

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6335260号  
(P6335260)

(45) 発行日 平成30年5月30日(2018.5.30)

(24) 登録日 平成30年5月11日(2018.5.11)

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

請求項の数 23 外国語出願 (全 52 頁)

(21) 出願番号	特願2016-240283 (P2016-240283)	(73) 特許権者	500586875
(22) 出願日	平成28年12月12日(2016.12.12)		フィリップ モリス プロダクツ エス
(62) 分割の表示	特願2014-514108 (P2014-514108)		アー
原出願日	平成24年6月11日(2012.6.11)		スイス国 2000 ヌーシャテル ケ
(65) 公開番号	特開2017-73161 (P2017-73161A)	(74) 代理人	100078282
(43) 公開日	平成29年4月13日(2017.4.13)		弁理士 山本 秀策
審査請求日	平成28年12月12日(2016.12.12)	(74) 代理人	100113413
(31) 優先権主張番号	61/495,824		弁理士 森下 夏樹
(32) 優先日	平成23年6月10日(2011.6.10)	(74) 代理人	100181674
(33) 優先権主張国	米国 (US)		弁理士 飯田 貴敏
(31) 優先権主張番号	61/525,700	(74) 代理人	100181641
(32) 優先日	平成23年8月19日(2011.8.19)		弁理士 石川 大輔
(33) 優先権主張国	米国 (US)	(74) 代理人	230113332
			弁護士 山本 健策

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) 【特許請求の範囲】

【請求項1】

作用物質に応答した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、前記方法は、

プロセッサにおいて、処置データである第1の処置データとコントロールデータである第2の処置データとを取得することと、

前記生物系における生物学的実体を表すノードとノード間の関係を表すエッジとを含むネットワークモデルを識別することであって、前記ノードは、前記第1の処置データと前記第2の処置データとの間の変化の予測される方向を表す方向値に関連付けられている、ことと、

前記ノードに対して、前記第1の処置データと前記第2の処置データとの間の差異を表す活性尺度を取得することと、

前記方向値および前記活性尺度に少なくとも部分的に基づいて、前記作用物質に対する前記生物系の攪乱を表す前記ネットワークモデルについてのスコアを生成することとを含む、方法。

【請求項2】

前記第1の処置データは、前記作用物質に対する前記生物系の前記応答に対応し、前記第2の処置データは、前記作用物質に対して曝露されない前記生物系に対応する、請求項1に記載の方法。

【請求項3】

前記活性尺度は、倍率変化値を含み、それぞれのノードに対する各それぞれの倍率変化は、前記それぞれのノードによって表される前記生物学的実体に対する前記第 1 の処置データと前記第 2 の処置データとの間の差異の対数を含む、請求項 1 に記載の方法。

【請求項 4】

前記活性尺度は、倍率変化値を含み、前記方法は、前記倍率変化がゼロ倍率変化の基礎となる帰無仮説からの逸脱を表す確率に少なくとも部分的に基づいて重み値を決定することをさらに含み、前記スコアは、前記重み値に少なくとも部分的に基づいて生成される、請求項 1 に記載の方法。

【請求項 5】

前記スコアを生成することは、  
前記活性尺度に重み値および前記方向値を掛けることにより、一組の重み付けられた活性尺度を取得すること  
 を含む、請求項 1 に記載の方法。

10

【請求項 6】

前記スコアを生成することは、  
 前記重み付けられた活性尺度を足し合わせることにより、合計を取得することと、  
 前記ネットワークモデルにおいて表される前記生物学的実体の数に基づくファクターによって前記合計を正規化すること  
 をさらに含み、請求項 5 に記載の方法。

【請求項 7】

前記スコアの符号は、プロセスが上方調節されているか、または、下方調節されているかを示す、請求項 6 に記載の方法。

20

【請求項 8】

前記スコアを生成することは、  
 前記活性尺度が前記方向値に整合している度合を表す正の活性化距離を取得することと、  
 前記活性尺度が前記方向値に整合していない度合を表す負の活性化距離を取得すること  
 を含む、請求項 1 に記載の方法。

【請求項 9】

前記スコアは、前記活性尺度が与えられるときに前記ネットワークモデルによって表される 1 以上の生物学的機構が活性化される確率を定量化したものである、請求項 8 に記載の方法。

30

【請求項 10】

前記方向値は、前記方向値のうち前記活性尺度に整合している値である正の方向値と、前記方向値のうち前記活性尺度に整合していない値である負の方向値とを含み、  
前記方向値は、前記正の方向値に関連付けられた、前記ネットワークモデルにおける前記生物学的実体を第 1 のサブセットの生物学的実体として識別し、  
前記方向値は、前記負の方向値に関連付けられた、前記ネットワークモデルにおける前記生物学的実体を第 2 のサブセットの生物学的実体として識別する、請求項 8 に記載の方法。

40

【請求項 11】

前記スコアを生成することは、  
前記第 1 のサブセットの生物学的実体に対する前記活性尺度に重み値および前記正の方向値を掛けることにより、第 1 の組の重み付けられた活性尺度を取得することと、  
前記第 2 のサブセットの生物学的実体に対する前記活性尺度に重み値および前記負の方向値を掛けることにより、第 2 の組の重み付けられた活性尺度を取得することと  
 を含む、請求項 10 に記載の方法。

【請求項 12】

前記正の活性化距離は、前記第 1 の組の重み付けられた活性尺度の合計に少なくとも部

50

分的に基づいて生成され、

前記負の活性化距離は、前記第2の組の重み付けられた活性尺度の合計に少なくとも部分的に基づいて生成される、請求項11に記載の方法。

【請求項13】

前記方向値は、前記方向値のうち前記活性尺度に整合している値である正の方向値と、前記方向値のうち前記活性尺度に整合していない値である負の方向値とを含み、前記スコアは、前記正の方向値を伴う前記ノードに対する第1の和および前記負の方向値を伴う前記ノードに対する第2の和に少なくとも部分的に基づいて生成される、請求項1に記載の方法。

【請求項14】

前記正の方向値を伴う前記ノードは、その活性が前記第2の処置に比較して前記第1の処置に应答して増加すると予想される生物学的実体に対応し、前記負の方向値を伴う前記ノードは、その活性が前記第2の処置に比較して前記第1の処置に应答して減少すると予想される生物学的実体に対応する、請求項13に記載の方法。

【請求項15】

前記正の方向値を伴う前記ノードに対して、

前記活性尺度を増加する順番に順序付けることにより、第1の順序付けられた組の活性尺度を取得することと、

前記第1の順序付けられた組の活性尺度における互いに隣接する活性尺度の間のそれぞれの差異を計算することと

をさらに含み、

前記第1の和は、前記それぞれの差異および前記正の方向値に少なくとも部分的に基づいて生成される、請求項14に記載の方法。

【請求項16】

前記負の方向値を伴う前記ノードに対して、

前記活性尺度を増加する順番に順序付けることにより、第2の順序付けられた組の活性尺度を取得することと、

前記第2の順序付けられた組の活性尺度における互いに隣接する活性尺度の間のそれぞれの差異を計算することと

をさらに含み、

前記第2の和は、前記それぞれの差異および前記負の方向値に少なくとも部分的に基づいて生成される、請求項15に記載の方法。

【請求項17】

前記第1の和は、前記第1の順序付けられた組の活性尺度におけるそれぞれの活性尺度と前記それぞれの活性尺度に隣接する他の活性尺度との間の差異によって乗算されたそれぞれの活性尺度に対応する値に対するものであり、

前記第2の和は、前記第2の順序付けられた組の活性尺度におけるそれぞれの活性尺度と前記それぞれの活性尺度に隣接する他の活性尺度との間の差異によって乗算されたそれぞれの活性尺度に対応する値に対するものである、請求項16に記載の方法。

【請求項18】

前記第1の順序付けられた組の活性尺度における第1の活性尺度は、前記第1の順序付けられた組の活性尺度における第2の活性尺度よりも低く、前記スコアは、前記第2の活性尺度が前記第1の活性尺度よりも前記スコアに対して大きな効果を有するように生成される、請求項15に記載の方法。

【請求項19】

前記スコアについての信頼区間を生成することをさらに含む、請求項1に記載の方法。

【請求項20】

前記信頼区間を生成することは、前記活性尺度に適用される一組の係数が活性尺度の関数であるか否かを決定することを含む、請求項19に記載の方法。

【請求項21】

10

20

30

40

50

前記信頼区間を生成することは、前記活性尺度の分散に基づいて前記スコアの分散を評価することをさらに含む、請求項20に記載の方法。

【請求項22】

生物系に対する攪乱の影響を決定するためのコンピュータシステムであって、前記コンピュータシステムは、請求項1～21のいずれか一項に記載の方法を実行するように構成されたプロセッサを含む、コンピュータシステム。

【請求項23】

プログラムコードが記録されたコンピュータ読み取り可能な記録媒体であって、前記プログラムコードは、請求項1～21のいずれか一項に記載の方法を実行することをコンピュータに行わせるためのものである、コンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【背景技術】

【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るかどうかに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。

【0002】

この10年間のうちに、従来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文献からの生物学的経路に関する知識と組み合わせることで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シグネチャ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シグネチャ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シグネチャを識別し、スコア化することを伴った。これらの表現型派生シグネチャは、著しい分類能力を備えるが、単一の特定の攪乱とシグネチャとの間の機械的関係または因果関係を欠いている。したがって、これらのシグネチャは、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。

【0004】

10

20

30

40

50

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。遺伝子などの個々の実体は複数の生物学的過程に関わり得るので（例えば、炎症および細胞増殖）、該遺伝子の活性を測定することは、その活性を誘導する基礎をなす（*underlying*）生物学的過程を識別するには不十分である。現在の技法のどれも、生物学的実体の活性に関与する基礎をなす機構をマイクロスケールで識別するために適用されたこともなければ、潜在的に有害な作用物質および実験条件に応答した、これらの実体が役割を果たす様々な生物学的機構の活性化の定量的評価を行うために適用されたこともない。したがって、生物学的機構を考慮して系全体にわたる生物学的データを分析し、その系が作用物質または環境の変化に応答するときの生物系の変化を定量化するためのシステムおよび方法の改善が必要とされている。

10

【発明の概要】

【課題を解決するための手段】

【0005】

要旨

一態様では、本明細書に記載のシステムおよび方法は、作用物質に応答した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、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。

【0006】

上記コンピュータ化された方法は、一態様では、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、ここで、該生物系が複数の生物学的実体を含む（*include*）かまたは備え（*comprise*）、それぞれの生物学的実体が、該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用するステップと、第2のプロセッサで、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るステップと、第3のプロセッサで、該生物系を表すと共に、該生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および該コントロールデータと該処置データの間の変化の予測方向を該ノードについて表す方向値（*direction value*）を含むかまたは備える、計算因果関係ネットワークモデルを提供するステップと、第4のプロセッサにより、該処置データと該コントロールデータの間の差異を表す活性尺度を該ノードについて計算するステップと、第5のプロセッサにより、少なくとも1つの重み値が少なくとも1つの他の重み値と異なる複数の重み値を該ノードについて計算するステップと、第6のプロセッサにより、該作用物質に対する該生物系の攪乱を表す計算モデルについてのスコアを生成するステップであって、ここで、該スコアは該方向値、該重み値および該活性尺度に基づくステップとを含む。上記生物系は、少なくとも1つの機構仮説によって表すことができる。上記生物系は、複数の計算因果関係ネットワークモデルによって、または複数の機構仮説を備える少なくとも1つの計算因果関係ネットワークモデルによって表すことができる。この方法はさらに、それぞれの上記計算モデルにおける測定可能なノードの数に基づいて上記スコアを正規化するステップを含み得る。

20

30

【0007】

上記重み値により、処置データおよびコントロールデータの上記組の少なくとも1つにおける信頼度を表すことができる。上記重み値は、局所偽非発見率（*local false non-discovery rate*）を含むかまたは備えてよい。この方法はさらに、第7のプロセッサにより、ノードの活性尺度の近似分布をモデルまたはモデルにおける機構仮説に関して計算するステップと、第8のプロセッサにより、該近似分布に関する活性尺度の予測値を計算するステップと、第9のプロセッサにより、該作用物質に対する生物系のサブセットの攪乱を表す計算モデルごとに、予測値に基づくスコアを生成するステップとを含み得る。上記近似分布は上記活性尺度に基づいてよい。いくつかの実装では、予測値を計算するステップは長方形近似を実施するステップを含み得る。この方法はさらに、第10のプロセッサにより、上記モデルに関して上記活性尺度と上記方向値の間

40

50

の一致性および不一致性をそれぞれ表す正の活性化距離および負の活性化距離を該活性尺度に基づいて計算するステップと、第11のプロセッサにより、上記作用物質に対する上記生物系のサブセットの攪乱を表す計算モデルごとに、正および負の活性化スコアに基づくスコアを生成するステップとを含み得る。上記正の活性化距離、負の活性化距離、または両方は、局所偽非発見率に基づいてよい。上記活性尺度は倍率変化値でよく、ノードごとの倍率変化値は、それぞれの該ノードによって表された上記生物学的実体についての上記処置データと上記コントロールデータとの間の差異の対数を含むかまたは備える。上記生物系のサブセットは、細胞増殖機構、細胞性ストレス機構、細胞炎症機構、およびDNA修復機構のうちの少なくとも1つを含むかまたは備えてよい。上記作用物質は、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、または紙巻きタバコの煙のうちの少なくとも1つを含むかまたは備え得る。上記作用物質は、上記生物系に存在しないし由来もしない分子もしくは実体が含まれる異物を含むかまたは備え得る。上記作用物質は、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、および食物を含むかまたは備え得る。各測定可能ノードが第1の確率分布によって定義される複数の倍率変化値と、第2の確率分布によって定義される複数の重み値とを含むかまたは備えるように、処置データの組は、処置データの複数の組を含むかまたは備えてよい。各測定可能ノードが複数の倍率変化値およびその対応する重み値を含むかまたは備えるように、処置データの組は、処置データの複数の組を含むかまたは備えてよい。上記スコアを生成する上記ステップは、上記活性尺度、上記重み値および上記方向値の線形結合または非線形結合と、スケール係数による該結合の正規化とを含み得る。上記結合は算術結合でよく、該スケール係数は、その測定データが受け取られる生物学的実体の数の平方根である。上記スコアは、幾何学的攪乱指標スコア化技法、確率論的攪乱指標スコア化技法、または予測攪乱指標スコア化技法によって生成することができる。この方法はさらに、上記スコアについての信頼区間をパラメトリックまたは非パラメトリック計算ブートストラッピング技法に基づいて決定するステップを含み得る。

【0008】

別の態様でまた、作用物質に応答した生物系の攪乱を定量化するためのコンピュータシステムが記載される。このシステムは、少なくとも1つのプロセッサを備え、該プロセッサは、作用物質に対する生物系の応答に対応する処置データの組を受け取るように構成または適合され、ここで、該生物系は、複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体は、該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用し、該プロセッサは、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るように構成または適合され、該プロセッサは、該生物系を表す計算因果関係ネットワークモデルを提供するように構成または適合され、かつ、該生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および該コントロールデータと該処置データの間の変化の予測方向を該ノードについて表す方向値を含むかまたは備え、該プロセッサは、該処置データと該コントロールデータとの間の差異を表す活性尺度を該ノードについて計算するように構成または適合され、該プロセッサは、少なくとも1つの重み値が少なくとも1つの他の重み値と異なる複数の重み値を該ノードについて計算するように構成または適合され、かつ該プロセッサは、該作用物質に対する該生物系の攪乱を表す該計算モデルについてのスコアであって、該スコアは該方向値、該重み値および該活性尺度に基づくスコアを生成するように構成または適合される。上記生物系は、少なくとも1つの機構仮説によって表すことができる。上記生物系は、複数の計算因果関係ネットワークモデルによって、または複数の機構仮説を備える少なくとも1つの計算因果関係ネットワークモデルによって表すことができる。このコンピュータシステムはさらに、それぞれの上記計算モデルにおけるスコア化可能な(scorable)ノードの数に基づいて上記スコアを正規化するステップを含み得る。上記重み値により、処置データおよびコントロールデータの上記組の少なくとも1つにおける信頼度を表すことができる。上記重み値は、局所偽非発見率を含むかまたは備えてよい。いくつかの実装では、このコンピュータシステムはさらに、ノードの活性尺度の近似分布をモデルまたはモデルにおける機構仮説に関し

