

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7130630号

(P7130630)

(45)発行日 令和4年9月5日(2022.9.5)

(24)登録日 令和4年8月26日(2022.8.26)

(51)国際特許分類

F I

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00 6 3 0

G 0 1 N 33/50 (2006.01)

G 0 1 N 33/50 Z

G 0 1 N 33/15 (2006.01)

G 0 1 N 33/15 Z

C 1 2 Q 1/02 (2006.01)

C 1 2 Q 1/02

C 1 2 M 1/00 (2006.01)

C 1 2 M 1/00 A

請求項の数 12 (全27頁) 最終頁に続く

(21)出願番号 特願2019-517016(P2019-517016)

(86)(22)出願日 平成29年10月2日(2017.10.2)

(65)公表番号 特表2019-537102(P2019-537102
A)

(43)公表日 令和1年12月19日(2019.12.19)

(86)国際出願番号 PCT/EP2017/075003

(87)国際公開番号 WO2018/060523

(87)国際公開日 平成30年4月5日(2018.4.5)

審査請求日 令和2年8月14日(2020.8.14)

(31)優先権主張番号 62/402,272

(32)優先日 平成28年9月30日(2016.9.30)

(33)優先権主張国・地域又は機関
米国(US)

前置審査

(73)特許権者 511227336

モレキュラー デバイシーズ, エルエル
シーアメリカ合衆国 カリフォルニア 9 5 1
3 4, サン ノゼ, エヌ. ファースト
ストリート 3 8 6 0

(74)代理人 100078282

弁理士 山本 秀策

(74)代理人 100113413

弁理士 森下 夏樹

(72)発明者 フェンテス, エマニュエル・イスラエル
アメリカ合衆国、ウィスコンシン州・5
3 1 8 8-1 6 7 7、ウォキシオー、ノ
ース・グランドビュー・ブルバード、
3 0 0 0

最終頁に続く

(54)【発明の名称】 最適候補化合物を検出するためのコンピュータ装置およびその方法

(57)【特許請求の範囲】

【請求項1】

複数のサンプルとプレートマップ配置とに基づいて最適候補化合物を検出するためのコンピュータ装置(100)のための方法であって、前記複数のサンプルは、セル株と1つ以上のバイオマーカとを含み、前記プレートマップ配置は、1つ以上のバイオマーカに曝露されたセル株を含むサンプルの位置と、少なくとも1つの濃度勾配を形成する異なる濃度の候補化合物とを提供しており、前記候補化合物は、複数の候補化合物に含まれており、前記方法は、

複数の連続的な時点における前記複数の候補化合物のそれぞれの各濃度勾配の表現型プロファイルを生成することにより、複数の化合物プロファイルを形成すること(310)であって、前記表現型プロファイルを生成することは、前記濃度勾配に含まれる各サンプルを表す画像データを取得するステップ(312)と、前記画像データに基づいて前記サンプルの各セルについてクラスラベルおよびクラスを生成するステップ(314)とを含む、こと(310)と、

前記複数の化合物プロファイルについて比較基準を評価することによって、前記最適候補化合物を検出すること(320)と

を含み、

前記最適候補化合物を検出すること(320)は、

1つ以上の基準化合物プロファイルを取得すること(322)と、

前記1つ以上の基準化合物プロファイルに基づいて、前記複数の化合物プロファイルの

各化合物プロファイルについて多次元差分値を計算すること（３２４）と、

前記複数の化合物プロファイルについて比較基準を評価することによって、前記最適候補化合物を検出すること（３２０）であって、前記比較基準は、前記多次元差分値に基づいて評価され、前記比較基準は、損失関数、費用関数、または、最小化基準のうちの少なくとも１つである、こと（３２０）と

をさらに含む、方法。

【請求項２】

前記方法は、

各クラスラベルおよび／またはクラスについて前記サンプルの少なくとも１つのセルを表す画像データの典型的なサブセットを選択すること（３１５）と、

前記画像データの典型的なサブセットとそれぞれのクラスラベルおよび／またはクラスとをユーザに表示すること（３１６）と、

少なくとも１つのクラスについての操作を示すユーザ入力データを前記ユーザから受け取ること（３１７）と、

前記ユーザ入力データに基づいて、前記サンプルのセルの前記クラスラベルおよび／またはクラスについての前記操作を実行すること（３１８）と

をさらに含む、請求項１に記載の方法。

【請求項３】

前記少なくとも１つのクラスについての前記操作は、クラスの追加、クラスの削除、クラスの分割、または、クラスの併合から選択される、請求項２に記載の方法。

【請求項４】

前記画像データは、前記濃度勾配に含まれる各サンプルを複数の視野から表しており、および／または、前記画像データは、複数の画像フィルタを用いて処理された前記濃度勾配に含まれる各サンプルを表している、請求項１～３のいずれか１項に記載の方法。

【請求項５】

複数のサンプルとプレートマップ配置とに基づいて最適候補化合物を検出するためのコンピュータ装置（１００）であって、前記複数のサンプルは、セル株と１つ以上のバイオマーカとを含み、前記プレートマップ配置（２３０）は、１つ以上のバイオマーカに曝露されたセル株を含むサンプルの位置と、少なくとも１つの濃度勾配を形成する異なる濃度の候補化合物とを提供しており、前記候補化合物は、複数の候補化合物に含まれており、

前記コンピュータ装置（１００）は、

プロセッサ（１１２）と、

メモリ（１１５）と

を備え、

前記メモリ（１１５）は、前記プロセッサによって実行可能な命令を含み、これにより、前記コンピュータ装置は、

複数の連続的な時点における前記複数の候補化合物のそれぞれの各濃度勾配の表現型プロファイルを生成することにより、複数の化合物プロファイルを形成することであって、前記表現型プロファイルを生成することは、前記濃度勾配に含まれる各サンプルを表す画像データを取得するステップと、前記画像データに基づいて前記サンプルの各セルについてクラスラベルおよびクラスを生成するステップとを含む、ことと、

前記複数の化合物プロファイルについて比較基準を評価することによって、前記最適候補化合物を検出することと

を行うように動作可能であり、

前記コンピュータ装置は、前記最適候補化合物を検出するときに、

１つ以上の基準化合物プロファイルを取得することと、

前記１つ以上の基準化合物プロファイルに基づいて、前記複数の化合物プロファイルの各化合物プロファイルについて多次元差分値を計算することと、

前記複数の化合物プロファイルについて比較基準を評価することによって、前記最適候補化合物を検出することであって、前記比較基準は、前記多次元差分値に基づいて評価さ

10

20

30

40

50

れ、前記比較基準は、損失関数、費用関数、または、最小化基準のうちの少なくとも1つである、ことと

を行うようにさらに動作可能である、コンピュータ装置（100）。

【請求項6】

前記コンピュータ装置（100）は、入力装置（117）と表示装置（118）とをさらに備え、

前記コンピュータ装置（100）は、

各クラスラベルおよび／またはクラスについて前記サンプルの少なくとも1つのセルを表す画像データの典型的なサブセットを選択することと、

前記画像データの典型的なサブセットとそれぞれのクラスラベルおよび／またはクラスとをユーザに対して前記表示装置（118）上に表示することと、

少なくとも1つのクラスについての操作を示すユーザ入力データを前記入力装置（117）から受け取ることであって、前記ユーザ入力データは、前記ユーザによって示される、ことと、

前記ユーザ入力データに基づいて、前記サンプルのセルの前記クラスラベルおよび／またはクラスについての前記操作を実行することと

を行うようにさらに動作可能である、請求項5に記載のコンピュータ装置（100）。

【請求項7】

前記コンピュータ装置（100）は、前記表現型プロファイルおよび／または化合物プロファイルを前記表示装置（118）上に表示するようにさらに動作可能である、請求項6に記載のコンピュータ装置（100）。

【請求項8】

候補化合物の前記濃度勾配は、複数の別々のウェルを含み、各ウェルは、1つ以上のバイオマーカに曝露された前記セル株のサンプルと異なる濃度の候補化合物とを含み、各ウェルは、前記プレートマップ配置に従って配置されている、請求項5～7のいずれか1項に記載のコンピュータ装置（100）。

【請求項9】

コンピュータ実行可能な命令を含むコンピュータプログラムであって、前記コンピュータプログラムは、前記コンピュータ実行可能な命令がコンピュータ装置（100）に含まれる処理装置上で実行されると、

複数の連続的な時点における複数の候補化合物のそれぞれの各の濃度勾配の表現型プロファイルを生成することにより、複数の化合物プロファイルを形成することであって、前記表現型プロファイルを生成することは、前記濃度勾配に含まれる各サンプルを表す画像データを取得するステップと、前記画像データに基づいて前記サンプルの各セルについてクラスラベルおよびクラスを生成するステップとを含む、ことと、

前記複数の化合物プロファイルについて比較基準を評価することによって、前記複数の候補化合物の中から最適候補化合物を検出することと

を含む方法を実行することを前記コンピュータ装置（100）に行わせ、

前記最適候補化合物を検出することは、

1つ以上の基準化合物プロファイルを取得することと、

前記1つ以上の基準化合物プロファイルに基づいて、前記複数の化合物プロファイルの各化合物プロファイルについて多次元差分値を計算することと、

前記複数の化合物プロファイルについて比較基準を評価することによって、前記最適候補化合物を検出することであって、前記比較基準は、前記多次元差分値に基づいて評価され、前記比較基準は、損失関数、費用関数、または、最小化基準のうちの少なくとも1つである、ことと

をさらに含む、コンピュータプログラム。

【請求項10】

前記方法は、

各クラスラベルおよび／またはクラスについて前記サンプルの少なくとも1つのセルを

10

20

30

40

50

表す画像データの典型的なサブセットを選択することと、

前記画像データの典型的なサブセットとそれぞれのクラスラベルおよび／またはクラスとをユーザに表示することと、

少なくとも1つのクラスについての操作を示すユーザ入力データをユーザから受け取る

ことと、
前記ユーザ入力データに基づいて、前記サンプルのセルの前記クラスラベルおよび／またはクラスについての前記操作を実行することと

をさらに含む、請求項9に記載のコンピュータプログラム。

【請求項11】

前記少なくとも1つのクラスについての前記操作は、クラスの追加、クラスの削除、クラスの分割、または、クラスの併合から選択される、請求項10に記載のコンピュータプログラム。

【請求項12】

コンピュータ読み取り可能な記憶媒体を含むコンピュータプログラム製品であって、前記コンピュータ読み取り可能な記憶媒体には、請求項9～11のいずれか1項に記載のコンピュータプログラムが具現化されている、コンピュータプログラム製品。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、最適候補化合物を検出するためのコンピュータ装置に関する。さらに、本発明は、対応する方法、コンピュータプログラム、およびコンピュータプログラム製品に関する。

