

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7336384号

(P7336384)

(45)発行日 令和5年8月31日(2023.8.31)

(24)登録日 令和5年8月23日(2023.8.23)

(51)国際特許分類

F I

G 1 6 B 45/00 (2019.01)

G 1 6 B 45/00

G 0 1 N 15/14 (2006.01)

G 0 1 N 15/14

B

請求項の数 20 (全28頁)

(21)出願番号	特願2019-531469(P2019-531469)	(73)特許権者	517389757
(86)(22)出願日	平成29年12月13日(2017.12.13)		フロージョー エルエルシー
(65)公表番号	特表2020-513630(P2020-513630 A)		FlowJo, LLC
(43)公表日	令和2年5月14日(2020.5.14)		アメリカ合衆国 オレゴン アッシュラン
(86)国際出願番号	PCT/US2017/065987		ド ウィリアムソン ウェイ 385
(87)国際公開番号	WO2018/111982		385 Williamson Way,
(87)国際公開日	平成30年6月21日(2018.6.21)		Ashland, Oregon 975
審査請求日	令和2年12月11日(2020.12.11)		20, United States of
(31)優先権主張番号	62/433,930	(74)代理人	America
(32)優先日	平成28年12月14日(2016.12.14)		100094569
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	弁理士 田中 伸一郎
前置審査		(74)代理人	100103610
		(74)代理人	弁理士 吉 田 和彦
		(74)代理人	100109070
			弁理士 須田 洋之

最終頁に続く

(54)【発明の名称】 大きいマルチパラメータ・データセットにおけるパラメータの管理、合成、可視化、及び探索のための応用コンピュータ技術

(57)【特許請求の範囲】

【請求項1】

マルチパラメータ・データセットを可視化する方法であって、

プロセッサが、前記マルチパラメータ・データセットの散布図を第1及び第2の軸にわたって生成するステップを含み、前記データセットが複数のデータ項目を含み、各データ項目が、複数のパラメータに関連付けられ、関連付けられた前記パラメータの各々に対するデータ値を呈示し、前記第1及び第2の軸が前記データセット内のユーザ選択パラメータであり、前記散布図が複数のドットを含み、各ドットが、前記データセットからのデータ項目に対応し、第1及び第2の軸に沿って該軸に対応する前記パラメータに対する前記データ値に対応する場所にある位置において前記散布図上に配置され、

前記散布図が細胞ビュー散布図を含み、

プロセッサが、前記細胞ビュー散布図から、複数の異なる遺伝子を複数の細胞集団にわたってプロットする遺伝子ビュー散布図にピボットするステップをさらに含み、

前記複数の細胞集団のサブセットが、前記細胞ビュー散布図におけるゲーティングにより2つの細胞集団にグループ化される場合、これら2つの細胞集団をピボットして、ピボット表を作成し、前記ピボット表における行は、サンプル内の複数の細胞に対して、細胞毎の遺伝子発現を示す表では列であった遺伝子であり、前記ピボット表における列は、前記2つの細胞集団であり、各表セルに、各細胞集団の細胞における関連付けられた遺伝子の発現に対する遺伝子カウントの値に対する平均の値がポピュレートされ、前記2つの細胞集団が前記遺伝子ビュー散布図にピボットされることを特徴とする方法。

10

20

【請求項 2】

前記マルチパラメータ・データセットが細胞遺伝子発現データを含み、前記データ項目が複数の細胞を含み、前記パラメータが複数の遺伝子を含み、ここで前記データ値が、各細胞内の対応する遺伝子の発現に対するカウントを示すデータを含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記生成するステップは、プロセッサが、前記細胞ビュー散布図のための軸パラメータとしての少なくとも 1 つの遺伝子セットのユーザ選択に基づいて、前記細胞遺伝子発現データから前記細胞ビュー散布図を生成することを含むことを特徴とする請求項 2 に記載の方法。

10

【請求項 4】

プロセッサが、細胞集団を定義するために前記細胞ビュー散布図内のドットのグループをゲーティングするステップと、

プロセッサが、定義された前記細胞集団を代表するデータオブジェクトをワークスペース内で作成するステップと

をさらに含むことを特徴とする請求項 3 に記載の方法。

【請求項 5】

プロセッサが、前記ゲーティングするステップ及び作成するステップを繰り返して、複数の細胞集団を表す複数のデータオブジェクトを前記ワークスペース内で作成することを特徴とする請求項 4 に記載の方法。

20

【請求項 6】

プロセッサが、複数の細胞集団をブール演算に従って組み合わせて、前記ワークスペース内のデータオブジェクトとして含めるための新たな細胞集団を生成するステップをさらに含むことを特徴とする請求項 4 ~ 請求項 5 のいずれかに記載の方法。

【請求項 7】

前記細胞ビュー散布図を変更して、ゼロに対応するドットを前記散布図の定義されたゼロスペースにわたって展開するステップをさらに含むことを特徴とする請求項 3 ~ 請求項 6 のいずれかに記載の方法。

【請求項 8】

プロセッサが、遺伝子セットを定義するために遺伝子ビュー散布図内のドットのグループをゲーティングするステップと、

プロセッサが、定義された前記遺伝子セットを代表する合成パラメータ・データオブジェクトをワークスペース内で作成するステップと

をさらに含むことを特徴とする請求項 2 ~ 請求項 7 のいずれかに記載の方法。

30

【請求項 9】

プロセッサが、前記細胞遺伝子発現データを、前記細胞遺伝子発現データ内の細胞に関連付けられた合成パラメータとしての前記定義された遺伝子セットで増補するステップをさらに含むことを特徴とする請求項 8 に記載の方法。

【請求項 10】

プロセッサが、前記定義された遺伝子セットに基づいて前記細胞ビュー散布図内のドットのグループをゲーティングして、別の細胞集団を定義するステップと、

プロセッサが、前記定義された別の細胞集団の代表であるデータオブジェクトを前記ワークスペース内で作成するステップと

をさらに含むことを特徴とする請求項 8 に記載の方法。

40

【請求項 11】

プロセッサが、複数の遺伝子セットをブール演算に従って組み合わせて、前記ワークスペース内のデータオブジェクトとして含めるための新たな遺伝子セットを生成するステップをさらに含むことを特徴とする請求項 8 ~ 請求項 10 のいずれかに記載の方法。

【請求項 12】

前記散布図が、前記細胞遺伝子発現データからの遺伝子ビュー散布図を含み、前記生成

50

するステップは、プロセッサが、複数の細胞集団のユーザ選択に基づいて遺伝子ビュー散布図を生成することを含むことを特徴とする請求項 2 ～ 請求項 11 のいずれかに記載の方法。

【請求項 13】

プロセッサが、複数の細胞集団の前記ユーザ選択に応答して、前記細胞遺伝子発現データの細胞ビュー散布図から遺伝子ビュー散布図にピボットするステップをさらに含むことを特徴とする請求項 12 に記載の方法。

【請求項 14】

プロセッサが、前記遺伝子ビュー散布図内の複数の遺伝子をゲーティングすることによって遺伝子セットを生成することをさらに含むことを特徴とする請求項 12 ～ 請求項 13 のいずれかに記載の方法。

10

【請求項 15】

プロセッサが、細胞遺伝子発現データの細胞ビュー散布図の領域をゲーティングして細胞集団を作成するステップであって、前記細胞遺伝子発現データが、複数のパラメータに対する複数のデータ値を含み、前記細胞ビュー散布図が、前記細胞遺伝子発現データの第 1 のパラメータ及び第 2 のパラメータに対応する第 1 及び第 2 の軸を含む、ステップと、

プロセッサが、前記細胞遺伝子発現データの遺伝子ビュー散布図にピボットするステップであって、前記遺伝子ビュー散布図が、作成された前記細胞集団に対応する少なくとも 1 つの軸を含む、ステップと、

プロセッサが、遺伝子セットを作成するために、前記細胞遺伝子発現データの前記遺伝子ビュー散布図の領域をゲーティングするステップと

20

を含み、
プロセッサが、作成された前記遺伝子セットが前記細胞遺伝子発現データの合成パラメータとして利用可能となるように、作成された前記遺伝子セットで前記細胞遺伝子発現データを増補するステップをさらに含み、

前記細胞集団のサブセットが、前記細胞ビュー散布図におけるゲーティングにより 2 つの細胞集団にグループ化される場合、これら 2 つの細胞集団をピボットして、ピボット表を作成し、前記ピボット表における行は、サンプル内の複数の細胞に対して、細胞毎の遺伝子発現を示す表では列であった遺伝子であり、前記ピボット表における列は、前記 2 つの細胞集団であり、各表セルに、各細胞集団の細胞における関連付けられた遺伝子の発現に対する遺伝子カウントの値に対する平均の値がポピュレートされ、前記 2 つの細胞集団が前記遺伝子ビュー散布図にピボットされることを特徴とする方法。

30

【請求項 16】

プロセッサが、作成された前記遺伝子セットに対応する少なくとも 1 つの軸を含む、前記細胞遺伝子発現データの別の細胞ビュー散布図にピボットするステップをさらに含むことを特徴とする請求項 15 に記載の方法。

【請求項 17】

マルチパラメータ・データセットを格納するように構成されたメモリと、
請求項 1 ～ 請求項 16 のいずれかに記載の方法を行うように構成された、前記メモリと協働するためのプロセッサと、

40

を含むことを特徴とする装置。

【請求項 18】

請求項 1 ～ 請求項 16 のいずれかに記載の散布図をユーザに図形的に提示するための、前記プロセッサと協働するディスプレイをさらに含むことを特徴とする請求項 17 に記載の装置。

【請求項 19】

非一時的なコンピュータ可読ストレージ媒体上に常駐する複数のプロセッサ実行可能命令を含み、前記命令が、プロセッサによって実行されると、請求項 1 ～ 請求項 16 のいずれかに記載の方法を前記プロセッサに実行させることを特徴とするコンピュータプログラム製品。

50

【請求項 20】

前記命令が、前記プロセッサによって実行されると、請求項 1～請求項 16 のいずれかに記載の散布図をユーザに図形的に提示するようディスプレイに命令するようにさらに構成されることを特徴とする請求項 19 に記載の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】

関連特許出願の相互参照及び優先権主張

本出願は、引用によりその開示全体が組み入れられる 2016 年 12 月 14 日出願の「
Applied Computer Technology for Management, Synthesis, Visualization, and Exploration of Parameters in Large Multi-Parameter Data Sets」と題する米国仮特許出願第 62/433,930 号に対する優先権を主張する。

10

【背景技術】

【0002】

緒言

バルク集団及び個々の細胞の両方に対して利用可能な遺伝及び遺伝子発現情報の量は、もはや研究者の手に余るところまで増大してきている。例えば、細胞遺伝子発現データは、今や個々の細胞について測定することができる、数千もの遺伝子（例えば、10,000 - 30,000 又はそれより多くの遺伝子）を含むことがあり、1 サンプルあたり千個もの細胞が測定され得る。このことは、細胞遺伝子発現データの可視化、分析、探索、及び理解の分野において、甚だしい技術的課題を提示している。

20

【先行技術文献】

【特許文献】

【0003】

【文献】米国特許第 9,567,645 号明細書

米国特許第 8,835,358 号明細書

米国特許出願公開第 2016/0328249 号明細書

【発明の概要】

30

【発明が解決しようとする課題】

【0004】

例えば、細胞遺伝子発現データの可視化を促進するのにコンピュータを使用する従来の手法では、可視化が最終的な終点であり、可視化は、ユーザが R プログラミング言語を用いて手動でスクリプトを記述した結果として達成され、それはそのユーザがデータ入力を行うための異なるライブラリ、再フォーマット、操作、計算及びグラフ化の知識を有することを要する。これらのスクリプトは、通常、特定のデータセットに合わせてカスタマイズしなければならず、その作成には、プログラミング言語、既存のライブラリ、及び結果を生成するための要求される入力の専門的な知識を要する。そのうえ、このような従来の手法は、異種細胞集団の深い探索を妨げる。

40

【課題を解決するための手段】

【0005】

この技術的課題に対する解決策として、本発明者らは、細胞が個々のデータポイントとして可視化される細胞（又は細胞集団）ビュー散布図（例えば、細胞の、遺伝子対遺伝子散布図）と、遺伝子が個々のデータポイントとして可視化される遺伝子ビュー散布図（例えば、遺伝子の、細胞集団対細胞集団散布図）とを含む、細胞発現データの種々の次元にわたる革新的な散布図表示を用いたコンピュータ技術の適用を開示する。これらの散布図内でゲーティング（gating）を行って細胞集団及び遺伝子セットをそれぞれ作成することができ、それは、細胞遺伝子発現データを増補し、有意義な研究のための新たな道を拓く際に使用するためにワークスペース内に新規データオブジェクトとして追加さ