10

20

30

40

50

て計算するステップと、第8のプロセッサにより、該近似分布に関する活性尺度の予測値を計算するステップと、上記作用物質に対する上記生物系のサブセットの攪乱を表す計算モデルごとに、予測値に基づくスコアを生成するステップとを含む。上記近似分布は上記活性尺度に基づいてよい。上記コンピュータシステムのいくつかの実装では、コンピュータシステムはさらに、予測値を計算するステップを含み得、長方形近似を実施するステップを含む。このシステムはさらに、上記モデルに関して上記活性尺度と上記方向値の間の一貫性および不貫性をそれぞれ表す上記正の活性化距離および負の活性化距離を該活性尺度に基づいて計算するステップと、上記作用物質に対する上記生物系のサブセットの攪乱を表す計算モデルごとに、正および負の活性化スコアに基づくスコアを生成するステップとを含み得る。上記正の活性化距離、負の活性化距離、または両方は、局所偽非発見率に基づいてよい。上記活性尺度は倍率変化値でよく、ノードごとの該倍率変化値は、それぞれの該ノードによって表された上記生物学的実体についての上記処置データと上記コントロールデータとの間の差異の対数を含むかまたは備えてよい。上記生物系のサブセットは、細胞増殖機構、細胞性ストレス機構、細胞炎症機構、およびDNA修復機構のうち少なくとも1つを含むかまたは備えてよい。上記作用物質は、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、または紙巻きタバコの煙のうち少なくとも1つを含むかまたは備え得る。上記作用物質は、上記生物系に存在しないし由来もしない分子もしくは実体が含まれる異物を含むかまたは備え得る。上記作用物質は、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、および食物を含むかまたは備え得る。各測定可能ノードが第1の確率分布によって定義される複数の倍率変化値と、第2の確率分布によって定義される複数の重み値とを含むかまたは備えるように、処置データの組は、処置データの複数の組を含むかまたは備えてよい。各測定可能ノードが複数の倍率変化値およびその対応する重み値を含むかまたは備えるように、処置データの組は、処置データの複数の組を含むかまたは備えてよい。上記スコアを生成する上記ステップは、上記活性尺度、上記重み値および上記方向値の線形結合または非線形結合と、スケール係数による結合の正規化とを含み得る。上記結合は算術結合でよく、上記スケール係数は、その測定データが受け取られる生物学的実体の数の平方根である。上記スコアは、幾何学的攪乱指標スコア化技法、確率論的攪乱指標スコア化技法、または予測攪乱指標スコア化技法によって生成することができる。このシステムはさらに、上記スコアについての信頼区間をパラメトリックまたは非パラメトリック計算ブートストラッピング技法に基づいて決定するステップを含み得る。

#### 【0009】

いくつかの態様では、上記コンピュータ化された方法は、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、該生物系が複数の生物学的実体を含み、それぞれの生物学的実体が、該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用するステップと、第2のプロセッサで、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るステップとを含み得る。上記コンピュータ化された方法は、第3のプロセッサで、上記生物系を表す計算因果関係ネットワークモデルを提供するステップを含み得る。この計算モデルは、生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および上記コントロールデータと上記処置データの間の変化の予測方向を該ノードについて表す方向値を含むかまたは備えることができる。上記コンピュータ化された方法はさらに、第4のプロセッサにより、上記処置データと上記コントロールデータとの間の差異を表す活性尺度を上記ノードについて計算するステップと、第5のプロセッサにより、少なくとも1つの重み値が少なくとも1つの他の重み値と異なる複数の重み値を上記ノードについて計算するステップとを含み得る。上記コンピュータ化された方法はまた、第6のプロセッサにより、上記作用物質に対する上記生物系の攪乱を表す上記計算モデルについてのスコアを生成するステップであって、該スコアは上記方向値、上記重み値および上記活性尺度に基づくステップを含むこともできる。いくつかの実装では、上記コンピュータ化された方法はさらに、それぞれの上記計算モデルにおけるノードの数に基づいて上記スコアを正規化するステップを含

10

20

30

40

50

む。いくつかの実装では、上記第1から第6のプロセッサのそれぞれが、単一のプロセッサまたは単一の計算デバイスの中に含まれるか、または備えられる。他の実装では、上記第1から第6のプロセッサのうちの1つまたは複数、複数のプロセッサまたは計算デバイスにまたがって分布している。

**【0010】**

いくつかの実装では、上記計算因果関係ネットワークモデルは、潜在的原因を表すノードと測定量を表すノードとの間に存在する因果関係の組を含むかまたは備える。このような実装では、上記活性尺度は倍率変化を含み得る。上記倍率変化は、ノード測定値がコントロールデータと処置データの間で初期値から最終値までどれだけ変化するかを示す数としてよい。上記倍率変化数は、コントロール条件と処置条件の間における上記生物学的実体の活性の倍率変化の対数としてよい。各ノードについての活性尺度は、それぞれの該ノードによって表された上記生物学的実体についての上記処置データと上記コントロールデータとの間の差異の対数を含むかまたは備え得る。このような実装では、上記重み値により、上記ノードの倍率変化値に与えられるべき重みを表すことができる。上記重み値により、目的の特徴または結果に関して、測定された上記ノードの公知の生物学的有意性を表すことができる（例えば、がん研究における公知の発がん物質）。上記重み値により、攪乱データとコントロールデータの組の少なくとも1つにおける信頼度を表すことができる。より具体的には、上記重み値は、局所偽非発見率を含むかまたは備え得る。このような実装では、上記コンピュータ化された方法により上記計算モデルについてのスコアを生成することが、上記活性尺度に上記重み値および上記方向値を掛け、上記ノード全部について合計することによって可能である。いくつかの実装では、上記コンピュータ化された方法は、生成された上記スコアのそれぞれについての信頼区間をプロセッサにより生成するステップを含むかまたは備える。この信頼区間は、生成されたスコアの分布を近似することを含み得る。

**【0011】**

別の態様では、本明細書に記載のシステムおよび方法は、作用物質に応答した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を対象とする。このコンピュータ化された方法は、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、ここで、この生物系が複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体が該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用するステップと、第2のプロセッサで、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るステップとを含み得る。上記コンピュータ化された方法は、第3のプロセッサで、上記生物系を表す計算因果関係ネットワークモデルを提供するステップを含み得る。この計算モデルは、上記生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および上記コントロールデータと上記処置データの間の変化の予測方向を該ノードについて表す方向値を含むかまたは備えることができる。上記コンピュータ化された方法はさらに、第4のプロセッサにより、上記処置データと上記コントロールデータとの間の差異を表す活性尺度を上記ノードについて計算するステップと、第5のプロセッサにより、この活性尺度の近似分布を該ノードにわたって計算するステップとを含み得る。上記コンピュータ化された方法はまた、第6のプロセッサにより、この近似分布の予測値を計算するステップを含むかまたは備え得る。上記コンピュータ化された方法はまた、第7のプロセッサにより、上記作用物質に対する上記生物系のサブセットの攪乱を表す計算モデルごとに、上記予測値に基づくスコアを生成するステップを含み得る。いくつかの実装では、上記第1から第7のプロセッサのそれぞれが、単一のプロセッサまたは単一の計算デバイスの中に含まれるか、または備えられる。他の実装では、上記第1から第7のプロセッサのうちの1つまたは複数、複数のプロセッサまたは計算デバイスにまたがって分布している。

**【0012】**

いくつかの実装では、上記計算因果関係ネットワークモデルは、潜在的原因を表すノードと測定量を表すノードとの間に存在する因果関係の組を含むかまたは備える。このよう

10

20

30

40

50



な実装では、上記活性尺度は倍率変化を含むかまたは備え得る。上記倍率変化は、ノード測定値がコントロールデータと処置データの間で初期値から最終値までどれだけ変化するかを示す数としてよい。上記倍率変化数は、コントロール条件と処置条件の間における上記生物学的実体の活性の倍率変化の対数を表し得る。上記コンピュータ化された方法は、プロセッサにより、上記処置条件下の上記生物系で上記倍率変化値が取ることのできる値の組の近似値を表し得る該倍率変化密度についての値域を生成するステップを含むかまたは備え得る。上記プロセッサは、倍率変化値の近似確率分布を含むかまたは備え得る近似倍率変化密度 (approximate fold-change density) を生成することができる。このような実装では、上記コンピュータ化された方法はさらに、上記近似倍率変化密度の近似予測値を計算するステップを含むかまたは備え得る。上記コンピュータ化された方法により、計算された上記予測値に基づいて上記計算モデルについての上記スコアを生成することができる。

10

**【0013】**

いくつかの実装では、上記近似分布は一般に、上記活性尺度に基づいてよい。加えて、また必要に応じて、上記予測値は長方形近似値を含んでよい。いくつかの実装では、上記コンピュータ化された方法は、プロセッサにより、生成された上記スコアごとに信頼区間を生成するステップを含むかまたは備える。上記信頼区間を生成するステップは、パラメトリックブートストラッピング技法を実施するステップを含み得る。

**【0014】**

さらに別の態様では、本明細書に記載のシステムおよび方法は、作用物質に応答した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を対象とする。このコンピュータ化された方法は、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、ここで、この生物系が複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体が該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用するステップと、第2のプロセッサで、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るステップとを含み得る。上記コンピュータ化された方法は、第3のプロセッサで、上記生物系を表す計算因果関係ネットワークモデルを提供するステップを含み得る。この計算モデルは、上記生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および上記コントロールデータと上記処置データの間の変化の予測方向を該ノードについて表す方向値を含むかまたは備えることができる。上記コンピュータ化された方法はさらに、第4のプロセッサにより、上記処置データと上記コントロールデータとの差異を表す活性尺度を上記ノードについて計算するステップと、第5のプロセッサにより、上記活性尺度と上記方向値の間の一致性および不一致性をそれぞれ表す正の活性化スコアおよび負の活性化スコアを該活性尺度に基づいて計算するステップとを含むことができる。上記コンピュータ化された方法はまた、第6のプロセッサにより、上記作用物質に対する上記生物系のサブセットの攪乱を表す計算モデルごとに、上記正および負の活性化スコアに基づくスコアを生成するステップを含み得る。いくつかの実装では、上記第1から第6のプロセッサのそれぞれが、単一のプロセッサまたは単一の計算デバイスの中に含まれるか、または備えられる。他の実装では、上記第1から第6のプロセッサのうちの1つまたは複数が、複数のプロセッサまたは計算デバイスにまたがって分布している。

20

30

40

**【0015】**

いくつかの実装では、上記計算因果関係ネットワークモデルは、潜在的原因を表すノードと測定量を表すノードとの間に存在する因果関係の組を含むかまたは備える。このような実装では、上記活性尺度は倍率変化を含むかまたは備え得る。上記倍率変化は、ノード測定値がコントロールデータと処置データの間で初期値から最終値までどれだけ変化するかを示す数としてよい。その倍率変化数は、コントロール条件と処置条件の間における上記生物学的実体の活性の倍率変化の対数を表し得る。上記コンピュータ化された方法は、プロセッサにより、上記処置条件下の上記生物系でその倍率変化値が取ることのできる値の組の近似値を表し得る倍率変化密度の値域を生成するステップを含むかまたは備え得る

50

。上記コンピュータ化された方法は、プロセッサにより、上記倍率変化値および上記方向値に基づく正の活性化スコアを計算するステップを含み得る。上記正および負の活性化スコアは、生物学的実体の観測された活性化/阻害が、変化の予測方向と一致しているか、それとも一致していないかを示すことができる。一例では、上記正の活性化スコアは、上記方向値が上記活性尺度と一致している確率である。上記負の活性化スコアは、上記方向値が上記活性尺度と一致していない確率であり得る。上記コンピュータ化された方法はさらに、上記正の活性化スコアと負の活性化スコアを結合することによって上記計算モデルについてのスコアを生成するステップを含むかまたは備え得る。いくつかの実装では、このスコアは局所偽非発見率に基づく。

#### 【0016】

いくつかの実装では、上記生物系のサブセットは、細胞増殖機構、細胞性ストレス機構、細胞炎症機構、およびDNA修復機構のうち少なくとも1つを含むかまたは備える。上記作用物質は、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、または紙巻きタバコの煙のうち少なくとも1つを含むかまたは備え得る。上記作用物質は、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(NN)、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))を含み得る。いくつかの実装では、上記作用物質は、ニコチン置換療法に使用される生成物を含むかまたは備え得る。上記作用物質は、上記生物系に存在しないし由来もしない分子もしくは実体が含まれる異物を含むかまたは備え得る。上記作用物質はまた、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、および食物を含むかまたは備え得る。いくつかの実装では、処置データの組は、生物学ネットワークモデルのいくつかのノードに対応する処置データの複数の組を含むかまたは備え、ここで、それぞれのこのようなノードは、第1の確率分布によって定義される複数の倍率変化値と、第2の確率分布によって定義される複数の重み値とに対応する。

#### 【0017】

さらに別の態様では、本明細書に記載のシステムおよび方法は、作用物質に応答した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、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。このコンピュータ化された方法は、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、ここで、この生物系が複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体が該複数の生物学的実体のうち少なくとも1つの他の実体と相互作用するステップと、第2のプロセッサで、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るステップとを含み得る。上記コンピュータ化された方法は、第3のプロセッサで、上記生物系を表す計算因果関係ネットワークモデルを提供するステップを含み得る。この計算モデルは、上記生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および上記コントロールデータと上記処置データの間の変化の予測方向を該ノードについて表す方向値を含むかまたは備えることができる。上記コンピュータ化された方法はさらに、第4のプロセッサにより、上記処置データと上記コントロールデータの間の差異を表す活性尺度を上記ノードについて計算するステップを含み得る。上記コンピュータ化された方法はまた、第5のプロセッサにより、上記作用物質に対する上記生物系の攪乱を表す上記計算モデルについてのスコアを生成するステップであって、該スコアは上記方向値および上記活性尺度に基づくステップを含むこともできる。いくつかの実装では、上記コンピュータ化された方法はさらに、それぞれの上記計算モデルにおけるノードの数に基づいて上記スコアを正規化するステップを含む。上記コンピュータ化された方法はまた、第6のプロセッサにより、生成された上記スコアのそれぞれについての信頼区間を生成するステップを含み得る。この信頼区間は、生成された上記スコアの分布を近似することを含むことができ、生成されたスコアの近似分布の分散からt統計量を導き出すこと

10

20

30

40

50

ができる。いくつかの実装では、第1から第6のプロセッサのそれぞれが、単一のプロセッサまたは単一の計算デバイスの中に含まれるか、または備えられる。他の実装では、第1から第6のプロセッサのうちの1つまたは複数、複数のプロセッサまたは計算デバイスにまたがって分布している。

【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の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含むかまたは備え、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コア化エンジンを含むかまたは備える。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ンピュータ化されたシステムを含むかまたは備えうる。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。さらなる態様では、本明細書に記載の方法を実施するように適合されたプログラムコードを含むコンピュータプログラム製品が記載される。さらなる態様では、上記コンピュータプログラム製品を含むコンピュータまたはコンピュータ記録可能な媒体もしくはデバイスが記載される。

例えば、本願発明は以下の項目を提供する。

(項目1)

作用物質に応答した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、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、ここで、該生物系が複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体が、複数の該生物学的実体のうちの少なくとも1つの他の実体と相互作用するステップと、

第2のプロセッサで、該作用物質に曝露されない該生物系に対応するコントロールデータの組を受け取るステップと、

第3のプロセッサで、該生物系を表すと共に、

該生物学的実体を表すノード、

該生物学的実体間の関係を表すエッジ、および

該コントロールデータと該処置データの間の変化の予測方向を該ノードについて表す方向値

を含むかまたは備える、計算因果関係ネットワークモデルを提供するステップと、

第4のプロセッサにより、該処置データと該コントロールデータの間の差異を表す活性尺度を該ノードについて計算するステップと、

第5のプロセッサにより、少なくとも1つの重み値が少なくとも1つの他の重み値と異なる複数の重み値を該ノードについて計算するステップと、

第6のプロセッサにより、該作用物質に対する該生物系の攪乱を表す該計算モデルについてのスコアを生成するステップであって、該スコアが該方向値、該重み値および該活性尺度に基づくステップと

を含む、コンピュータ化された方法。

(項目2)

前記生物系が少なくとも1つの機構仮説によって表される、項目1に記載のコンピュータ化された方法。

(項目3)

前記生物系が複数の計算因果関係ネットワークモデルによって、または複数の機構仮説を備える少なくとも1つの計算因果関係ネットワークモデルによって表される、項目1または項目2に記載のコンピュータ化された方法。

(項目4)

それぞれの前記計算モデルにおける測定可能なノードの数に基づいて前記スコアを正規化するステップをさらに含む、前記項目のいずれかに記載のコンピュータ化された方法。

(項目5)