【背景技術】

【0002】

例えば医学的用途を意図した化合物など、実験における新規化合物の試験および評価は、典型的には、生物学的文脈においてセル機構を研究して、複数の候補化合物から候補化合物を検出および／または選択することを含み得る。例えばウェルプレートの所定の位置にプレートマップ配置に従って配置された複数のサンプルを含む実験を、純粋に、例えば研究対象のセルの観察可能な特徴または形質などの生物学的表現型への新たな洞察を発見または説明するために、実行することができる。実験から得られる大量の高次元（高密度）のサイトメトリデータセットを解釈することは、例えば統計的な実験の洞察の生成およびセル集団の分類の実行によって、研究者および科学者にとって問題を引き起こす。とくに、サンプル中のセルなどの生物学的な対象物の自動分類のための従来からのツールまたはシステムの欠点は、ユーザが、一変数および二変数可視化を利用して、クラスおよび／またはセル集団領域の間のカットオフ領域および／または決定領域を決定しなければならないことである。さらに、単一の指標を使用し、潜在的には単一の視野（FOV）の反復分析を実行することによって視覚的に実験のサイトメトリデータセットを分析するとき、セル集団間の決定領域の決定は、線形で明確な輪郭であることがまれであり、集団の粒度を理解することが困難であり得る。さらなる手作業のサイトメトリセル指標の調査は、労働集約的であり、どのサイトメトリセル指標が分析において適当であり、どれが集団の分割において最良の能力を含み、したがってヒューマンエラーに対して脆弱であるかを選択することは、困難である。いくつかの従来からのシステムは、視覚によるセルのラベリングおよび教師付きのセル分類を含む2段階の手動プロセスを適用する。これらのシステムの使用は、ユーザがセルをクリックして、それらに手作業でのテキスト入力によって注釈を付けることにより、セルオブジェクトをそれぞれのセル集団の見本として分類することを必要とする。これは、比較的時間がかかる労働集約的なプロセスであるという欠点を有する。これは、通常は、単一の視野を表す画像データを眺めることによって行われ、そのような単一の視野は、用量反応が存在する場合にセル集団の全体を表すことはまれであるため、予測モデリングの可能性が低くなり、したがって入手可能な大規模かつ高密度のデータセットに本来備わっている潜在力が制限されるという欠点を有する。従来からのシス

10

20

30

40

50

テムにおける傾向は、特徴の抽出および／またはサイトメトリの特徴を、それらが人間の理解に束縛されるがゆえに回避し、代わりに、例えばサイトメトリの特徴よりも少ない労力でコンピュータによって抽出することができる指標および／または特徴などの特徴の機械学習に頼ることを試みる。特徴の機械学習の使用の欠点は、それらをユーザが理解または解釈することが困難または不可能であり、したがってユーザによる入力および／または中間データの有効性および／または偏差の評価および排除が、不可能なことである。

【0003】

したがって、従来からの解決策の欠点および問題を軽減または解決する解決策を提供する必要がある。

【発明の概要】

【0004】

本発明の実施形態の目的は、従来からの解決策の欠点および問題を軽減または解決する解決策を提供することである。上記の目的およびさらなる目的は、独立請求項の主題によって達成される。本発明のさらなる好都合な実装形態が、従属請求項によって定められる。

【0005】

本発明の第1の態様によれば、上記の目的および他の目的は、セル株および1つ以上のバイオマーカを含む複数のサンプルに基づいて最適候補化合物を検出するためのコンピュータ装置によって達成される。最適候補化合物の検出は、プレートマップ配置にさらに基づく。プレートマップ配置は、1つ以上のバイオマーカおよび少なくとも1つの濃度勾配を形成する異なる濃度の1つ以上の候補化合物に曝露されたセル株を含むサンプルの位置を提供しており、候補化合物は、複数の候補化合物に含まれており、コンピュータ装置は、プロセッサおよびメモリを備え、前記メモリが、前記プロセッサによって実行することができる命令を含むことにより、このコンピュータ装置は、複数の連続的な時点における複数の候補化合物の各々の各々の濃度勾配の表現型プロファイルを生成して、複数の化合物プロファイルを生成するように動作することができる。コンピュータ装置は、濃度勾配に含まれる各サンプルを表す画像データをさらに取得することによって表現型プロファイルをさらに生成し、画像データに基づいてサンプルの各セルについてクラスラベルおよびクラスを生成するようにさらに動作することができる。さらに、コンピュータ装置は、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するように動作することができる。

【0006】

この態様の利点は、例えば最適候補化合物を検出するための統計的実験の洞察が、化合物プロファイルを生成し、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出することにより、効率的なやり方でプレートマップ配置に関する大きな高次元（高密度）のサイトメトリデータセットを処理することによって改善されることである。さらなる利点は、必要とされる時間および労力を、人間の相互作用を最小限で済ませ、あるいは全く必要とせずに、クラスラベルおよびクラスおよび／またはセル集団領域を生成することによって、低減できることである。

【0007】

第1の態様によるコンピュータ装置の第1の可能な実装形態において、コンピュータ装置は、入力装置および表示装置をさらに備え、コンピュータ装置は、各々のクラスラベルおよび／またはクラスについてサンプルの少なくとも1つのセルを表す画像データの典型的なサブセットを選択し、画像データの典型的なサブセットならびにそれぞれのクラスラベルおよび／またはクラスをユーザへと表示装置上に表示し、少なくとも1つのクラスラベルおよび／またはクラスについての操作を表すユーザによって示されるユーザ入力データを入力装置から受信し、ユーザ入力データに基づいてサンプルのセルのクラスラベルおよび／またはクラスについての操作を実行するようにさらに動作することができる。第1の実装形態の一実施形態において、コンピュータ装置は、クラス追加、クラス削除、クラス分割、またはクラス併合のうちの1つから選択された動作を実行するようにさらに動作可能である。

【0008】

この実装形態の利点は、視覚化の精通および概念的な理解をユーザに提供し、反復的なユーザ修正を可能にし、したがって分類の品質および／または信頼性を向上させることである。

【0009】

第1の態様の第1の実装形態または第1の態様自体によるコンピュータ装置の第2の可能な実装形態において、コンピュータ装置は、さらには、1つ以上の基準化合物プロファイルを取得し、1つ以上の基準化合物プロファイルに基づいて複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算し、複数の化合物プロファイルについて比較基準を多次元差分値に基づいて評価することによって最適候補化合物を検出することによって、最適候補化合物を検出するように動作することができる。この実装形態の一実施形態において、多次元差分値は、好ましくはN次元であり、Nは2以上である。より好ましくは、多次元差分値は、N次元であり、Nは3より大きい。

10

【0010】

この実装形態の利点は、候補化合物の検出の品質が、化合物プロファイルを含む高次元（高密度）データセットにわたる評価によって信頼性および／または品質を改善することによって改善されることである。

【0011】

第1の態様の第1または第2の実装形態あるいは第1の態様自体によるコンピュータ装置の第3の可能な実装形態において、コンピュータ装置は、複数の視野から濃度勾配に含まれる各サンプルを表す画像データを取得するように動作可能であり、さらには／あるいは画像データは、複数の画像フィルタで処理された濃度勾配に含まれる各サンプルを表している。

20

【0012】

この実装形態の利点は、セルが複数の視野を使用して表され、さらには／あるいは複数の画像フィルタを使用して表されるため、表現型プロファイルの品質、したがって化合物プロファイルの品質が、サンプルのセルのクラスラベルおよびクラスを生成するときの信頼性および／または品質の改善によって改善されることである。

【0013】

第1の態様の上述のいずれかの実装形態あるいは第1の態様自体によるコンピュータ装置の第4の可能な実装形態において、コンピュータ装置は、さらには、画像データに基づいて各々のセルについてサイトメトリパラメータセットを生成して、サイトメトリパラメータセットの第1の集合を形成し、パラメータ類似度関数および第1の集合に基づいてサイトメトリパラメータセットをクラスおよび／または表現型クラスにマッピングするように構成された表現型分類モデルを生成することによって、表現型プロファイルを生成するように動作することができ、パラメータ類似度関数は、第1の集合のサイトメトリパラメータセットを互いに相関付けるように構成される。第4の実装形態の一実施形態において、コンピュータ装置は、パラメータ類似度関数および第1の集合に基づいてサイトメトリパラメータセットをクラスにマッピングするように構成された表現型分類モデルをさらに生成することによって表現型プロファイルを生成するように動作することができ、パラメータ類似度関数は、第1の集合のサイトメトリパラメータセットを互いに相関付けるように構成される。

30

40

【0014】

この実装形態の利点は、ユーザによって選択された単一のサイトメトリパラメータおよび／またはサイトメトリセル指標に依存する従来からの方法と対照的に、例えば高次元または高密度のサイトメトリパラメータセットが分類に使用されるため、ヒューマンエラーに対する脆弱性が低減されることである。この実装形態のさらなる利点は、表現型分類モデルの生成されたクラスを連続した実験で再利用できるため、予測モデリングの可能性がさらに向上し、必要な時間および労力がさらに削減されることである。

【0015】

50

第1の態様の上述のいずれかの実装形態あるいは第1の態様自体によるコンピュータ装置の第5の可能な実装形態において、コンピュータ装置は、表現型プロファイルおよび／または化合物プロファイルを表示装置上に表示するようにさらに動作することができる。

【0016】

この実装形態の利点は、中間および最終結果をユーザへと視覚化することによって、ユーザの理解が改善され、したがって検出に必要な時間が短縮されることである。

【0017】

第1の態様の上述のいずれかの実装形態あるいは第1の態様自体によるコンピュータ装置の第6の可能な実装形態において、候補化合物の濃度勾配は、複数の別々のウェルを含み、各々のウェルは、1つ以上のバイオマーカおよび異なる濃度の候補に曝露されたセル株のサンプルを含み、1つ以上のウェルプレート上にプレートマップ配置に従って配置される。

10

【0018】

本発明の第2の態様によれば、上記の目的および他の目的は、複数のサンプルに基づいて最適候補化合物を検出するためのコンピュータ装置のための方法によって達成される。サンプルは、セル株および1つ以上のバイオマーカを含むことができる。最適候補化合物の検出は、プレートマップ配置にさらに基づく。プレートマップ配置は、1つ以上のバイオマーカと、少なくとも1つの濃度勾配を形成する異なる濃度の1つ以上の候補化合物とに曝露されるセル株を含むサンプルの位置を提供している。候補化合物は、複数の候補化合物に含まれる。本方法は、複数の連続的な時点における複数の候補化合物の各々の各々の濃度勾配の表現型プロファイルを生成し、複数の化合物プロファイルを形成することを含み、表現型プロファイルの生成は、濃度勾配に含まれる各々のサンプルを表す画像データを取得するステップと、画像データに基づいてサンプルの各々のセルについてクラスラベルおよびクラスを生成するステップと、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するステップとを含む。

20

【0019】

この態様の利点は、例えば最適候補化合物を検出するための統計的実験の洞察が、化合物プロファイルを生成し、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出することにより、効率的なやり方でプレートマップ配置に関する大きな高次元（高密度）のサイトメトリデータセットを処理することによって改善されることである。さらなる利点は、必要とされる時間および労力を、人間の相互作用を必要とせずにクラスラベルおよびクラスおよび／またはセル集団領域を生成することによって、低減できることである。

30

【0020】

第2の態様による方法の第1の可能な実装形態においては、各々のクラスラベルおよび／またはクラスについてサンプルの少なくとも1つのセルを表す画像データの典型的なサブセットを選択するステップと、画像データの典型的なサブセットならびにそれぞれのクラスラベルおよび／またはクラスをユーザへに表示するステップと、少なくとも1つのクラスについての操作を表すユーザからのユーザ入力データを受け付けるステップと、ユーザ入力データに基づいてサンプルの少なくとも1つのセルのクラスラベルおよび／またはクラスについての操作を実行するステップとがさらに含まれる。第1の実装形態の一実施形態において、少なくとも1つのクラスについての操作は、クラスの追加、クラスの削除、クラスの分割、またはクラスの併合から選択される。

40

【0021】

この実装形態の利点は、視覚化の精通および概念的な理解をユーザに提供し、反復的なユーザ修正を可能にし、したがって分類の品質および／または信頼性を向上させることである。

【0022】

第2の態様の第1の実装形態または第2の態様自体による方法の第2の可能な実装形態において、最適候補化合物を検出するステップは、1つ以上の基準化合物プロファイルを

50

取得することと、1つ以上の基準化合物プロファイルに基づいて複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算することと、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出することとをさらに含み、比較基準は、多次元差分値に基づいて評価される。

