

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2022-501011

(P2022-501011A)

(43) 公表日 令和4年1月6日(2022.1.6)

(51) Int.Cl.			F I			テーマコード (参考)		
<b>C 1 2 Q</b>	<b>1/06</b>	<b>(2006.01)</b>	C 1 2 Q	1/06				4 B O 2 9
<b>C 1 2 Q</b>	<b>1/6869</b>	<b>(2018.01)</b>	C 1 2 Q	1/6869		Z		4 B O 6 3
<b>C 1 2 M</b>	<b>1/00</b>	<b>(2006.01)</b>	C 1 2 M	1/00		A		
<b>C 1 2 M</b>	<b>1/34</b>	<b>(2006.01)</b>	C 1 2 M	1/34		A		

審査請求 未請求 予備審査請求 未請求 (全 79 頁)

(21) 出願番号 特願2021-501001 (P2021-501001)  
 (86) (22) 出願日 令和1年7月16日 (2019.7.16)  
 (85) 翻訳文提出日 令和3年1月12日 (2021.1.12)  
 (86) 国際出願番号 PCT/US2019/041976  
 (87) 国際公開番号 W02020/018519  
 (87) 国際公開日 令和2年1月23日 (2020.1.23)  
 (31) 優先権主張番号 62/805,888  
 (32) 優先日 平成31年2月14日 (2019.2.14)  
 (33) 優先権主張国・地域又は機関 米国 (US)  
 (31) 優先権主張番号 62/805,884  
 (32) 優先日 平成31年2月14日 (2019.2.14)  
 (33) 優先権主張国・地域又は機関 米国 (US)

(71) 出願人 520445473  
 フラッグシップ パイオニアリング イノベーションズ シックス, エルエルシー  
 アメリカ合衆国マサチューセッツ州02142, ケンブリッジ, ケンブリッジ・パークウェイ 55, エイス・フロア  
 (74) 代理人 100079049  
 弁理士 中島 淳  
 (74) 代理人 100084995  
 弁理士 加藤 和詳

最終頁に続く

(54) 【発明の名称】 細胞を分析する方法

(57) 【要約】

第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差を表す単一細胞の移行シグネチャーにアクセスする。変化した状態は、第1の細胞状態から変化した細胞状態への細胞の移行を経て生じる。移行シグネチャーは、複数の細胞成分の同定、並びに、各このような細胞成分について、それぞれの細胞成分の発現の変化と、第1の細胞状態と変化した細胞状態との間の細胞状態の変化との間の関連性を定量する対応する第1の重要度スコアを含む。移行シグネチャーは、非摂動細胞と摂動に曝された摂動細胞との間の細胞成分の発現の差を表す摂動シグネチャーと比較される。摂動シグネチャーは、各それぞれの細胞成分について、(i) 非摂動細胞と摂動細胞との間の細胞成分の発現の変化と、(ii) 非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第2の重要度スコアを含む。

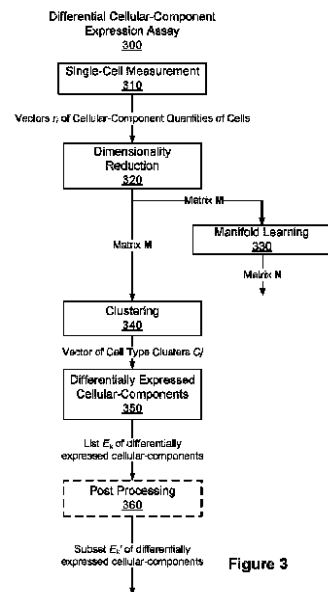


Figure 3

## 【特許請求の範囲】

## 【請求項 1】

摂動が細胞の移行に影響を与えるかどうかを予測するための方法であって、前記方法が

メモリ及び1つ以上のプロセッサを含むコンピュータシステムにおいて：

第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーに電子的形態でアクセスする工程であって、前記変化した細胞状態は、前記第1の細胞状態から前記変化した細胞状態への細胞の移行を経て生じ、前記単一細胞の移行シグネチャーは、複数の細胞成分の同定及び、前記複数の細胞成分中の各それぞれの細胞成分について、前記それぞれの細胞成分の発現の変化と、前記第1の細胞状態と前記変化した細胞状態との間の細胞状態の変化との間の関連性を定量する対応する第1の重要度スコアを含む、工程と；