50

れる生物学的に妥当な次元 (biologically-relevant dimension) として役立つことができる。対比のために、このような分析を孤立化し、サイロ化 (siloe d) した方式で個々の遺伝子ベースで速やかに行うと手に余るものとなるのに対し、細胞ビュー散布図と遺伝子ビュー散布図との間でピボットする能力は、遺伝子の生物学的に妥当なグループ分けを、ユーザが見いだすことを可能にし、次いでこれを細胞ビュー散布図の合成パラメータとしてさらに調査することができる。

【0006】

上記のように、当該分野の従来の可視化システムでは、可視化は、プロセスの終点としての役割を果たすものであり、さらなる調査のためのさらなる可視化の洗練をさらに創出するための開始点としての役割を果たすことはできない。一例として、転移性黒色腫の患者由来の免疫細胞のサンプルはT細胞を含有し得るが、当該分野の従来の可視化システムでは、このサブセットを免疫細胞内で単に同定することができるに過ぎない。しかしながら、本明細書で説明する革新的なコンピュータシステムは、例示的な実施形態を参照してより詳しく後述するように、T細胞サブセットの深い探索及び分析を可能にして、T細胞内の複数のサブセット、例えば「消耗した」T細胞を同定し、この状態を個々の遺伝子まで追跡し、次いでそれを標的として、この消耗を逆転させ、T細胞を活性化し、かくして可能であれば免疫応答を刺激して転移を根絶する。

【0007】

したがって、本明細書に記載された革新的な可視化技術を通じて、コンピュータ技術を細胞遺伝子発現データに適用して、細胞と遺伝子との間の新たな関係を見だし、これらの関係を表す新たな結合的データ構造 (associative data structure) を細胞遺伝子発現データ内に作り出すことができる。

【0008】

これら及び他の特徴を通じて、本発明の例示的な実施形態は、応用バイオインフォマティクス技術における著しい技術的前進をもたらす。

【図面の簡単な説明】

【0009】

【図1】本明細書で説明する革新的なデータ処理及び可視化技術をサポートするために用いることができる例示的なコンピュータシステムを開示する。

【図2A】細胞遺伝子発現データセットの例を示す。

【図2B】例示的な細胞遺伝子発現データの表ビューを示す。

【図3】例示的な細胞ビュー・グラフウインドウ散布図を示す。

【図4】細胞ビュー散布図を作成するための実行のための例示的なプロセスフローを示す。

【図5】パラメータ選択メニューを含む例示的な細胞ビュー散布図ユーザインタフェースを示す。

【図6】細胞集団を作成するために細胞ビュー散布図ユーザインタフェース内でどのようにゲーティングを行うことができるかの一例を示す。

【図7】ユーザが細胞ビュー散布図のゼロを展開すること (spread) を可能にする例示的なユーザインタフェースを示す。

【図8】遺伝子以外のパラメータが軸パラメータとして選択された、例示的な細胞ビュー散布図を示す。

【図9】細胞集団を作成するために図8の細胞ビュー散布図内で行うことができるゲーティングのやり方の例を示す。

【図10】散布図提示においてディスプレイ設定を調整するための例示的なユーザインタフェースを示す。

【図11】細胞集団選択メニューを含む、例示的な遺伝子ビュー・グラフウインドウ散布図の例を示す。

【図12】遺伝子ビュー散布図のための細胞集団データを作成するためにどのように細胞遺伝子発現データをピボットすることができるかの例を示す。

【図13A】遺伝子ビュー散布図で使用するためのワークスペースにおいて相補的細胞集

10

20

30

40

50

団をどのように作成し及び定義することができるかの例を示す。

【図 1 3 B】遺伝子ビュー散布図で使用するためのワークスペースにおいて相補的細胞集団をどのように作成し及び定義することができるかの例を示す。

【図 1 4】遺伝子セットを作成するために遺伝子ビュー散布図ユーザインタフェース内でゲーティングをどのように行うことができるかの例を示す。

【図 1 5】細胞遺伝子発現データを合成パラメータとしての遺伝子セットでどのように増補することができるかの例を示す。

【図 1 6】遺伝子セットがパラメータ選択メニューにおけるオプションとして提示される細胞ビュー散布図の例を示す。

【図 1 7 A】ワークスペース内で閲覧、編集及び新たな遺伝子セットを他の遺伝子セットから作成するための例示的なユーザインタフェースを示す。

10

【図 1 7 B】ワークスペース内で閲覧、編集及び新たな遺伝子セットを他の遺伝子セットから作成するための例示的なユーザインタフェースを示す。

【図 1 8 A】ユーザ選択された 3 次元を散布図上にどのようにオーバーレイすることができるかの例を示す。

【図 1 8 B】ユーザ選択された 3 次元を散布図上にどのようにオーバーレイすることができるかの例を示す。

【図 1 8 C】ユーザ選択された 3 次元を散布図上にどのようにオーバーレイすることができるかの例を示す。

【図 1 8 D】ユーザ選択された 3 次元を散布図上にどのようにオーバーレイすることができるかの例を示す。

20

【図 1 9】図 1 8 A - 図 1 8 D の 3 D 散布図内でどのようにゲーティングを行うことができるかの例を示す。

【図 2 0 A】システムを通じて例示的なレポートを作成することができることを示す。

【図 2 0 B】システムを通じて例示的なレポートを作成することができることを示す。

【図 2 0 C】システムを通じて例示的なレポートを作成することができることを示す。

【図 2 0 D】システムを通じて例示的なレポートを作成することができることを示す。

【図 2 1】細胞ビューモードと遺伝子ビューモードとの間で切り替えるためにシステムによって実行される例示的なプロセスフローを示す。

【図 2 2】細胞データの研究及びリサーチをサポートするために細胞ビューモードと遺伝子ビューモードとの間で切り替えるためにシステムをどのように動作させることができるかの例を示す。

30

【発明を実施するための形態】

【0010】

図 1 は、本明細書で説明する革新的なデータ処理及び可視化技術をサポートするために用いることができる例示的なコンピュータシステム 100 を開示する。例示的なコンピュータシステム 100 は、バス 110 などの相互接続技術を介して互いに通信することができる、プロセッサ 102 と、メモリ 104 と、データベース 106 と、ディスプレイ 108 とを含む。

【0011】

40

プロセッサ 102 は、本明細書で説明する動作を行うのに適した任意のプロセッサの形態をとることができる。例えば、ラップトップ又はワークステーションの CPU はプロセッサ 102 としての使用に適するであろう。プロセッサ 102 は、ネットワーク上で互いに通信して本明細書で説明するタスクを実行する分散型プロセッサを含む、複数のプロセッサを含むことができることを理解されたい（例えば、クラウドコンピューティング処理リソース）。メモリ 104 は、本明細書で説明するタスクの実行の際にプロセッサ 102 と協働するのに適した任意のコンピュータメモリの形態をとることができる。メモリ 104 は、ネットワーク全体にわたって分散されたメモリを含む、複数のメモリデバイスの形態をとることができることを理解されたい。同様に、データベース 106 は、プロセッサ 102 にアクセス可能な任意のデータリポジトリの形態をとることができる（例えば、コン

50

ピュータ上のファイルシステム、リレーショナルデータベース等）、また、データベース 106 は、複数の分散したデータベース（例えば、クラウドストレージ）の形態をとることができることを理解されたい。ディスプレイ 108 は、本明細書で説明する可視化を生成することが可能なコンピュータモニタ又はスクリーンの形態をとることができる。

【0012】

本明細書で説明する革新的なデータ分析及び可視化技術は、細胞遺伝子発現データ 112 に対して行うことができる。細胞遺伝子発現データ 112 は、次世代シーケンシング（例えば、他のシーケンシング手法の中でもとりわけ RNA シーケンシング（RNA Seq）及びシングルセル（単一細胞）RNA シーケンシング（scRNA-Seq）の測定用）によって生成することができる。しかしながら、これは単なる一例であり、細胞遺伝子発現データ 112 を生成するための他の技術を使用することができる。追加例として、デジタルドロップレット及び逆転写酵素を含む、ポリメラーゼ連鎖反応手法が挙げられる。さらなる例として、とりわけフローサイトメトリによる RNA 測定、及びマイクロアレイが挙げられ、これは DNA 及び / 又は RNA の定量を含むデータファイルを生成し、又は生リードデータを処理するソフトウェアプログラム（一次及び二次分析）を通じて遺伝子発現データファイルを生成する。さらに別の例は、サンプルの確率的標識化（stochastic labeling）から誘導された遺伝子発現データである。確立的標識化された遺伝子発現データの例は、特許文献 1 及び特許文献 2 並びに 2017 年 9 月 25 日出願の「Measurement of Protein Expression Using Reagents with Barcoded Oligonucleotide Sequences」と題する米国特許出願第 15 / 715,028 号に見いだすことができ、これらの各々の全開示は引用により組み入れられる。

【0013】

さらに、本明細書で説明する革新的なデータ分析及び可視化技術は、様々な手段で単一細胞から生成されるデータに適用することができる。単一細胞分析は、核酸若しくはタンパク質又はタンパク質と核酸との任意の組合せの確率的標識化を含むことができる。一例として、本明細書で説明する革新的なデータ分析及び可視化技術は、タンパク質密度若しくは遺伝子発現又はそれらの任意の組合せの定量的特徴を分析するために用いることができる。本明細書で説明する革新的なデータ分析及び可視化技術はまた、単一細胞における様々なタンパク質又は核酸の確率的標識化によって生成される定量的データの改善された可視化を提供することもできる。個々の細胞内の生物学的集団（核酸、タンパク質等）の定量的値を、他の個々の細胞と比較することができ、又は細胞型の間で比較することができ、又はデータを生成する方法間で比較することさえ可能である。例えば遺伝子発現値を、データセットを生成するために用いた方法の関数として可視化することができる。本明細書で説明する革新的なデータ分析及び可視化技術はまた、本明細書で説明する種々の手段によって生成された定量的データを、そうしたデータの生成方法とは独立に定量的生物学的集団データの可視化を提供する方式で比較するために用いることもできる。

【0014】

このデータ 112 は、大きいマルチパラメータ・データセットとして特徴づけることができ、これは、特に生物学的に妥当な情報が視覚的方式でユーザに有意味に提示されるように根底にある生物学に関して考えたときに、有意味な可視化の作成において特別な技術的課題を提起する。例えば、細胞遺伝子発現データは、多数の個々の細胞及び細胞集団に対するデータを含むことがあり、各細胞又は細胞集団に対するパラメータは 10,000 ~ 30,000 パラメータ以上に及ぶことがある。細胞遺伝子発現データ 112 をデータベース 106 内のファイルから読み出し、複数のデータ構造 116 としてメモリ 104 内にロードすることができ、このデータ構造が分析及び可視化プログラム 114 の実行中にプロセッサ 102 によって操作されることになる。プログラム 114 は、メモリ 104 のような非一時的コンピュータ可読ストレージ媒体上に常駐する複数のプロセッサ実行可能命令の形態の、プロセッサ実行可能コンピュータコードを含むことができる。

【0015】

図 2 A は、細胞遺伝子発現データセットの例を示し、ここで各細胞（又は細胞集団）は、細胞 ID で識別され、複数のパラメータに関連付けられ、各パラメータは、細胞 ID に関連して ID 及び値を有する。上で示したように、細胞の遺伝子発現データは高次元であり、各細胞に対するパラメータの数は、1 細胞当たり又は 1 細胞集団当たり 10,000 ~ 30,000 パラメータ以上に達し得る。細胞データにおけるパラメータの例は、多数の遺伝子についての対象細胞内の遺伝子発現のカウントを含む。それゆえ細胞 1 (Cell 1) のパラメータ 1 (Parameter 1) は遺伝子 1 (Gene 1) に対応するものとしてことができ、その値は、細胞 1 内の遺伝子 1 についての発現のカウントとすることができる。同様に、細胞 1 のパラメータ 2 (Parameter 2) は、遺伝子 2 (Gene 2) に対応するものとしてことができ、その値は、細胞 1 内の遺伝子 2 についての発現のカウントとすることができる。図 2 B は、例示的な細胞遺伝子発現データ 112 の表ビューを示す。表 200 内の各行は、異なる細胞に対応し（「細胞」列参照）、遺伝子 1、遺伝子 2 等と表示された種々の列は、異なる遺伝子に対応し、表のセルは、各対象細胞における対応遺伝子についての遺伝子発現のカウントを特定する。この表は、遺伝子以外のパラメータを含むこともできる。例えば、細胞遺伝子発現データ 112 は、t 分布型確率的近傍埋め込み (t-distributed stochastic neighbor embedding (tSNE))、主成分分析 (PCA)、線形判別分析 (LDA) 等のようなパラメータに対するデータ値を各表セル中に含むことができ、これらのデータ値は、その値が n パラメータにわたって個々のセルに対する差異を捕捉する、分析計算を表す。細胞遺伝子発現データ 112 は、任意の幾つかの形式（例えば、CSV ファイル、データベース表（例えば、リレーショナルデータベースにおけるリレーショナルデータ）、スペアデータ表現、バイナリ形式、及びその他）で格納することができる。【0016】