【0023】

第2の態様の第1または第2の実装形態あるいは第2の態様自体による方法の第3の可能な実装形態において、画像データは、複数の視野から濃度勾配に含まれる各サンプルを表しており、さらには／あるいは画像データは、複数の画像フィルタで処理された濃度勾配に含まれる各サンプルを表している。

【0024】

この実装形態の利点は、セルが複数の視野を使用して表され、さらには／あるいは複数の画像フィルタを使用して表されるため、表現型プロファイルの品質、したがって化合物プロファイルの品質が、サンプルのセルのクラスラベルおよびクラスを生成するときの信頼性および／または品質の改善によって改善されることである。

【0025】

第2の態様の上述のいずれかの実装形態あるいは第2の態様自体による方法の第4の可能な実装形態において、表現型プロファイルの生成は、画像データに基づいて各々のセルについてサイトメトリパラメータセットを生成して、サイトメトリパラメータセットの第1の集合を形成することと、パラメータ類似度関数および第1の集合に基づいてサイトメトリパラメータセットをクラスおよび／または表現型クラスにマッピングするように構成された表現型分類モデルを生成することとをさらに含み、パラメータ類似度関数は、第1の集合のサイトメトリパラメータセットを互いに相関付けるように構成される。第4の実装形態の一実施形態において、表現型プロファイルの生成は、パラメータ類似度関数および第1の集合に基づいてサイトメトリパラメータセットをクラスにマッピングするように構成された表現型分類モデルを生成することとをさらに含み、パラメータ類似度関数は、第1の集合のサイトメトリパラメータセットを互いに相関付けるように構成される。

【0026】

利点は、ユーザによって選択された単一のサイトメトリパラメータおよび／またはサイトメトリセル指標に依存する従来からの方法と対照的に、サイトメトリパラメータセットが分類に使用されるため、分類時のヒューマンエラーに対する脆弱性が低減されることである。この実装形態のさらなる利点は、表現型分類モデルの生成されたクラスを連続した実験で再利用できるため、予測モデリングの可能性がさらに向上し、必要な時間および労力がさらに削減されることである。

【0027】

本発明の第3の態様によれば、上記の目的および他の目的は、コンピュータ実行可能命令を含んでおり、コンピュータ実行可能命令は、コンピュータ装置に含まれる処理装置において実行されたときに、コンピュータ装置に、複数の連続的な時点における複数の候補化合物の各々の各々の濃度勾配の表現型プロファイルを生成して、複数の化合物プロファイルを形成するステップであって、表現型プロファイルの生成は、濃度勾配に含まれる各サンプルを表す画像データを取得するステップと、画像データに基づいてサンプルの各セルについてクラスラベルおよびクラスを生成するステップとを含むステップと、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するステップとを実行させる、コンピュータプログラムによって達成される。

【0028】

第3の態様によりコンピュータプログラムの第1の可能な実装形態においては、各々のクラスラベルおよび／またはクラスについてサンプルの少なくとも1つのセルを表す画像データの典型的なサブセットを選択するステップと、画像データの典型的なサブセットならびにそれぞれのクラスラベルおよび／またはクラスをユーザへと表示するステップと、少なくとも1つのクラスラベルおよび／またはクラスについての操作を表すユーザからのユーザ入力データを受け付けるステップと、ユーザ入力データに基づいてサンプルのセル

10

20

30

40

50

のクラスラベルおよび／またはクラスについての操作を実行するステップとを、コンピュータ装置にさらに実行させる。

【0029】

第3の態様の第1の実装形態または第3の態様自体によるコンピュータプログラムの第2の可能な実装形態においては、1つ以上の基準化合物プロファイルを取得するステップと、1つ以上の基準化合物プロファイルに基づいて複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算するステップと、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するステップとをさらに実行することによって、最適候補化合物を検出するステップをコンピュータ装置にさらに実行させ、比較基準は多次元差分値に基づいて評価される。

10

【0030】

本発明の第4の態様によれば、上記の目的および他の目的は、コンピュータ可読記憶媒体を含むコンピュータプログラム製品によって達成され、コンピュータ可読記憶媒体は、第3の態様の第1の実装形態または第3の態様自体によるコンピュータプログラムの可能ないずれかの実装形態によるコンピュータプログラムを有する。

【0031】

第2～第4の態様の利点は、第1の態様による対応する方法と同じである。

【0032】

本発明の実施形態のさらなる用途および利点は、以下の詳細な説明から明らかになるであろう。本発明の範囲は、参照によってこの部分に組み込まれる特許請求の範囲によって定義される。本発明の実施形態のより完全な理解、ならびにそのさらなる利点の実現が、1つ以上の実施形態の以下の詳細な説明を検討することによって、当業者にもたらされるであろう。添付の図面を、最初に簡単に説明し、参照する。

20

本発明は、例えば、以下を提供する。

(項目1)

セル株および1つ以上のバイオマーカを含む複数のサンプル、ならびにプレートマップ配置(230)に基づいて、最適候補化合物を検出するためのコンピュータ装置(100)のための方法であって、前記プレートマップ配置(230)は、1つ以上のバイオマーカおよび少なくとも1つの濃度勾配(660)を形成する異なる濃度の候補化合物に曝露されたセル株を含むサンプルの位置を提供しており、前記候補化合物は、複数の候補化合物に含まれており、当該方法は、

30

複数の連続的な時点における前記複数の候補化合物の各々の各々の濃度勾配(660)の表現型プロファイル(710)を生成して、複数の化合物プロファイルを形成するステップ(310)であって、表現型プロファイル(710)の生成は、前記濃度勾配(660)に含まれる各サンプルを表す画像データを取得するステップ(312)と、前記画像データに基づいて前記サンプルの各セルについてクラスラベルおよびクラス(810, 820, 830, 840)を生成するステップ(314)とを含むステップ(310)と、前記複数の化合物プロファイルについて比較基準を評価することによって前記最適候補化合物を検出するステップ(320)と

を含む、方法。

40

(項目2)

各々のクラスラベルおよび／またはクラス(810, 820, 830, 840)について前記サンプルの少なくとも1つのセルを表す画像データの典型的なサブセット(850, 860, 870, 880)を選択するステップ(315)と、

前記画像データの典型的なサブセット(850, 860, 870, 880)ならびにそれぞれのクラスラベルおよび／またはクラス(810, 820, 830, 840)をユーザへと表示するステップ(316)と、

少なくとも1つのクラス(810, 820, 830, 840)についての操作を表すユーザからのユーザ入力データを受け付けるステップ(317)と、

前記ユーザ入力データに基づいて前記サンプルのセルの前記クラスラベルおよび／また

50

はクラス（８１０，８２０，８３０，８４０）についての前記操作を実行するステップ（３１８）と

をさらに含む、項目１に記載の方法。

（項目３）

前記少なくとも１つのクラス（８１０，８２０，８３０，８４０）についての操作は、クラスの追加、クラスの削除、クラスの分割、またはクラスの併合から選択される、項目２に記載の方法。

（項目４）

前記最適候補化合物を検出するステップ（３２０）は、

１つ以上の基準化合物プロファイルを取得するステップ（３２２）と、

前記１つ以上の基準化合物プロファイルに基づいて前記複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算するステップ（３２４）と、

前記複数の化合物プロファイルについて比較基準を前記多次元差分値に基づいて評価することによって前記最適候補化合物を検出するステップ（３２０）と

をさらに含む、項目１乃至３のいずれか１項に記載の方法。

（項目５）

前記画像データは、複数の視野（６５０）から前記濃度勾配（６６０）に含まれる各々のサンプルを表しており、さらには／あるいは前記画像データは、複数の画像フィルタによって処理された前記濃度勾配（６６０）に含まれる各々のサンプルを表している、項目１乃至４のいずれか１項に記載の方法。

（項目６）

セル株を含む複数のサンプル、１つ以上のバイオマーカ、およびプレートマップ配置（２３０）に基づいて、最適候補化合物を検出するためのコンピュータ装置（１００）であって、前記プレートマップ配置（２３０）は、１つ以上のバイオマーカおよび少なくとも１つの濃度勾配（６６０）を形成する異なる濃度の候補化合物に曝露されたセル株を含むサンプルの位置を提供しており、前記候補化合物は、複数の候補化合物に含まれており、当該コンピュータ装置（１００）は、

プロセッサ（１１２）と、

メモリ（１１５）とを備え、前記メモリ（１１５）が、前記プロセッサ（１１２）によって実行することができる命令を含むことにより、当該コンピュータ装置（１００）は、

複数の連続的な時点における前記複数の候補化合物の各々の各々の濃度勾配（６６０）の表現型プロファイル（７１０）を、前記濃度勾配（６６０）に含まれる各々のサンプルを表す画像データを取得するステップと、前記画像データに基づいて前記サンプルの各セルについてクラスラベルおよびクラス（８１０，８２０，８３０，８４０）を生成するステップとを含むステップによって生成して、複数の化合物プロファイルを形成し、

前記複数の化合物プロファイルについて比較基準を評価することによって前記最適候補化合物を検出する

ように動作することができる、コンピュータ装置（１００）。

（項目７）

入力装置（１１７）および表示装置（１１８）をさらに備え、

各々のクラスラベルおよび／またはクラス（８１０，８２０，８３０，８４０）について前記サンプルの少なくとも１つのセルを表す画像データの典型的なサブセット（８５０，８６０，８７０，８８０）を選択し、

前記画像データの典型的なサブセット（８５０，８６０，８７０，８８０）ならびにそれぞれのクラスラベルおよび／またはクラス（８１０，８２０，８３０，８４０）をユーザへと前記表示装置（１１８）上に表示し、

少なくとも１つのクラス（８１０，８２０，８３０，８４０）についての操作を表すユーザによって示されるユーザ入力データを前記入力装置（１１７）から受け付け、

前記ユーザ入力データに基づいて前記サンプルのセルの前記クラスラベルおよび／またはクラス（８１０，８２０，８３０，８４０）についての前記操作を実行する

10

20

30

40

50

ようにさらに動作することができる、項目 6 に記載のコンピュータ装置（１００）。
（項目 8）

前記表現型プロファイル（７１０）および／または化合物プロファイルを前記表示装置（１１８）上に表示するようにさらに動作することができる、項目 6 または 7 に記載のコンピュータ装置（１００）。

（項目 9）

前記最適候補化合物を検出するときに、

1 つ以上の基準化合物プロファイルを取得し、

前記 1 つ以上の基準化合物プロファイルに基づいて前記複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算し、

前記複数の化合物プロファイルについて比較基準を前記多次元差分値に基づいて評価することによって前記最適候補化合物を検出する

ようにさらに動作することができる、項目 6 乃至 8 のいずれか 1 項に記載のコンピュータ装置（１００）。

（項目 10）

候補化合物の前記濃度勾配（６６０）は、複数の別々のウェル（６２０）を含み、各々のウェル（６２０）は、1 つ以上のバイオマーカーおよび異なる濃度の候補化合物に曝露された前記セル株のサンプルを含み、前記プレートマップ配置（２３０）に従って配置されている、項目 6 乃至 9 のいずれか 1 項に記載のコンピュータ装置（１００）。

（項目 11）

コンピュータ実行可能命令を含んでおり、該コンピュータ実行可能命令は、コンピュータ装置（１００）に含まれる処理装置において実行されたときに、該コンピュータ装置（１００）に、

