitt, Constanze

2

Form PCT/ISA/210 (second sheet) (April 2005)

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2012/003760

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	STEPHEN J. HAGGARTY ET AL: "Chemical Genomic Profiling of Biological Networks Using Graph Theory and Combinations of Small Molecule Perturbations", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 125, no. 35, 1 September 2003 (2003-09-01), pages 10543-10545, XP055071504, ISSN: 0002-7863, DOI: 10.1021/ja035413p.p.10544, left-hand col., paragraph 2 -----	1-17
A	Y. YAMANISHI ET AL: "Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis", BIOINFORMATICS, vol. 19, no. Suppl 1, 3 July 2003 (2003-07-03), pages i323-i330, XP055071582, ISSN: 1367-4803, DOI: 10.1093/bioinformatics/btg1045 abstract -----	1-17
X,P	FLORIAN MARTIN ET AL: "Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks", BMC SYSTEMS BIOLOGY, BIOMED CENTRAL LTD, LO, vol. 6, no. 1, 31 May 2012 (2012-05-31), page 54, XP021116941, ISSN: 1752-0509, DOI: 10.1186/1752-0509-6-54 the whole document -----	1-17
X,P	JULIA HOENG ET AL: "A network-based approach to quantifying the impact of biologically active substances", DRUG DISCOVERY TODAY, vol. 17, no. 9-10, 1 May 2012 (2012-05-01), pages 413-418, XP055071052, ISSN: 1359-6446, DOI: 10.1016/j.drudis.2011.11.008 the whole document -----	1-17
A	LEE EUNJUNG ET AL: "Analysis of AML genes in dysregulated molecular networks", BMC BIOINFORMATICS, BIOMED CENTRAL, LONDON, GB, vol. 10, no. Suppl 9, 17 September 2009 (2009-09-17), page S2, XP021055797, ISSN: 1471-2105, DOI: 10.1186/1471-2105-10-S9-S2 the whole document -----	1-17
	----- -/--	

2

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2012/003760

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TEGNER ET AL: "Perturbations to uncover gene networks", TRENDS IN GENETICS, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, vol. 23, no. 1, 23 December 2006 (2006-12-23), pages 34-41, XP005734738, ISSN: 0168-9525, DOI: 10.1016/J.TIG.2006.11.003 the whole document -----	1-17
A	F. R.K. Chung: "Chapter 1: Eigenvalues and the Laplacian of a graph" In: "Spectral Graph Theory", 1 January 1997 (1997-01-01), American Mathematical Society, Providence, XP055071554, pages 1-22, the whole document -----	1-17
A	US 2006/177827 A1 (DEJORI MATHAUS [DE] ET AL) 10 August 2006 (2006-08-10) the whole document -----	1-17

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No

PCT/EP2012/003760

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006177827 A1	10-08-2006	US 2006177827 A1	10-08-2006
		WO 2005003368 A2	13-01-2005
-----			

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

Continuation of Box II.2

Claims Nos.: 1-17(partially)

Reasons for the limitation of the search:

1.1 The claimed subject-matter is unclear (Article 6 PCT) to such an extent that no meaningful opinion application with regard to novelty, inventive step or industrial applicability can presently be given. The major reasons are the following: 1.2 Claim 1 : It is unclear what the relationship between the "first set of biological entities" and the "second set of biological entities" is. Furthermore, it is unclear how the "first set of nodes" and the "second set of nodes" are related. It is not clear what is meant by "corresponding nodes". How is this correspondence defined? The term "second set of activity values" is unclear, as there is only this one set defined. Moreover, it is unclear how the "activity measures" and "activity values" are obtained: How are "activity values" calculated "based on the first computational network model" and "the first set of activity measures"? Furthermore, it is not clear how "a difference" between sets of treatment data is computed (this also applies to claim 8). It is not clear which entities are connected by edges in "edges connecting nodes and representing relationships between the biological entities". It is not clear what "direction values" are and how they are computed. It is also not clear how "providing ... a first computational causal network model" is defined in terms of technical features. How is this causal network obtained? These objections also apply to the terms as used in claims 8 and 11. 1.2 Claim 2 : It is not clear how the score for the first computational causal network model is generated "based on the first computational causal network model and the second set of activity values". 1.3 Claim 3 : It is unclear what is meant by "a difference statement" and how this "difference statement depends on the activity values of each node in the second set of nodes". Furthermore, it is unclear how a difference is computed between "the activity value of the particular node and the activity value or activity measure of nodes to which the particular node is connected", i.e., between one measure and a set of measures. 1.4 Claim 5 : It is unclear what is meant by "a variation estimate", as the term has no established meaning in the art. Furthermore, it is unclear how "a variation estimate for each activity value of the second set of activity values" is computed by "forming a linear combination of variation estimates" and how variation of a single value ("each activity value") can be computed as such. How is this variation estimate used in combination with the features of claim 1? 1.5 Claim 6 : The terms "contributing vector" and "non-contributing vector" are unclear. To what are they contributing? The phrase "decomposing the first activity value vector into a first contributing vector and a first non-contributing vector, such that the sum of the first contributing and non-contributing vectors is the first activity value vector" describes a result to be achieved, which merely amounts to a statement of the underlying problem, without providing the technical features necessary for achieving this result. This reasoning also applies to these terms as used in claim 8. 1.6 Claim 7 : The term "a quadratic function based on a signed Laplacian" is vague and unclear, as it is not clear how the quadratic function is defined. 1.7 In so far as the present application can be understood, it seems to be directed

International Application No. PCT/ EP2012/ 003760

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

to a method for determining the degree of perturbation of a biological system using different data sets describing different perturbations on the biological system. The method seems to map the perturbation data onto a set of nodes of a precomputed causal biological network with edge "direction values" (denoting the regulatory relationship between nodes) and then to determine "activity measures", e.g. fold changes (claim 13), for each node in the set of nodes according to its activity in the different data sets. Networks seem to be scored w.r.t. "activity values", which depend on the "activity measures" and denote an inferred activity of a second node set of the causal network. The activity values seem to be represented as a vector and split into two linear components according to their "contribution" (i.e. a "contributing" and a "non-contributing" vector) to the "activity value" (claim 6), using matrix decomposition of the signed Laplacian associated with the causal network (claim 7). The method seems to be applied to a second biological system, further sets of treatment data and a second precomputed causal biological network (claim 11) and the resulting "contributing" vector compared to the one obtained on the first biological system, e.g. via kernel canonical correlation analysis (claim 12).

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.

## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC

(72)発明者 マルティン , フロリアン  
スイス国 ツェーハー - 2 0 3 4 プスー , シュマン ドゥ ロレー 1