10

20

図 3 は、ディスプレイ 108 を介してユーザに提示するための、プログラム 114 の実行を通じて生成することができる例示的なグラフウインドウ (GW) を示し、ここではグラフウインドウは、細胞集団に関して遺伝子 1 対遺伝子 2 を可視化している。この可視化を、細胞遺伝子発現データ 112 の細胞ビューと称することができる。後述するように、細胞ビューにおける個々の細胞の選択に対するゲーティングが、細胞集団を作成する。このグラフウインドウは、ユーザ選択ファイル 306 内の細胞集団に関して、ユーザが選択した 2 つの遺伝子 302 及び 304（この例ではそれぞれ TME216 及び MMP2）についての遺伝子発現の散布図 300 を提示する。各ファイルは、細胞集団の単一サンプル、単一細胞集団（例えば、フローサイトメトリ活性細胞ソートによってソートされ、分析された 1 つの集団）、又は場合によっては連結された複数のサンプル（異なる患者由来、又は異なる時点での 1 人の患者由来）に対応し得る。散布図内の各ドット 308 は、対象ファイル 306 の細胞集団内の細胞を表す。この例の X 軸及び Y 軸のスケールは、対応する遺伝子に対するカウントを特定する。したがって、水平の X 軸上のドット 308 は、対象ドット 308 に対応する細胞内に遺伝子 MMP2 がどれだけ存在するかのカウントを特定し、垂直の Y 軸上のドット 308 の位置は、対象ドット 308 に対応する細胞内に遺伝子 TME216 がどれだけ存在するかのカウントを特定する。したがって、散布図 300 の右上の象限内のドット 308 は、TME216 及び MMP2 の両方が高度に発現した細胞に対応するのに対し、散布図 300 の左下の象限内のドットは、TME216 及び MMP2 の両方が低レベルで発現した細胞に対応する。同様に、左上の象限は、TME216 は高度に発現しているが MMP2 がそうではない細胞に対応するのに対し、右下の象限は、MMP2 は高度に発現しているが TME216 がそうではない細胞に対応する。y = x の対角線上は、細胞 308 が選択された遺伝子の両方を等しく発現する場合にそれら細胞が配置される場所である。したがってこの対角線からいずれかの方向に細胞 308 が遠ざかった距離は、所与の細胞における選択された遺伝子の示差的な発現の程度を示す。多くの細胞集団について、選択された遺伝子の発現がそれらの細胞についてはゼロレベルである多数の細胞が存在することが予期される。このことは、ドット 308 がそれぞれ X 軸及び Y 軸上のゼロレベル 310 及び 312 において大きいクラスタを形成する

30

40

50

ことにつながる。3 1 0 及び 3 1 2 においてカラーコード化を用いて、このようなゼロレベルの遺伝子発現の細胞の密度を示すことができる。

【0 0 1 7】

図 4 は、散布図 3 0 0 をどのようにして生成することができるかを説明する、プログラム 1 1 4 の一部として実行するための例示的なプロセスフローを示す。ステップ 4 0 0 において、プロセッサは、メモリワークスペース内でデータ構造を作成する（図 1 の 1 1 6 参照）。このデータ構造は、細胞ビューデータを保持するために用いることができる。

【0 0 1 8】

第 1 の軸に対して、プロセッサは、ユーザ入力に基づいて細胞遺伝子発現データ 1 1 2 内の遺伝子を選択する（ステップ 4 0 2 ）。例えば、プロセッサは、ユーザ入力に回答して表 2 0 0 の中の遺伝子列を選択することができる。図 5 は、ユーザインタフェースがユーザに対してどのように各軸のための遺伝子のリスト提示することができるかを示す。選択メニューにアクセスするために、ユーザは、それぞれ Y 軸及び X 軸用の遺伝子セクタ 3 0 2 又は 3 0 4 を選択することができる。この例における第 1 の軸が Y 軸であるとする、3 0 2 を選択したとき、Y 軸に関する選択のために利用可能なパラメータのリストを提示するパラメータ選択メニュー 5 0 0 が示される。このリストに、細胞遺伝子発現データ 1 1 2 からのパラメータをポピュレートすることができる。図 5 に示すように、Y 軸に対するユーザ選択 5 0 2 は T M E M 2 1 6 である。ステップ 4 0 4 において、プロセッサは、データ構造に、細胞遺伝子発現データ 1 1 2 からの細胞のリスト（例えば、表 2 0 0 の細胞列内の細胞）をポピュレートする。リストに挙げられた各細胞は、選択された第 1 の軸の遺伝子に対するそのカウント値に関連付けられる。

【0 0 1 9】

第 2 の軸に対して、プロセッサは、ユーザ入力に基づいて細胞遺伝子発現データ 1 1 2 内の別の遺伝子を選択する（ステップ 4 0 6 ）。例えば、プロセッサは、ユーザ入力に回答して表 2 0 0 の中の遺伝子列を選択することができる。図 5 は、X 軸パラメータ選択メニュー 5 0 6 にアクセスするために 3 0 4 が選択された（その結果、M M P 2 である選択 5 0 8 が得られる）ユーザインタフェースの例を示す。ステップ 4 0 8 において、プロセッサは、細胞リストを増補して、各細胞に、選択された第 2 の軸の遺伝子に対するその関連付けられたカウント値を付加する。したがって、この時点で、データ構造は、第 1 及び第 2 の軸の選択された遺伝子についてのカウント対に関連付けられた細胞のリストを含む。例えば、リストは、細胞遺伝子発現データ 1 1 2 内の各細胞に対するベクトル { 細胞 I D 1 , 遺伝子 1 カウント , 遺伝子 2 カウント } の組を含むことができる。

【0 0 2 0】

ステップ 4 1 0 において、プロセッサは、リストを構文解析して、各選択遺伝子に対して最大値（最高カウント）を見いだす。次いでプロセッサがこれら最大値を用いて、散布図における X 軸及び Y 軸に対する適切なスケールを定める（ステップ 4 1 2 ）。例えば、X 軸遺伝子に対する最大値が 1 0 であるならば、X 軸スケールは、0 - 1 0 までとすることができる。ステップ 4 1 4 において、プロセッサは、細胞リスト及び定められたスケールに基づいて、細胞の関連付けられたカウント値を散布図における X , Y 座標として用いて、散布図を描画する。結果は、図 3 に示されるような散布図 3 0 0 である。

【0 0 2 1】

図 3 に戻ると、ユーザは、ユーザインタフェース内のゲート作成ツール 3 2 0 にアクセスすることができ、これを通じて子細胞集団が作成される。例えば、ユーザは、ツール 3 2 0 にアクセスして、細胞 3 0 8 のサブセットを取り囲む形状を散布図 3 0 0 内に描画することができる。図 6 は、2 つの選択遺伝子の非ゼロ発現を有する細胞を捕捉するために描画されたゲート 6 0 0 を示す。描画された形状 6 0 0 の内側に入る細胞 3 0 8 は、ゲーティングされたそれ自身の子細胞集団になり、この細胞集団を明確に区別されたオブジェクト 6 0 2 としてワークスペースに追加することができる。細胞集団を作成するために、プロセッサは、ゲート 6 0 0 を、ゲーティングされる細胞集団に対する複数の境界条件に翻訳することができる。例えば、ゲート 6 0 0 によって示されるような矩形に関して、境

10

20

30

40

50

界条件は、(1) X軸の値が1と10との間、かつ(2) Y軸の値が1と8との間のすべての細胞とすることができる。細胞リストデータ構造(例えば、ベクトル{細胞ID 1, 遺伝子1カウント, 遺伝子2カウント}の組)をトラバースして、これらの基準を満たすすべての細胞IDを見いだすことができ、これらの細胞IDに対するデータをワークスペース内の新たな子細胞集団データ構造にポピュレートすることができる。また、子細胞集団に対する遺伝子発現の対応する散布図300を次いで提示することができる(ここで細胞集団は1つ以上の細胞308を含む)。このゲーティングは、ユーザが散布図300内の細胞308の生物学的に興味深いクラスタに焦点を合わせることが可能にすることができる。

【0022】

また、図3の細胞ビューモードは、ユーザがナビゲーションツール322を使用して、子細胞集団及び親細胞集団を含む異なるサンプルを視覚的に比較することを可能にする。例えば、ツール322の「戻る」(back)ボタン及び「次へ」(next)ボタンにより、ユーザは、ワークスペース内の次のサンプル及び以前のサンプルの散布図300に進むことができる。ツール322におけるダウボタンにより、ユーザは、分析階層内の子細胞集団に(例えば、図6のゲート600により作成された子細胞集団に)進むことができる。また、アップボタン(図示せず)により、ユーザは、分析階層内の親細胞集団に進むことができる。

【0023】

さらに、細胞ビューモードの場合、多くの細胞集団について、選択された遺伝子の発現がそれらの細胞についてはゼロレベルである多数の細胞が存在することが予想される。このことは、ドット308が図3に示される散布図300のそれぞれX軸及びY軸上のゼロレベル310及び312において大きいクラスタを形成することにつながる。310及び312においてカラーコード化を用いて、このようなゼロレベルの遺伝子発現の細胞の密度を示すことができる。しかしながら、本発明者らは、2つの選択遺伝子(又は後述するような遺伝子セット)についてゼロレベルの可視化を拡張することによって、特定の状況においてユーザに生物学的に妥当な情報を提供するために表示を強化することができる。これを達成するために、ユーザインタフェースは、図7に示すように「展開ゼロ(spread zero)」ユーザ制御部700(例えばチェックボックス、ボタン等)を含むことができる。この制御部700は、軸パラメータを選択するためのメニュー上に設けることができるが、実務者(practitioner)が所望するならば、制御部700は、どこか別のところ、例えば図3によって示されるユーザインタフェース上のどこかに配置することができる。図7の上半分は、ゼロが展開されていない散布図を示す。図7の下半分は、入力機構700を介して展開ゼロ・オプションが選択されたときの散布図を示す。

【0024】

図7の下散布図によって示されるように、ボックス702は、X軸遺伝子(この例ではMMP2)の発現に関してゼロ値を示す細胞の拡張されたビューを提供する。ボックス704は、Y軸遺伝子(この例ではTMM216)の発現に関してゼロ値を示す細胞の拡張されたビューを提供する。X軸及びY軸に沿ったゼロスペースの深さは、各遺伝子に対してゼロレベルにあるドット/細胞が幾つあるかの関数として定めることができ、ドット/細胞は、ボックス702及び704によって定められるゼロスペースにわたって各ゼロ位置における細胞の密度の関数として展開することができるが、他の分布技術を用いることもできる。理解されるように、図7の下散布図の左下象限(ボックス702と704との重なりによって定められる)は、X軸遺伝子及びY軸遺伝子の両方に対してゼロ発現を有する細胞を含む。この散布図の右下象限は、Y軸遺伝子のゼロ発現を有するがX軸遺伝子の正の発現を有する細胞を含み、この散布図の左上象限は、X軸遺伝子のゼロ発現を有するがY軸遺伝子の正の発現を有する細胞を含む。この散布図の右上象限は、両方の遺伝子の正の発現を有する細胞を含む(事実上、図7の上部に示す散布図である)。図7の下散布図から分かるように、右上象限の細胞集団は、他の象限よりも疎である。本発

10

20

30

40

50

明者らは、この展開方式でゼロレベルを可視化する能力は、ユーザに対して、生物学的に妥当な情報（例えば、予期せぬ分布が生じた場合、例えば右上象限が一般に予期されるように疎ではない場合に、遺伝子の組合せを評価する能力）を提供することができると信じる。

【0025】

細胞ビュー散布図はまた、遺伝子以外の、細胞遺伝子発現データ112内の選択されたパラメータに対する細胞情報を表示することもできる。上記で示したように、細胞遺伝子発現データ112は、tSNE、LDA、PCA等の結果得られるような次元の低減（dimensionality reduction）からのパラメータ、及び品質制御パラメータ（閾値を上回る（above threshold）、リボソームRNA（rRNA）存在量に関連したパラメータ、等）を含むことができる。これらのパラメータは、パラメータ選択メニュー500及び506のオプションとして提示することができる。図8は、Y軸のための選択800がパラメータtSNE軸2、X軸のための選択506がパラメータtSNE軸1である例を示す。結果として得られる図8の散布図は、図4のプロセスフローによって作成することができる。ユーザは、図6に関して上で説明したように、図8の散布図内の所望の細胞集団をゲーティングすることもできる（図9のゲート900参照）。