前記重み値が、処置データおよびコントロールデータの前記組の少なくとも1つにおける信頼度を表す、前記項目のいずれかに記載のコンピュータ化された方法。

(項目6)

前記重み値が局所偽非発見率を含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

(項目7)

第7のプロセッサにより、ノードの前記活性尺度の近似分布をモデルまたはモデルにおける機構仮説に関して計算するステップと、第8のプロセッサにより、該近似分布に関する活性尺度の予測値を計算するステップと、第9のプロセッサにより、前記作用物質に対する前記生物系のサブセットの攪乱を表す計算モデルごとに、予測値に基づくスコアを生成するステップとをさらに含む、前記項目のいずれかに記載のコンピュータ化された方法。

(項目8)

前記近似分布が前記活性尺度に基づく、項目7に記載のコンピュータ化された方法。

(項目9)

予測値を計算するステップが長方形近似を実施するステップを含む、項目7または項目8に記載のコンピュータ化された方法。

(項目10)

第10のプロセッサにより、前記モデルに関して前記活性尺度と前記方向値の間の一致性および不一致性をそれぞれ表す正の活性化距離および負の活性化距離を該活性尺度に基づいて計算するステップと、第11のプロセッサにより、前記作用物質に対する前記生物系のサブセットの攪乱を表す計算モデルごとに、正および負の活性化スコアに基づくスコアを生成するステップとをさらに含む、前記項目のいずれかに記載のコンピュータ化された方法。

(項目11)

前記正の活性化距離、負の活性化距離、または両方が局所偽非発見率に基づく、前記項目のいずれかに記載のコンピュータ化された方法。

(項目12)

前記活性尺度が倍率変化値であり、ノードごとの該倍率変化値が、それぞれの前記ノードによって表された前記生物学的実体についての前記処置データと前記コントロールデータとの間の差異の対数を含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

(項目13)

前記生物系のサブセットが、細胞増殖機構、細胞性ストレス機構、細胞炎症機構、およびDNA修復機構のうちの少なくとも1つを含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

(項目14)

前記作用物質が、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、または紙巻きタバコの煙のうちの少なくとも1つを含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された

10

20

30

40

50

方法。

(項目 15)

前記作用物質が、前記生物系に存在しないし由来もしない分子もしくは実体が含まれる異物を含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

(項目 16)

前記作用物質が、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、および食物を含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

(項目 17)

処置データの前記組が処置データの複数の組を含むかまたは備え、各測定可能ノードが第1の確率分布によって定義される複数の倍率変化値と、第2の確率分布によって定義される複数の重み値とを含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

10

(項目 18)

処置データの前記組が処置データの複数の組を含むかまたは備え、各測定可能ノードが複数の倍率変化値およびその対応する重み値を含むかまたは備える、前記項目のいずれかに記載のコンピュータ化された方法。

(項目 19)

前記スコアを生成する前記ステップが、前記活性尺度、前記重み値および前記方向値の線形結合または非線形結合と、スケール係数による該結合の正規化とを含む、前記項目のいずれかに記載のコンピュータ化された方法。

20

(項目 20)

前記結合が算術結合であり、前記スケール係数が、その測定データが受け取られる生物学的実体の数の平方根である、項目 19 に記載のコンピュータ化された方法。

(項目 21)

前記スコアが、幾何学的攪乱指標スコア化技法、確率論的攪乱指標スコア化技法、または予測攪乱指標スコア化技法によって生成する、前記項目のいずれかに記載のコンピュータ化された方法。

(項目 22)

前記スコアについての信頼区間をパラメトリックまたは非パラメトリック計算ブートストラッピング技法に基づいて決定するステップをさらに含む、前記項目のいずれかに記載のコンピュータ化された方法。

30

(項目 23)

作用物質に応答した生物系の攪乱を定量化するための、少なくとも1つのプロセッサを備えるコンピュータシステムであって、該プロセッサが、

作用物質に対する生物系の応答に対応する処置データの組を受け取るように構成または適合され、ここで、該生物系が複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体が、該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用し、該プロセッサが、

該作用物質に曝露されない該生物系に対応するコントロールデータの組を受け取るように構成または適合され、該プロセッサが、計算因果関係ネットワークモデルであって、

40

該生物系を表すと共に、

該生物学的実体を表すノード、

該生物学的実体間の関係を表すエッジ、および

該コントロールデータと該処置データの間の変化の予測方向を該ノードについて表す方向値

を含むかまたは備える計算因果関係ネットワークモデルを提供するように構成または適合され、該プロセッサが、

該処置データと該コントロールデータ間の差異を表す活性尺度を該ノードについて計算するように構成または適合され、該プロセッサが、

少なくとも1つの重み値が少なくとも1つの他の重み値と異なる複数の重み値を該ノード

50

ドについて計算するように構成または適合され、かつ該プロセッサが、

該作用物質に対する該生物系の攪乱を表す該計算モデルについてのスコアであって、該スコアが該方向値、該重み値および該活性尺度に基づくスコアを生成するように構成または適合された、コンピュータシステム。

(項目 2 4)

項目 1 から 2 2 のいずれかに記載の方法を実施するように適合されたプログラムコードを含む、コンピュータプログラム製品。

(項目 2 5)

項目 2 4 に記載のコンピュータプログラム製品を備えるコンピュータ、またはコンピュータ記録可能媒体もしくはデバイス。

10

【 0 0 1 9 】

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ともに、以下の詳細な記載を考察した後、明らかになる。

【図面の簡単な説明】

【 0 0 2 0 】

【図 1】図 1 は、攪乱に対する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システムのブロック図である。

【図 2】図 2 は、ネットワーク攪乱振幅 ( N P A ) スコアを計算することによって攪乱に対する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

20

。【図 3】図 3 は、2 つの作用物質、2 つのパラメータ、N 個の生物学的実体についての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グラフ表示である。

【図 4】図 4 は、いくつかの生物学的実体およびこれらの関係を有する生物学的ネットワークの計算モデルの図である。

【図 5】図 5 は、幾何学的攪乱指標 ( G P I ) スコアを生成するための例示的なプロセスの流れ図である。

【図 6】図 6 は、確率論的攪乱指標 ( P P I ) スコアを生成するための例示的なプロセスの流れ図である。

【図 7】図 7 は、予測攪乱指標 ( E P I ) スコアを生成するための例示的なプロセスの流れ図である。

30

【図 8】図 8 は、幾何学的攪乱指標 ( G P I ) スコアについての信頼区間を生成するための例示的なプロセスの流れ図である。

【図 9】図 9 は、本明細書に開示されたシステムおよび方法を用いて分析された生物学的ネットワークモデルの図である。

【図 1 0】図 1 0 ~ 1 4 は、ネットワークをベースとする生物学的機構についてのネットワーク攪乱振幅 ( N P A ) スコア化結果を示す図である。

【図 1 1】図 1 0 ~ 1 4 は、ネットワークをベースとする生物学的機構についてのネットワーク攪乱振幅 ( N P A ) スコア化結果を示す図である。

【図 1 2】図 1 0 ~ 1 4 は、ネットワークをベースとする生物学的機構についてのネットワーク攪乱振幅 ( N P A ) スコア化結果を示す図である。

40

【図 1 3】図 1 0 ~ 1 4 は、ネットワークをベースとする生物学的機構についてのネットワーク攪乱振幅 ( N P A ) スコア化結果を示す図である。

【図 1 4】図 1 0 ~ 1 4 は、ネットワークをベースとする生物学的機構についてのネットワーク攪乱振幅 ( N P A ) スコア化結果を示す図である。

【図 1 5】図 1 5 は、生物学的攪乱の影響を定量化するための例示的な分散型コンピュータ化システムのブロック図である。

【図 1 6】図 1 6 は、本明細書に記載のいずれかのコンピュータ化システムのあらゆる構成要素 ( c o m p o n e n t ) を実施するのに使用できる例示的な計算デバイスのブロック図である。

50

## 【発明を実施するための形態】

## 【0021】

## 詳細な説明

「含む (including)」、「備える (comprising)」という語は、他の要素またはステップを除外せず、不定冠詞の「1つの (a)」または「1つの (an)」は複数を除外しない。本明細書では、生物系が作用物質によって攪乱されるとき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を記載する。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か、または備える。この計算には入力として、作用物質によって上記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が使用される。次に上記データは、上記生物系の特徴のネットワークモデルに適用される。この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として使用され、上記生物系において目的の特徴を可能なものにする生物学的な機構および経路を表す。この特徴、またはその機構および経路の一部は、疾患の病状および上記生物系の有害な健康への作用の一因となり得る。データベースで表される上記生物系についての従来の知識を用いて上記ネットワークモデルが構築され、このネットワークモデルには、正常状態下および作用物質による攪乱下を含む様々な状態のもとでの、多数の生物学的実体の状態についてのデータが格納される。使用される上記ネットワークモデルは、攪乱にตอบสนองした様々な生物学的実体の状態の変化を表すという点で動的であり、上記生物系に対する作用物質の影響についての定量的で客観的な評価を生み出すことができる。これらの計算方法を実行するためのコンピュータシステムもまた提供される。

10

20

## 【0022】

本発明のコンピュータ化された方法によって生成される数値を使用して、(安全性評価または比較に関して) 製造物、(効果の判定または健康上の利益に関して) 栄養補助剤を含む治療化合物、および(長期間曝露の危険ならびに有害作用および疾患の発症との関係の予測に関して) 環境的活性物質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ことができる。

## 【0023】

一態様では、本明細書に記載のシステムおよび方法により、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が、攪乱された生物学的機構のネットワークモデルに基づいて得られる。本明細書でネットワーク攪乱振幅 (NPA) スコアと呼ばれる数値を使用して、規定された生物学的機構における様々な実体の状態変化を簡潔に表すことができる。様々な作用物質または様々な種類の攪乱について得られる数値は、それ自体生物系の特徴として機能できるまたは生物系の特徴として顕在化する生物学的機構に対する様々な作用物質または攪乱の影響を相対的に比較するのに使用することができる。すなわち、NPA スコアは、様々な攪乱に対する生物学的機構の応答を測定するのに使用することができる。「スコア」という用語は、本明細書では一般に、生物系の変化の大きさの定量的尺度を与える値または値の組を指す。このようなスコアは、当技術分野で公知の様々な数学的アルゴリズムおよび計算アルゴリズムのいずれかを使用することによって、また、サンプルまたは被験体から得られる1つまたは複数のデータセットを使用する本明細書で開示された方法により、計算される。上記NPAスコアは、診断、実験の設計、治療決定およびリスク評価を改善する上で、研究者および臨床医の助けになり得る。例えば、上記NPAスコアを用いて、毒物学分析の際に候補の生物学的機構の組を選別し、それによって、潜在的に有害な作用物質への曝露で最も影響を受ける可能性のあるものを識別することができる。攪乱に対するネットワーク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、分子事象(実験データによって測定される)と、細胞、組織、器官または生物のレベルで生じる表現型または生物学的結果との相関を取ることができる。臨床医はNPA値を用いて、作用物質によって影響を受ける生物学的機構を患者の生理学的状態と比較し、それによって、その作用物質に曝露されたときにどのような健康上のリスクまたは利益をその患者が最も受ける可能性があるかを判定することができる(例えば、免疫無防備状態

30

40

50

である患者は、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る可能性がある)。

#### 【0024】

図1は、攪乱に対する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システム100のブロック図である。詳細には、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を含むか、または備える。上記エンジン110、112および114は時々相互接続され、さらに、攪乱データベース102、測定可能物データベース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と時々接続される。本明細書でエンジンは、コンピュータ、マイクロプロセッサ、論理デバイスなどの1つまたは複数の処理デバイス、または図14に関して記載される1つまたは複数の他のデバイスを含むか、または備え、この処理デバイスは、1つまたは複数の計算作業を実行するためのハードウェア、ファームウェアおよびソフトウェアで構成される。

10

#### 【0025】

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に対する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。上記プロセス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、また適切な順序に配置され得るか、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。このデータは、生物学的システムが攪乱される実験からのデータ、ならびにコントロールデータを含む。ステップ212で、上記SRPエンジン110はシステム応答プロファイル(SRP)を生成し、このプロファイルは、ある生物系内の1つまたは複数の実体が、その生物系に作用物質が与えられることに応答して変化する程度を表す。ステップ214で、ネットワークモデリングエンジン112は、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し、これらネットワークモデルのうちの1つが上記作用物質と、または目的の特徴と関連があるものとして選択される。この選択は、上記系の生物学的機能の基礎をなす機構についての従来知識に基づいて行うことができる。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、上記システム応答プロファイル、上記データベースにおけるネットワーク、および以前に文献に記載されているネットワークを使用して上記系内の実体間の因果関係を抽出し、それによってネットワークモデルを生成、洗練または拡張することができる。ステップ216で、上記ネットワークスコア化エンジン114は、ステップ214で上記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と、ステップ212で上記SRPエンジン110によって生成したSRPとを使用して、攪乱ごとのNPAスコアを生成する。NPAスコアにより、攪乱または処置に対する生物学的応答(SRPによって表される)が、上記生物学的実体間の基礎をなす関係(上記ネットワークによって表される)に即して定量化される。以下の記載は、限定するためではなく開示を明確にするために小区分に分割されている。

20

30

40

#### 【0026】

##### A. 生物系

本発明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、およびブタを含むかまたは備える。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

50



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、トランスジェニック動物であるか、遺伝子組み換え動物であるか、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、オスマ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が無症候性である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、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。

#### 【0027】

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。生物系の例は、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、および生殖器系を含むか、それらを備える。生物系の他の例は、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を含むか、それらを備える。これらの細胞のうちのいくつかは、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。細胞機能の例は、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を含むか、それらを備える。生物系として考えることができる細胞成分の例は、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、および抗体を含むか、それらを備える。

10

20

30

40

50

## 【 0 0 2 8 】

## B . 攪乱

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に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は、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を含むか備える。物理的作用物質の非限定的な例は、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、あるいは外傷を含むか備える。

10

20

## 【 0 0 2 9 】

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る。

30

## 【 0 0 3 0 】

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、または紙巻きタバコの煙、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むかまたは備えうる。作用物質のさらなる非限定的な例は、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))、およびニコチン置換療法のために使用される生成物を含むかまたは備えうる。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

40

50

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。

#### 【0031】

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。

10

#### 【0032】

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、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かまたは備えることができる。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。

20

#### 【0033】

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ตอบสนอง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、ネットワーク振幅(NPA)値を生成することができる。

#### 【0034】

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1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。

30

#### 【0035】

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ArrayExpress(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

40

50

、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かまたは備えるものとしてよい。*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）を含むかまたは備えうる。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  
10  
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。

#### 【0036】

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  
20  
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、とりわけ含むかまたは備える。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、ビオチン化、ステアロイル化、ホルミル化、ミリストイル化、パルミトイル化、ゲラニルゲラニル化、ペグ化、リン酸化、硫酸化、グリコシル化、糖修飾、脂質化、脂質修飾、ユビキチン化、スモイル化、ジスルフィド結合、システイニル化、酸化、グルタチオン化、カルボキシル化、グルクロン酸化、および脱アミドを含むかまたは備える  
30  
ことができる。それに加えて、タンパク質は、アマドリ反応、シッフ塩基反応、および糖化タンパク質生成物を生じるメイラード反応などの一連の反応によって翻訳後修飾されうる。

#### 【0037】

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むかまたは備えうる。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むかまたは備えうる。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  
40  
、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。

#### 【0038】

このデータは、システム応答プロファイルを決定する際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、または表現型に関する情報を含むかまたは備える。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かまたは備える。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  
50

ータを含むかまたは備えうる。

【0039】

ステップ212で、上記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、有意性判定、および差次的応答(例えば、差次的に発現した遺伝子)の識別のうちの1つまたは複数を含むか、あるいは備えてよい。SRPは、生物系における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)に応答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、上記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)に対する測定結果の組を収集する。図3は、2つのSRP、つまり、可変パラメータ(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かまたは備える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かまたは備える類似のSRP304を示している。SRPに含まれるかまたは備えられるデータは、生実験データ、処理済み実験データ(例えば、域外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。上記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。

10

20

【0040】

前のステップで導出されたSRPは、上記実験データからネットワーク攪乱の大きさが決定される、その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、上記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層(a layer of mechanistic understanding)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、該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。

30

【0041】

このようなネットワークの構築は、反復プロセスである。上記ネットワークの境界の描写は、目的のプロセス(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、以前の知識から抽出されてネットワークの核をなす。文献ベースのネットワークは、関連する表現型エンドポイント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を使用して上記データセットを解析することができ、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。