複数の非摂動細胞と摂動に曝された複数の摂動細胞との間の細胞成分の発現の差の尺度を表す摂動シグネチャーに電子的形態でアクセスする工程であって、前記摂動シグネチャーは、前記複数の細胞成分の全て又は一部の同定及び、前記複数の細胞成分の全て又は一部の中の各それぞれの細胞成分について、(i)前記複数の非摂動細胞と前記複数の摂動細胞との間のそれぞれの細胞成分の発現の変化と、(ii)前記複数の非摂動細胞と前記複数の摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第2の重要度スコアを含む、工程と；

前記単一細胞の移行シグネチャー及び前記摂動シグネチャーを比較し、それによって、前記摂動が前記細胞の移行に影響を与えるかどうかを決定する工程とを含む、方法。

## 【請求項 2】

前記単一細胞の移行シグネチャーにアクセスする工程が、

(i)第1の複数の第1の単一細胞細胞成分の発現データセット、及び(ii)第2の複数の第2の単一細胞細胞成分の発現データセットに基づいて、前記単一細胞の移行シグネチャーを決定する工程を含み、ここで：

前記第1の複数の第1の単一細胞細胞成分の発現データセット中の各それぞれの第1の単一細胞細胞成分の発現データセットが、前記第1の細胞状態の第1の複数の細胞の対応する単一細胞から得られ、

前記第2の複数の第2の単一細胞細胞成分の発現データセット中の各それぞれの第2の単一細胞細胞成分の発現データセットが、前記変化した細胞状態の第2の複数の細胞の対応する単一細胞から得られる、請求項1に記載の方法。

## 【請求項 3】

前記第1の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第1の複数の細胞成分ベクトル中の、対応する細胞成分ベクトルを含み、

前記第2の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第2の複数の細胞成分ベクトル中の、対応する細胞成分ベクトルを含み、

前記第1及び第2の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルが、複数の要素を含み、前記それぞれの細胞成分ベクトル中の各それぞれの要素が、前記複数の細胞成分中の対応する細胞成分に関連しており、前記第1及び第2の複数の単一細胞細胞成分の発現データセットの前記それぞれのデータセットによって表される前記対応する単一細胞についての前記対応する細胞成分の量を表す対応する値を含む、請求項2に記載の方法。

## 【請求項 4】

前記第1及び/又は前記第2の複数の単一細胞細胞成分の発現データセットに対して次元削減を行って、複数の次元削減成分を生成する工程と；

前記第1及び第2の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルについて、前記複数の次元削減成分を、前記それぞれの細胞成分ベクトルに適用して、前記複数の次元削減成分中の各それぞれの次元削減成分についての次元削減成分値を含む対応する次元削減ベクトルを形成し、それによって、対応する第1及び第2の複数の次元削減ベク

トルを形成する工程と；

クラスタリングを行って、クラスターのセット  $C_j$  を生成する工程であって、各クラスターが、前記第 1 及び第 2 の複数の次元削減ベクトルのサブセットに対応する複数の点を含む、工程と；

前記クラスターのセット  $C_j$  の第 1 のクラスターから前記第 1 の複数の細胞を同定する工程と；

前記クラスターのセット  $C_j$  の第 2 のクラスターから前記第 2 の複数の細胞を同定する工程とをさらに含み、

前記方法が、任意に、前記対応する第 1 及び第 2 の複数の次元削減ベクトルを用いて多様体学習を行って、前記第 1 及び第 2 の複数の細胞における各細胞の他の各細胞に対する相対的な細胞状態を同定する工程をさらに含む、請求項 3 に記載の方法。