複数の連続的な時点における複数の候補化合物の各々の各々の濃度勾配（６６０）の表現型プロファイル（７１０）を生成して、複数の化合物プロファイルを形成するステップであって、表現型プロファイル（７１０）の生成は、前記濃度勾配（６６０）に含まれる各サンプルを表す画像データを取得するステップと、前記画像データに基づいて前記サンプルの各セルについてクラスラベルおよびクラス（８１０，８２０，８３０，８４０）を生成するステップとを含むステップと、

前記複数の化合物プロファイルについて比較基準を評価することによって前記最適候補化合物を検出するステップと

を実行させる、コンピュータプログラム。

（項目 12）

各々のクラスラベルおよび／またはクラス（８１０，８２０，８３０，８４０）について前記サンプルの少なくとも 1 つのセルを表す画像データの典型的なサブセット（８５０，８６０，８７０，８８０）を選択するステップと、

前記画像データの典型的なサブセット（８５０，８６０，８７０，８８０）ならびにそれぞれのクラスラベルおよび／またはクラス（８１０，８２０，８３０，８４０）をユーザへと表示するステップと、

少なくとも 1 つのクラス（８１０，８２０，８３０，８４０）についての操作を表すユーザからのユーザ入力データを受け付けるステップと、

前記ユーザ入力データに基づいて前記サンプルのセルの前記クラスラベルおよび／またはクラス（８１０，８２０，８３０，８４０）についての前記操作を実行するステップと
を前記コンピュータ装置（１００）にさらに実行させる、項目 11 に記載のコンピュータプログラム。

（項目 13）

前記少なくとも 1 つのクラス（８１０，８２０，８３０，８４０）についての操作は、クラスの追加、クラスの削除、クラスの分割、またはクラスの併合から選択される、項目 12 に記載のコンピュータプログラム。

（項目 14）

前記最適候補化合物を検出するステップを、
1つ以上の基準化合物プロファイルを取得するステップと、
前記1つ以上の基準化合物プロファイルに基づいて前記複数の化合物プロファイルの各
々の化合物プロファイルについて多次元差分値を計算するステップと、
前記複数の化合物プロファイルについて比較基準を前記多次元差分値に基づいて評価す
ることによって前記最適候補化合物を検出するステップと
をさらに実行することによって前記コンピュータ装置（100）にさらに実行させる、
項目11乃至13のいずれか1項に記載のコンピュータプログラム。
（項目15）

コンピュータ可読記憶媒体を含むコンピュータプログラム製品であって、前記コンピュ
ータ可読記憶媒体には項目11乃至14のいずれか1項に記載のコンピュータプログラム
が具現化されているコンピュータプログラム製品。

【0033】

添付の図面は、本発明の種々の実施形態の明確化および説明を意図している。

【図面の簡単な説明】

【0034】

【図1】本発明の実施形態によるコンピュータ装置の図を示している。

【図2】本発明の実施形態によるコンピュータ装置を含む候補化合物検出システムの図を示している。

【図3】本発明の実施形態による複数のサンプルに基づいて最適候補化合物を検出するためのコンピュータ装置のための方法のフロー図を示している。

【図4】本発明のさらなる実施形態による方法のフロー図を示している。

【図5】本発明のさらなる実施形態による方法のフロー図の詳細を示している。

【図6】本発明の実施形態に従って画像生成部によって画像データをどのように生成できるのかを概略的に示している。

【図7】本発明の実施形態に従って表現型プロファイルがどのように表示されるかの一例を示している。

【図8】本発明の実施形態による画像データの典型的なサブセットおよびそれぞれのクラスを表示の一例を示している。

【図9】本発明による方法によって生成された表現型プロファイルと手作業による方法によって生成された表現型プロファイルとの間の比較の一例を示している。

【図10】本発明の実施形態による複数のサンプルに基づいて最適候補化合物を検出するためのコンピュータ装置のための方法のフロー図を示している。

【発明を実施するための形態】

【0035】

本開示の実施形態およびそれらの利点は、以下の詳細な説明を参照することによって、最もよく理解される。1つ以上の図に示されている同等の要素を特定するために、同等の参照番号が使用されることを理解されたい。

【0036】

図1が、本発明の実施形態によるコンピュータ装置100の図を示している。さらなる実施形態において、コンピュータ装置100は、セル実験評価装置、サーバ、据え置き型コンピュータ装置、ラップトップコンピュータ、タブレット型コンピュータ、ハンドヘルドコンピュータ、手首装着型コンピュータ、スマートウォッチ、PDA、スマートフォン、スマートTV、電話機、メディアプレーヤー、ゲーム機、車載コンピュータシステム、またはナビゲーション装置の形態である。いくつかの実施形態において、コンピュータ装置100は、プロセッサおよび／または処理手段112を備え、プロセッサおよび／または処理手段112は、例えば処理回路、中央処理装置、プロセッサモジュール、または互いに協働するように構成された複数のプロセッサなど、プロセッサ／処理ユニットを備え、さらには／あるいはプロセッサ／処理ユニットの形態である。プロセッサおよび／または処理手段に、例えばメモリにて、本明細書に記載の方法の1つ以上の実施形態のステッ

10

20

30

40

50

ブおよび機能を実行するために処理手段112を制御するように構成された特別設計のプログラミングまたはプログラムコード部分を提供することができる。コンピュータ装置100は、プログラミングまたはプログラムコード部分、ならびに／あるいは処理手段112から書き込み信号として受信されたデータ値またはパラメータを記憶するように構成された少なくとも1つのメモリ115またはデータストレージをさらに備えることができる。少なくとも1つのメモリ115またはデータストレージを、データ値および／またはパラメータを取り出し、読み出し信号として処理手段112へと送信するようにさらに構成することができる。コンピュータ装置100は、例えば無線トランシーバおよび／または通信ネットワークアダプタなど、データ値またはパラメータを信号として処理手段112あるいはデータベースまたは画像生成部などの他の外部ノードへと送信し、さらには／あるいはデータ値またはパラメータを信号として処理手段112あるいはデータベースまたは画像生成部などの他の外部ノードから受信するように構成された通信インターフェース116をさらに備えることができる。一実施形態において、通信インターフェース116は、ノード間で直接通信し、あるいは通信ネットワークを介して通信する。1つ以上の実施形態において、コンピュータ装置100は、ユーザから入力または指示を受け付け、ユーザの入力または指示を表すユーザ入力信号を処理手段112へと送信するように構成された入力装置117をさらに備えることができる。1つ以上の実施形態において、コンピュータ装置は、テキストまたはグラフィカルユーザ入力オブジェクトなどのレンダリングされたオブジェクトを表す表示信号を処理手段112から受け取り、受け取った信号をテキストまたはグラフィカルユーザ入力オブジェクトなどのオブジェクトとして表示するように構成された表示装置118をさらに含む。一実施形態において、表示装置118は、ユーザ入力装置117と統合され、テキストまたはグラフィカルユーザ入力オブジェクトなどのレンダリングされたオブジェクトを表す表示信号を処理手段112から受け取り、受け取った信号をテキストまたはグラフィカルユーザ入力オブジェクトなどのオブジェクトとして表示するように構成され、さらには／あるいはユーザから入力または指示を受け付け、ユーザの入力または指示を表すユーザ入力信号を処理手段112へと送信するように構成される。一例において、入力データは、ユーザによって示され、少なくとも1つのクラスラベルおよび／またはクラスに対する操作を表す。いくつかの実施形態において、処理手段112は、メモリ115および／または通信インターフェースおよび／または入力装置および／または表示装置および／または随意によるセンサ（図示せず）に通信可能に接続される。いくつかの実施形態において、通信インターフェース116は、有線および／または無線通信技術を使用して通信する。いくつかの実施形態において、1つ以上のメモリ115は、ハードRAM、ディスクドライブ、フロッピーディスクドライブ、磁気テープドライブ、光ディスクドライブ、CDまたはDVDドライブ（RまたはRW）、あるいは他のリムーバブルまたは固定メディアドライブの選択を含むことができる。

【0037】

一実施形態において、コンピュータ装置100は、例えば実験において最適候補化合物を検出するために提供される。一例において、実験は、セルにおいて、コンピュータ装置によって画像化され、調査中の生物学的プロセスへの洞察を提供すると予測される蛍光タグで標識された1つ以上のバイオマーカを、測定するように構成される。最適候補化合物の検出は、セル株および1つ以上のバイオマーカを含む複数のサンプルに基づくことができる。一例において、複数のサンプルは、哺乳動物起源のセル株を含むことができる。さらに別の例において、セル株は、他の種に由来し、あるいはヒト組織などの組織から抽出されたセルまたは上述の目的でマイクロタイタプレートへともたすために適した任意の他のコンテンツを含むことができる。

【0038】

1つ以上のバイオマーカは、オワンクラゲ（jellyfish Aequorea Victoria）からの緑色蛍光タンパク質（GFP）などの遺伝子組み換えセル株によって生成された蛍光タンパク質、または天然タンパク質を検出するために使用される蛍光標識抗体であってよい。さらなる例においては、「4', 6-ジアミジノ-2-フェニルイ

10

20

30

40

50

ンドール」(DAPI)などの蛍光セル染色剤が、核または核酸に使用され、種々の他のセル浸透性蛍光染色剤を、他のセル構造を標識するために使用することができる。

【0039】

最適候補化合物の検出は、プレートマップ配置230にさらに基づくことができる。プレートマップ配置230は、例えばウェルプレート上に配置され、1つ以上のバイオマーカと、少なくとも1つの濃度勾配660を形成する異なる濃度の1つ以上の候補化合物とに曝露されるセル株を含むサンプルの位置を提供している。一例において、プレートマップ配置は、メモリ115および／または外部ノードおよび／または内部もしくは外部データベースに記憶され、例えばウェルプレートのアイデンティティならびに行および列のペアを表すデータ構造の形態である。さらなる例において、プレートマップ配置は、1つ以上のバイオマーカについて蛍光標識され、さらには／あるいは1つ以上のバイオマーカおよび種々の濃度の1つ以上の候補処理／化合物に曝露されたセル株を含むサンプルの位置を提供しており、候補処理／化合物は、予測された効果または未知の効果を有する天然または合成由来の種々の化合物、ならびに少なくとも1つの濃度勾配を形成する形態で配列され得る化合物を含むことができる。さらに、プレートマップ配置は、ウェルプレート上での候補化合物の位置、および溶媒中の溶質などの候補化合物の濃度を表すことができる。さらなる例において、プレートマップ配置は、複数のウェルプレートにわたるサンプルの位置を表すことができる。1つ以上の候補化合物のサンプルは、プレートマップ配置230に従ってウェルプレート上に配置されてよく、さらには／あるいはプレートマップ配置230に従ったウェルプレート上の位置を有することができる。候補化合物は、例えば特定の目的または効果について試験される多数の候補化合物など、複数の候補化合物に含まれてよい。コンピュータ装置100は、プロセッサ112およびメモリ115をさらに備えることができる。さらに上述したように、前記メモリ115は、前記プロセッサ112によって実行可能な命令を含む。前記プロセッサ112は、処理回路、中央処理装置、プロセッサモジュール、または互いに協働するように構成された複数のプロセッサであってよい。コンピュータ装置100は、複数の連続的な時点における複数の候補化合物の例えば各々の候補化合物の各々の濃度勾配660の表現型プロファイルを生成し、複数の化合物プロファイルを形成するように動作可能であってよい。一例において、濃度勾配の表現型プロファイルは、濃度勾配に含まれる各々のサンプルにおける各クラスおよび／または表現型クラスのセル数を含むことができ、ここで各々のサンプルは異なる濃度の候補化合物を含んでいる。表現型プロファイルは、図9に関連してさらに説明される。さらに、コンピュータ装置100は、濃度勾配に含まれる各サンプルを表す画像データを取得するステップと、画像データに基づいて濃度勾配に含まれるサンプルおよび／または各サンプルの各セルについてクラスラベルおよびクラスを生成するステップと、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するステップとをさらに実行することによって表現型プロファイルを生成するように動作可能であってよい。比較基準は、当業者にとって利用可能な損失関数および／または費用関数および／または最小化基準であってよい。一例において、表現型プロファイルは、候補化合物の濃度勾配660のサンプル内の各クラスについて、セル数を絶対数またはサンプル内のセルの総数のパーセンテージ%として表すデータを含む。さらなる一例において、濃度プロファイルは、クラスについてのセル数および候補化合物の濃度勾配660の濃度を表すデータを含む。さらなる一例において、表現型プロファイルは、複数の濃度プロファイルを含むことができる。一例において、濃度勾配に含まれる各サンプルを表す画像データの取得は、メモリ115からの画像データの取り出しおよび／または受信、外部ノードからの画像データの取り出しおよび／または受信、内部および／または外部データベースからの画像データの取り出しおよび／または受信、あるいは画像生成部からの画像データの取り出しおよび／または受信を含む。一例において、クラスラベルおよびクラスの生成は、セルオブジェクトとして各セルの画像データをセグメント化すること、および／または画像データに基づいてセルオブジェクトとして各セルのサイトメトリパラメータセットを生成することを含む。セルオブジェクトを、データ構造として実装することができる。