40

【0042】

C. ネットワーク

図2を再び参照すると、ステップ214で、上記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共に上記SRPエンジン110からの上記システム応答プロファイルを使用する。ある態様では、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含むかまたは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はまた、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

50

クモデルを生成することによってネットワーク生成のプロセスを反復することができる。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。ネットワークポロジが維持され、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。

#### 【0043】

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トワーク400を示している。これらのノードは、それだけには限らないが、化合物、DNA、RNA、遺伝子、タンパク質、ペプチド、抗体、細胞、組織、器官、および細胞過程または分子過程などの、生物系内の生物学的実体または生物学的過程を表すことができる。この生物学的実体は、処置データまたはコントロールデータが受け取られるか、または入手可能である生物学的実体に必ずしも限定されない。すなわち、上記生物学的実体を表すノードは該複数の生物学的実体を含むか、または備えることができ、かつ1つまたは複数のさらなる生物学的実体を含むか、または備えてもよい。上記ノードの少なくとも一部はスコア化可能であり、上記スコアは該ノード（1つまたは複数）の活性レベルを表すことができる。上記ノードの多くが、上記活性レベルが測定可能である生物学的実体を表す。しかし、いくつかの実装（implantation）では、上記コンピュータ化された方法において、すべてのこのような測定可能ノードについてのデータを必ずしも受け取る必要がない。すなわち、上記ノードはスコア化可能および/または測定可能である。いくつかの実装では、ノードのほとんどが測定可能である。測定可能ノードは、測定されたデータを含むか、または備えてよい。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、一対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。

#### 【0044】

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

10

20

30

40

50

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。

#### 【0045】

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かまたは備えさせ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係するデータ、および他の任意の種類生命科学関係の情報を含むかまたは備えることも可能である。

#### 【0046】

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かまたは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、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、そこでは、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

10

20

30

40

50

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。上記生物系は、計算因果関係ネットワークモデルを含む複数のネットワークモデルによって表すことができる。

【0047】

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係るか、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。

【0048】

本発明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(measurable104)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。

【0049】

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。複数の機構仮説など、少なくとも1つの機構仮説を組み立てることができる。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

10

20

30

40

50



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センブルして、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。

#### 【0050】

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。

#### 【0051】

機構仮説は、例えば遺伝子発現データである測定データの組と、それらの遺伝子の公知のコントローラである生物学的実体との間の関係を表す。加えて、これらの関係は、上流の実体と下流の遺伝子の差次的発現との間の影響の符号（正または負）を含むか、または備える。ある仮説の下流の遺伝子は、文献精選した因果的生物学知識のデータベースから引き出される。計算可能な因果関係ネットワークモデルの形態で上記上流の実体を下流の遺伝子に結合する機構仮説の因果関係は、上記NPASコア化法によってネットワーク変化を計算するための基盤になる。上記生物系は、複数の機構仮説など、少なくとも1つの機構仮説によって表すことができる。少なくとも1つの上記計算因果関係ネットワークモデルは、複数の機構仮説を備えてよい。

#### 【0052】

上記モデルにおける実体を表す個々の機構仮説を集め、すべての上記下流の遺伝子の接続部を、複合因果関係ネットワークモデル全体を表す単一の上流の過程に再編成することによって、生物学的実体のスコア化可能複合因果関係ネットワークモデルを単一因果関係ネットワークモデルに変換することが可能であり、この変換は本質的に、基礎となるグラフ構造の平坦化である。このようにして、上記ネットワークモデルによって示された上記生物学的実体の活性変化は、その個々の機構仮説の組合せによって、基礎となる遺伝子発現測定が全体として上記ネットワークに寄与するように評価することができる。

#### 【0053】

本発明の方法で使用するためのスコア化可能ネットワークを生成するために、参照ノードが最初に、開始（通常は複合）因果関係ネットワークモデルから選択される。上記参照ノードは、上記ネットワークにおける、そのレベルまたは活性が全体として該ネットワークの活性と正の関係がある任意の実体とすることができる（対照的には、例えば、その活性が上記ネットワーク活性と負の関係があるインヒビター）。次に、上記モデルにおける各ノードと上記参照ノードの間の因果関係が決定される。この決定は、上記モデルが「因果的に一貫している」ことを最初に要求することによって行うことができる。上記モデルにおけるノードごとに下流測定可能実体（この例では遺伝子発現）の調節の符号は、そのモデルノードと上記参照ノードの関係に基づいて調整される。例えば、上記参照ノードに対して正の因果関係があるモデルノード（すなわち、このノードは、該参照ノードが増加するときに正に調節されると予測される）の下流の遺伝子発現の符号は維持される。一方で、上記参照ノードに対して負の因果関係があるモデルノード（すなわち、このノードは、該参照ノードが増加するときに負に調節されると予測される）の下流の遺伝子発現の符号は反転される。次に、上記下流の遺伝子発現およびその符号すべてが単一の機構仮説に組み合わされ、（複数のモデルノードからの）相反する符号をもつ下流の遺伝子発現が機構仮説から除外される。

#### 【0054】

因果的に一貫しているべきネットワークモデルでは、上記モデルにおけるどのノードの増加に対しても、該モデルにおける1つおきのノードに「正の調節」または「負の調節」の符号を明白にマッピングすることが、該ノードを接続する因果関係を追跡することによって可能でなければならない。どのプロセスが上記機構仮説によってスコア化されているか、またどの符号で各ノードが効果的に上記参照ノードと関連づけられるかを考慮することによって、生物学的解釈を用いてあいまいさを解消し、それによって因果的に一貫しているモデルを構築することができる。例えば、負のフィードバックが元のモデルとつながるノードは、スコア化されているプロセスと特別な関係があり、該負のフィードバックによりこのノードを調節できるが、この関係は変わらないはずである。したがって、負のフィードバックループとこのノードの間の接続を上記モデルから取り除いて、公知の事実と適合するようにして因果一貫性を得ることができる。上記の手法の変形は、米国特許出願公開第2007/0225956号および第2009/0099784号で論じられており、これらの特許文献は、参照によりその全体が本明細書に組み込まれる。例示的な因果関係ネットワークモデルは、Westra 7W、Schlage WK、Frushour BP、Gebel S、Catlett NL、Han W、Eddy SF、Hengstermann A、Matthews AL、Mathis C、Construction of a Computable Cell Proliferation Network Focused on Non-Diseased Lung Cells、BMC Syst Biol、2011年、5巻：105頁に記載されており、同文献は、参照によりその全体が本明細書に組み込まれる。

10

20

## 【0055】

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。

30

## 【0056】

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係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むかまたは備えうる。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。

40

## 【0057】

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

50

ン 1 1 2 は、1 つまたは複数のルールを含むかまたは備えることができる。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 1 1 2 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 1 0 6 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 1 1 2 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 1 1 2 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して行うことができる。

10

## 【 0 0 5 8 】

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 1 1 2 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）。

## 【 0 0 5 9 】

上記ネットワークモデリングエンジン 1 1 2 はさらに、識別された上記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か、または備えてよい。ノードは、その関連づけられた生物学的実体、関連づけられた該生物学的実体が測定可能な量であるかないかを表示するもの、または他の任意の該生物学的実体の記述子によって記述することができる。上記ノードの一部はスコア化可能であり、このスコアは、該ノード（1 つまたは複数）の活性レベルを表すことができる。上記ノードの多くは、その活性レベルを測定することが可能な生物学的実体を表す。しかし、いくつかの実装では、上記コンピュータ化された方法が、すべてのこのような測定可能ノードについてのデータを受け取ることを必ずしも必要としない。すなわち、上記ノードはスコア化可能および/または測定可能である。いくつかの実装では、上記ノードのほとんどが測定可能である。測定可能ノードは、測定されたデータを含むか、または備えてよい。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 1 1 2 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 1 1 2 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 1 1 2 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實體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、またはその特定の實體の測定結果が破棄されうる。いくつかの実装形態では、上記ノードに関する方向値により、上記コントロールデータと上記処置データの間の値の予測変化（*expected change in value*）を表すことができる。いくつかの実装形態では、上記ノードに関する方向値により、コントロールデータと処置データの間の予測される値の変化を表すことができる。いくつかの実装形態では、ノードに関する方向値により、上記コントロ

20

30

40

50

ールデータおよび上記処置データの値の予測される増加または減少を表すことができる。この変化は適切に、処置後の変化を表す。

【0060】

D. ネットワーク攪乱振幅

本明細書に提示された計算方法およびシステムにより、SRPをNPAスコアに変換する。ネットワークモデルにおける攪乱の下流の作用として識別される各実験測定値は、ネットワーク固有の応答スコアに統合される。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、ステップ214で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と、ステップ212でSRPエンジン110によって生成されたSRPとを使用して、攪乱ごとのNPAスコアを生成する。NPAスコア化により、定義された1つまたは複数のアルゴリズムが一連の処置とコントロールの比較からなる実験データセットに適用され、この実験データは、定義された生物学的ネットワークモデルに即して生物学の識別の範囲（例えば、遺伝子発現関係の識別の組）を表すようにフィルタリングされる。NPAスコアにより、処置に対する生物学的応答（SRPによって表される）が、生物学的実体間の基礎をなす関係（識別されたネットワークによって表される）に即して定量化される。ネットワークスコア化エンジン114は、ネットワークモデリングエンジン112内に収容された、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ごとにNPAスコアを生成するためのハードウェアおよびソフトウェアの構成要素を含むか、または備える。

10

【0061】

ネットワークスコア化エンジン114は、いくつかのスコア化技法のどれでも実施するように構成することができる。このような技法には、スカラ値のスコアを生成するものが含まれる。このような技法にはまた、ベクトル値のスコアを生成するものも含まれる。ベクトル値のスコアは、攪乱に対するネットワークの応答の大きさおよび位相分布を表す。

20

【0062】

記載されるスコア化技法の1つは、強度スコア化技法である。強度スコアはスカラ値のスコアであり、活性の平均値である。強度スコアは、SRPで表される様々な実体の活性観測結果の平均値である。ネットワーク応答の強度は次式により計算される。

【0063】

【化1】

$$\text{強度} = \frac{\sum_i d_i \beta_i}{N}$$

(1)

30

ここで $d_i$ は、ノード $i$ と関連づけられた実体の活性変化の予測方向を表し、 $\beta_i$ は、処置とコントロール条件の間の活性の倍率変化のログ（すなわち、量が初期値から最終値まで動いてどれだけ変化するかを表現する数）を表し、 $N$ は、測定された生物学的実体と関連づけられたノードの数である。正の強度スコアは、SRPが、識別されたネットワークから導かれた予測活性変化に一致することを示し、負の強度スコアは、SRPが、予測活性変化に一致しないことを示す。

40

【0064】

上記スコアは、幾何学的攪乱指標スコア化技法、確率論的攪乱指標スコア化技法、または予測攪乱指標スコア化技法によって生成することができる。スコア化技法の1つは、幾何学的攪乱指標（GPI）スコア化技法である。図5は、ネットワークスコア化エンジン114によって実施できるGPIスコア化技法の流れ図500である。ステップ502で、ネットワークスコア化エンジンにより倍率変化ベクトルをアセンブルする。倍率変化は、攪乱条件とコントロール条件の間などの様々な条件下で、測定可能要素（measurable）が初期値から最終値まで動いてどれだけ変化するかを表現する数である。この倍率変化ベクトルは、測定される生物学的実体と関連づけられたネットワーク内のノ

50

ドの数に対応するN個の成分を有する。いくつかの実装では、倍率変化ベクトルのi番目の成分  $\beta_i$  は、攪乱条件とコントロール条件の間のi番目に測定された生物学的実体の活性の倍率変化の対数(例えば、底2)を表す(すなわち、2つの条件の間で実体の活性を変化させる要素のログ)。そのため、 $\beta_i$  のゼロの値は、攪乱条件とコントロール条件の間に活性の変化が観測されなかったことを示す。対数演算は含まれる必要がなく、あるいは他の任意の線形関数または非線形関数に置き換えてもよい。例えば、いくつかの実装では、 $\beta_i$  は、対数演算を用いずに攪乱条件間の活性の倍率変化を表し、このような実装では、 $\beta_i$  の1の値は、攪乱条件とコントロール条件の間に活性の変化が観測されなかったことを示す。倍率変化は、本明細書に記載のネットワークスコア化技法と共に使用するための、活性を定量化する単なる1つの実行可能な手法であり、測定可能要素の変化を表現する他の任意の従来技法が用いられてもよいことを理解されたい。いくつかの実施形態では、上記スコアを生成する上記ステップは、上記活性尺度、上記重み値および上記方向値の線形または非線形の結合、ならびにこの結合をスケール係数によって正規化することを含んでよい。上記結合は算術結合でよく、該スケール係数は、その測定データが受け取られる生物学的実体の数の平方根でよい。いくつかの実施形態では、スコアはスカラ値のスコアではない。

#### 【0065】

ステップ504で、ネットワークスコア化エンジン114は、重みベクトル  $r$  を生成する。重みベクトル  $r$  もまた、倍率変化ベクトル  $\beta$  の成分のそれぞれに1つの、N個の成分を有する。重みベクトル  $r$  のそれぞれの成分  $r_i$  は、i番目に観測された倍率変化  $\beta_i$  に与えられるべき重みを表す。いくつかの実施形態では、重みは、対象(例えば、がん研究で公知の発がん物質)の特徴または結果に関して、i番目に測定された実体の公知の生物学的重要性を表す。いくつかの実装では、重みは、ノードと関連づけられた生物学的実体についての活性測定の信頼度を表す。信頼度推定値を用いてログ倍率変化を重みづけすることによって、信頼度が低い倍率変化  $\beta_i$  は、GPIスコアにあまり寄与しない。改善された実験室条件、増加した生物学的再現数、再現性がよりよいこと、分散がより小さいこと、および信号がより強いことがすべて、識別の  $\beta_i$  においてより高い信頼度に寄与し得る。

#### 【0066】

重みづけに有利に使用されうる1つの値は、局所的偽不発見率(local false non-discovery rate)  $fndr_i$  (つまり、場合によっては、観察されたp値を条件として、倍率変化値  $\beta_i$  がゼロ倍率変化の基礎となる帰無仮説からの逸脱を表す確率)であり、これはStrimmerら、「A general modular framework for gene set enrichment analysis」、BMC Bioinformatics 10:47、2009年およびStrimmer、「A unified approach to false discovery rate estimation」、BMC Bioinformatics 9:303、2008年に記載されているとおりであり、それぞれ参照によりその全体が本明細書に組み込まれている。いくつかの実施形態では、 $fndr_i$  は次式により計算され、

#### 【0067】

#### 【化2】

$$fndr_i(\beta_1, \dots, \beta_N) = 1 - 2v_i(\beta_1, \dots, \beta_N) \int_{|\beta_i/S_i(\beta_1, \dots, \beta_N)|}^{\infty} t_{df}(x) dx, \quad (2)$$

ここで  $fndr_i$  は、局所偽発見率(すなわち、倍率変化値  $\beta_i$  がゼロ倍率変化の基本的帰無仮説からの逸脱を表さない確率)、 $v_i$  はBenjamini-Hochberg調整係数であり、これは、Benjaminiら、「Controlling the false discovery rate: a practical and powe

10

20

30

40

50

「A robust approach to multiple testing」、Journal of the Royal Statistical Society、B巻57号：289頁、1995年、に記載されており、同文献は、参照によりその全体が本明細書に組み込まれる。pは、実際に観測された倍率変化  $\beta_i$  と少なくとも同程度に極端な倍率変化が得られる確率であり（ゼロ倍率変化の帰無仮説が真であると仮定して）、 $t_{df}$  は、dfの自由度をもつt分布である。pは  $\beta_i$  および標準偏差  $S_i$  の関数であり、その結果としてすべての  $\beta_i$  に基づくことに留意されたい。一代替実装では、複数の試験に対し調整が行われず、したがって、 $v_i (i = 1, \dots, N)$  は1に等しく、重みベクトル  $r_i = 1 - p(\beta_i, S_i (i = 1, \dots, N))$  になる。

【0068】

10

ステップ506で、ネットワークスコア化エンジン114は、重みベクトル  $r$  を用いて倍率変化ベクトル  $\beta$  をスケーリングする。その結果は、スケーリングされた倍率変化ベクトルになり、各成分  $\beta_i$  に、その関連づけられた重み成分  $r_i$  が掛けられている。このようなスケーリングを計算的に実現する1つの方法は、式3に示されるように、重み成分  $r_i$  が対角にある  $N \times N$  対角行列を生成し、この行列に  $N \times 1$  ベクトル  $\beta$  を掛けることである。