10

【請求項 5】

前記複数の非摂動細胞が、摂動に曝されていない対照細胞であり、又は前記非摂動細胞が、摂動に曝された非関連摂動細胞の平均である、請求項 1 に記載の方法。

【請求項 6】

前記単一細胞の移行シグネチャー及び前記摂動シグネチャーをブルーニングして、前記複数の細胞成分を、任意に RNA レベルで測定される転写因子に制限する工程をさらに含む、請求項 1 に記載の方法。

【請求項 7】

前記単一細胞の移行シグネチャーを決定する工程が、

20

母平均の差の検定、ウィルコクソンの順位和検定、t 検定、ロジスティック回帰、及び一般化線形モデルのうちの一つを用いて、(i) 前記第 1 の複数の第 1 の単一細胞細胞成分の発現データセットと、前記第 2 の複数の第 2 の単一細胞細胞成分の発現データセットとの間の前記複数の細胞成分にわたる細胞分量の差を決定する工程を含む、請求項 2 に記載の方法。

【請求項 8】

細胞成分の発現の差の尺度が、ウィルコクソンの順位和検定、t 検定、ロジスティック回帰、及び一般化線形モデルのうちの一つを用いて、(i) 第 3 の複数の第 3 の単一細胞細胞成分の発現データセットと、(ii) 第 4 の複数の第 4 の単一細胞細胞成分の発現データセットとの間の細胞分量の差を定量し、ここで：

30

前記第 3 の複数の第 3 の単一細胞細胞成分の発現データセット中の各それぞれの第 3 の単一細胞細胞成分の発現データセットが、前記複数の非摂動細胞中の対応する単一細胞から得られ、

前記第 4 の複数の第 4 の単一細胞細胞成分の発現データセット中の各それぞれの第 4 の単一細胞細胞成分の発現データセットが、摂動に曝された前記複数の摂動細胞中の第 4 の複数の細胞の対応する単一細胞から得られる、請求項 1 に記載の方法。

【請求項 9】

前記単一細胞の移行シグネチャー及び前記摂動シグネチャーをフィルタリングして、前記単一細胞の移行シグネチャー及び前記摂動シグネチャーに含まれる細胞成分の数を減少させる工程をさらに含み、

40

任意に、前記単一細胞の移行シグネチャー及び前記摂動シグネチャーをフィルタリングする工程が、閾値 p 値にしたがって、又は細胞成分の閾値数にしたがって、前記単一細胞の移行シグネチャー及び前記摂動シグネチャーに含まれる細胞成分の数を減少させる工程を含む、請求項 1 に記載の方法。

【請求項 10】

それぞれの細胞成分についての前記対応する第 2 の重要度スコアを決定する工程が、

前記複数の細胞成分中の各それぞれの細胞成分について、前記それぞれの細胞成分についての前記重要度スコアを、前記それぞれの細胞成分についての対応するマッチングスコアで置き換える工程と；

前記複数の細胞成分についての前記マッチングスコアを組み合わせ、前記摂動につい

50

てのマッチングスコアを生成する工程と；

前記摂動が、前記それぞれの摂動についての前記マッチングスコアに基づいて、前記第1の細胞状態と前記変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とを含み、

任意に、前記対応するマッチングスコアが、別個の又は連続したスコアを含む、請求項1に記載の方法。

【請求項11】

前記重要度スコアを置き換える工程が、

前記それぞれの細胞成分についての前記単一細胞の移行シグネチャーからの前記細胞成分量及び前記それぞれの細胞成分についての前記摂動シグネチャーからの前記細胞成分量が両方とも上方制御される場合、前記重要度スコアを第1のスコアで置き換える工程と；

前記それぞれの細胞成分についての前記単一細胞の移行シグネチャーからの前記細胞成分量が上方制御され、前記それぞれの細胞成分についての前記摂動シグネチャーからの前記細胞成分量が下方制御される場合、前記重要度スコアを第2のスコアで置き換える工程と；