【0026】

図3の細胞ビュー・ユーザインタフェースはまた、ユーザがそれを通じて散布図300内の情報の表示を変更することができる、ユーザ制御部330（例えば、図3の「T」ボタン）を含むこともできる。Tボタン330を選択すると、図10のユーザインタフェースを提示することができる。このインタフェースを通じて、ユーザは、散布図データのビンニング／表示を対話方式で、ライブで変更することができる。図10の右側のリストは、散布図のX軸に対して利用可能な選択パラメータを示す。ヒストグラム1000は、対象細胞集団に関して、選択パラメータに対する値のビンニング（binning）を示す。ゼロの率が高いと仮定すると、ヒストグラム1000は、ゼロレベルにおいて大きなスパイクを示し、この例のスケールは、ヒストグラムにおける非ゼロレベルを不明瞭にする。図10の上部近くに示された左／右矢印により、ユーザは、異なるサンプルに対する表示をすばやくプレビューすることができる。図10の中ほどの＋／－制御部により、ユーザは、ヒストグラムのズームレベルを容易に変更することができる。

【0027】

スケール制御部1002により、ユーザは、散布図のX軸を様々な方式で調整することができる。例えば、X軸スケールを、線形スケール又は他の何らかのスケール（例えば、log2スケール）を示すように定めることができる（制御部1004参照）。また、最小（min）／最大（max）制御部1006により、ユーザは、X軸の最小境界及び最大境界を定めることができる。例えば、これらの制御部をより、最小値をゼロより大きい値に定めることができ、これによりゼロスパイクがヒストグラムから除去されて新たなスケールのヒストグラムが再提示され、そこではX軸にわたる非ゼロ値の分布をより明確に見ることができる。スライダ1008は、ユーザに変換変数の容易な制御を与えることができる。また、ユーザ入力によってその変数を調整することができるユーザ提供の変換を含む、付加的な変換のオプションを提供することができることを理解されたい（例えば、引用によりその全開示が本明細書に組み入れられる「Plugin Interface and Framework for Integrating External Algorithms with Sample Data Analysis Software」と題する特許文献3参照）。

【0028】

本明細書で開示される特に革新かつ強力な本発明のシステムの態様は、散布図表示を細胞ビューモードから遺伝子ビューモードにピボットする能力である。図11は、遺伝子ビューモードの散布図1100の例を示す。本明細書で用いる場合、「遺伝子ビュー」可視化は、遺伝子が、軸次元に対して測定される個々のデータ点であるような可視化を意味す

10

20

30

40

50

る（例えば、2つの細胞集団に対して測定された散布図におけるドットとしての遺伝子）。散布図1100では、軸パラメータは細胞集団である（B細胞の集団を特定するX軸パラメータ1102及び「非B」細胞の集団を特定するY軸パラメータ1104を参照のこと）。散布図1100内のドット1106は、（図3の散布図300の場合の個々の細胞ではなく）特定の遺伝子を表す。集団選択メニュー1110は、X軸上で用いられる選択のために利用可能な細胞集団のリストを提供し、集団選択メニュー1112は、Y軸上で用いられる選択のために利用可能な細胞集団のリストを提供する。上記のように、この例において、選択1114及び1116は、B細胞の集団及び「非B」細胞の集団にそれぞれ対応する。

【0029】

このピボットにより、図2の表200のような細胞遺伝子発現データ112を参照すると、遺伝子が表の行になり、細胞のサブセットが2つの細胞集団にグループ化され、それが表の列になるように、表200をピボットすることができる。さらに、2つの細胞集団の細胞について表2における遺伝子カウントに対する計算を行って、ピボット表のセルにポピュレートする値を決定することができる。図11の制御部1120により、ユーザは、ピボット表のセルについての値を計算するために遺伝子データに対して行う計算を定義することができる。図11の例においては、正規化平均計算が選択されている。しかしながら、正規化メジアン、正規化モード、直接（straight）平均、直接メジアン、直接モード等のようなその他の計算オプションを利用することができることを理解されたい。したがって、ピボット表データに対して行われる計算は、ユーザにとって生物学的に妥当と考えられる任意のユーザ定義関数とすることができることを理解されたい。発明者らは、計算は、2つの細胞集団の細胞カウントにおける潜在的な不一致を考慮したある種の平均化に基づくこと、及び/又はサンプル間の変動を考慮した正規化（例えば、測定すべき既知量のRNAを含むExternal RNA Controls Consortium（ERCC）コントロールのスパイクイン（spike-in）に対する正規化）に基づくことが望ましい場合があることに注目する。例えば、ピボット表の値が2つの細胞集団における各遺伝子の直接カウントである場合であって、2つの細胞集団の細胞のカウントに有意な差がある場合、2つの細胞集団にわたって集約した遺伝子カウントの合計は、比較の意味ではそれほど有益（informative）であるとは言えないであろう。しかしながら、細胞集団がおおまかに同じようなサイズである場合、2つの細胞集団内の各遺伝子の直接遺伝子カウントは、やはり有益であるだろう。

【0030】

図12は、細胞ビューから遺伝子ビューへの例示的なピボットを示す。図12の表200は、サンプル内の多数の細胞に対して、細胞毎の遺伝子発現を示す。各行は異なる細胞に対応し、列は、関連付けられた細胞における異なる遺伝子についての発現カウントに対応する。これらの細胞のサブセットが、図12に示すように（例えば、細胞ビュー散布図300におけるゲーティングにより）2つの細胞集団1210及び1212にグループ化される場合、これら2つの細胞集団を図示したようにピボットして、ピボット表1200を作成することができる。ピボット表1200において、行は、表200では列であった遺伝子である。表1200における列は、2つの細胞集団1210及び1212である。各表セルに、各細胞集団の細胞における関連付けられた遺伝子の遺伝子カウントの接続（concatenation）がポピュレートされる。図12の例において、これらの値に対して行われる計算は、ピボット表1200aに示されるような平均を計算するための直接平均である。やはりまた、実務者によって所望される場合、他の計算を行うことができることを理解されたい。次に、図11の散布図1100によって示されるタイプの遺伝子ビュー散布図をピボット表1200aから（表200ではなくピボット表1200aを用いる以外は）図4で説明される幾つかの基本技術を用いて作成することができる。

【0031】

図11に戻ると、遺伝子ビュー散布図1100は、このようにして、極めて大きいサイズであり得る遺伝子のリストにわたって2つの細胞集団間の示差的な遺伝子発現を可視化

10

20

30

40

50

する強力な方法をユーザに提供する。本発明者らは、散布図 1 1 0 0 が、細胞遺伝子発現データを、実務者に数多くの広範な新しい研究オプションを開放するように可視化する、先駆的な新たな方式を表すものと信じる。図 1 1 の例示的な遺伝子ビュー散布図 1 1 0 0 は、 x/y 軸集団の指定におけるフレキシビリティ及び集団が変化したときの遺伝子発現グラフの動的な更新におけるフレキシビリティをユーザに与える。例示的な遺伝子ビュー散布図 1 1 0 0 はまた、有り余るほどのグラフ・オプション及び解像度をユーザに与える。重要なことに、分析はこのグラフで終わらない。後述するように、それは新たな遺伝子セットを作成し、データをより深く掘り下げる能力を伴って続く。特に新しいシングルセル法によって明らかにされる細胞の不均一性及びサンプル/患者間の差異の検査は、この手法なくしては不可能である。

10

【0032】

$y = x$ の対角線は、遺伝子 1 1 0 6 が、1 1 2 0 を介して定義されたメトリックに従って両方の細胞集団内で等しく発現される場合に配置される場所である。したがって、この対角線からいずれかの方向に遺伝子 1 1 0 6 が遠ざかった距離は、2 つの選択された細胞集団間での対象遺伝子 1 1 0 6 の示差的な発現の程度を示す。大部分の遺伝子 1 1 0 6 は多くの細胞集団において発現されない（又はごくわずかに発現する）ことが予想されると仮定すると、典型的には散布図 1 1 0 0 の左下象限内に遺伝子 1 1 0 6 の大きいクラスタが存在することが予想される。

【0033】

さらに、ユーザは、図 6 に関連して説明したワークスペース内のデータオブジェクトとして存在する複数の階層的に関連した細胞集団を作成することができるので、ユーザは、遺伝子ビューモードにある間にこれらの細胞集団のいずれかを選択することができる。さらに、ワークスペースを用いて、遺伝子ビューモードにおける探索のための細胞集団のブール演算式を作成する（例えば「非 B 細胞」であるすべての細胞集団の OR 結合は相補的な細胞集団を作成する）ことができることを理解されたい。図 1 3 A 及び図 1 3 B は、ワークスペース内のこのような相補的ブール細胞集団の例を示す。したがって、ユーザは、B 細胞の集団と「非 B」細胞の集団との間で示差的な遺伝子発現を示す散布図 1 1 0 0 を作成することができる。

20

【0034】

本明細書で説明する遺伝子ビューの別の強力かつ革新的な態様は、ユーザが、遺伝子ビューモードにある間に遺伝子をゲーティングすることにより遺伝子セットを作成する能力である。図 1 4 は、遺伝子ビュー散布図内にゲート 1 4 0 0 が作成され、ユーザがさらなる検査に値すると考える遺伝子のセットをキャプチャする例を示す。ゲート 1 4 0 0 は、図 6 に関連して上述した技術を用いて散布図上に描くことができる。このゲーティングは、ゲート 1 4 0 0 によって定められた遺伝子セットに対応するデータ構造をワークスペース内に作成する（遺伝子セットは、ユーザが定義したゲート 1 4 0 0 によって幾つの遺伝子 1 1 0 6 が取り囲まれるかに応じて 1 以上の遺伝子を含む）。次いでこのような遺伝子セットは、細胞ビューモードにある間に可視化のために選択することができる細胞遺伝子発現データ 1 1 2 に関連した合成パラメータとして役立つことができる。図 1 4 の例において、ゲート 1 4 0 0 は、台形であり、B 細胞対非 B 細胞において正の示差的発現を有する遺伝子 1 1 0 6 をキャプチャする。正の示差的発現の範囲は、ユーザによって、散布図の $y = x$ 対角線からの台形の対角線の距離により制御される。

30

40

【0035】

図 1 5 は、細胞遺伝子発現データ 1 1 2 を合成パラメータとしての遺伝子セットでどのように増補することができるかの例を（例示的な増補された表 2 0 0 の形で）示す。この表 2 0 0 は、パラメータとしての 2 つの遺伝子セットで増補されている（遺伝子セット 1 及び遺伝子セット 2 の列を参照）。この例において、遺伝子セット 1 は、{ 遺伝子 2 , 遺伝子 3 } から成り、遺伝子セット 2 は { 遺伝子 2 , 遺伝子 4 } から成る。次いでこれらの遺伝子セットのための表セル（各々が表の行内の異なるセルに対応する）を、対象の遺伝子セットを構成する遺伝子に対する遺伝子カウントの合計でポピュレートすることができ

50

る。したがって、ユーザは、細胞ビューモードに戻って、X軸及びY軸のパラメータの一方又は両方が遺伝子セットである細胞ビュー散布図を作成することができる。この例を図16によって示す。図16の各軸に対するパラメータ選択メニューは、利用可能な遺伝子セットオプションを一覧にしたセクションを含む。図16の細胞ビュー散布図は、このようにして、X軸パラメータが「B__GeneTable」と表示された遺伝子セットであり、Y軸パラメータが「B__GeneTable」と表示された遺伝子セットである細胞308の散布図を示す。このように、遺伝子セットを作成するための遺伝子ビュー散布図におけるゲーティング能力を、生成された遺伝子セットを細胞ビュー散布図において合成パラメータとして使用する能力と組み合わせ、ユーザに、細胞遺伝子発現データのマルチパラメータ・データスペースをインテリジェントに削減する予期せぬ能力を与えることを理解されたい。この組合せは、ユーザが、集団内の共通遺伝子の同定及び集団内の示差的遺伝子の同定という極めて困難な課題を克服することを可能にする。換言すれば、ユーザがどの遺伝子が共通であるのかを探索している場合又はどの遺伝子が異なるかを探索している場合のどちらであっても、新たな遺伝子セットを作成する能力及びそれらを単一パラメータとして見る／分析することは、ユーザが細胞集団と遺伝子セットとの間の重要な関係に焦点を合わせることを可能にする。さらに、ユーザが細胞集団の種々の比較に基づいて生物学的に興味深い遺伝子セットを作成すると、これらの遺伝子セットを（含まれる遺伝子の定義されたリストとして）他のユーザに渡すことによって、これらの遺伝子セットを他のユーザと共有することができるので、それら他のユーザは、その遺伝子セットを自身の細胞データのサンプルで評価することができる。異なる細胞データセットにわたるこのような共有される研究及び独立した研究に基づいて、細胞に関して遺伝子挙動へのより深い洞察を得ることが期待される。

【0036】

図17Aは、ワークスペースの例示的なビューを示し、これはワークスペースを、ワークスペース内に存在する種々の遺伝子セット1700、サンプル1702、及び細胞集団1704に分割する。図示したように、遺伝子セットを一覧にしたセクション1700は、各対象遺伝子セットに関するメタデータ（例えば、名称、遺伝子セット内の遺伝子のカウント、及び遺伝子セットの説明）を提供する種々の表示フィールドを含むことができる。名称及び説明は、ユーザによって編集可能なものとすることができ、その遺伝子セットの関連の特性に関してユーザに知らせるのに有用である。また、遺伝子セットは、遺伝子セット内に存在する何らかの階層関係を示すような方式で一覧にすることができる。