【0069】

【化3】

$$\begin{bmatrix} r_1\beta_1 \\ r_2\beta_2 \\ \vdots \\ r_N\beta_N \end{bmatrix} = \underbrace{\begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & r_N \end{bmatrix}}_{diag(r)} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}}_{\beta}$$

20

(3)

ステップ508で、ネットワークスコア化エンジン114は、倍率変化ベクトル  $\beta$  の成分ごとに変化の予測方向を識別する。ネットワークスコア化エンジン114がそのようにすることは、ネットワークモデリングエンジン112を検索して変化の予測方向を因果生物学的ネットワークモデルから取り出すことによって可能である。次に、ネットワークスコア化エンジン114は、これらの変化の予測方向をアセンブルしてN成分のベクトル  $d$  を作ることができ、この場合ベクトル  $d$  の  $i$  番目の成分  $d_i$  は、 $i$  番目に測定された生物学的実体の変化の予測方向を表す（例えば、活性の増大は + 1、活性の減少は - 1）。

30

【0070】

ステップ510で、ネットワークスコア化エンジン114は、スケーリングされた倍率変化ベクトルの成分（ステップ506で生成）を、成分ごとの変化の予測方向（ステップ508で識別）と結合する。いくつかの実装では、上記結合は算術結合であり、該スケーリングされた倍率変化  $r_i \beta_i$  のそれぞれに、その対応する変化の予測方向  $d_i$  が掛けられ、その結果がN個の生物学的実体すべてについて合計される。数学的には、このステップ510の実装を次式で表すことができる。

40

【0071】

【化4】

$$\sum_i d_i r_i \beta_i$$

(4)

他の実装では、ベクトル  $d$ 、 $r$  および  $\beta$  は、任意の線形または非線形の方法で結合することができる。

【0072】

50

ステップ512で、ネットワークスコア化エンジン114は、ステップ510で結合したものを正規化する。いくつかの実装では、正規化は、所定のスケール係数を掛けることからなる。このようなスケール係数の1つは、生物学的実体の数Nの平方根である。この実装では、GPIスコアは次式で表すことができる。

【0073】

【化5】

$$GPI = \frac{\sum_i d_{r_i} \beta_i}{\sqrt{N}}$$

(5)

10

所定のものであってもなくてもよい他のスケール係数もまた用いることができる。いくつかの実施形態では、因果関係ネットワークモデル（例えば、機構仮説）は、N次元の下流の測定可能な空間内（各次元が因果関係ネットワークモデルの、ここでは遺伝子発現である、下流の測定可能要素を表す）の単位符号ベクトル  $s = (1, 1, -1, 1, \dots, -1) / \sqrt{N}$  と理解することができる。下流の遺伝子発現に対する攪乱の観測された作用もまた、この空間内のベクトルになる。したがって、因果関係ネットワークモデルにおける攪乱の振幅は幾何学的に、差次的  $\log_2$  発現ベクトルを仮説単位ベクトルの上に投影することによって定量化することができる。しかし、因果関係ネットワークモデルの下流測定値は総称モデルから来る。NPAスコアを支持するデータの識別性を明確に処理するために、各下流には、局所偽非発見率 ( $fndr_i = (1 - fdr_i)$ ) になるように設定される活性化の信念が割り当てられる。これは、各差次的発現の信念により下流の遺伝子発現空間の次元に重みづけすることに等しく、したがって、重みづけされたスカラー積を考慮して遺伝子発現空間の幾何形状を定義することに等しい： $\langle s | \cdot \rangle_w = s^T \cdot \text{diag}(fndr) \cdot \cdot$ 。それゆえ、 $GPI = (s_i \cdot fndr_i \cdot \cdot)_i / \sqrt{N}$  になる。差次的  $\log_2$  発現を偽非発見率で重みづけすることによって、信頼度がほとんどない個々の差次的発現値はゼロ（変化なし）により近く移動される一方で、より強い信頼度がある値は最小限の低減がされる。正のGPIスコアは、機構仮説によって記述された処理の上方調節を示し、ゼロのGPIスコアは、その処理が機構仮説の方向  $s$  に沿って変えられないことを示し、負のGPIスコアは、その処理が下方調節されることを示す。

20

30

【0074】

図6は、ネットワークスコア化エンジン114によって実施できる確率論的攪乱指標（PPI）スコア化技法の流れ図600である。SRPエンジン110（図1）、およびプロセス200のステップ212（図2）に関して前に論じたように、各SRPは、ある処置条件下で測定された生物学的実体の活性（または活性の変化）を表す。次に、各SRPは、いくつかの測定された活性（測定された生物学的実体それぞれに1つ）と関連づけられる。PPIは、目的のネットワークによって表された生物学的機構が活性化される確率を、観測されたSRPを前提として定量化したものである。

【0075】

ステップ602で、ネットワークスコア化エンジン114は、倍率変化ベクトル  $\cdot$  をアセンブルする。この倍率変化ベクトルは、N個の測定された生物学的実体の活性について観測された倍率変化を表し、図5に示された幾何学的攪乱指標（GPI）スコア化技法のステップ502に関して前述されたようにアセンブルすることができる。ステップ604で、ネットワークスコア化エンジン114は、倍率変化密度の値域を生成する。倍率変化密度の値域は、処置条件下の生物系で倍率変化値が取ることのできる値の組の近似値を表し、値域  $[-W, W]$  で近似値を求めることができ、ここでWは、理論的に予測される  $\log_2$  倍率変化の最大絶対値である。このようにWを選ぶことによって、観測されるすべての倍率変化が値域  $[-W, W]$  に入る。例えば、遺伝子チップの予測される最大信号（例えば、 $\log_2$  スケールで16）を値Wとして用いることができる。

40

【0076】

50

ステップ606で、ネットワークスコア化エンジン114は、倍率変化ベクトル の成分ごとの変化の予測方向を識別する。このステップは、図5に示されたGPIスコア化技法のステップ508に関して前述したように実施することができ、それによって、観測された倍率変化  $d_i$  に対応する変化の予測方向  $d_i$  の組が得られる。

【0077】

ステップ608で、ネットワークスコア化エンジン114は、正の活性化距離を生成する。いくつかの実装では、正の活性化距離は、生物学的実体の観測された活性化/抑制が  $d_i$  で表された変化の予測方向と一致していることのSRPが示す程度を表す。一致した挙動は、本明細書では「正の活性化」と呼ばれる。用いることができる正の活性化距離の1つは、1つまたは複数のネットワークが正の活性化をされる確率である。このような確率は  $PPI^+$  と呼ばれ、次式により計算することができる。

10

【0078】

【化6】

$$PPI^+ = \Pr(\text{正の活性化}) = \frac{1}{W} \int_0^W \Pr(\text{正の活性化} | \varphi) d\varphi, \quad (6)$$

ここで

【0079】

【化7】

$$\Pr(\text{正の活性化} | \varphi) = \frac{1}{N} \sum_{0 < d_i \beta_i < \varphi} f n d r_i \quad (7)$$

20

ここで  $f n d r_i$  は、式1に関して前に論じた偽非発見率である。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、0とWの間の  $d_i$  の値を表すビンの組を用いて式6の式を数値積分するように構成される。用いることができるビンの組の1つはビン  $[d_{(i-1)}, d_{(i)}]$  であり、ここで  $(\cdot)$  の下付き文字は、最小倍率変化から最大倍率変化まで順に取られる値を表し、慣習で  $d_{(0)} = 0$  である。このような実装では、ネットワークスコア化エンジン114は、正の活性化距離  $PPI^+$  の近似値を次式により計算する。

30

【0080】

【化8】

$$PPI^+ \approx \frac{1}{WN} \sum_{0 < d_i \beta_i} f n d r_i d_i \beta_i \quad (8)$$

ステップ610で、ネットワークスコア化エンジン114は、負の活性化距離を生成する。いくつかの実装では、負の活性化距離は、生物学的実体の観測された活性化/抑制が  $d_i$  で表された変化の予測方向と一致していないことの、SRPが示す程度を表す。一致していない挙動は、本明細書では「負の活性化」と呼ばれる。用いることができる負の活性化距離の1つは、1つまたは複数のネットワークが負の活性化をされる確率である。このような確率は  $PPI^-$  と呼ばれ、次式により計算することができる。

40

【0081】

【化9】

$$PPI^- = \Pr(\text{負の活性化}) = \frac{1}{W} \int_{-W}^0 \Pr(\text{負の活性化} | \varphi) d\varphi, \quad (9)$$

50



ここで

【 0 0 8 2 】

【 化 1 0 】

$$\Pr(\text{負の活性化} \mid \varphi) = \frac{1}{N} \sum_{\varphi < d_i \beta_i < 0} f n d r_i \quad (10)$$

ここで  $f n d r_i$  は、式 1 および式 7 に関して前に論じた偽非発見率である。正の活性化距離に関して前述したように、いくつかの実装では、ネットワークスコア化エンジン 1 1 4 は、 $-W$  と  $0$  の間の  $\varphi$  の値を表すピンの組を用いて式 9 の式を数値積分するように構成される。用いることができるピンの組の 1 つはピン  $[d_{(i-1)}, d_{(i)}]$  であり、ここで  $(\cdot)$  の下付き文字は、最小倍率変化から最大倍率変化まで順に取られる値を表し、慣習で  $d_{(0)} = 0$  である。このような実装では、ネットワークスコア化エンジン 1 1 4 は、負の活性化距離  $P P I^-$  の近似値を次式により計算する。

【 0 0 8 3 】

【 化 1 1 】

$$P P I^- \approx \frac{1}{W N} \sum_{d_i \beta_i < 0} f n d r_i d_i \beta_i \quad (11)$$

ステップ 6 1 2 で、ネットワークスコア化エンジンは、正の活性化距離（ステップ 6 0 8 で生成）と負の活性化距離（ステップ 6 1 0 で生成）を結合して、確率論的攪乱指標または  $P P I$  と呼ばれる合成距離を生成する。ステップ 6 1 2 の結合は、任意の線形結合または非線形結合とすることができる。いくつかの実装では、 $P P I$  は、正の活性化距離と負の活性化距離との重みづけ線形結合である。例えば、ネットワークスコア化エンジン 1 1 4 は、次式により  $P P I$  を生成するように構成することができる。

【 0 0 8 4 】

【 化 1 2 】

$$P P I = \frac{1}{2} (P P I^+ + P P I^-) \quad (12)$$

ここで  $P P I^+$  および  $P P I^-$  は、前述の正および負の活性化距離である。式 1 2 により生成された  $P P I$  は、式 5 により計算された  $G P I$  と次式のように関連づけられる。

【 0 0 8 5 】

【 化 1 3 】

$$G P I = \frac{W}{\sqrt{N}} (P P I^+ - P P I^-) \quad (13)$$

加えて、ネットワークスコア化エンジン 1 1 4 は、 $i$  番目の成分が次式で定義されるベクトルの  $L 1$  ノルムを計算することによって式 1 2 の  $P P I$  を算出できるように構成することができる。

【 0 0 8 6 】

【化 1 4】

$$\left[ \frac{1}{2WN} fndr_i d_i \beta_i \right] \quad (14)$$

図 7 は、ネットワークスコア化エンジン 1 1 4 によって実施できる予測攪乱指標 (EPI) スコア化技法の流れ図 7 0 0 である。SRP エンジン 1 1 0 (図 1) およびプロセス 2 0 0 のステップ 2 1 2 (図 2) に関して前に論じたように、各 SRP は、ある処置条件下で測定された生物学的実体の活性 (または活性の変化) を表す。次に、各 SRP は、いくつかの測定された活性 (測定された生物学的実体ごとに 1 つ) と関連づけられる。EPI は、SRP で表されるすべての生物学的実体についての平均の活性変化を定量化したものである。一般に、SRP で表される測定された活性は、測定された活性の分布から無作為に引き当てたものでよく、EPI がその分布の予測値を表す。倍率変化  $\beta_i$  のそれぞれが分布  $p(\cdot)$  から引き出される場合、その分布の予測値は次式となる。

【0 0 8 7】

【化 1 5】

$$EPI = \int \varphi \cdot p(\varphi) \cdot d\varphi \quad (15)$$

真の理論的分布  $p(\cdot)$  は容易には分からないので、ネットワークスコア化エンジン 1 1 4 は、下記のステップを実行するように構成して、観測された活性、およびシステム 1 0 0 から引き出された他の情報に基づいた EPI 値の近似値を求めることができる。

【0 0 8 8】

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、倍率変化ベクトル をアセンブルする。この倍率変化ベクトルは、N 個の測定された生物学的実体の活性について観測された倍率変化を表し、図 5 に示された幾何学的攪乱指標 (GPI) スコア化技法のステップ 5 0 2 に関して、または図 6 に示された確率論的攪乱指標 (PPI) スコア化技法のステップ 6 0 2 に関して前述されたようにアセンブルすることができる。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、倍率変化密度の値域を生成する。ネットワークスコア化エンジン 1 1 4 は、図 6 に示された PPI スコア化技法のステップ 6 0 4 に関して前述されたように、倍率変化密度の値域を生成することができる。

【0 0 8 9】

ステップ 7 0 6 で、ネットワークスコア化エンジン 1 1 4 は、倍率変化ベクトル の成分ごとの変化の予測方向を識別する。このステップは、図 5 に示された GPI スコア化技法のステップ 5 0 8 に関して前述したように実施することができ、それによって、観測された倍率変化  $\beta_i$  に対応する変化の予測方向  $d_i$  の組が得られる。

【0 0 9 0】

ステップ 7 0 8 で、ネットワークスコア化エンジン 1 1 4 は、近似の倍率変化密度を生成する。倍率変化  $\beta_i$  のそれぞれが分布  $p(\cdot)$  から引き出される場合、分布  $p(\cdot)$  は近似的に次式で表すことができる。

【0 0 9 1】

10

20

30

40

【化 1 6】

$$\hat{p}(\varphi) \propto \begin{cases} \frac{1}{N} \sum_{d_i \beta_i > \varphi} \frac{|\beta_i|}{W}, & \varphi > \varepsilon \\ \frac{1}{N} \sum_{d_i \beta_i < \varphi} \frac{|\beta_i|}{W}, & \varphi < \varepsilon \end{cases} \quad (16)$$

ステップ 7 1 0 で、ネットワークスコア化エンジン 1 1 4 は、近似の倍率変化密度の近似予測値を生成し、それによって E P I スコアが得られる。いくつかの実装では、ネットワークスコア化エンジン 1 1 4 は、計算補間技法（例えば、線形または非線形補間技法）を適用して式 1 6 の分布から近似連続分布を生成し、次に、式 1 5 の式を用いてその分布の予測値を計算する。他の実装では、ネットワークスコア化エンジン 1 1 4 は、連続分布の長方形近似として式 1 6 の離散型分布を用いるように、かつ次式により E P I を計算するように構成される。

【 0 0 9 2】

【化 1 7】

$$EPI \approx \frac{1}{WN} \left[ \sum_{d_i \beta_i > 0} (d\beta)_{(i)} \left( \sum_{j=1}^{n_+} (d\beta)_{(j)} \right) \left( (d\beta)_{(i)} - (d\beta)_{(i-1)} \right) - \sum_{d_i \beta_i < 0} -(d\beta)_{(i)} \left( \sum_{j=1}^{n_-} -(d\beta)_{(j)} \right) \left( -(d\beta)_{(i)} - (-(d\beta)_{(i-1)}) \right) \right] \quad (17)$$

式 1 7 で (・) の下付き文字は、最小倍率変化から最大倍率変化まで順に取られる値を表し、 $n^+$  は、処置に回答して活性が増加すると予測された実体の数であり ( $d_i \beta_i > 0$ ) (ステップ 7 0 6 で)、 $n^-$  は、処置に回答して活性が減少すると予測された実体の数である ( $d_i \beta_i < 0$ ) (ステップ 7 0 6 で)。E P I スコアでは、高い値の倍率変化が、低い値のものよりも多く考慮に入れられて、識別性が高い活性測定値が得られる。