前記それぞれの細胞成分についての前記摂動シグネチャーからの前記細胞成分量が、有意に上方制御又は下方制御されない場合、前記重要度スコアを第3のスコアで置き換える工程とを含む、請求項10に記載の方法。

【請求項12】

前記重要度スコアを置き換える工程が、

前記それぞれの細胞成分についての前記単一細胞の移行シグネチャーからの前記細胞成分量及び前記細胞成分についての前記摂動シグネチャーからの前記細胞成分量が両方とも下方制御される場合、前記重要度スコアを第1のスコアで置き換える工程と；

前記それぞれの細胞成分についての前記単一細胞の移行シグネチャーからの前記細胞成分量が下方制御され、前記細胞成分についての前記摂動シグネチャーからの前記細胞成分量が上方制御される場合、前記重要度スコアを第2のスコアで置き換える工程と；

前記細胞成分についての前記摂動シグネチャーからの前記細胞成分量が、有意に上方制御又は下方制御されない場合、前記重要度スコアを第3のスコアで置き換える工程とを含む、請求項10に記載の方法。

【請求項13】

前記複数の細胞成分が、任意にRNAレベルで測定される複数の遺伝子を含む、請求項1に記載の方法。

【請求項14】

前記第1の複数の第1の単一細胞細胞成分の発現データセット及び前記第2の複数の第2の単一細胞細胞成分の発現データセット中の各単一細胞細胞成分の発現データセットが、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスポーズでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seq、及びそれらの任意の組合せ、並びに単一細胞細胞成分の発現データセットにおける活性化された経路を表す線形組合せなどの組合せを含む、上記のものの集合からなる群から選択される方法を用いて生成される、請求項2に記載の方法。

【請求項15】

前記比較に基づいて、前記変化した細胞状態を促進するものとして前記摂動を同定する工程、又は

前記比較に基づいて、前記変化した細胞状態を阻害するものとして前記摂動を同定する工程をさらに含む、請求項1に記載の方法。

【請求項16】

前記細胞の移行シグネチャー及び前記摂動シグネチャーが、異なるタイプの細胞成分を用いて生成される、請求項1に記載の方法。

10

20

30

40

50

## 【請求項 17】

前記細胞の移行シグネチャー及び前記摂動シグネチャーが、同じタイプの細胞成分を用いて生成される、請求項 1 に記載の方法。

## 【請求項 18】

前記電子的形態でアクセスする工程が、複数の摂動における各それぞれのシグネチャーについて行われ、それによって、複数の摂動シグネチャーを得て、

前記比較する工程が、前記単一細胞の移行シグネチャー及び前記摂動シグネチャーを、複数の摂動シグネチャーにおける各それぞれのシグネチャーと比較し、それによって、前記第 1 の細胞状態と前記変化した細胞状態との間の細胞の移行に関連する前記複数の摂動のサブセットを決定する、請求項 1 に記載の方法。

10

## 【請求項 19】

1 つ以上のプロセッサ及びメモリを含むコンピュータシステムであって、前記メモリが、摂動が細胞の移行に影響を与えるかどうかを予測するための方法を実行するための命令を保存し、前記方法が、

第 1 の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーに電子的形態でアクセスする工程であって、前記変化した細胞状態は、前記第 1 の細胞状態から前記変化した細胞状態への細胞の移行を経て生じ、前記単一細胞の移行シグネチャーは、複数の細胞成分の同定及び、前記複数の細胞成分中の各それぞれの細胞成分について、前記それぞれの細胞成分の発現の変化と、前記第 1 の細胞状態と前記変化した細胞状態との間の細胞状態の変化との間の関連性を定量する対応する第 1 の重要度スコアを含む、工程と；