10

20

30

40

50

【0040】

使用事例の実施形態において、本開示を、セルの実験的動揺に起因して既知のセル表現型において生じ得る変化を調査するセルのベースライン表現型検査に使用することができる。実験を、純粋に、生物学的表現型、すなわち研究対象のセルの観察可能な特徴または形質への新たな洞察を発見または説明するために行うことができる。実験的動揺の種類として、多くの場合に遺伝子組み換え技術、生物活性分子を発見するための大規模薬物スクリーニングキャンペーン、およびセルベースの毒性試験を用いて、セル内の酵素活性を阻害または促進する効果を観察することを挙げることができる。これらの研究の明確な特徴は、既知の表現型の存在であり、その目的は、セル株中のセルの被処理サンプルにおける表現型プロファイルを、同じセル株の未処理対照サンプルと比較することである。これらの実験は、一般に、マルチウェルプレートにおいて行われるため、物理的な実験のレイアウトを、例えばプレートマップ配置として、観察結果にマッピングする必要がある。この例において、サンプルの表現型プロファイリングは、単一のセルタイプをセル周期状態などの異なるサブクラスへと分類することを必要とし得る。また、セルタイプの混合が存在する可能性があり、セルタイプに特有の分類ツールが必要とされる組織のセルモデルを開発する傾向も高まっている。本開示は、より迅速かつより正確な分類など、ユーザ目的のワークフローの単純化を目的とする半教師付きパイプラインを提供する。本開示は、そのようなデータ特性に基づいてクラスラベルおよび／またはクラスを生成するのに好適なコンピュータ装置およびその方法を活用する。さらなる例においては、画像データに基づいてサンプルの各セルについてクラスラベルおよびクラスを生成することを、k平均などの教師なしクラスタリングおよび／または分類アルゴリズムを使用して、N次元のセルごとのデータおよび／またはサイトメトリパラメータセットの全体における内在クラスを導出することによって、実行することができる。さらなる例においては、例えばワードクラスタリング、階層的クラスタリング、または自己組織化マップなど、教師なしクラスタリング技術のファミリーからの任意のクラスタリングおよび／または分類アルゴリズムを使用することができる。

【0041】

本開示において、サンプルの各セルについてのクラスラベルおよびクラスの生成は、例えば表現ベースのモデル (expression-based model) など、表現型分類モデルの生成を含むことができる。さらに、本開示は、反復的なユーザ修正を可能にするとともに、デフォルトでは最小限のユーザインタラクションしか必要とせず、もしくはユーザインタラクションを必要とせずに、表現ベースのモデルをもたらすことができる。本開示は、候補化合物から製品までの時間の短縮、より正確なデータ操作、ユーザエラーの排除、事前のユーザ知識の必要性の排除、機械学習の一部としての継続的成長および学習、表現型プロファイルライブラリ／レポジトリを作成する能力、知識／リソースを外挿して顧客にとって手頃なソリューションを生み出す能力、およびより高い忠実度のワークフローの提供など、多数の利点を有する。問題の複雑さのさらなる例として、高次元サイトメトリデータに基づくセルの分類が挙げられる。画像由来のサイトメトリデータは、当然ながら、高次元であり、そして体積、量、および／またはストレージ空間が大きい。例えば色などの波長帯など、取得された画像データの各々の画像データチャネルについて、抽出することができる画像の特徴の数はますます増えている。2つの画像データチャネルによる簡単な実験は、約30の測定値を有する。問題の複雑さは、急速に「次元の呪い」の領域にある。控えめな予備的研究において、典型的には、数テラバイト程度のデータセットが生じる。画像の取得および／または捕捉および／または生成において使用される倍率に基づき、単一の視野を含む画像データに、約300～500個のセルオブジェクトおよび／またはセルオブジェクトが含まれる。ウェル内のサンプルを表す約3つの視野を含む画像データ、最も一般的なマイクロタイタープレートおよび／またはプレートマップ配置および／またはウェルプレート実験における96個のウェル、ならびにスクリーンごとに数百のプレートマップ配置および／またはウェルプレートにおいて、全体としてのセルオブジェクトの数は、数百万に達し得る。したがって、候補化合物を検出するために

必要な時間および労力の要求を低減する必要性は、きわめて高い。

【0042】

一実施形態においては、入力装置117および表示装置118をさらに備えるコンピュータ装置100が提供され、コンピュータ装置100は、各々のクラスラベルおよび／またはクラスについてサンプルの少なくとも1つのセルを表す画像データの典型的なサブセットを選択し、画像データの典型的なサブセットならびにそれぞれのクラスラベルおよび／またはクラスを表示装置118においてユーザへと表示し、少なくとも1つのクラスラベルおよび／またはクラスについての動作を表すユーザによって示される入力装置117からのユーザ入力データを受け取り、ユーザ入力データに基づいてサンプルのセルのクラスラベルおよび／またはクラスについての動作を実行するようにさらに動作することができる。第1の実装形態の一実施形態において、コンピュータ装置100は、クラス追加、クラス削除、クラス分割、またはクラス併合のうちの1つから選択される動作を実行するようにさらに動作可能である。一例においては、例えば「死滅」などの同じクラスのセルが、第1および第2のグループおよび／またはクラスとして分類されている。表示装置において画像データの典型的なサブセットを眺めるユーザは、それらが実際には同じクラスに属するべきであることに気付く、第1および第2のグループおよび／またはクラスを併合すべきであると判断し、これを入力装置によって示すことができる。次いで、コンピュータ装置100は、第2のグループ／クラスのクラスおよび／またはクラスラベル情報を第1のグループ／クラスと同一になるように設定し、2つのクラスのセルおよび／またはセルオブジェクトを単一のクラスに効果的に合併させることができる。さらなる一例において、少なくとも1つのセルを表す画像データのサブセットは、2Dサムネイルの形態で表示され、それぞれのクラスラベルおよび／またはクラスは、図8に関してさらに説明される1つ以上のサムネイルを囲む矩形として表示される。

【0043】

一実施形態においては、表現型プロファイルおよび／または化合物プロファイルを表示装置118に表示するようにさらに動作することができるコンピュータ装置が提供される。一例において、表現型プロファイルは、線グラフとして表示され、化合物プロファイルは、ボックスプロットとして表示される。さらなる例において、表現型プロファイルおよび／または化合物プロファイルは、化合物濃度および／または濃度勾配に関し、さらには複数の連続的な時点および／または時間に関して、各々のそれぞれのクラスについてセル数を表示する線グラフプロットとして、表示装置に表示される。表現型プロファイルおよび／または化合物プロファイルを、線プロット、回帰プロット、または表面プロットとして表示することができる。

【0044】

一実施形態においては、最適候補化合物を検出するときに、1つ以上の基準化合物プロファイルを取得し、1つ以上の基準化合物プロファイルに基づいて複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算し、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するようにさらに動作することができるコンピュータ装置が提供され、比較基準は、多次元差分値に基づいて評価される。比較基準は、損失関数および／または費用関数および／または最小化基準であってよい。一例において、比較基準は、観察またはデータが平均から相違する標準偏差の符号付きの数を表す三次元Z値および／または基準である。さらなる例において、基準化合物プロファイルは、90%の「死んでいない」、3%の「死んでいる」、および7%の「その他」のセル数を指示する。化合物プロファイルは、異なる濃度および時間に関してそれぞれのセル数を示す。多次元差分値を、三次元Z値として計算することができ、かつ／または各クラスの差セル数の幾何平均として計算することができる。

【0045】

一実施形態においては、候補化合物の濃度勾配が複数の別々のウェル620を含んでおり、各々のウェル620が、1つ以上のバイオマーカおよび異なる濃度の候補に曝露されたウェルプレート上にプレートマップ配置230に従って配置されたセル株のサンプルを

含むコンピュータ装置が提供される。

【0046】

図2が、本発明の実施形態によるコンピュータ装置100を含む候補化合物検出システム200の図を示している。システム200は、図1に関してさらに説明したように、コンピュータ装置100を含むことができる。システムは、例えばリレーショナルデータベース、オブジェクトリレーショナルデータベース、XMLデータベース、NoSQLデータベース、またはNewSQLデータベースなどのデータベース210をさらに含むことができる。データベース210を、画像データなどのデータを格納し、直接または通信ネットワーク240を介してコンピュータ装置100へと送信するように構成することができる。通信ネットワークは、任意の有線または無線通信ネットワークであってよい。通信ネットワーク240を、これらに限られるわけではないが、ローカルエリアネットワーク（LAN）、メトロポリタンエリアネットワーク（MAN）、グローバル移動体ネットワークシステム（GSM）、拡張データGSM環境（EDGE）、高速ダウンリンクパケットアクセス（HSDPA）、広帯域符号分割多元接続（W-CDMA）、符号分割多元接続（CDMA）、時分割多元接続（TDMA）、Bluetooth（登録商標）、Zigbee（登録商標）、Wi-Fi、ボイスオーバーインターネットプロトコル（VoIP）、LTE Advanced、IEEE 802.16m、WirelessMAN-Advanced、エボルブド高速パケットアクセス（HSPA+）、3GPPロングタームエボリューション（LTE）、モバイルWiMAX（IEEE 802.16e）、ウルトラモバイルブロードバンド（UMB）（以前のEvolution-Data Optimized（E-DO）Rev. C）、シームレスハンドオフ直交周波数分割多重による高速低レイテンシアクセス（Flash-OFDM）、大容量空間分割多元接続（iBurst（登録商標））およびモバイルブロードバンド無線アクセス（MBWA）（IEEE 802.20）システム、高性能無線メトロポリタンエリアネットワーク（HIPERMAN）、ビーム分割多元接続（BDMA）、マイクロ波アクセス用ワールドインターオペラビリティ（Wi-MAX）、ならびに超音波通信、などから選択することができる。システムは、画像生成部220をさらに含むことができる。画像生成部220を、画像データを生成し、直接または通信ネットワーク240を介してコンピュータ装置100へと送信するように構成することができる。システムは、1つ以上のプレートマップ配置230および／またはそれぞれのウェルプレートにさらに含むことができる。画像生成部220を、プレートマップ配置230および／またはウェルプレートに従って配置され、さらには／あるいはプレートマップ配置230および／またはウェルプレート上に配置された1つ以上のサンプルを表す画像データを生成するように構成することができる。プレートマップ配置230および／またはウェルプレートは、サンプルを受け入れるように構成された複数の別々のウェルまたはレセプタクルを表すことができ、さらには／あるいは含むことができる。各々のウェルは、1つ以上のバイオマーカおよび濃度勾配の候補化合物の決定された濃度に曝露されたセル株を含むサンプルを含むことができ、例えばウェルプレート上にプレートマップ配置に従って配置される。