【 0 0 9 3】

ネットワークスコア化エンジン 1 1 4 はまた、ネットワークスコアまわりで信頼区間を決定するように構成することもできる。これらの信頼区間は、ネットワークスコアに反映される実験結果を評価するために臨床医または研究者によって使用されてよく、また別のデータ処理ステップでシステム 1 0 0 の他の構成要素によって（例えば、集合エンジン 1 1 0 によって）使用されてもよい。信頼区間を決定する有効な方法の 1 つは、所与のタイプ I (偽陽性) エラーリスク（例えば、 $\alpha = 0.05$ ）に対しゼロであるネットワークスコアの帰無仮説（または、処置条件とコントロール条件の間に活性の差がないことを表す他の適切なナル値）を評価することである。いくつかの実装では、ネットワークスコア化エンジン 1 1 4 は、パラメトリックまたは非パラメトリックブートストラッピング技法などの計算ブートストラッピング技法を用いて、計算されたメトリクスの分布を見積もる。多くのこのようなブートストラッピング技法が当技術分野で公知である。基礎をなす分布についての仮定をすることがほとんどできない場合、非パラメトリック技法を有利に用いることができる。基礎をなす分布が仮定される場合には、パラメトリック技法を有利に用いることができる。下記で論じられる例では、 $t_i$  は、 $t_{df}$  自由度に基づいた平均ゼロおよびサンプル分散  $S_i^2$  で、帰無仮説のもとで正規分布から生じると仮定される。ネットワークスコア化エンジンは、これらの量、ならびに「リマ (L i m m a)」R パッケージの線形モデル手法によって生成された  $t$  統計量および調整  $t$  統計量などの  $t_i$  を表す  $t$  統計量および調整  $t$  統計量を、統計的見積試験手順を用いることによって生成すること

10

20

30

40

50

ができる。この線形モデル手法は差次的遺伝子発現の分析において一般に用いられ、参照によりその全体が本明細書に組み込まれる Smyth、「Linear models and empirical Bayes methods for assessing differential expression in microarray experiments」、Statistical Applications in Genetics and Molecular Biology、3巻：3頁、2004年に記載されている。例えば、EPIスコアの信頼区間を決定するために（図7に関して前に論じたように）、ネットワークスコア化エンジン114は、パラメトリックブートストラッピング技法を実施するように構成して  $s_i$  の分布を、 $s_i$  が基本的正規分布から生じると仮定して見積もることができる。百分位数ブートストラッピング技法の適用に関する仮説に反するよう見える、EPIを含むかまたは備えてよい実装では、ネットワークスコア化エンジン114はさらに、Efron、「The jackknife, the bootstrap, and other resampling plans」、SIAM、1982年、およびDiciuccioら、「A review of bootstrap confidence intervals」、Journal of the Royal Statistical Society、50巻：338頁、1988年、に記載のバイアス補正百分位数法を適用することができる。これらの文献それぞれは、参照によりその全体が本明細書に組み込まれる。

【0094】

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、ブートストラッピング技法の代替の、またはブートストラッピング技法と組み合わせた分析的手法を用いて、信頼区間を決定することができる。分析的に信頼区間を決定するためにネットワークスコア化エンジン114によって実施される識別の技法は、用いられる識別のネットワークスコア化技法と、 $s_i$  の基礎をなす統計的分布についての仮定とに依存する。

【0095】

例えば、ネットワークスコア化エンジン114が（式1により）強度スコアを計算するように構成されている場合、ネットワークスコア化エンジン114は強度スコアを、独立した、ほぼ正規確率変数の重みづけされた合計からなる確率変数として処理する。結果として、強度スコアの分布は、ゼロ平均で分散が次式で計算されるほぼ正規確率変数になる。

【0096】

【化18】

$$S_{strength}^2 = \frac{1}{N^2} \sum_i S_i^2 \quad (18)$$

ネットワークスコア化エンジン114は、分散  $S_{strength}^2$  を用いて次式により  $t$  統計量を導出することができる。

【0097】

【化19】

$$t = \frac{\text{強度}}{S_{strength}} \quad (19)$$

その自由度  $df$  は、Satterthwaite、「An approximate distribution of estimates of variance components」、Biometrics、2巻：110頁、1946年、およびWelch、「The generalization of student's problems when several different population variances are involved」、Biometrika、34巻：28

10

20

30

40

50

頁、1947年、に記載されているWelch-Satterthwaite式を用いて近似値が求められる。これらの文献それぞれは、その全体が参照により本明細書に組み込まれる。これらの量を用いて、ネットワークスコア化エンジン114は、強度スコアの(1 - ) - 信頼区間を次式により生成することができる。

【0098】

【化20】

$$\text{強度} \pm t_{df}^{\alpha/2} S_{strength} \quad (20)$$

別の例として、ネットワークスコア化エンジン114が(図5に関して前に論じたように) GPIスコアを計算するように構成されている場合、ネットワークスコア化エンジン114はまた、GPIスコアの信頼区間を図8の流れ図800のステップに従って計算するように構成することもできる。ステップ802で、ネットワークスコア化エンジン114は、式5で表されたGPIスコアの1次テイラー展開を  $\beta_i$  の関数として次式により行い、

【0099】

【化21】

$$GPI(\beta_1, \dots, \beta_N) = GPI(\hat{\beta}_1, \dots, \hat{\beta}_N) + \sum_i \frac{\partial GPI}{\partial \beta_i} \Big|_{\hat{\beta}_i} (\beta_i - \hat{\beta}_i) + O(N^2) \quad (21)$$

ここで  $\hat{\beta}_i$  ハットは、測定された倍率変化値である。GPIスコアの1次テイラー近似では、最初の2項を保持し、 $O(N^2)$ 項を削除する。

【0100】

ステップ804で、ネットワークスコア化エンジン114は、GPI計算における  $\beta_i$  項の係数が  $\beta_i$  の関数であるかどうかを評価する。これらの係数は、予測方向項  $d_i$  および重み  $r_i$  を含むか、または備える。これらの係数が  $\beta_i$  の値に依存しない場合、式21の1次項は  $\beta_i$  に対して定数値になり、ネットワークスコア化エンジン114はステップ808に進む。しかし、係数が  $\beta_i$  の値に依存する場合、ネットワークスコア化エンジン114は、ステップ806に進んで式21の1次項の近似値を求める。特に、重みベクトル  $r$  が  $\beta_i$  の関数であり、予測方向項  $d_i$  が  $\beta_i$  の関数ではない場合、1次項は次式で表すことができる。

【0101】

【化22】

$$\frac{\partial GPI}{\partial \beta_i} = \frac{1}{\sqrt{N}} \left( d_i r_i + d_i \beta_i \frac{\partial r_i}{\partial \beta_i} \right) \quad (22)$$

特に、重みベクトル  $r$  が、式2および図5のステップ504に関して前に論じたように、偽非発見率の値  $fndr_i$  のベクトルである場合、ネットワークスコア化エンジン114は、式22の導関数項に対して次式を用いることができる。

【0102】

10

20

30

40

【化 2 3】

$$\frac{\partial}{\partial \beta_i} fndr_i(\beta_1, \dots, \beta_N) = -2 \underbrace{\frac{\partial v_i(\beta_1, \dots, \beta_N)}{\partial \beta_i}}_{\text{項 1}} \underbrace{\int_{\left| \frac{\beta_i}{S_i} \right|}^{\infty} t_{df}(x) dx}_{\text{項 2}}$$

$$- 2v_i(\beta_1, \dots, \beta_N) \frac{\partial}{\partial \beta_i} \left( \int_{\left| \frac{\beta_i}{S_i} \right|}^{\infty} t_{df}(x) dx \right) \quad (23)$$

式 2 3 で「項 1」と標識された導関数は、Benjamini - Hochberg 調整係数の導関数を表し、「項 2」と標識された積分は、i 番目の生物学的実体の倍率変化に対する p 値を表す。Benjamini - Hochberg 項は p 値が小さいときに最も関連性があるので、ネットワークスコア化エンジン 1 1 4 は、ステップ 8 0 6 で項 1 と項 2 の積をゼロに近似するように構成されてよい。結果として、ネットワークスコア化エンジン 1 1 4 は、計算の基礎定理を適用し、式 2 3 の導関数項の次の近似式を用いることができる。

【0 1 0 3】

【化 2 4】

$$\frac{\partial}{\partial \beta_i} fndr_i(\beta_1, \dots, \beta_N) \approx 2 \operatorname{sgn}(\beta_i) \frac{v_i(\beta_1, \dots, \beta_N)}{S_i} t_{df} \left( \left| \frac{\beta_i}{S_i} \right| \right) \quad (24)$$

式 2 4 の近似式を式 2 1 の式に含めると、GPI スコアの次の近似式が得られる。

【0 1 0 4】

【化 2 5】

$$GPI(\beta_1, \dots, \beta_N) \approx GPI(\hat{\beta}_1, \dots, \hat{\beta}_N) + (\beta_i - \hat{\beta}_i) \cdot \sum_i \left( d_i fndr_i + d_i \left| \hat{\beta}_i \right| \left[ 2 \frac{v_i(\hat{\beta}_1, \dots, \hat{\beta}_N)}{\hat{S}_i} t_{df} \left( \left| \frac{\hat{\beta}_i}{\hat{S}_i} \right| \right) \right] \right) \frac{1}{\sqrt{N}} \quad (25)$$

ステップ 8 0 8 で、ネットワークスコア化エンジン 1 1 4 は、前のステップで生成された GPI スコアの近似値を用いて、GPI スコアの近似分散を決定する。GPI スコアが（式 2 1 のように）確率変数  $\beta_i$  のアフィン関数として近似されている場合、近似の分散は、次式で与えられる  $\beta_i$  の各分散の重みづけされた合計になる。

【0 1 0 5】

【化 2 6】

$$S_{GPI}^2 = \sum_i \left( \frac{\partial GPI}{\partial \beta_i} \right)^2 S_i^2, \quad (26)$$

ここで  $S_i^2$  は i 番目の倍率変化  $\beta_i$  の分散である。したがって、式 2 5 の近似式の分散は次式のように書き表すことができる。

【0 1 0 6】

10

20

30

40

50

【化27】

$$S_{GPI}^2 \approx \sum_i \left( fndr_i + |\beta_i| \left[ 2 \frac{v_i(\beta_1, \dots, \beta_N)}{\hat{S}_i} t_{df} \left( \left| \frac{\beta_i}{\hat{S}_i} \right| \right) \right] \right)^2 S_i^2 \frac{1}{N}, \quad (27)$$

ここで  $d_i$  項は、 $d_i^2 = 1$  であるので  $d_i = + / - 1$  のとき削除される。

【0107】

ステップ810で、ネットワークスコア化エンジン114は、観測された倍率変化値でのGPIスコアの分散（例えば、式27で表されたもの）を評価する。ステップ812で、ネットワークスコア化エンジン114は、GPIスコアの信頼区間を次式により生成する。

【0108】

【化28】

$$GPI \pm t_{df}^{\alpha/2} S_{GPI}, \quad (28)$$

ここで、 $S_{GPI}$  は、式26および式27に関して前述したように計算される。式28は、観測された倍率変化値でのPPIスコアの分散を決定するように必要に応じて適合されてよい。

【0109】

ネットワークスコア化エンジン114は、上述のスカラー値のスコアに加えて、またはその代わりに、ベクトル値のスコアを生成することができる。ベクトル値のスコアの1つは、測定されたノードごとの活性の倍率変化または絶対変化のベクトルである。

【0110】

いくつかの実装では、攪乱（例えば、既知または未知の作用物質に曝露すること）のそれぞれについて、ネットワークスコア化エンジン114は、複数のNPAスコアを生成することができる。例えば、ネットワークスコア化エンジン114は、識別のネットワーク、識別の投与量の作用物質、および曝露された識別の時間に対するNPAスコアを生成することができる。

【0111】

## E. 実験結果

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に対する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を、腫瘍壊死因子（TNF）で処置された正常ヒト気管支上皮（NHBE）細胞をいくつかの因果関係ネットワークモデルを使用して分析するために用いた。上記ストレス応答および免疫応答転写因子NF- $\kappa$ B（活性化B細胞の核因子カッパ軽鎖エンハンサー）の活性化が、様々な系における腫瘍壊死因子アルファ（TNF $\alpha$ ）誘導シグナル伝達の主要なメディエーターとして詳細に明らかにされている。正常ヒト気管支上皮（NHBE）細胞を4つの異なる用量のTNF（0.1、1、10、および100 ng/ml）で処理し、処理後の4つの異なる時間（30分、2時間、4時間、および24時間）に全RNAをマイクロアレイ測定のために集めた。すべての処置物を、時間を一致させた偽の処理コントロールと比較して16の対照（contrast）（4用量 $\times$ 4時点）を得た。正常ヒト気管支上皮細胞（Lonza Walkersville, Inc.）を標準成長培地（Clonetics培地、Lonza Walkersville, Inc.）で培養した。細胞をTNF（シグマ）またはビヒクルコントロール（HBSS）でのいずれか処理し、次に、所望の攪乱期間の後に回収した。直ちに細胞を氷の上に置き、3つの技術的反復物（technical replicate）に分割し、これらから全RNAをRNeasy Microkit（Qiagen）を使用して抽出した。次に、処理されたRNA試料をAffymetrix U133 Plus 2.0マイクロアレイとハイブリッド形成させる。細胞生存率

10

20

30

40

50

および細胞数をすべての条件について、24時間後にCellTiter-Glo(登録商標)アッセイ(Promega)で制御した。NF-kB核移行をCellomics NF-kB Activation HCS Reagent Kit(Thermo Scientific)を使用して測定した。データ処理およびNPA法をR統計環境において実施した。生のRNA発現データを、R統計環境で使用可能なマイクロアレイ分析ツールのBioconductor一式のアフィおよびリマパッケージを使用して分析した(Gentleman R: Bioinformatics and computational biology solutions using R and Bioconductor, New York: Springer Science+Business Media, 2005年、Gentleman RC、Carey VJ、Bates DM、Bolstad B、Dettling M、Dudoit S、Ellis B、Gautier L、Ge Y、Gentry JR、Bioconductor: open software development for computational biology and bioinformatics、Genome Biol、2004年、5巻:R80頁)。Robust Microarray Analysis(RMA)バックグラウンド補正および分位正規化を用いてプローブセット発現値を生成した(Irizarryら、Exploration, normalization, and summaries of high density oligonucleotide array probe level data、Biostatistics、2003年、4巻:249~264頁)。全体線形モデルを複製物のすべての群についてのデータに適合させ、目的の識別的な対照(「処置された」条件と「コントロール」条件の比較)を評価して、発現アレイにおけるプローブセットごとに生のp値を生成した。その後、Benjamini-Hochberg偽発見率(FDR)を用いて生のp値を複数の試験効果(testing effect)に対して補正した。

#### 【0112】

プローブセットは、HG-U133\_\_Plus\_\_2.na30プローブセットマッピングおよび以下の基準を使用して、Selventa KnowledgebaseのRNA Abundanceノードと一致させた。第1に、「at」または「s\_\_at」プローブセットだけを考慮に入れた。第2に、複数の遺伝子に対して位置決めしたプローブセットを廃棄した。第3に、複数のプローブセットが同一の遺伝子に対して位置決めした場合、「s\_\_at」プローブセットよりも「at」プローブセットを優先させた。最後に、同一の遺伝子に対して位置決めした複数のプローブセットが依然として残った場合、目的のすべての対照にわたって最小幾何平均FDR補正p値をもつプローブセットを選択した。次に、線形モデルを複製物のすべての群について、知識ベース(knowledgebase)におけるRNA Abundanceノードに対して位置決めするプローブセットのみに再適合させ、FDR補正p値を再計算した。上記Selventa Knowledgebaseは、150万個を超えるノード(生物学的な概念および実体)および750万個を超えるエッジ(ノード間の因果関係および非因果関係についての主張)を含むリポジトリである。上記Selventa Knowledgebaseにおける主張は、同等者により検閲された(peer-reviewed)科学文献、ならびに他の公開データベースおよび専有のデータベースに由来する。具体的には、各主張は、ヒト、マウス、およびラット種の環境についてin vitroまたはin vivoのいずれかで実施された実験の個々の実験観察結果を記載する。主張はまた、参照情報源についての情報(例えば、MEDLINEに列挙されている学術論文についてのPubMed ID(PMID))、ならびに上記実験観測結果が導き出された種(ヒト、マウス、またはラット)および組織または細胞系を含む重要な関連情報も捕捉する。例示的な因果主張としては、NFkB(活性化B細胞の核因子カッパ軽鎖エンハンサー)の転写活性の増大がCXCL1(ケモカイン(C-X-Cモチーフ)リガンド1)のmRNA発現の増加を引き起こすことである[HeLa cell line、Human、PMID1641498