20

複数の非摂動細胞と摂動に曝された複数の摂動細胞との間の細胞成分の発現の差の尺度を表す摂動シグネチャーに電子的形態でアクセスする工程であって、前記摂動シグネチャーは、前記複数の細胞成分の全て又は一部の同定及び、前記複数の細胞成分の全て又は一部の中の各それぞれの細胞成分について、( i ) 前記複数の非摂動細胞と前記複数の摂動細胞との間のそれぞれの細胞成分の発現の変化と、( i i ) 前記複数の非摂動細胞と前記複数の摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第 2 の重要度スコアを含む、工程と；

前記単一細胞の移行シグネチャー及び前記摂動シグネチャーを比較し、それによって、前記摂動が前記細胞の移行に影響を与えるかどうかを決定する工程とを含む、コンピュータシステム。

30

## 【請求項 20】

摂動が細胞の移行に影響を与えるかどうかを予測するための、コンピュータによって実行可能な 1 つ以上のコンピュータプログラムを保存する非一時的なコンピュータ可読媒体であって、前記コンピュータが、1 つ以上のプロセッサ及びメモリを含み、前記 1 つ以上のコンピュータプログラムが、

第 1 の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーに電子的形態でアクセスする工程であって、前記変化した細胞状態は、前記第 1 の細胞状態から前記変化した細胞状態への細胞の移行を経て生じ、前記単一細胞の移行シグネチャーは、複数の細胞成分の同定及び、前記複数の細胞成分中の各それぞれの細胞成分について、前記それぞれの細胞成分の発現の変化と、前記第 1 の細胞状態と前記変化した細胞状態との間の細胞状態の変化との間の関連性を定量する対応する第 1 の重要度スコアを含む、工程と；

40

複数の非摂動細胞と摂動に曝された複数の摂動細胞との間の細胞成分の発現の差の尺度を表す摂動シグネチャーに電子的形態でアクセスする工程であって、前記摂動シグネチャーは、前記複数の細胞成分の全て又は一部の同定及び、前記複数の細胞成分の全て又は一部の中の各それぞれの細胞成分について、( i ) 前記複数の非摂動細胞と前記複数の摂動細胞との間のそれぞれの細胞成分の発現の変化と、( i i ) 前記複数の非摂動細胞と前記複数の摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第 2 の重要度スコアを含む、工程と；

50

前記単一細胞の移行シグネチャー及び前記摂動シグネチャーを比較し、それによって、前記摂動が前記細胞の移行に影響を与えるかどうかを決定する工程とを含む方法を実行するコンピュータで実行可能な命令を集合的に符号化する、非一時的なコンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、2018年7月16日に提出された米国仮特許出願第62/698,701号、2019年2月14日に提出された同第62/805,884号、及び2019年2月14日に提出された同第62/805,888号に対する優先権の利益を主張するものであり、これらの仮特許出願の全内容が、参照により本明細書に援用される。

10

【0002】

本発明は、一般に、細胞を分析するためのシステム及び方法に関する。より具体的には、本発明は、摂動が細胞の移行に影響を与えるかどうかを予測することに関する。

【背景技術】

【0003】

細胞機構の研究は、疾患を理解するために重要である。

【0004】

組織は、個々の細胞の複雑な生態系であり、そこでは、細胞状態の調節異常が疾患の基礎となる。既存の創薬努力は、細胞を健康な状態から疾患状態へと移行させる分子機構を特徴付け、これらの移行を食い止めるか又は阻害する薬理的な手法を特定しようとしている。過去の努力はまた、これらの移行を特徴付ける分子シグネチャーを同定し、これらのシグネチャーを食い止める薬理的な手法を特定しようとしてきた。

20

【0005】

表面マーカーによって豊かになった組織又は細胞における大量の細胞コレクションについての分子データは、集団中の個々の細胞の表面型多様性及び分子多様性を隠している。これらの大量の細胞コレクション中の細胞の不均一性により、疾患促進機構を解明することを目的とした現在の取り組みの結果が、誤解を招くか又は完全に誤ったものにさえなる。単一細胞RNAシーケンシングなどの新たな手法は、分子レベルで個々の細胞を特徴付けることができる。これらのデータは、様々な細胞状態をより高い解明度(resolution)で理解するための基盤を提供し、細胞が有する状態の豊かで且つ驚くべき多様性を明らかにする。