【0047】

1つ以上のプレートマップ配置230および／またはウェルプレートに含まれる複数のウェルは、同じ1つ以上のバイオマーカであってよい1つ以上のバイオマーカおよび少なくとも1つの濃度勾配を形成する候補化合物の異なる濃度に曝露されたセル株のサンプルを含むことができる。一例においては、セル株のセルがウェルに含まれ、さらには／あるいはプレートマップ配置および／またはウェルプレートの第1列に配置されたサンプルが、同じ1つ以上のバイオマーカならびに第1の候補化合物の第1の濃度勾配および／または異なる濃度に曝露される。さらに、プレートマップ配置および／またはウェルプレートの第2の列に配置された各々のウェルに含まれるサンプルのセルは、同じ1つ以上のバイオマーカならびに第2の候補化合物の第2の濃度勾配および／または異なる濃度に曝露される。さらなる例においては、候補化合物の異なる濃度が、複数の列にわたって分布する。さらなる例においては、複数の候補化合物が、複数のプレートマップ配置および／また

10

20

30

40

50

はウェルプレートにわたって分布する。画像生成部220を、例えば候補化合物の異なる濃度を含む各々のウェルに含まれる各々のサンプルを表す各々の濃度勾配を表す画像データを生成および／または捕捉するように構成することができる。一例において、画像生成部220は、可視光デジタルカメラ、赤外線デジタルカメラ、あるいはX線、ラジオグラフィ、超音波、コンピュータ断層撮影（CT）、陽電子放出断層撮影（PET）を含む核医学、および磁気共鳴イメージング（MRI）システムなどの放射線システムの群から選択される。さらなる例において、画像生成部220は、蛍光顕微鏡法および／または伝統的な明視野技術を使用する。画像データを、ピクセル、ボクセル、またはベクトルデータとして表すことができる。画像データは、各々のウェルに含まれるサンプルのセルの各々のサンプルの複数の視野を含むことができ、さらには／あるいは画像データは、複数の画像フィルタで処理された濃度勾配に含まれる各々のサンプルを表すソース画像の複数のバージョンを含むことができる。

10

【0048】

本発明のさらなる態様においては、本開示による方法が提供される。

【0049】

一実施形態においては、複数の連続的な時点における複数の候補化合物の各々の各々の濃度勾配660の表現型プロファイルを生成し、複数の化合物プロファイルを形成すること（310）と、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出すること（320）とを含む方法が提供される。

【0050】

20

図3が、本発明の実施形態による複数のサンプルに基づいて最適候補化合物を検出するためのコンピュータ装置100のための方法のフロー図を示している。サンプルは、少なくともセル株、1つ以上のバイオマーカを含むことができる。サンプルを、プレートマップ配置230上に配置および／またはプレートマップ配置230に従って配置することができ、さらには／あるいはウェルプレート上に配置することができる。プレートマップ配置230は、1つ以上のバイオマーカと、例えばプレートマップ配置230に従って配置された少なくとも1つの濃度勾配660を形成する異なる濃度の1つ以上の候補化合物とに曝露されるセル株を含むサンプルの位置を提供することができる。候補化合物は、例えば実験の多数の化合物部分など、複数の候補化合物に含まれてよい。一実施形態において、本方法は、複数の連続的な時点における複数の候補化合物の各々の各々の濃度勾配660の表現型プロファイルを生成し、複数の化合物プロファイルを形成すること（310）を含み、表現型プロファイルの生成は、濃度勾配に含まれる各々のサンプルを表す画像データを取得するステップ312と、画像データに基づいてサンプルの各々のセルについてクラスラベルおよびクラスを生成するステップ314と、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するステップ320とを含む。

30

【0051】

図4が、本発明のさらなる実施形態による方法のフロー図を示している。この方法は、各々のクラスラベルおよび／またはクラスについてサンプルの少なくとも1つのセルを表す画像データの典型的なサブセットを選択するステップ315と、画像データの典型的なサブセットならびにそれぞれのクラスラベルおよび／またはクラスをユーザへと表示するステップ316と、少なくとも1つのクラスについての操作を表すユーザからのユーザ入力データを受け付けるステップ317と、ユーザ入力データに基づいてサンプルのセルのクラスラベルおよび／またはクラスについての操作を実行するステップ318とをさらに含む。第1の実装形態の一実施形態において、少なくとも1つのクラスについての操作は、クラスの追加、クラスの削除、クラスの分割、またはクラスの併合から選択される。一例において、画像データの典型的なサブセットは、画像データがクラスをどの程度良好に表しているかについての信頼度を計算し、典型的なサブセットを最も高い信頼度を有する画像データとして選択することによって選択される。

40

【0052】

図5が、本発明のさらなる実施形態による方法のフロー図の詳細を示している。一実施

50

形態において、この方法の最適候補化合物を検出するステップ320は、1つ以上の基準化合物プロファイルを取得するステップ322と、1つ以上の基準化合物プロファイルに基づいて複数の化合物プロファイルの各々の化合物プロファイルについて多次元差分値を計算するステップ324と、複数の化合物プロファイルについて比較基準を評価することによって最適候補化合物を検出するステップ320とをさらに含み、比較基準は、多次元差分値に基づいて評価される。

【0053】

一実施形態において、画像データは、複数の視野650からの濃度勾配660に含まれる各サンプルを表しており、さらには／あるいは画像データは、複数の画像フィルタで処理された濃度勾配660に含まれる各サンプルを表している。

10

【0054】

一実施形態において、表現型プロファイルの生成は、画像データに基づいて各々のセルについてサイトメトリパラメータセットを生成して、サイトメトリパラメータセットの第1の集合を形成することをさらに含む。一実施形態において、表現型プロファイルの生成は、パラメータ類似度関数および第1の集合に基づいてサイトメトリパラメータセットをクラスにマッピングするように構成された表現型分類モデルを生成することをさらに含み、パラメータ類似度関数は、第1の集合のサイトメトリパラメータセットを互いに関連付けるように構成される。

【0055】

一例において、類似度関数は、相関関数であり、より好ましくは2つ以上のサイトメトリパラメータセットにわたって動作するように構成された多次元相関関数である。さらなる例において、類似度関数は、例えば自己組織化マップ、自動エンコーダ、Wardクラスタリング、K平均クラスタリング、t-SNE次元削減など、機械学習技術から選択される。

20

【0056】

一実施形態において、表現型プロファイルの生成は、各々のセルについてサイトメトリパラメータセットに基づいて偏差度および／または除外基準を表示することと、ユーザ入力データを受け取ることと、画像データが有効(viable)でなく、あるいは残りの画像データセットから実質的に逸脱していることを示すユーザ入力データに基づいてセルを表す画像データを除外することとをさらに含む。一例において、偏差度および／または除外基準は、サイトメトリパラメータセットに含まれるサイトメトリパラメータの平均に対する標準偏差または距離指標である。この実施形態の利点は、入力および／または中間データの有効性および／または偏差をユーザが評価できることである。

30

【0057】

一例においては、画像データおよび／または蛍光画像が、サイトメトリ測定値および／またはサイトメトリパラメータセットに関連するセルオブジェクトへと変換および／または生成される。次いで、ユーザは、手動でデータを検索して、実験、ならびにセル株、化合物、および異なる濃度などのさまざまな変化に起因して存在する反応について、ベースラインの理解を構築する。加えて、視野を含む特定の画像データについて、焦点が合っておらず、あるいは画像データにアーチファクトが存在し、画像処理特徴抽出および／または各セルについてのサイトメトリパラメータセットの生成に影響が及ぶ場合がある。セルについてのクラスラベルおよびクラスを生成する前に、有効でなく、かつ／または残りのデータセットから実質的に逸脱している実験の局面を、排除することが重要である。これらの考察を念頭に、本発明の発明者は、主成分分析、分散しきい値を用いた特徴選択、例えばピアソン(Pearson)およびスピアマン(Spearman)などの相関統計、制御基準正規化、オブジェクト特徴ベクトルプロファイル、階層的クラスタリング、およびt分布確率的近隣埋め込み(t-SNE)の選択に基づいて、一組の分析を開発した。

40

【0058】

図6が、画像生成部220によって画像データをどのように生成できるのかを概略的に示している。一実施形態においては、複数のウェル620が、プレートマップ配置230

50

および／またはウェルプレートに含まれる。複数のウェル620の組の各々は、同じ1つ以上のバイオマーカであってよい1つ以上のバイオマーカおよび少なくとも1つの濃度勾配660を形成する候補化合物の異なる濃度に曝露されたセル株のサンプルを含むことができる。画像生成部220を、1つ以上の濃度勾配660の各々を表し、例えば少なくとも1つの濃度勾配660を形成する候補化合物の異なる濃度を含む各々のウェルに含まれるサンプルのセルの各々のサンプルを個別に表す画像データを生成および／または捕捉するように構成することができる。一例において、ウェル620を表す画像データは、複数の視野650からセルのサンプルを捕捉し、さらには／あるいは表す複数の表現および／または画像を含むことができる。さらなる一例において、ウェル620を表す画像データは、セルのサンプルを表す1つ以上のソース表現および／またはソース画像と、ローパス、ハイパス、平均化、またはカラー画像フィルタ、あるいは当業者にとって利用可能な任意の他の画像フィルタなどの複数の画像フィルタによる処理によって生成された複数の処理済み表現および／または処理済み画像とを含むことができる。さらなる一例において、ウェル620を表す画像データは、超音波などの音、あるいはラジオ、マイクロ波、赤外線、可視光、紫外線、X線、およびガンマ線などの電磁放射線の異なる波長を表す異なる画像データチャネル610内のセルのサンプルを捕捉し、さらには／あるいは表す複数の表現および／または画像を含むことができる。

【0059】

図7が、本発明の実施形態に従って表現型プロファイルが例えば円グラフトレリス (pie chart trellis) としてどのように表示されるかの一例を示している。一実施形態においては、表現型プロファイルが、プレートマップ配置230および／またはウェルプレートのレイアウト上に重ねて表示される。一実施形態においては、プレートマップ配置230および／またはウェルプレートの各々のウェルの各々のサンプルの表現型プロファイル710が表示される。一例において、濃度Aを有する化合物1の円グラフトレリスは、「死んでいる」として分類された8個のセル、「死んでいない」として分類された12個のセル、および「その他」として分類された6個のセルを示す円グラフとして示されている。さらなる実施形態において、表現型プロファイル710は、列720および行730に配置される。さらなる実施形態においては、濃度勾配が、プレートマップ配置230および／またはウェルプレートの列720のウェルに沿って配置される。

【0060】

一例において、表現型プロファイルは、プレートのレイアウトを横切る円グラフトレリスとしての視覚化ダッシュボードとして表示される。利点は、プレートマップ上に配置されたサンプル間で集団がどのように変化したかをユーザがすばやく理解できることである。粒度を上げるために、ウェルごとの比較が、分析の良好なソースとなり得る。列を横切るウェルは、異なる濃度および／または薬物投与量において異なる傾向にあり、行を横切るウェルは、候補化合物および／または薬物の種類において異なる傾向にあるからである。これらの比較は、分類後の分析において絶対的に重要である。本発明の発明者は、生物学者が、或るクラスのセルが特定の薬物の下で生き延びたか否か、あるいはそれらの何パーセントが特定の用量の下で生き延びたかを知りたいと思うことに気が付いた。さらなる例において、ユーザは、表現型プロファイルを眺め、画像データが有効でなく、さらには／あるいは残りのデータセットから実質的に逸脱していることを示すユーザ入力データをもたらすことができ、したがって価値のない特徴を除外する能力をユーザに提供することができる。さらなる一例においては、ユーザが特定の集団を分析しており、その集団がエリアに関してどのように変化したかを知りたい場合に、その特定の傾向を調査するために、特徴のないすべてのエリアを除外することができる。