10

20

30

40

50



5 ]。この知識ベースは、健康な組織、ならびに炎症、代謝疾患、心血管傷害、肝臓傷害およびがんなどの疾患領域から導き出された因果関係を包含する。

【 0 1 1 3 】

上記 G P I、E P I および P P I スコア化法は最初に、N F - k B 活性化の識別の尺度になるように創製した因果関係ネットワークモデル ( N F - k B 直接モデル ) を使用して調査した。このモデルは、(異なる 2 4 7 の参考文献から精選され、一部の遺伝子は複数の参考文献によって支持されている) N F - k B によって直接調節されることが公知である 1 5 5 個の遺伝子 ( N F - k B に依存するようにして発現が調節され、N F - k B によってプロモータ配列が直接結合される遺伝子) からなる。両スコア化法で T N F に対する同じパターンの応答が示されて、常時の用量依存性応答と、概してもっと後の時間に飽和した時間依存性応答とが実証された ( 図 1 0 a 参照 )。上記 E P I 法は、E P I スコアが 2 時間から 4 時間まで、さらに 2 4 時間まで連続して増加するのに対して G P I スコアが 4 時間から 2 4 時間まで横ばい状態であるという点で、G P I 法と定性的に異なった。また、上記 E P I 法では、0 . 1 n g / m L の T N F でゼロに近いスコアが生じた。一般に、E P I スコアは、他の方法で比較的低い傾向があったスコアが 0 ( または 0 近く ) まで低減するようであった。上記 E P I 法についての 2 時間の時点以外のすべてについての最少用量は、上記 N F - B 直接ネットワークに特異的ではないことが見出された。

10

【 0 1 1 4 】

次に、N F - B 直接モデルスコアを N F - B 核移行と比較した。活性化すると N F - B は、それが多くの遺伝子の発現を調節するように働く核の中に輸送される。次に、一連のフィードバックループにより、N F - B が元の細胞質へと続いて移行することになり、この振動性のサイクル ( o s c i l l a t t o r y c y c l e ) が数回続く。N F - B 振動が集団内の別々の細胞においてわずかに異なる周期で起こるので、その 1 番目の振動は、N F - B 活性化の最も信頼性の高い集団基準 ( p o p u l a t i o n - m e a s u r e ) になり得る。上記 1 番目の振動の時間は用量に依存するが、T N F 処理後の 3 0 分間は、使用された用量での N F - B 核移行を測定するための現実的な時間であり得る。3 つのすべてのスコア化法により、スコアと核移行の間に単調で、場合によってほぼ線形の関係が得られ、上記 G P I および E P I スコア化法についてのピアソン相関係数は 0 . 8 5 と 0 . 9 8 の間にあった ( 図 1 1 )。図 1 1 は、3 0 分での N F - B 核移行に対してグラフ化した、3 0 分での N F - B 直接 N P A スコアを示す。N F - B 核移行のエラーバーは、同じ細胞集団の異なる 3 つの視野での平均核移行の標準偏差を表す。興味深いことに、この用量依存関係は、T N F 処理後の別々の時間でも保存された ( 図 1 3 )。これらの発見により、上記因果関係ネットワークモデルをベースとする N P A スコアにより N F - B 転写活性を定量化できることが実証される。

20

30

【 0 1 1 5 】

本発明の N P A スコア化法に関する因果関係ネットワークモデルの範囲および構成の効果もまた調査した。まず、N F - B で調整されることが公知である測定値の組を特に T N F に依存するようにして、手作業で選択することの効果の評価した。逆転写酵素ポリメラーゼ連鎖反応 ( R T - P C R ) によってあらかじめ測定した 2 0 個の遺伝子の組から副モデルを構築して、3 T 3 マウス線維芽細胞の T N F 処理に反応する N F - B 活性を評価した ( 直接のヒトオルソログをもたない 2 つの遺伝子は除外 )。これらの遺伝子を、T N F ( 範囲が 1 0 0 n g / m l から 0 . 0 0 5 n g / m l の異なる 1 0 種類の濃度 ) を 1 2 時間の時間経過にわたって投与することで、3 T 3 細胞において T N F によって攪乱されたものとして測定した。この副モデルでは、上記 N F - B 直接モデルと非常に類似した活性化のパターンが生じ ( 図 1 4 )、それによって、T N F 依存性発現が直接検証されていない遺伝子を含むことが上記スコアの質に悪影響を及ぼすことはないと言われた。図 1 4 は、T N F 処理された N H B E 細胞からのトランスクリプトームデータの結果を示し、このデータは、( a ) N F - B 直接モデル、( b ) マウス 3 T 3 線維芽細胞において T N F 応答性であると報告された 2 0 個の N F - B により調節される遺伝子 ( N F K B I A、C A S P 4、C C L 5、T N F A I P 3、C C L 2、Z F P 3 6

40

50

、RIPK2、TNFSF10、NFKBIE、IL6、CCL20、ICAM1、TNFRSF1A、TNFRSF1B、SQSTM1、NRG1、SOD1、IL1RL1、HIF1A、ERBB2)からなる副モデル(Tayら、Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing、Nature、2010年、466巻:267~271頁)、についてGPIおよびEPIを使用してスコア化した。

#### 【0116】

次に、測定部のあまり近位にはない上流の生物学的実体に由来する因果関係ネットワークモデルを使用する効果を調査した。そうするために2つの追加モデル、すなわち、I 10  
Bキナーゼ(IKK)タンパク質からNF- $\kappa$ B活性化に至るシグナル伝達の因果関係ネットワークモデルにおいて、タンパク質の攪乱によって調整されることが公知である99  
2個の遺伝子(異なる414の参考文献から精選されたもの)からなるIKK/NF- $\kappa$ Bシグナル伝達モデルと(図9)、TNFで細胞を処理することによって調整される(modulated)ことが公知である1741個の遺伝子(異なる589の参考文献から精選されたもの)からなるTNFモデルとを構築した。上記NF- $\kappa$ B直接モデルは全体的に、単一の転写因子(NF- $\kappa$ B)によって発現が直接制御された遺伝子からなるの  
20  
に対し、これら2つのモデルそれぞれは、直接の転写コントローラが必ずしも公知ではない遺伝子を含む。これらの遺伝子の発現は、上記モデルの構築に関与しない転写因子によって制御することができる。例えば、上記IKK/NF- $\kappa$ Bシグナル伝達モデルの遺伝子は、上記IKK/NF- $\kappa$ Bシグナル伝達因果関係ネットワークモデルにおけるタンパク質の攪乱によって調整されることが公知であるが、これらの遺伝子の一部は、NF- $\kappa$ Bによって直接調製される遺伝子のより小さいサブセットの発現の変化によって生じる二次効果として調節することができる。また、TNFはリガンドであり、したがって、い  
かなる遺伝子の転写も直接媒介しない。TNFで細胞を処理することにより無数の転写因子が活性化することになり、これらの転写因子のどれもが直接または間接的に(例えば、自己分泌シグナル伝達により)上記TNFモデルの各遺伝子の発現を変化させ得る。

#### 【0117】

図9は、全因果関係ネットワークモデル(上段)を基本モデル構築様式の概略図(中段)と共に示す。CHUK、IKKB、およびIKBKは、NFKBIA、NFKBIB、およびNFKBIEのインヒビターとして働き、それらはさらにはNFKB1、NFKB2、およびRELAのインヒビターでもある。このモデルにおいて使用されるノードは、それぞれのセクションの下に列記されている。太字のノードは、上記知識ベースに下流遺伝子発現の測定可能要素を有するノードを表し、測定可能要素の数は角括弧の中に示されている(同一の下流が複数のノードの下に見出されることがあるので、これら1227個の下流の測定可能要素が992個の固有の測定可能要素に対応する)。使用される表記法は次の通りである。「CHUK P@S」はセリンでリン酸化されたCHUKを表し(その残基が既知であれば示される)、「CHUK P@ST」はセリンまたはスレオニンでリン酸化されたCHUKを表し(その正確な残基は公知ではない)、「kaof(CHUK)」はCHUKのキナーゼ活性を表し、「CHUK:IKKB」はCHUKタン  
40  
パク質とIKKBタンパク質の複合体を表し、「IkappaB kinase complex Hs」は、ホモサピエンス(Hs)の様々なI Bキナーゼ(CHUK、IKKB、およびIKBK)の集合体を表し、「degradation of(NFKBIA)」はNFKBIA分解の過程を表し、「taof(NFKB1)」はNFKB1の転写活性を表す。

#### 【0118】

上記IKK/NF- $\kappa$ Bシグナル伝達モデルおよびTNFモデルは、その測定値に対して異なるレベルの近さにおいて機構仮説の挙動への洞察を与える。上記IKK/NF- $\kappa$ Bシグナル伝達モデルは主として、NF- $\kappa$ Bによって(直接または間接的に)調節される遺伝子からなり(図9)、上記NF- $\kappa$ B直接モデルと非常に類似している応答パター 50

ンを生じる(図10(b))。この類似した応答パターンにより、転写因子によって直接調節されることが公知である遺伝子の集団レベルの挙動と、直接の調節についての知識が公知ではない遺伝子の挙動との間に大きな相違がないことが示唆される。上記NF-B直接モデルについて認められた時間および用量依存性応答は、上記TNFモデルでは、例えば30分の時点において、いくぶん頑強さが少ないようであるが(図10(c))、それでもまたこれらの方法により非常に類似した応答が得られた。こうして、一般的な応答パターンが上記モデル間でよく保たれたが、小さくても注目すべき応答の差異が、測定が行われた実体にあまり近くないモデルで観測され得る。

#### 【0119】

関連するTNFシグナル伝達攪乱に特に応答する上記因果関係ネットワークモデルの能力を評価するために、重要な細胞周期成分である転写因子E2F1についての別のモデルが、E2F1がNF-Bと比較してTNFシグナル伝達のあまり直接的ではないエフェクターであるという仮定を用いて、構築された。上記E2F1直接モデルは、E2F1によって直接調節される(E2F1によって発現が制御され、E2F1によってプロモータ配列が結合される)ことが公知である80個の遺伝子(異なる54の参考文献から精選されたもの)からなる。NF-Bシグナル伝達と直接関係がない生物学に関するNPA結果の比較を行うために、上記で導入された4つのモデル(model)(NF-B-直接、IKK/NF-シグナル伝達、TNF、およびE2F1-直接)のNPA応答を、CDKインヒビターによる細胞周期進行の阻害に応じて評価した。具体的には、CDKインヒビターR547の異なる3つの濃度によるHCT116結腸がん細胞の処理に関する、公に入手可能なマイクロアレイデータセットを使用した(GSE15395)(Berkofsky-Fesslerら、Preclinical biomarkers for a cyclin-dependent kinase inhibitor translate to candidate pharmacodynamic biomarkers in phase I patients、Mol Cancer Ther、2009年、8巻:2517~2525頁)(図12)。3つすべてのNPA法により、4時間、6時間および24時間の時点において、用量および時間に依存するE2F1-直接モデルスコアの減少が実証された。上記TNFモデルは、上記E2F1直接モデルと同様の応答パターンを示した。対照的に、上記NF-B-直接モデルおよびIKK/NF-Bシグナル伝達モデルのスコアは、これと同一の、用量および時間に依存するパターンを呈さず、それによって、これら注目したモデルは潜在的に、細胞周期で調節される遺伝子をほとんど含まないことが示された。

#### 【0120】

##### F. ハードウェア

図15は、生物学的攪乱の影響を定量化するための分散型のコンピュータ化されたシステム1500のブロック図である。上記システム1500のコンポーネントは、図1のシステム100におけるものと同じであるが、該システム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1510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。

#### 【0121】

図16は、図1~10を参照しつつ記載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13のシステム1300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。SRP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1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、1つまたは複数のコンピューティングデバイス1600に実装されうる。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、1つコンピューティン

10

20

30

40

50

グデバイス 1600 内に含めるかまたは備えることができる。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 1600 にまたがって 1 つのコンポーネントおよび 1 つのデータベースを実装することができる。

#### 【0122】

上記コンピューティングデバイス 1600 は、少なくとも 1 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 1610、システムメモリー、および 1 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 1 つのランダムアクセスメモリー (RAM 1602) および少なくとも 1 つのリードオンリーメモリー (ROM 1604) を含むかまたは備える。これらの要素はすべて、中央処理装置 (CPU 1606) と通信して、該コンピューティングデバイス 1600 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 1600 は、多くの異なる方法で構成されうる。例えば、上記コンピューティングデバイス 1600 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 1600 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 1600 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うる。図 10 では、上記コンピューティングデバイス 1600 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。

10

#### 【0123】

上記コンピューティングデバイス 1600 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 1608 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。

20

30

#### 【0124】

上記 CPU 1606 は、1 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 CPU 1606 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 1 つまたは複数の補助コプロセッサを備える。上記 CPU 1606 は、上記通信インターフェースユニット 1608 および上記入力/出力コントローラ 1610 と通信し、これを通じて該 CPU 1606 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 1608 および上記入力/出力コントローラ 1610 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含むかまたは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。

40

#### 【0125】

上記 CPU 1606 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、RAM 1602、ROM 1604、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を含むか備え得る。上記 CPU 1606 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

50

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USB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CPU1606は、上記通信インターフェースユニット1608を介して上記データ記憶デバイスに接続されうる。上記CPU1606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。

【0126】

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1600用のオペレーティングシステム1612、(ii)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CPU1606に関して詳しく記載されているプロセスにより、該CPU1606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1614（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(iii)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）1616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）は、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を含むかまたは備える。

10

【0127】

上記オペレーティングシステム1612およびアプリケーション1614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かまたは備えることができる。上記プログラムの命令は、上記ROM1604または上記RAM1602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CPU1606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、本発明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。

20

【0128】

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1612、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1610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かまたは備えることができる。

30

【0129】

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1600（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ーなどの集積回路メモリーが含まれるかまたは備えられる。揮発性媒体としては、典型的にはメインメモリーを構成するダイナミックランダムアクセスメモリー（DRAM）が含まれるかまたは備えられる。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ー）、FLASH-EEPROM、任意の他の

40

50

メモリー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含まれるかまたは備えられる。

【 0 1 3 0 】

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CPU1606（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ーにロードし、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1600（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上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ーに伝え、上記プロセッサはそのメインメモリーから命令を取り出して実行する。メインメモリーに入った命令は、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ー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。さらなる諸態様および諸実施形態を以下の項目に示す。

1．作用物質に応答した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、該方法は、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、ここで、該生物系が複数の生物学的実体を含むかまたは備え、それぞれの生物学的実体が、該複数の生物学的実体のうちの少なくとも1つの他の実体と相互作用するステップと、第2のプロセッサで、該作用物質に曝露されない生物系に対応するコントロールデータの組を受け取るステップと、第3のプロセッサで、該生物系を表すと共に、該生物学的実体を表すノード、該生物学的実体間の関係を表すエッジ、および該コントロールデータと該処置データの間の変化の予測方向を該ノードについて表す方向値を含むかまたは備える、計算因果関係ネットワークモデルを提供するステップと、第4のプロセッサにより、該処置データと該コントロールデータの間を差を表す活性尺度を該ノードについて計算するステップと、第5のプロセッサにより、少なくとも1つの重み値が少なくとも1つの他の重み値と異なる複数の重み値を該ノードについて計算するステップと、第6のプロセッサにより、該作用物質に対する該生物系の攪乱を表す計算モデルについてのスコアを生成するステップであって、該スコアは該方向値、該重み値および該活性尺度に基づくステップとを含む。

2．それぞれの上記計算モデルにおけるノードの数に基づいて上記スコアを正規化するステップをさらに含む、項目1に記載のコンピュータ化された方法。

3．上記重み値が、処置データおよびコントロールデータの上記組の少なくとも1つにおける信頼度を表す、上記項目のいずれかに記載のコンピュータ化された方法。

4．上記重み値が局所偽非発見率を含む、上記項目のいずれかに記載のコンピュータ化された方法。

5．第7のプロセッサにより、上記活性尺度の近似分布を上記ノードわたって計算するステップと、第8のプロセッサにより、該近似分布の予測値を計算するステップと、第9のプロセッサにより、上記作用物質に対する上記生物系のサブセットの攪乱を表す計算モデルごとに、予測値に基づくスコアを生成するステップとをさらに含む、項目1に記載のコンピュータ化された方法。

6．上記近似分布が上記活性尺度に基づく、項目5に記載のコンピュータ化された方法。

7．予測値を計算するステップが長方形近似を実施するステップを含む、項目5～6のいずれかに記載のコンピュータ化された方法。

8．第10のプロセッサにより、上記活性尺度と上記方向値の間の一貫性および不貫性をそれぞれ表す正の活性化スコアおよび負の活性化スコアを該活性尺度に基づいて計算するステップと、第11のプロセッサにより、上記作用物質に対する上記生物系のサブセッ

10

20

30

40

50

トの攪乱を表す計算モデルごとに、該正の活性化スコアおよび負の活性化スコアに基づくスコアを生成するステップとをさらに含む、項目 1 に記載のコンピュータ化された方法。