30

【0006】

単一細胞のデータを解釈する場合、重大な問題が存在し、すなわち、これらのデータのまばらさ(sparsity)により、細胞中に存在する分子の存在、及びノイズが見落とされ、これらの分子測定の正確性における不確かさが伴うことである。したがって、新たな手法が、個々の細胞状態を制御するための薬理的な手法への洞察を導き、それに応じて疾患を解明するために必要とされている。

【0007】

化学物質(小分子、細胞外リガンド、mRNA、siRNA、及びその他を含む)のコンピュータによるポジショニング及びリポジショニングは、創薬を加速させるのに非常に大きな可能性を有する。過去の手法は、小分子で摂動されるバルク細胞から生じる発現シグネチャーの差を、健康な状態と疾患状態との間の細胞の発現デルタにマッピングしていた。この手法は、可能性を有するが、その現在の形態では、バルク細胞の不均一性及び分子的に摂動される細胞及び病変細胞の著しい細胞型の差のため、限られた適用性を有する。

40

【0008】

上記の背景を考慮すると、当該技術分野において必要とされているのは、向上した細胞分析を可能にするシステム及び方法である。特に、摂動が細胞の移行に影響を与えるかど

50

うかの予測を可能にすることが必要とされている。

【発明の概要】

【課題を解決するための手段】

【0009】

本開示は、上に示される欠点に対処するものである。本開示は、単一細胞のデータ及び分子摂動のデータを主なデータ基盤として用いて、これらの欠点に、少なくとも部分的に対処し、機械学習を用いて、天然の多様な細胞状態の理解を精緻化し、細胞が別の状態を選択する主な移行状態を明らかにし、細胞状態の変化の根底にある分子機構の理解を促進し、これらの状態変化を制御するための薬理的な手法を発見するものである。

【0010】

本開示の一態様は、摂動が細胞の移行に影響を与える（例えば、移行を促進するか又は移行を阻害する）かどうかを予測するための方法を提供する。本方法は、単一細胞の移行シグネチャーに電子的形態でアクセスする工程を含む。移行シグネチャーは、第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す。変化した細胞状態は、第1の細胞状態から変化した細胞状態への細胞の移行を経て生じる。単一細胞の移行シグネチャーは、複数の細胞成分の同定を含む。複数の細胞成分中の各それぞれの細胞成分について、対応する第1の重要度スコアが、それぞれの細胞成分の発現の変化と、第1の細胞状態と変化した細胞状態との間の細胞状態の変化との間の関連性を定量する。実際に、いくつかの単一細胞の移行シグネチャーが、このように得られ、それぞれが、第1の細胞状態と異なる変化した細胞状態との間の細胞成分の発現の差の尺度を表す。したがって、いくつかの異なる変化した細胞状態が、本出願の開示内容を用いて同時に分析され得る。

【0011】

本方法は、摂動シグネチャーに電子的形態でアクセスする工程をさらに含む。ある実施形態において、摂動シグネチャーは、1つ以上の非摂動細胞と摂動に曝された1つ以上の摂動細胞との間の細胞成分の発現の差の尺度を表す。さらに、摂動シグネチャーは、複数の細胞成分の全て又は一部の同定を含む。複数の細胞成分の全て又は一部の中の各それぞれの細胞成分について、1つ以上の非摂動細胞と1つ以上の摂動細胞との間のそれぞれの細胞成分の発現の変化と、1つ以上の非摂動細胞と1つ以上の摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第2の重要度スコア。実際に、いくつかの摂動シグネチャーが、このように得られ、それぞれが、1つ以上の非摂動細胞と、複数の摂動における異なる摂動に曝された1つ以上の摂動細胞との間の細胞成分の発現の差の尺度を表す。さらに、本方法は、単一細胞の移行シグネチャー、又は複数の単一細胞の移行シグネチャー、及び摂動シグネチャー、又は複数の摂動シグネチャーを比較し、それによって、摂動（又は複数の摂動）が、変化した状態への細胞の移行に影響を与える（又は複数の変化した状態への複数の細胞の移行に影響を与える）かどうかを決定する工程を含む。ある実施形態において、2、3、4、10又はそれ以上、例えば、15、20、25、30、40、50、60、70、80、90、若しくは100、又はそれ以上の変化した状態が、このように同時に分析される。ある実施形態において、2、3、4、10又はそれ以上、例えば、15、20、25、30、40、50、60、70、80、90、若しくは100又はそれ以上の摂動が、このように同時に分析される。