【0037】

遺伝子セット制御部1710（図17Bでより詳細に示す）は、遺伝子セットをブール演算によって他の遺伝子セットから作成する能力をユーザに与える。例えば、論理和（union）ボタンにより、ユーザは、リスト1700内の2つ以上の選択された遺伝子セットの和集合である新たな遺伝子セットを作成することができる。論理積（intersection）ボタンにより、ユーザは、リスト1700内の2つ以上の選択された遺伝子セットの両方／すべてに存在する遺伝子から成る新たな遺伝子セットを作成することができる。相補（complement）ボタンにより、ユーザは、リスト1700内の2つの遺伝子セット（遺伝子セット1及び遺伝子セット2）から相補的な遺伝子セットを作成することができ、ここで第1の相補的な新たな遺伝子セットは、遺伝子セット1にあるが遺伝子セット2の中になくすべての遺伝子から成るが、第2の相補的な新たな遺伝子セットは、遺伝子セット2の中にあるが遺伝子セット1の中になくすべての遺伝子から成る。「全比較（All Comparisons）」ボタンにより、ユーザは、上記の技術の各々により、同時に幾つかの新たな遺伝子セット（論理和演算、論理積演算、及び相補演算による新たな遺伝子セット）を作成することができる。既存の遺伝子セットから新たな遺伝子セットを作成するこのフレキシブルな能力は、処置条件、患者、実験等の間での遺伝子セットの比較を可能にする。したがって、制御部1710を使用することによって、ユーザは、遺伝子セットの組合せを含む遺伝子コレクションを作成することができる。

【0038】

10

20

30

40

50

プログラム 116 によって提供される可視化の別の強力かつ革新的な態様は、細胞ビュー及び/又は遺伝子ビュー散布図上に第3の次元を重ね合わせる能力を含む。例えば、カラーコード化（例えば、ヒートマップ）を散布図内の細胞308又は遺伝子1106に適用して、データの提示に別の次元を提供することができる。この例を図18A～図18Dによって示す。図18Aは、例示的な細胞ビュー散布図表示において第3の次元をどのように提示するかを定めるために第3の次元の制御部1800をどのように用いることができるかを示す。制御部1804を通じて、ユーザは、ヒートマップに用いる統計量を定めることができる。図18Bは、例示的な実施形態においてこの統計量1804のために利用可能なオプションを示す（例えば、メジアン、平均、幾何平均、変動係数（CV）、ロバストCV、標準偏差（SD）、ロバストSD、等）。ヒートマップ統計量は、以下のよう

10

に計算することができる。グラフ内の各ドットは、1つ以上の細胞を表す（通常、複数の細胞）。ドット毎に、選択された統計量（例えば、平均、メジアン等）を細胞（1つ又は複数）に対して計算する。したがってグラフ内に400ドットがある場合、400個の統計量が計算されることになる。400個の統計値に対して、最小及び最大が決定され、これがカラーマップ（勾配が1つの色から別の色へ変化するカラー値のアレイ）にインデックス付けするための下限及び上限として用いられ、すなわち最小値はカラーアレイ内のインデックス0にマッピングされ、最大値はカラーマップ内の最後のインデックスにマッピングされる。次いでカラーマップ内の色をなんらかの方式（値に対する色の線形分布など）で最小値と最大値との間の値に適用することができる。

【0039】

図18Cは、ユーザが、第3の次元として使用するために細胞遺伝子発現データからどのようにパラメータ1808を選択することができるかを示す。パラメータリスト1810に、細胞遺伝子発現データに対するパラメータのリスト（例えば表200の列）をポピュレートすることができる。この例において、ユーザは、図18Dに示す通り、カラーコードによって細胞ビュー散布図の上にオーバーレイする第3の次元として、パラメータHighTHighBを選択している。この第3の次元のオーバーレイは、散布図を有意味に解釈するさらなる能力をユーザに与える。例えば、ユーザは、カラーコード化を用いて、注目する細胞集団を同定し、そのような細胞集団をゲーティングする（図19のゲート1900を参照のこと）ことにより、さらなる調査のためにワークスペース内で別の新たなオブジェクトを作成することができる。

20

【0040】

さらに、開示されるシステムの別の態様によれば、細胞集団をワークスペースからレポートエディタ内にドラッグすることによって、レポートエディタ内でレポートを作成することができる（図20A参照）。別の例として、1つの細胞集団を別の細胞集団の上にドラッグして細胞集団をオーバーレイして、遺伝子セットによる2Dの重ね合わせ及びヒートマップを作成することができる。細胞集団のオーバーレイを図20A～図20Dに2つの方式で示す。散布図において、異なる集団は、同じグラフ内で異なる色のドットプロットとして示される（どちらの集団を他方の上に置くかを選ぶことができる）。ヒートマップにおいて、異なる集団は、編成されて、ヒートマップ内の追加列として水平方向に付加される。すなわち集団1のすべての細胞が一緒に描画されて標識され、集団2のすべての細胞がそれに続く、等々。図20Bは、遺伝子セット及び細胞集団による例示的なヒートマップを示す。図20Cは、図20Bの左側部分の拡大図を示し、図20Dは、図20Cの右側部分の拡大図を示す。

30

40

【0041】

例示的な使用事例

上記のように、開示されるシステムは、細胞ビューモードと遺伝子ビューモードとの間で（又はその逆に）切り替えるとともにこれら2つのビューモードにおいてゲーティングを行って注目するデータに焦点を合わせることによって、細胞遺伝子発現データ112を研究するための強力な機構を提供する。

【0042】

10

20

30

40

50

図 2 1 は、プログラム 1 1 4 の一部として実行するための例示的なプロセスフローを示し、これは、ユーザが細胞遺伝子発現データの深い研究をサポートするために細胞ビューと遺伝子ビューとの間をどのように切り替えることができるかを説明する。ステップ 2 1 0 0 において、プロセッサはユーザ入力に応答してワークスペースからサンプルを選択する。ステップ 2 1 0 2 において、細胞ビューモード又は遺伝子ビューモードのどちらを作動させるかに関する決定がなされる。この決定は、ユーザ入力応答して行うことができる。細胞ビューモードが選択された場合、プロセスフローはステップ 2 1 0 4 に進む。遺伝子ビューモードが選択された場合、プロセスフローはステップ 2 1 2 2 に進む。ステップ 2 1 0 4 - 2 1 2 0 は細胞ビューの動作に対応し、ステップ 2 1 2 2 - 2 1 4 0 は遺伝子ビューモードの動作に対応する。

10

【 0 0 4 3 】

ステップ 2 1 0 4 において、プロセッサは、ユーザ入力に応答して細胞ビューの軸パラメータを選択する（例えば、遺伝子又は遺伝子セットといったパラメータの選択）。ステップ 2 1 0 6 において、プロセッサは、ステップ 2 1 0 4 で選択された軸パラメータに基づいて細胞遺伝子発現データ 1 1 2 から細胞ビュー散布図データ構造を生成する。このステップは、表 2 0 0 において、選択されたパラメータに対応する列を選択して、細胞のリストと、各々選択されたパラメータに対するそれらの関連付けられた値とを得ることを伴い得る。ステップ 2 1 0 8 において、プロセッサは、ユーザに提示するために、ステップ 2 1 0 6 において作成した細胞ビュー散布図データ構造から細胞ビュー散布図 3 0 0 を生成する。

20

【 0 0 4 4 】

ステップ 2 1 1 0 において、プロセッサは、ユーザからの入力に応答して細胞ビュー散布図に関するゲートの規格を受け取る。このゲーティングは、細胞集団を作成し（ステップ 2 1 1 2 ）、作成された細胞集団は、新たなデータ構造としてワークスペース内に保存される。この時点で、ユーザは、（ 1 ）新たなサンプルを扱う（ステップ 2 1 1 4 でステップ 2 1 0 0 に戻る進行を参照）、（ 2 ）細胞ビューモードにある間に現在のサンプルに関して 1 つ以上の新たな軸パラメータを定義する（ステップ 2 1 1 6 でステップ 2 1 0 4 に戻る進行を参照）、（ 3 ）細胞ビューモードにある間に現在のサンプルに関して新たなゲートを定める（ステップ 2 1 1 8 でステップ 2 1 1 0 に戻る進行を参照）、又は（ 4 ）遺伝子ビューモードに切り替える（ステップ 2 1 2 0 でステップ 2 1 2 2 に進む進行を参照）のいずれかを選択することができる。

30

【 0 0 4 5 】

ステップ 2 1 2 2 において、プロセッサは、上で論じたように細胞遺伝子発現データ 1 1 2 をピボットする。次いで、ステップ 2 1 2 4 において、プロセッサは、ユーザ入力に応答してワークスペースから細胞集団を選択する。ステップ 2 1 2 6 において、プロセッサは、ステップ 2 1 2 4 で選択された細胞集団に基づいて、ピボットされた細胞遺伝子発現データから遺伝子ビュー散布図データ構造を生成する。このステップは、ピボットバージョンの表 2 0 0 において、選択された細胞集団に対応するピボット列を選択して、遺伝子のリストと、各々選択された細胞集団に対するそれらの関連付けられたメトリックとを得ることを伴い得る。ステップ 2 1 2 8 において、プロセッサは、ユーザに提示するために、ステップ 2 1 2 6 において作成した遺伝子ビュー散布図データ構造から遺伝子ビュー散布図 1 1 0 0 を生成する。

40

【 0 0 4 6 】

ステップ 2 1 3 0 において、プロセッサは、ユーザからの入力に応答して遺伝子ビュー散布図に関するゲートの規格を受け取る。このゲーティングは、遺伝子セットを作成し（ステップ 2 1 3 2 ）、作成された遺伝子セットは、細胞遺伝子発現データとの関係のために新たな合成パラメータとしてワークスペース内に保存される。このようにして、細胞遺伝子発現データを、ステップ 2 1 3 2 において生成された遺伝子セットに対応する新たなデータ値で増補することができる。この時点で、ユーザは、（ 1 ）新たなサンプルを扱う（ステップ 2 1 3 4 でステップ 2 1 0 0 に戻る進行を参照）、（ 2 ）遺伝子ビューモード

50

にある間に現在のサンプルに関して1つ以上の新たな細胞集団を定義する(ステップ2136でステップ2124に戻る進行を参照)、(3)遺伝子ビューモードにある間に現在のサンプルに関して新たなゲートを定める(ステップ2138でステップ2130に戻る進行を参照)、又は(4)細胞ビューモードに切り替える(ステップ2140でステップ2104に進む進行を参照)のいずれかを選択することができる。

【0047】

このように、図21は、どのようにして、ユーザがこのシステムを用いて、細胞ビューモードと遺伝子ビューモードとの間ですばやく移行する一方で、モード間の切替後の可視化を補助するのに使用することができる細胞集団及び遺伝子セットをそれぞれゲーティングによって作成することができるかを示す。

【0048】

一例として、強力かつ革新的な動作モードが図22の例示的なプロセスフローによって示され、ここでユーザは、図21のプロセスフローによってプログラム114と対話し、(1)表示の細胞ビューモードにおいて1つ以上の細胞集団を作成し(ステップ2200)、(2)複数の細胞集団にわたって遺伝子発現を比較して表示する遺伝子ビューモードにピボットし(ステップ2202)、(3)表示の遺伝子ビューモードにおいて1つ以上の遺伝子セットを作成し(ステップ2204)、これは細胞遺伝子発現データを合成パラメータとしての新たな遺伝子セットで増補し(ステップ2206)、(4)これら遺伝子セットの1つ以上を細胞ビュー散布図のための軸パラメータとして用いて、細胞ビュー散布図にピボットして戻り(ステップ2208)、(5)これらの動作を所望により反復的に繰り返して、細胞遺伝子発現データ112内に存在し得る生物学的に妥当な関係を掘り下げる(ここで動作(1)及び(3)をユーザが定義した第3の次元を細胞ビュー又は遺伝子ビューデータ表示上にオーバーレイすることによって補助することができる)。例えば、開示されたシステムは、精密な/個別化された医学を研究するための強力なツールとしての役立つことができ、多数の患者由来の細胞データを評価して、種々のがん若しくは他の疾病/病理に対する生存及び治療応答との相関が高い遺伝子を有する患者を見いだすような作業を行うことができる。