【0061】

図8が、本発明の実施形態による画像データの典型的なサブセットおよびそれぞれのクラスの表示の一例を示している。一例においては、各々のクラスについてサンプルの少なくとも1つのセルを表す画像データの典型的なサブセットが選択されており、それぞれのクラスラベルおよび／またはクラス810～840と共にユーザへと表示される。とくに

は、第1のクラス810についてサンプルの3つのセル850を表す画像データの典型的なサブセットが選択されており、それぞれのクラスラベルおよび／またはクラス810と共にユーザへと表示される。次いで、ユーザは、少なくとも1つのクラス810～840についての操作を指示することができ、コンピュータ装置100は、本開示によれば、ユーザ入力データに基づいてセル株のセルのクラスラベルおよび／またはクラスについて操作を実行する。操作は、「クラスの追加」、「クラスの削除」、「クラスの分割」、または「クラスの併合」から選択されてよい。セル周期の動揺を含むさらに別の例においては、単一のよく特徴付けられたセル株が、セル分裂を妨害し、あるいはセルの死を引き起こすことによってセル周期の進行を動揺させる異なる濃度の2つの異なる候補化合物で処理され、さらには／あるいはそのような候補化合物に曝露されている。この研究は、セル周期状態による既知の部分集団へのセルの分類、ならびに例えば「死んでいる」、「死んでいない」、または「その他」の健康状態を示す死亡セルについての蛍光バイオマーカの存在に基づくセル健康状態の定義の両方を必要とする。本開示の方法の適用の結果として、画像データの第1の典型的なサブセット850が、それぞれの第1のクラスラベルおよび／またはクラス810と共に「死んでいない」と表示される。さらに、画像データの典型的な第2のサブセット860が、それぞれの第2のクラスラベルおよび／またはクラス820と共に「死んでいる」と表示される。さらに、画像データの典型的な第3のサブセット870が、それぞれの第3のクラスラベルおよび／またはクラス830と共に「死んでいない」と表示される。さらに、画像データの典型的な第4のサブセット880が、それぞれの第4のクラスラベルおよび／またはクラス840と共に「その他」と表示される。ユーザが「死んでいない」第1および第3のクラスを同じクラスとして分類すべきであると認識した後に、ユーザは、それら2つを結合／併合し、第1および第3のクラスラベルおよび／またはクラスのすべてのセルを第1のクラスラベルおよび／またはクラスを示すように更新することを、選択することができる。

【0062】

1つの使用事例においては、蛍光タンパク質および／またはフルオロフォアの形態の2つのバイオマーカを使用してセルを染色することができ、DAPIを核または核酸またはNucleiを染色するために使用し、Cy5を死んだセルを染色するために使用することができる。これらの染色に基づいて、サンプルの各セルについてクラスラベルおよびクラスを生成するために本発明を使用し、さらには／あるいは教師なしクラスタリングを使用したとき、1つの可能なクラスタリング結果は、以下の通りである。

【0063】

クラスタ／クラス1：DAPI：核の明るさ。丸み。Cy5：明るさなし。生きているセルを示している。

【0064】

クラスタ／クラス2～3：DAPI：ほのかな明るさ。核の不規則な形状。Cy5：きわめて明るい。死んだセルを示している。クラスタ4～5：DAPI：中程度の明るさ。長円形。Cy5：明るい。死んだセルを示している。

【0065】

クラスタ／クラス6：DAPI：きわめて明るい。小核の均一な明るさ。Cy5：明るさなし。生きているセルを示している。

【0066】

クラスタ／クラス7：DAPI：明るさの不均一な広がり。不均一なラベリング。Cy5：軽いラベリング。ノイズ。

【0067】

これらのクラスタの各々についての典型的なサブセットが、ユーザへと提示され、ユーザは、クラスタ1、6、7をそのまま維持し、クラスタ2、3、4、および5を組み合わせることを選択することができる。この決定は、画像データの典型的なサブセットに基づいて行われるが、再グループ化は、データセット全体に対して決定に基づいて適用される。

【0068】

10

20

30

40

50

図 9 が、本発明による方法によって生成された表現型プロファイルと手作業による方法によって生成された表現型プロファイルとの間の比較の一例を示している。表現型プロファイルは、濃度勾配の各々の濃度に曝露された各々のサンプルについて特定のクラスに属すると分類されたセルのセル数についてのデータ点を含むことができる。表現型プロファイルは、複数のクラスおよび／または複数の濃度勾配についてのデータ点を含むことができる。上方の図 9 10 に、本発明の方法によって第 1 のクラスに分類されたセルの数が、第 1 の曲線 9 14 にて示されており、手作業による方法によって第 1 のクラスに分類されたセルの数が、第 2 の曲線 9 12 にて示されている。下方の図 9 20 に、本発明の方法によって第 2 のクラスに分類されたセルの数が、第 1 の曲線 9 24 にて示されており、手作業による方法によって第 2 のクラスに分類されたセルの数が、第 2 の曲線 9 22 にて示されている。曲線は、クラスに分類されたセルのセル数を、候補化合物の濃度に対して示している。この例において、本発明の方法は、手作業による方法と 7 % 異なっていたが、分類を 24 倍速く達成した。図 7 に関する例において、各々が濃度勾配 [濃度 A、濃度 B、・・・、濃度 H] を有する 8 つの候補化合物 [化合物 1、化合物 2、・・・、化合物 8] による実験が、本開示の方法およびコンピュータ装置によって処理される。複数の連続的な時点のうちの第 1 の時点 T_0 において、表現型プロファイルが、各々の候補化合物の各々の濃度勾配について、例えば以下のように生成される。

【0069】

化合物 1、濃度 A、8 (死んだクラス)、12 (死んでいないクラス)

化合物 1、濃度 B、4 (死んだクラス)、16 (死んでいないクラス)

...

化合物 1、濃度 H、2 (死んだクラス)、18 (死んでいないクラス)

図 10 が、本発明の実施形態による複数のサンプルに基づいて最適候補化合物を検出するためのコンピュータ装置 100 のための方法のフロー図を示している。一実施形態において、この方法は、オブジェクトサンプリング 1010 および／または画像データからのセルオブジェクトのセグメント化を含む。一実施形態において、この方法は、各々のセルオブジェクトについてそれぞれの対応するサイトメトリパラメータを生成することをさらに含む。この方法は、教師なしクラスタリングアルゴリズム 1030 を実行して、内在グルーピング 1040 を生成すること、ならびに／あるいは画像データに基づいて各々のセルオブジェクトについてクラスラベルおよび／またはクラスおよび／またはグループを生成することを、さらに含むことができる。この方法は、各々のクラスラベルおよび／またはクラスについてサンプルの少なくとも 1 つのセルオブジェクトおよび／または 1 つのセルを表す画像データの典型的なサブセットを見本サンプリングおよび／または選択することをさらに含むことができる。この方法は、モデルグループ化および／または少なくとも 1 つのクラスについての操作を表すユーザからのユーザ入力データの受信、ならびにユーザ入力データに基づくサンプルのセルのクラスラベルおよび／またはクラスについての操作の実行をさらに含むことができ、少なくとも 1 つのクラスについての操作は、クラスの追加、クラスの削除、クラスの分割、またはクラスの併合から選択される。随意により、この方法は、例えば表現ベースのモデルなど、表現型分類モデルの生成をさらに含むことができる。

【0070】

さらに、本発明の実施形態による任意の方法を、処理手段によって実行されたときに処理手段に本方法のステップを実行させるコード手段を有するコンピュータプログラムにて実装することができる。コンピュータプログラムは、コンピュータプログラム製品のコンピュータ可読媒体および／またはメモリに含まれる。コンピュータ可読媒体は、ROM (読み出し専用メモリ)、PROM (プログラム可能読み出し専用メモリ)、EPROM (消去可能 PROM)、フラッシュメモリ、EEPROM (電氣的消去可能 PROM)、またはハードディスクドライブなど、基本的にあらゆるメモリを含むことができる。

【0071】

さらに、コンピュータ装置 100 が、本解決策を実行するために、例えば機能、手段、

ユニット、要素などの形で必要な通信機能を備えることを、当業者であれば理解できるであろう。他のそのような手段、ユニット、要素、および機能の例は、本解決策を実行するために互いに適切に配置されるプロセッサ、メモリ、バッファ、制御ロジック、エンコーダ、デコーダ、レートマッチャー、デレートマッチャー、マッピングユニット、乗算器、決定ユニット、選択ユニット、スイッチ、インターリーバ、デインターリーバ、変調器、復調器、入力、出力、アンテナ、増幅器、受信機ユニット、送信機ユニット、DSP、MSD、TCMエンコーダ、TCMデコーダ、電源ユニット、給電装置、通信インターフェース、通信プロトコル、などである。

【0072】

とくに、本コンピュータ装置100のプロセッサは、例えば、中央処理装置(CPU)、処理装置、処理回路、プロセッサ、特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)、マイクロプロセッサ、または命令を解釈し実行することができる他の処理ロジックのうちの1つ以上のインスタンスを含むことができる。したがって、表現「プロセッサ」は、例えば上述のものの中のいずれか、いくつか、またはすべてなど、複数の処理回路を備える処理回路を表すことができる。処理回路は、呼処理制御、ユーザインターフェース制御、などのデータバッファリングおよびデバイス制御機能を含むデータの入力、出力、および処理のためのデータ処理機能をさらに実行することができる。

10

【0073】

最後に、本発明が上述の実施形態に限定されず、添付の独立請求項の範囲内のすべての実施形態にも関連し、それらを包含することを、理解すべきである。

20

【符号の説明】

【0074】

- 100 コンピュータ装置
- 112 プロセッサ、処理手段
- 115 メモリ
- 116 通信インターフェース
- 117 ユーザ入力装置
- 118 表示装置
- 200 候補化合物検出システム
- 210 データベース
- 220 画像生成部
- 230 プレートマップ配置
- 240 通信ネットワーク
- 610 画像データチャネル
- 620 ウェル
- 650 視野
- 660 濃度勾配
- 710 表現型プロファイル
- 720 列
- 730 行
- 810 第1のクラスラベルおよび／またはクラス
- 820 第2のクラスラベルおよび／またはクラス
- 830 第3のクラスラベルおよび／またはクラス
- 840 第4のクラスラベルおよび／またはクラス
- 850 セル、第1の典型的なサブセット
- 860 第2のサブセット
- 870 第3のサブセット
- 880 第4のサブセット
- 912 第2の曲線

30

40

50

- 9 1 4 第 1 の曲線
- 9 2 2 第 2 の曲線
- 9 2 4 第 1 の曲線
- 1 0 1 0 オブジェクトサンプリング
- 1 0 3 0 クラスタリングアルゴリズム
- 1 0 4 0 内在グルーピング