9 . 上記スコアが局所偽非発見率に基づく、項目 8 に記載のコンピュータ化された方法。

10 . 上記活性尺度が倍率変化値であり、上記ノードごとの該倍率変化値が、それぞれの該ノードによって表された上記生物学的実体についての上記処置データと上記コントロールデータとの間の差異の対数を含む、項目 8 ~ 9 のいずれかに記載のコンピュータ化された方法。

11 . 上記生物系のサブセットが、細胞増殖機構、細胞性ストレス機構、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、上記項目のいずれかに記載のコンピュータ化された方法。

10

12 . 上記作用物質が、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、または紙巻きタバコの煙のうち少なくとも 1 つを含む、上記項目のいずれかに記載のコンピュータ化された方法。

13 . 上記作用物質が、上記生物系に存在しないし由来もしない分子もしくは実体が含まれる異物を含む、上記項目のいずれかに記載のコンピュータ化された方法。

14 . 上記作用物質が、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、および食物を含む、上記項目のいずれかに記載のコンピュータ化された方法。

15 . 上記処置データの組が処置データの複数の組を含み、各ノードが第 1 の確率分布によって定義される複数の倍率変化値と、第 2 の確率分布によって定義される複数の重み値とを含む、上記項目のいずれかに記載のコンピュータ化された方法。

20

#### 【 0 1 3 1 】

本発明の諸実装を特定の例に関して詳細に示し記載したが、添付の特許請求の範囲に定義された本発明の趣旨および範囲から逸脱することなく、形態および細部の様々な変更を本発明に加えられることが当業者には理解されるはずである。すなわち、本発明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物の意味および範囲に入るすべての変更が包含されるべきものである。

【図1】

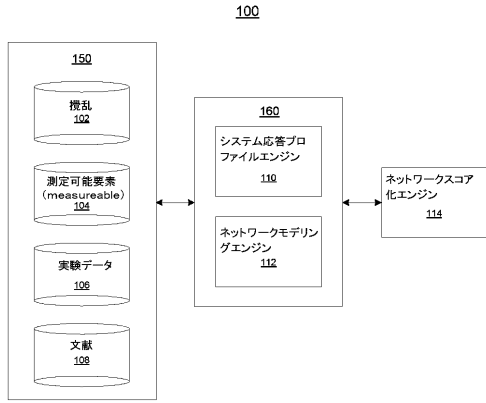


FIG. 1

【図2】

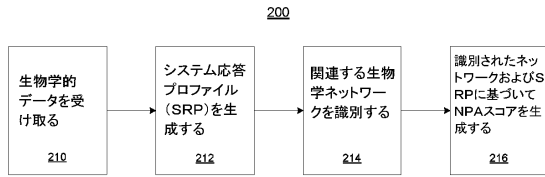


FIG. 2

【図3】

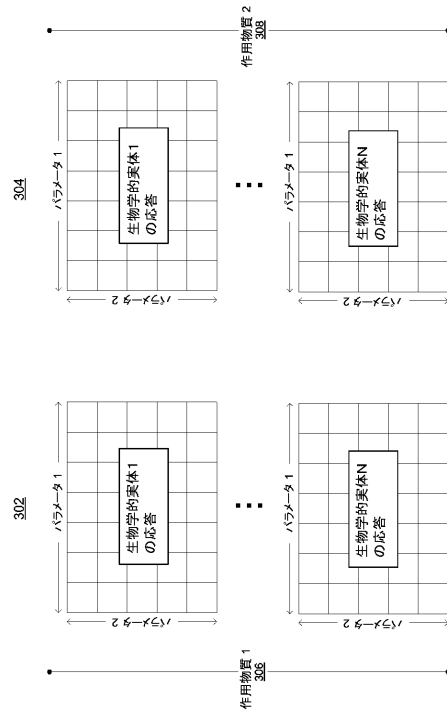


FIG. 3

【図4】

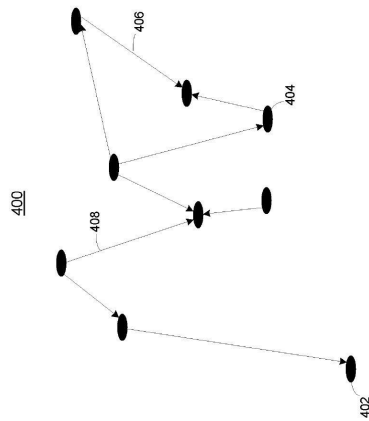


FIG. 4

【図5】

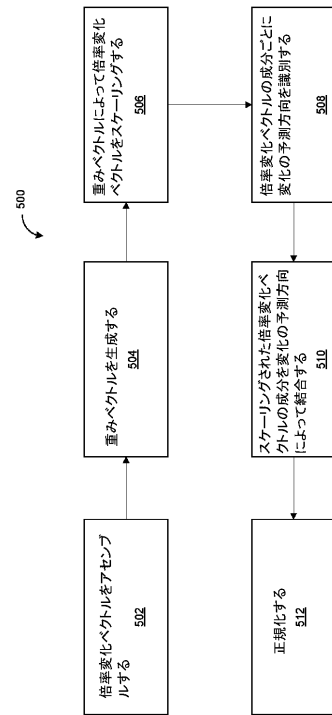


FIG. 5



【 図 6 】

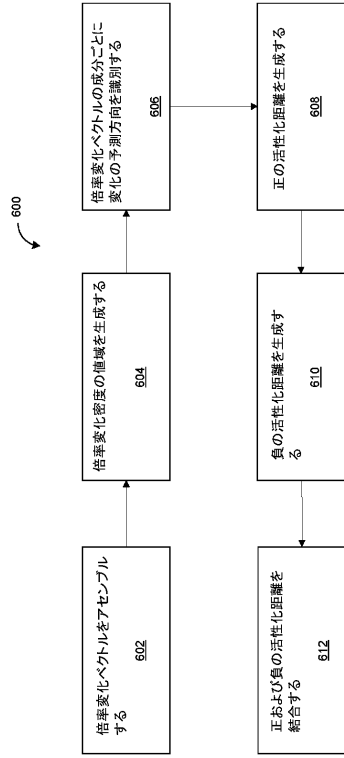


FIG. 6

【 図 7 】

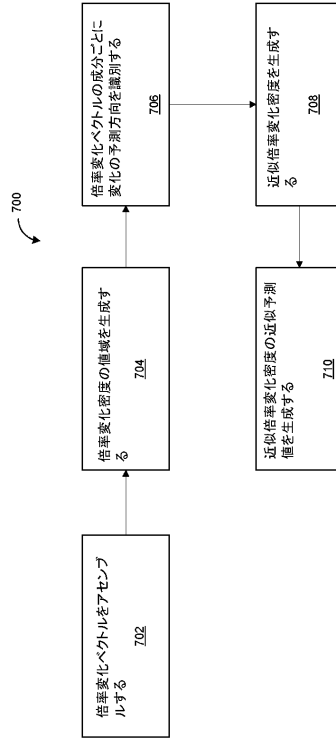


FIG. 7

【 図 8 】

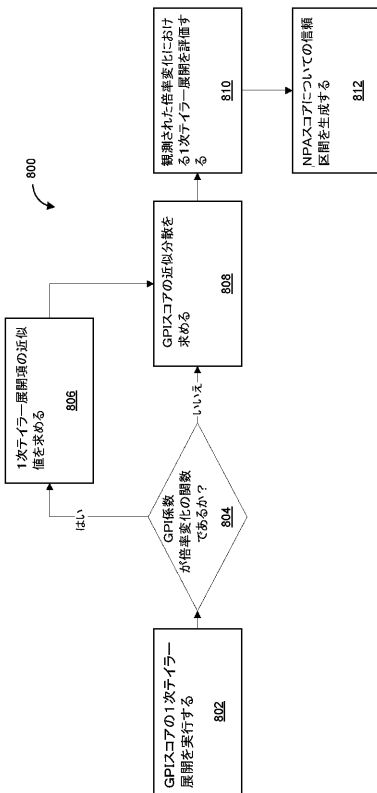


FIG. 8

【 図 9 】

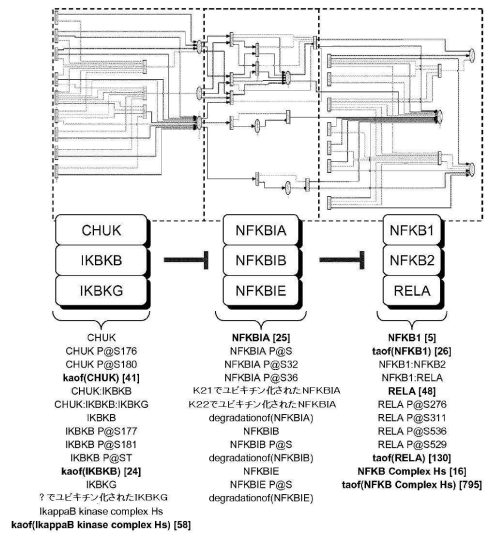


FIG. 9

【 図 1 0 】

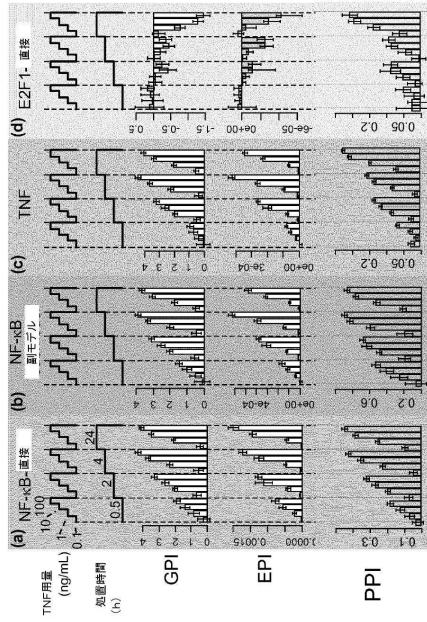


FIG. 10

【 図 1 1 】

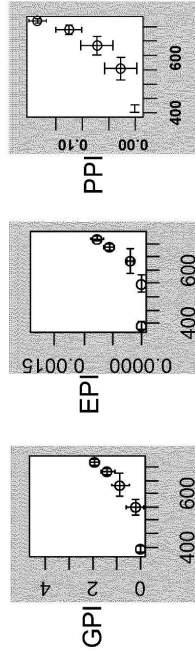


FIG. 11

【 図 1 2 】

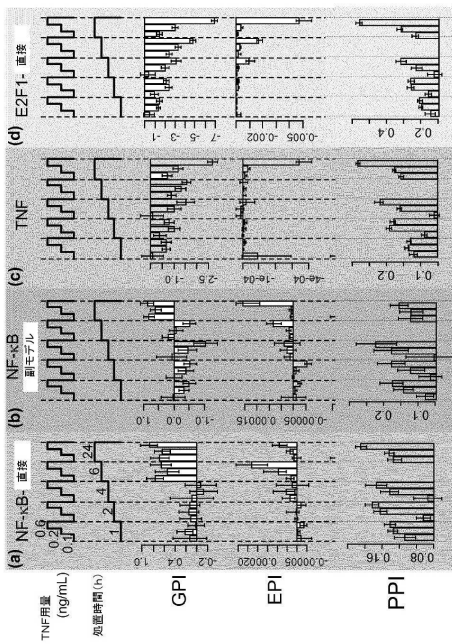


FIG. 12

【 図 1 3 】

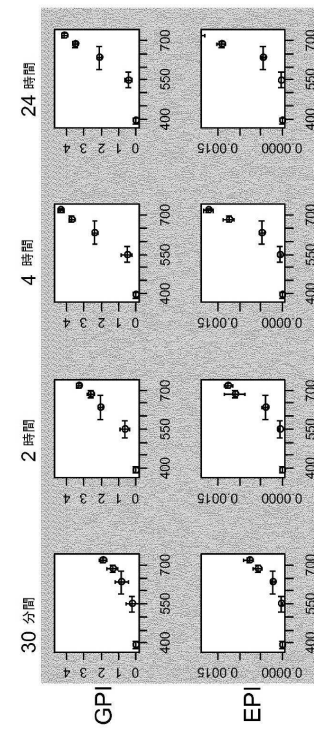


FIG. 13

【 図 1 4 】

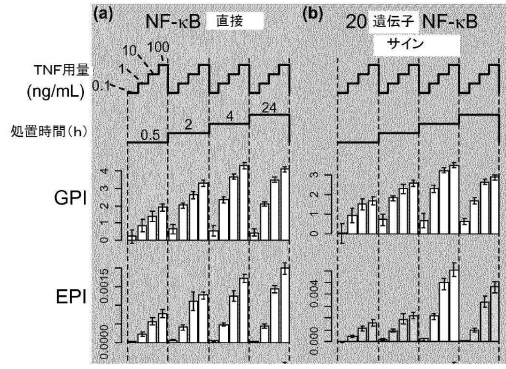


FIG. 14

【 図 1 5 】

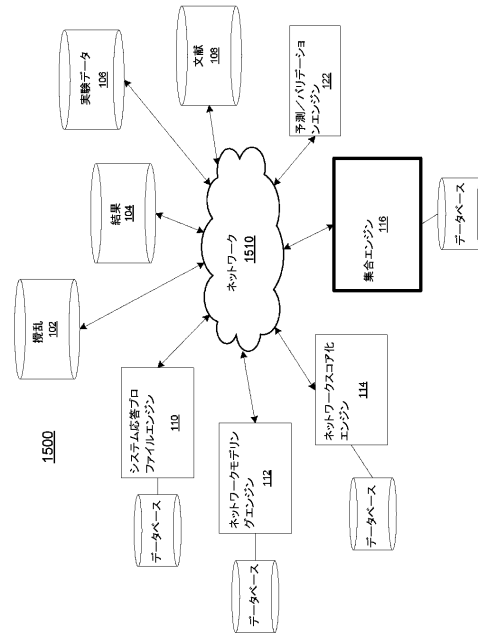


FIG. 15

【 図 1 6 】

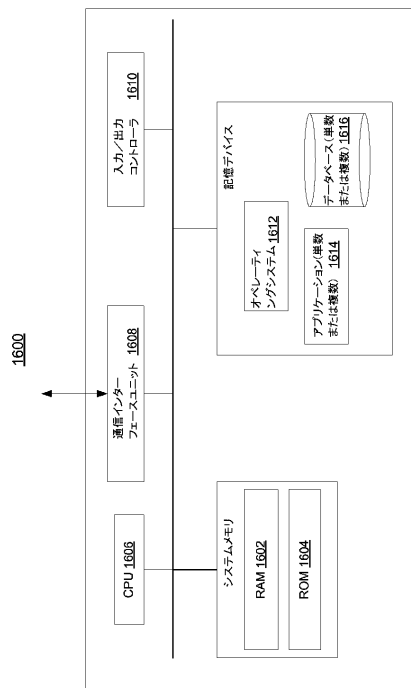


FIG. 16

## フロントページの続き

(31)優先権主張番号 11195417.8

(32)優先日 平成23年12月22日(2011.12.22)

(33)優先権主張国 欧州特許庁(EP)

(72)発明者 ユリア ヘンク

スイス国 ツェーハー - 2 0 0 0 ヌシヤテル, ケ ジャンルノー 3

(72)発明者 フロリアン マルティン

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

(72)発明者 マヌエル パイチ

スイス国 ツェーハー - 2 0 3 4 プスー, シュマン ガブリエル 5

(72)発明者 アラン セーヴェル

スイス国 ツェーハー - 3 9 7 7 グランジュ, リュ ドゥ ラ レシュ 3

審査官 塩田 徳彦

(56)参考文献 米国特許出願公開第2010/0299289(US, A1)

米国特許出願公開第2011/0119259(US, A1)

REVERSE CAUSAL REASONING METHODS - WHITE PAPER, SELVENTA.COM [ONLINE], 2011年 2月 4日, URL, [http://www.selventa.com/attachments/white\\_papers/reverse-causal-reasoning.pdf](http://www.selventa.com/attachments/white_papers/reverse-causal-reasoning.pdf)

PATRICIA GIMENEZ, LOCAL INFLUENCE ANALYSIS BASED ON THE PERTURBATION MANIFOLD IN FUNCTIONAL MEASUREMENT ERROR MODELS, CONGRESO MONTEIRO [ONLINE], 2009年 5月29日, URL, <http://www.matematica.uns.edu.ar/XCongresoMonteiro/actas.HTM>

(58)調査した分野(Int.Cl., DB名)

G 0 6 F 1 9 / 1 0 - 1 9 / 2 8