【0049】

例えば、本明細書で提供されるツールを通じて、ユーザは、細胞集団を分析して、特定のがんX(Cancer X)に関してより高い生存率(chances for survival)と相関する示差的に発現された遺伝子セット(これを「生存遺伝子セット(Survival Gene Set)」として標識することができる)を同定することが可能であり得る。同時に、ユーザは、細胞集団を分析して、がんXに関してより低い生存率と相関する示差的に発現された遺伝子セット(これを「非生存遺伝子セット(Not Survival Gene Set)」として標識することができる)を同定することが可能であり得る。さらにまた、ユーザは、細胞集団を分析して、がんXのための治療Y(Therapy Y)に対して良好に応答することと相関する示差的に発現された遺伝子セット(これを「治療応答遺伝子セット(Therapy Responsive Gene Set)」として標識することができる)を同定することが可能であり得る。次いでこれらの遺伝子セットをシステムの細胞ビューモードにおける合成パラメータとして用いて、がんXを生き延びるための治療Yによる処置に対して良好に応答する遺伝的な素因のある患者内の細胞集団を見いだすことができる。例えば、細胞ビュー散布図は、生存遺伝子セットを一方の軸パラメータ(例えばX軸パラメータ)として用い、非生存遺伝子セットを他方の軸パラメータ(例えばY軸パラメータ)として用いることができる一方で、治療応答遺伝子セットを第3の次元のオーバーレイとして用いることができる。得られた散布図は、生存及び治療応答の両方との相関が高い細胞集団(並びに生存又は治療応答との相関が低い細胞集団)を示すことができる。

【0050】

本発明をその例示的な実施形態に関連して説明してきたが、本発明の範囲内に入る種々の変更をこれに対して行うことができる。例えば、本明細書において示される例示的な散

10

20

30

40

50

布図は、X軸を水平軸、Y軸を垂直軸として提示されているが、実務家によっては散布図を傾けることが望ましいと気付く場合もある。一例は、対角線 $y = x$ が生物学的に重要だと思われるシナリオの場合である。このような場合、散布図を、 $y = x$ の対角線が45度の線ではなく水平線又は垂直線として提示されるように傾けることができ、それによりデータが $y = x$ の線からどれだけ離れているかということにユーザが集中することを助ける。したがって、本発明に対するこれら及び他の変更は、本明細書における教示を検討すると認識可能である。

【符号の説明】

【0051】

100	：コンピュータシステム	10
200	：細胞遺伝子発現データの表	
300	：細胞ビュー散布図	
300	：ユーザ制御部	
302、304	：遺伝子セクタ	
306	：ユーザ選択ファイル	
308	：ドット	
310、312	：ゼロレベル	
320	：ゲート作成ツール	
322	：ナビゲーションツール	
500、506	：パラメータ選択メニュー	20
600、900、1400	：ゲート	
700	：「展開ゼロ」ユーザ制御部	
1000	：ヒストグラム	
1100	：遺伝子ビューモードの散布図	
1110、1112	：集団選択メニュー	
1200	：ピボット表	
1700	：遺伝子セット	
1702	：サンプル	
1704	：細胞集団	
1710	：遺伝子セット制御部	30
1800	：第3の次元の制御部	

【図面】

【図 1】

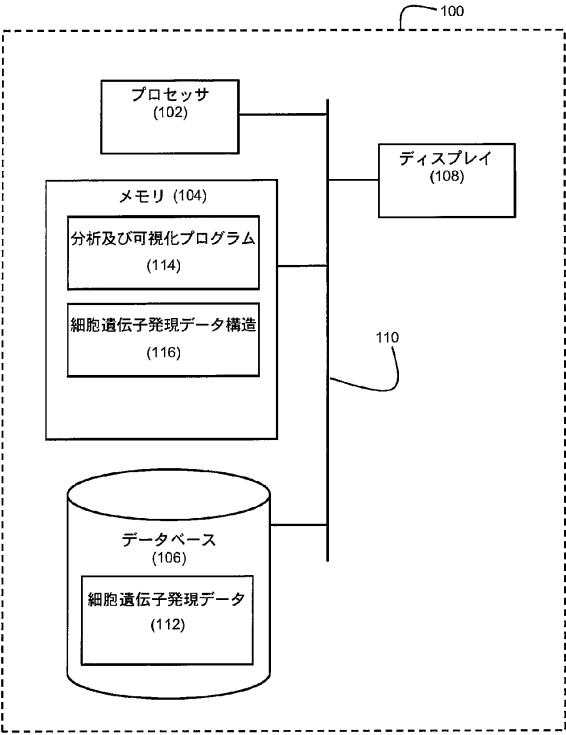


Figure 1

【図 2 A】

{細胞 ID 1, パラメータ 1 (ID, 値), パラメータ 2 (ID, 値), ..., パラメータ n (ID, 値)}
{細胞 ID 2, パラメータ 1 (ID, 値), パラメータ 2 (ID, 値), ..., パラメータ n (ID, 値)}

...

Figure 2A

10

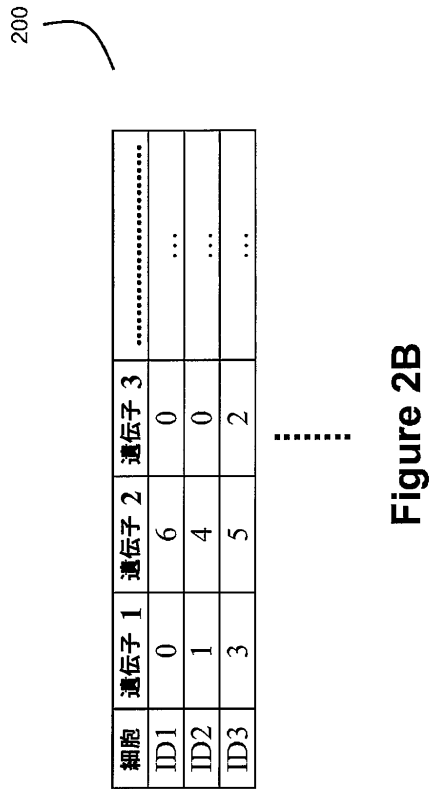
20

30

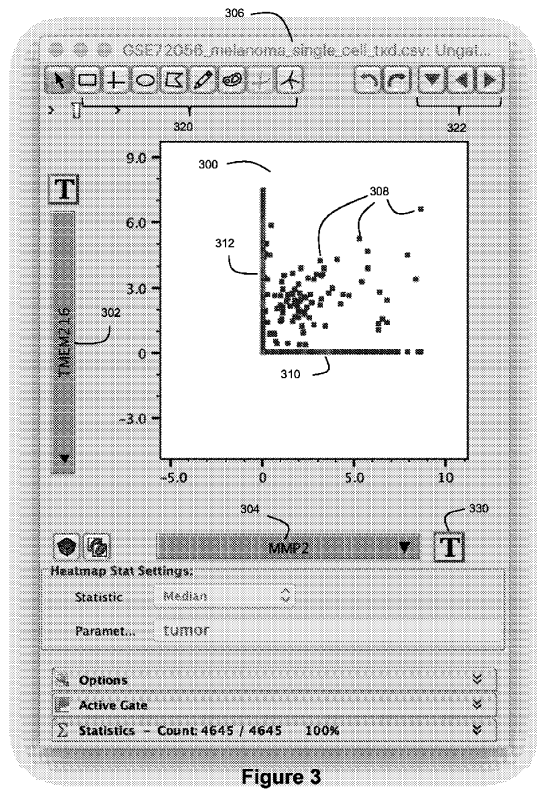
40

50

【図 2 B】



【図 3】



【図 4】

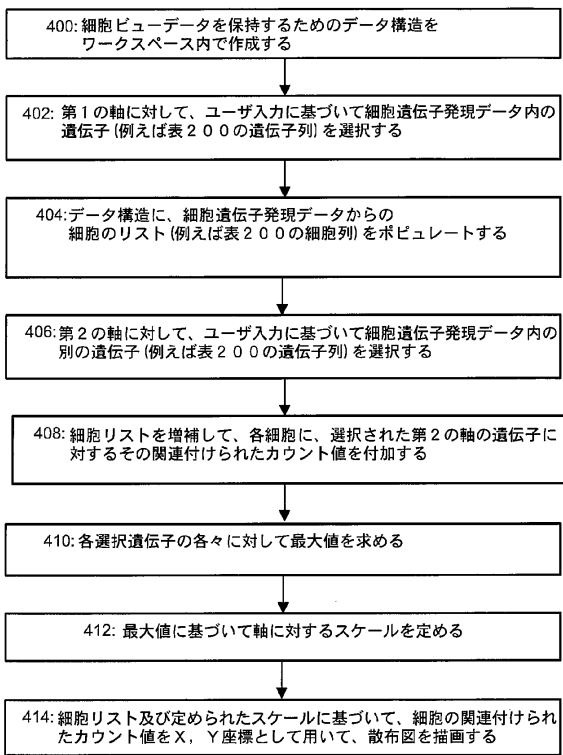
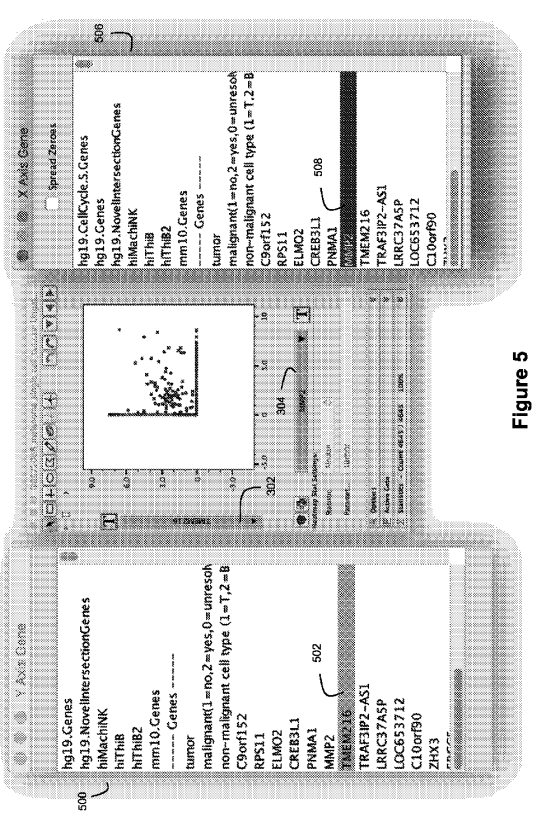