【図面】

【図 1】

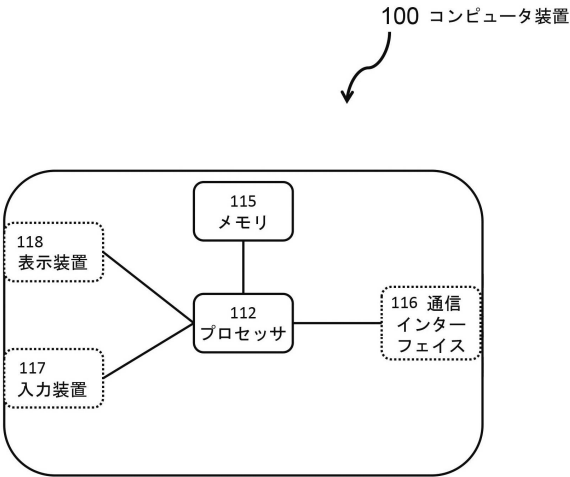


Fig. 1

【図 2】

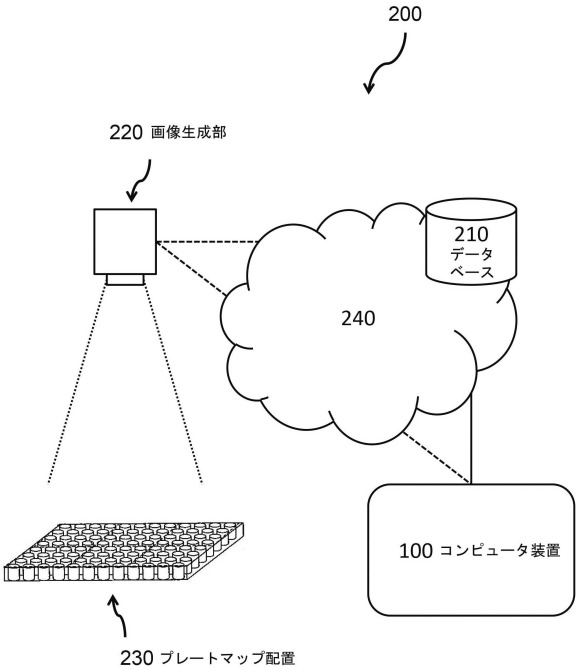


Fig. 2

10

20

30

40

50

【図 3】

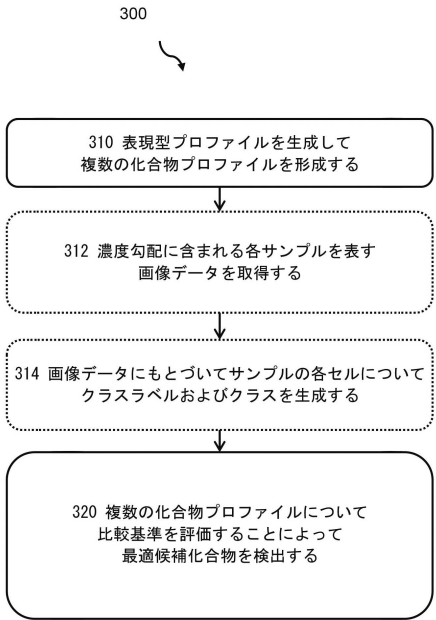


Fig. 3

【図 4】

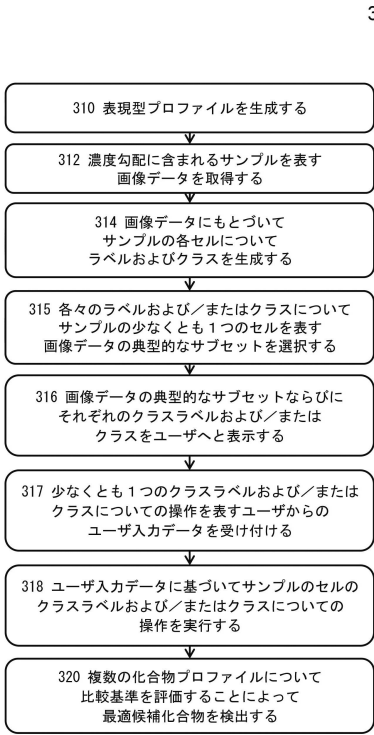


Fig. 4

【図 5】

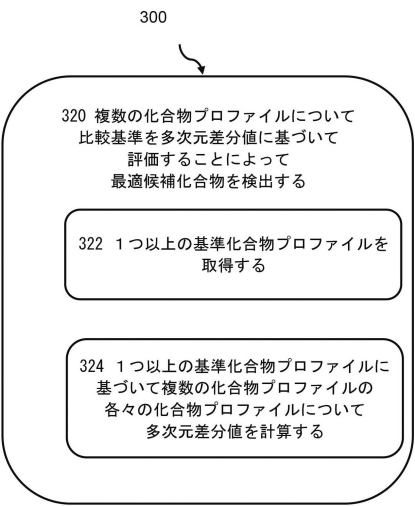


Fig. 5

【図 6】

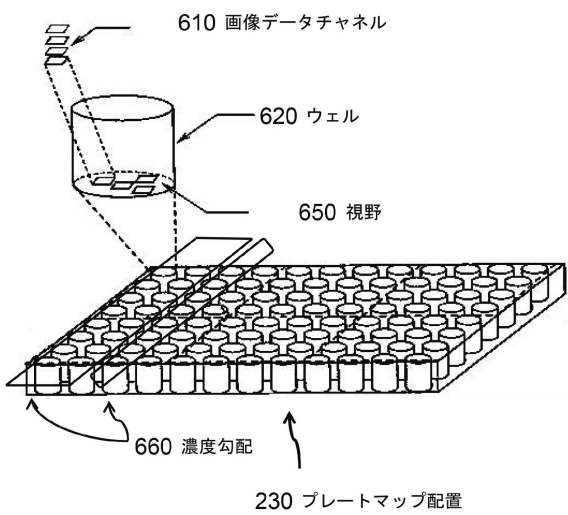


Fig. 6

10

20

30

40

50

【図 7】

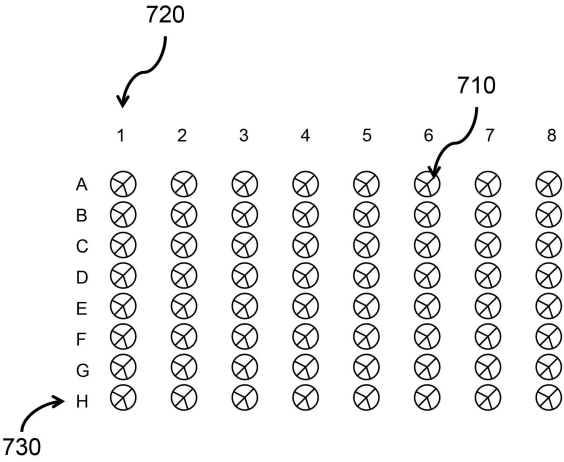


Fig. 7

【図 8】

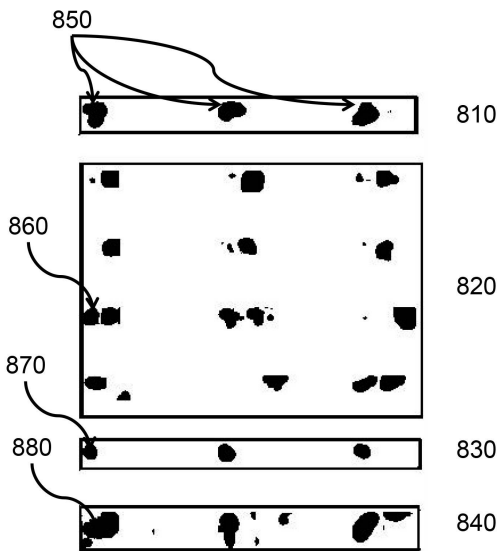


Fig. 8

【図 9】

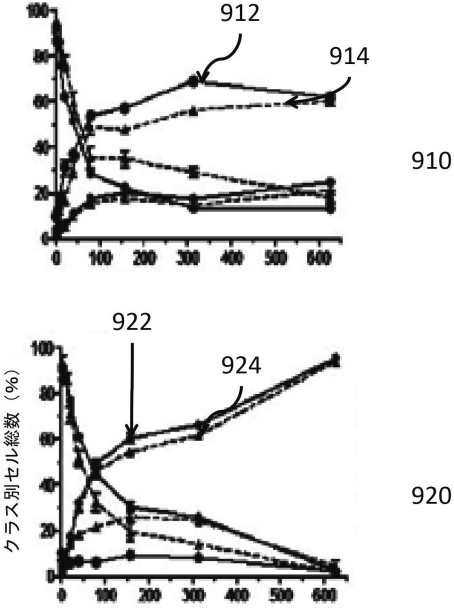


Fig. 9

【図 10】

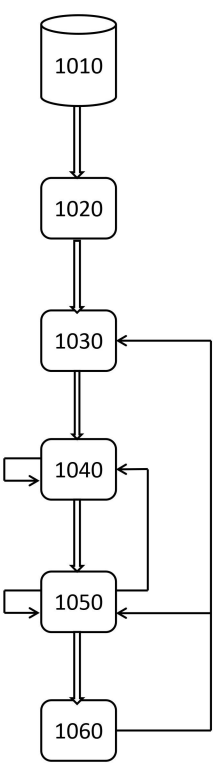


Fig. 10

10

20

30

40

50

フロントページの続き

(51)国際特許分類

C 1 2 M 1/34 (2006.01)

F I

C 1 2 M**1/34****Z**

(72)発明者 アヴィナシュ, ゴパル・ビリゲリ

アメリカ合衆国、カリフォルニア州・ 9 4 5 8 3、サン・ラモン、カミノ・ラモン、2 6 2 3、ジ
ーイー・ヘルスケア・デジタル

(72)発明者 グレイブス, ロバート・ジョン

アメリカ合衆国、カリフォルニア州・ 9 4 5 8 3、サン・ラモン、カミノ・ラモン、2 6 2 3、ジ
ーイー・ヘルスケア・デジタル

(72)発明者 ザット, アブヒジット・ヴィジェイ

アメリカ合衆国、カリフォルニア州・ 9 4 5 8 3、サン・ラモン、カミノ・ラモン、2 6 2 3、ジ
ーイー・ヘルスケア・デジタル

(72)発明者 コデシュ, アフェク

アメリカ合衆国、カリフォルニア州・ 9 4 5 8 3、サン・ラモン、カミノ・ラモン、2 6 2 3、ジ
ーイー・ヘルスケア・デジタル

(72)発明者 ジェフェリー, カロン

アメリカ合衆国、カリフォルニア州・ 9 4 5 8 3、サン・ラモン、カミノ・ラモン、2 6 2 3、ジ
ーイー・ヘルスケア・デジタル

(72)発明者 ダス, シャーミスタ

アメリカ合衆国、カリフォルニア州・ 9 4 5 8 3、サン・ラモン、カミノ・ラモン、2 6 2 3、ジ
ーイー・ヘルスケア・デジタル

審査官 真木 健彦

(56)参考文献 特表 2 0 0 9 - 5 2 6 5 1 9 (J P, A)

特開 2 0 1 4 - 1 3 4 5 2 4 (J P, A)

特開 2 0 0 9 - 0 6 3 5 0 8 (J P, A)

米国特許第 0 6 6 1 5 1 4 1 (U S, B 1)

川谷 誠, 長田 裕之, ケミカルバイオロジーと新規分子標的薬, 日本臨牀, 日本, 2015年08
月01日, 第73巻第8号 (2015-8), pp. 1273-1280

(58)調査した分野 (Int.Cl., D B名)

G 0 6 T 7 / 0 0

G 0 6 T 1 / 0 0

G 0 1 N 1 5 / 0 0

G 0 1 N 2 1 / 0 0

G 0 1 N 3 3 / 0 0

C 1 2 M 1 / 0 0 - 3 / 1 0

C 1 2 Q 1 / 0 0 - 3 / 0 0