Figure 4

【図 5】



10

20

30

40

50

【図 6】

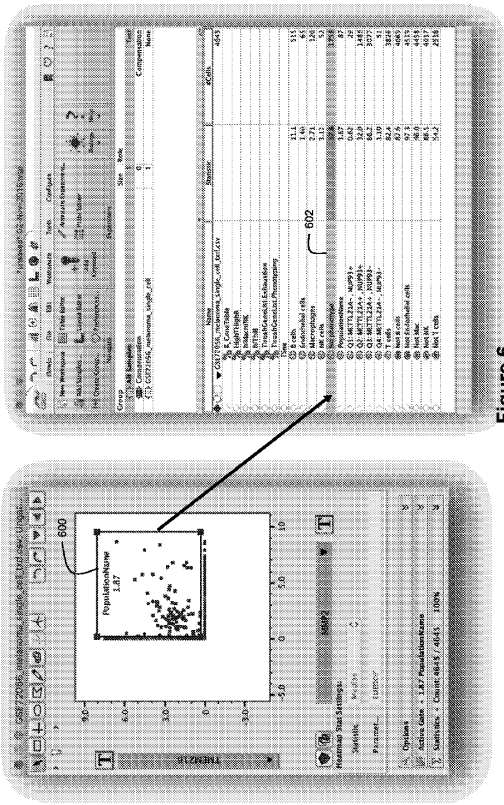


Figure 6

【図 7】

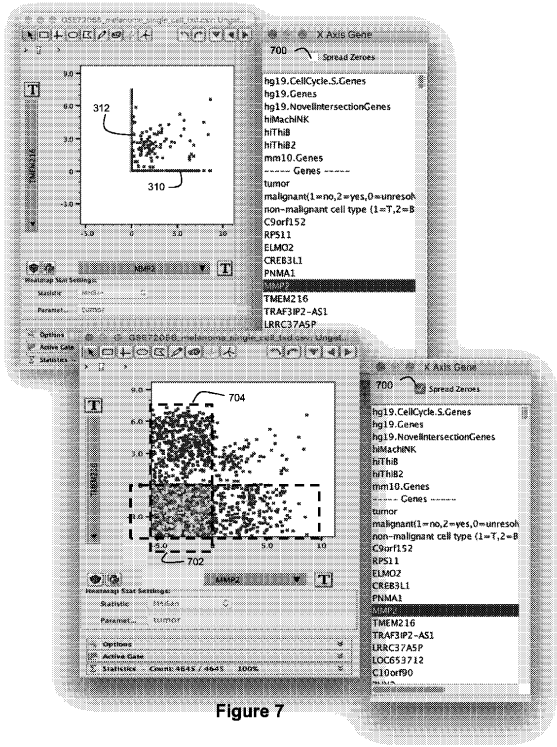


Figure 7

【図 8】

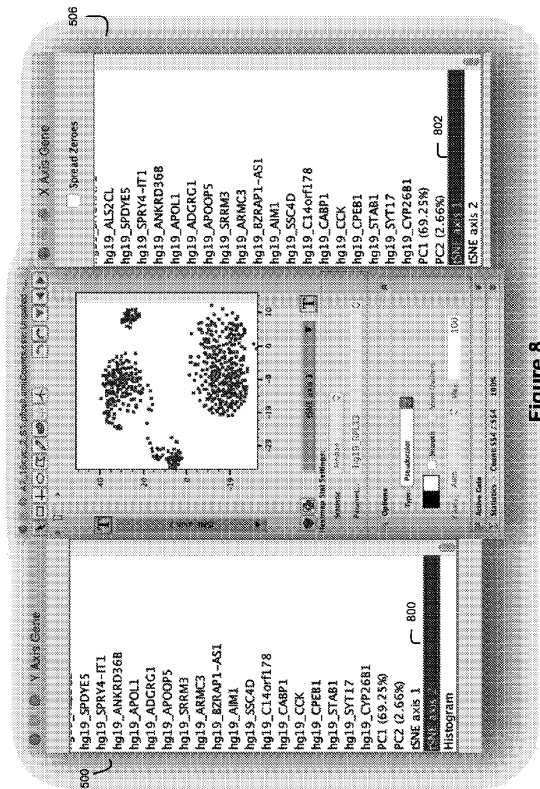


Figure 8

【図 9】

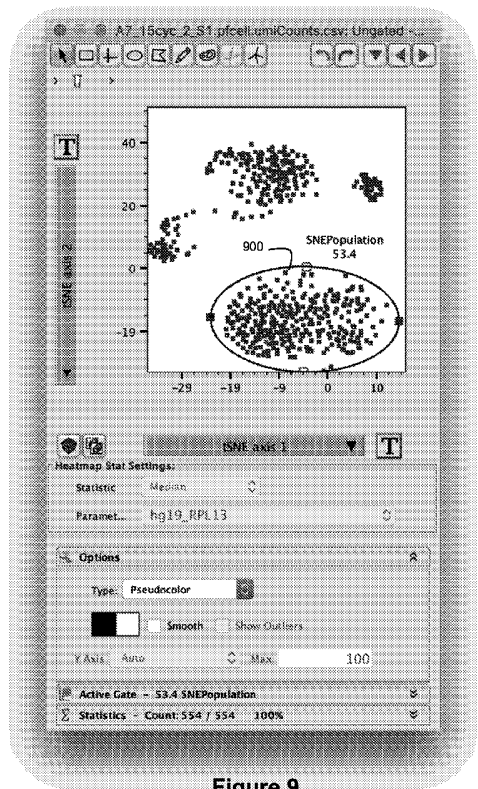


Figure 9

10

20

30

40

50

【 図 1 0 】

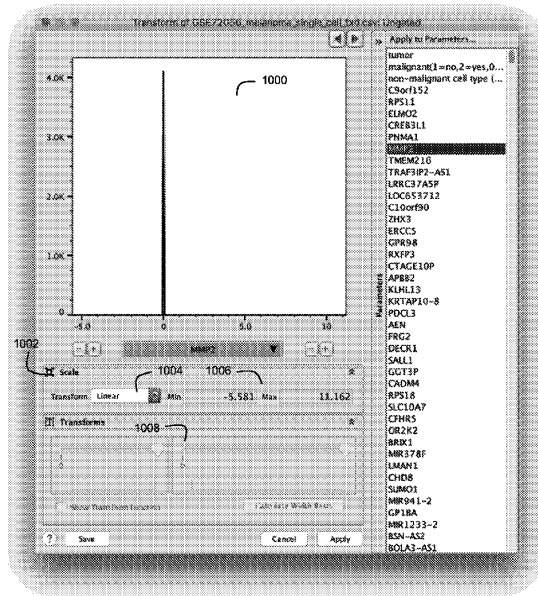


Figure 10

【 図 1 1 】

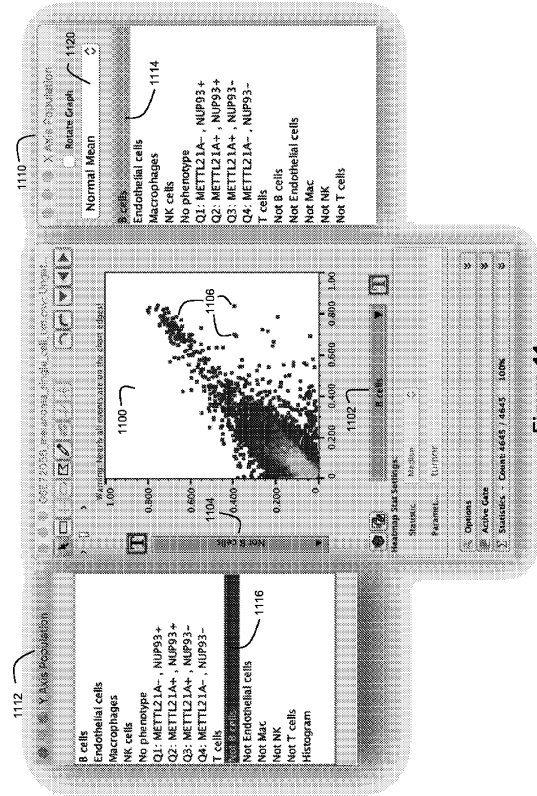


Figure 11

【 図 1 2 】

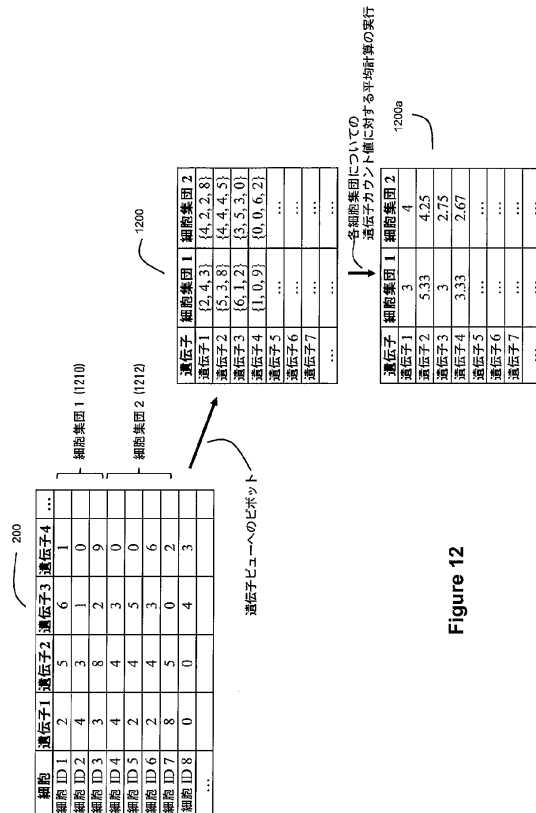


Figure 12

【 図 1 3 A 】

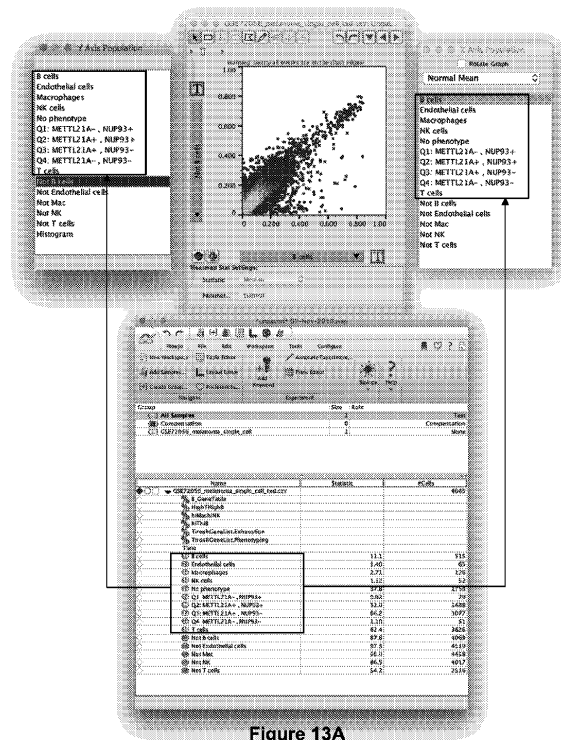


Figure 13A

【図 13 B】

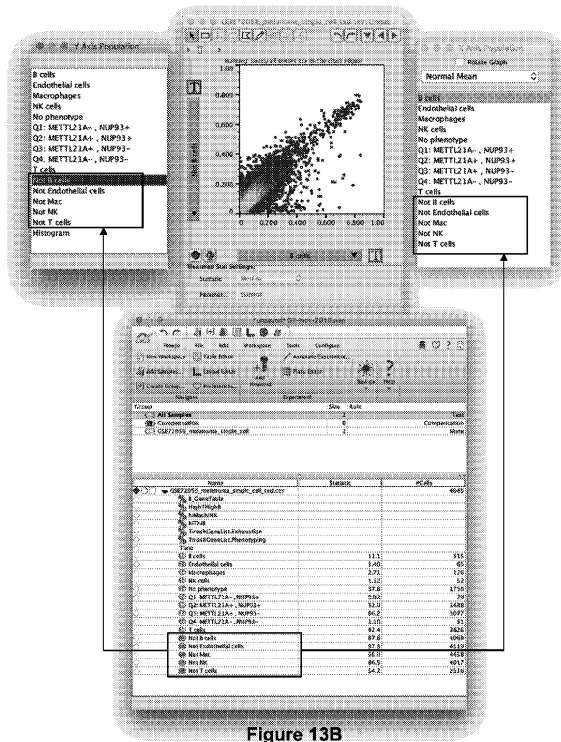


Figure 13B

【図 14】

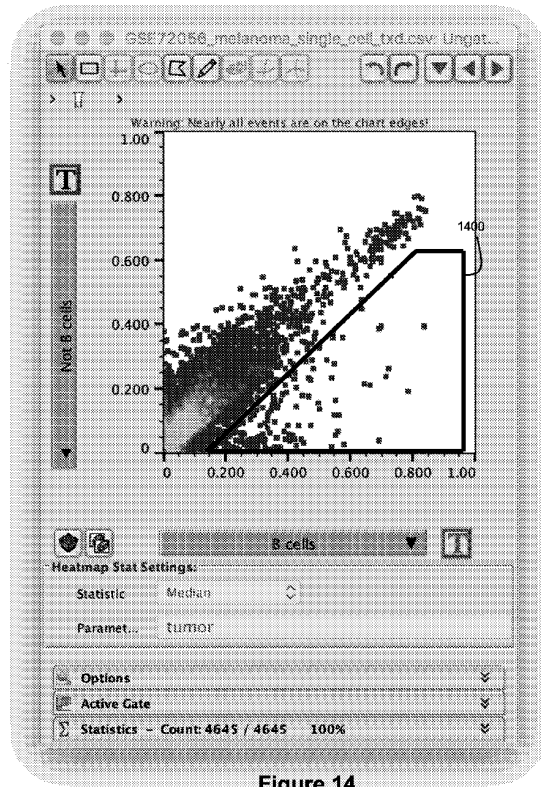


Figure 14

【図 15】

細胞	遺伝子1	遺伝子2	遺伝子3	遺伝子4	遺伝子セット1	遺伝子セット2	...
細胞 ID 1	2	5	6	1	Sum(5,6)=11	Sum(2,1)=3	...
細胞 ID 2	4	3	1	0	Sum(3,1)=4	Sum(4,0)=4	...
細胞 ID 3	3	8	2	9	Sum(8,2)=10	Sum(3,9)=12	...
細胞 ID 4	4	4	3	0	Sum(4,3)=7	Sum(4,0)=4	...
細胞 ID 5	2	4	5	0	Sum(4,5)=9	Sum(2,0)=2	...
細胞 ID 6	2	4	3	6	Sum(4,3)=7	Sum(2,6)=8	...
細胞 ID 7	8	5	0	2	Sum(5,0)=5	Sum(8,2)=10	...
細胞 ID 8	0	0	4	3	Sum(0,4)=4	Sum(0,3)=3	...

Figure 15

【図 16】

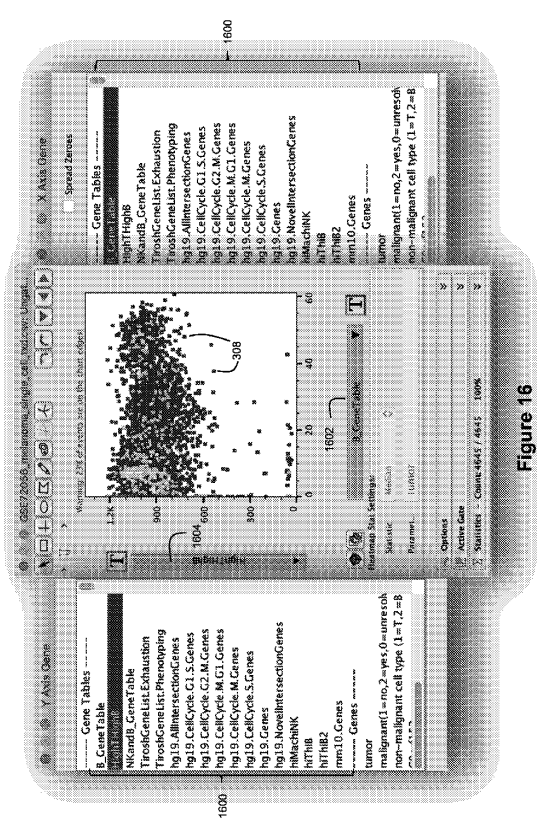


Figure 16

10

20

30

40

50

【図 17 A】

Name	# Genes	Description
Apoptosis Collection		
B-Cell Collection	18	VaccinatedRT36_d24_TStudyMemory x Naive
GeneSet1	290	VaccinatedRT36_d24_TStudyMemory2 x Naive2
GeneSet2		
T-Cell Collection	54	user description
Activation	12	VaccinatedRT36_d24_TStudyMemory x Naive
GeneSet3	20	pathological/source file.csv?(user editable description)
Immunophenotype	105	user description
Regulatory	110	Activation OR Regulatory Description

Samples

1) All Samples

2) VaccinatedRT36_d24_TStudy

3) VaccinatedRT36_d24_TStudy

Populations

1) VaccinatedRT36_d24_TStudy

2) Above threshold

3) T cells

4) Memory

5) Naive

Figure 17A

【図 17 B】

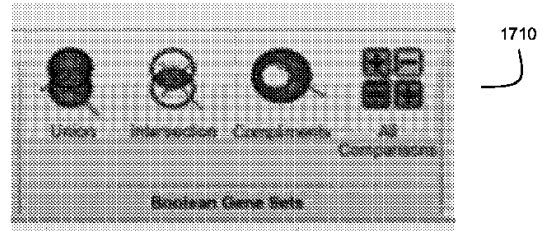


Figure 17B

【図 18 A】

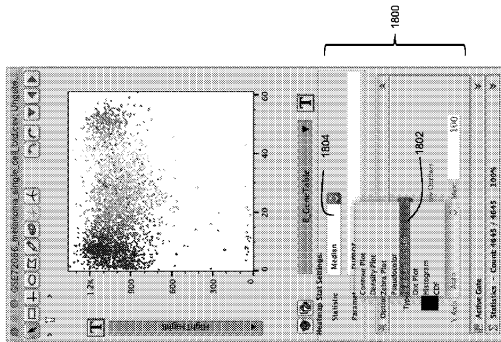


Figure 18A

【図 18 B】

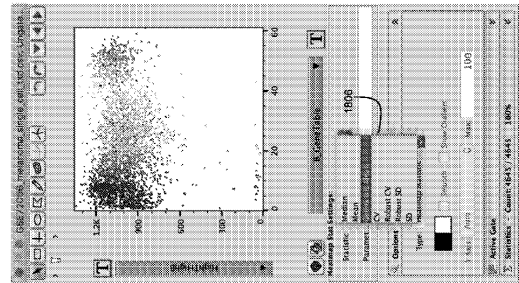


Figure 18B

【 図 1 8 C 】

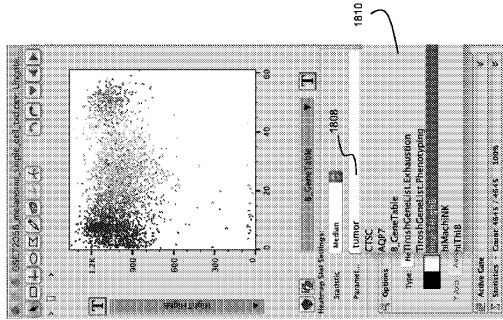


Figure 18C

【 図 1 8 D 】

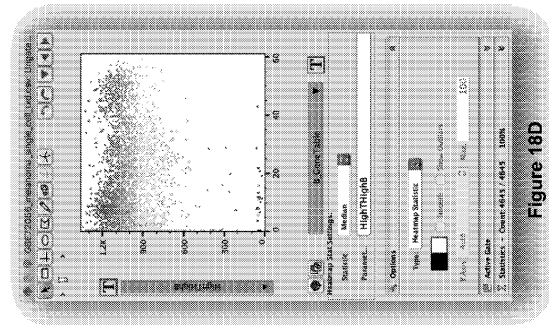


Figure 18D

10

【 図 1 9 】

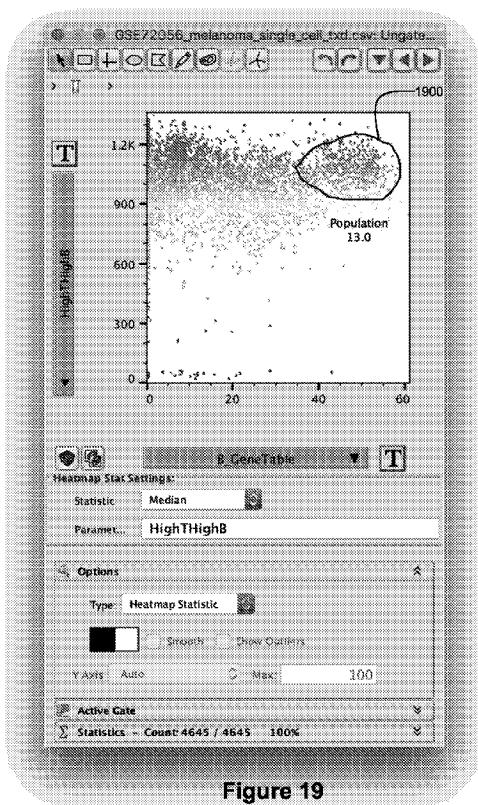


Figure 19

【 図 2 0 A 】

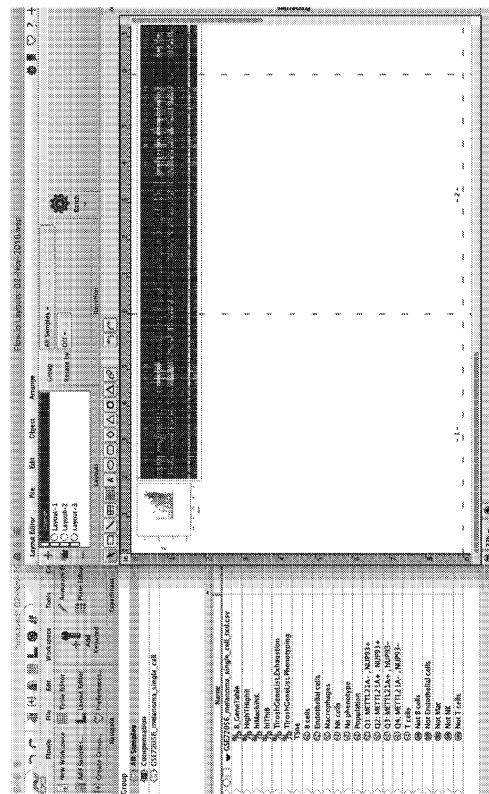


Figure 20A

20

30

40

50

【 図 2 0 B 】



Figure 20B

【図 20 C】

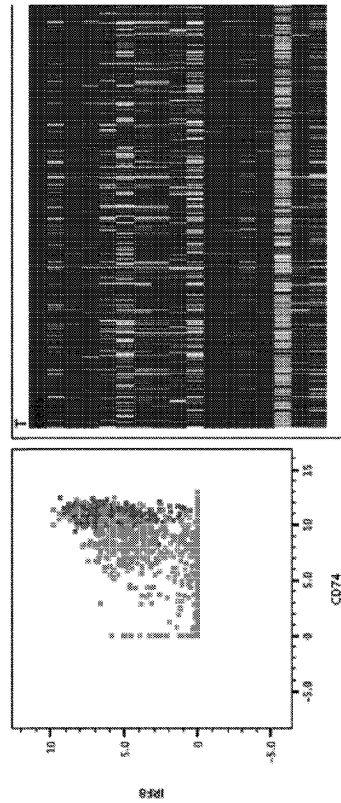


Figure 20C

【 図 2 0 D 】

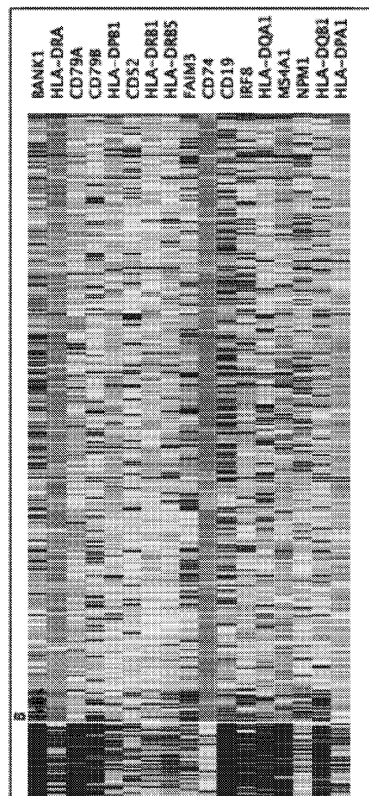


Figure 20D

【 図 2 1 】

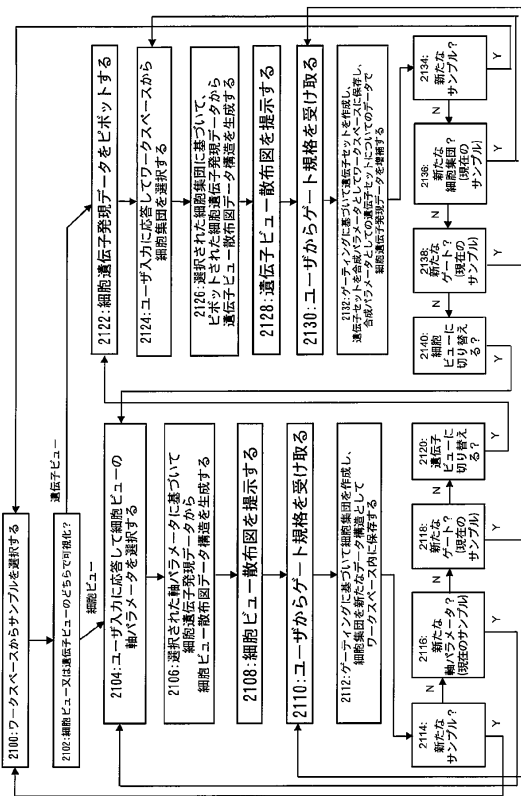


Figure 21

【 図 2 2 】

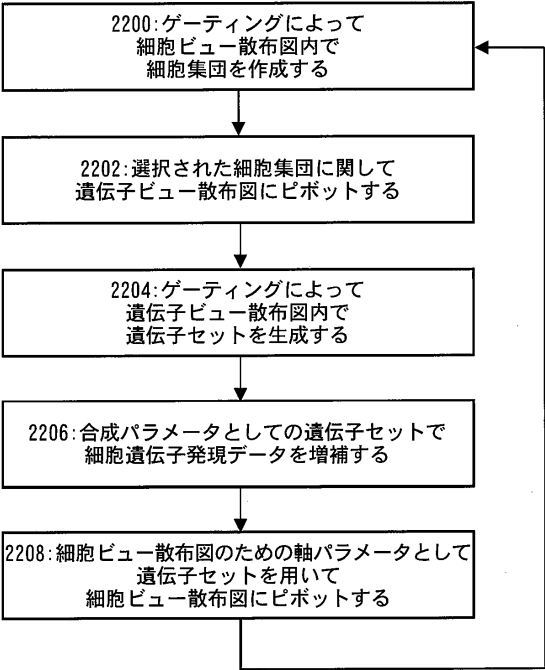


Figure 22

フロントページの続き

(74)代理人 100067013
弁理士 大塚 文昭

(74)代理人 100109335
弁理士 上杉 浩

(74)代理人 100120525
弁理士 近藤 直樹

(74)代理人 100139712
弁理士 那須 威夫

(74)代理人 100176418
弁理士 工藤 嘉晃

(72)発明者 アルマロード ジェイムズ エドワード
アメリカ合衆国 オレゴン州 97355 レバノン サウスウェスト ラーソン ロード 31011

(72)発明者 シュビードレン ヨーゼフ
カナダ ヴィクトリア 2E3 ブリティッシュ コロンビア コキットラム ランネル ドライヴ #
20-2561

(72)発明者 スタドニスキー マイケル デイヴィッド
アメリカ合衆国 オレゴン州 97520 アッシュランド サウス マウンテン アベニュー 755

審査官 加舎 理紅子

(56)参考文献 特開2005-352771(JP,A)
特表2001-511546(JP,A)
米国特許出願公開第2002/0067358(US,A1)
米国特許出願公開第2010/0070904(US,A1)
米国特許出願公開第2016/0130574(US,A1)
国際公開第2004/072866(WO,A1)

(58)調査した分野 (Int.Cl., DB名)
G16B 5/00 - 99/00
G01N 15/14