【0012】

ある実施形態において、単一細胞の移行シグネチャーにアクセスする工程は、第1の複数の第1の単一細胞細胞成分の発現データセット、及び第2の複数の第2の単一細胞細胞成分の発現データセットに基づいて、単一細胞の移行シグネチャーを決定する工程を含む。第1の複数の第1の単一細胞細胞成分の発現データセット中の各それぞれの第1の単一細胞細胞成分の発現データセットが、第1の細胞状態の第1の複数の細胞の対応する単一細胞から得られる。さらに、第2の複数の第2の単一細胞細胞成分の発現データセット中の各それぞれの第2の単一細胞細胞成分の発現データセットが、変化した細胞状態の第2の複数の細胞の対応する単一細胞から得られる。

10

20

30

40

50

## 【 0 0 1 3 】

ある実施形態において、第 1 の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第 1 の複数の細胞成分ベクトル中の、対応する細胞成分ベクトルを含む。さらに、第 2 の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第 2 の複数の細胞成分ベクトル中の、対応する細胞成分ベクトルを含む。さらに、第 1 及び第 2 の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルが、複数の要素を含む。それぞれの細胞成分ベクトル中の各それぞれの要素が、複数の細胞成分中の対応する細胞成分に関連しており、第 1 及び第 2 の複数の単一細胞細胞成分の発現データセットのそれぞれのデータセットによって表される対応する単一細胞についての対応する細胞成分の量を表す対応する値を含む。

10

## 【 0 0 1 4 】

ある実施形態において、次元削減が、第 1 及び / 又は第 2 の複数の単一細胞細胞成分の発現データセットにおいて行われて、複数の次元削減成分を生成する。したがって、第 1 及び第 2 の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルについて、複数の次元削減成分が、それぞれの細胞成分ベクトルに適用されて、複数の次元削減成分中の各それぞれの次元削減成分についての次元削減成分値を含む対応する次元削減ベクトルを形成する。これにより、対応する第 1 及び第 2 の複数の次元削減ベクトルが形成される。本方法は、クラスタリングを行って、クラスターのセット  $C_j$  を生成する工程を含む。各クラスターは、第 1 及び第 2 の複数の次元削減ベクトルのサブセットに対応する複数の点を含む。第 1 の複数の細胞は、クラスターのセット  $C_j$  の第 1 のクラスターから同定され、第 2 の複数の細胞は、クラスターのセット  $C_j$  の第 2 のクラスターから同定される。

20

## 【 0 0 1 5 】

ある実施形態において、多様体学習が、対応する第 1 及び第 2 の複数の次元削減ベクトルを用いて行われて、第 1 及び第 2 の複数の細胞における各細胞の他の各細胞に対する相対的細胞状態が同定される。

## 【 0 0 1 6 】

ある実施形態において、複数の非摂動細胞は、摂動に曝されていない対照細胞であり、又は非摂動細胞は、摂動に曝された非関連摂動細胞の平均である。

## 【 0 0 1 7 】

ある実施形態において、本方法は、単一細胞の移行シグネチャー及び摂動シグネチャーをブルーニングして、複数の細胞成分を転写因子に制限する工程をさらに含む。

30

## 【 0 0 1 8 】

ある実施形態において、単一細胞の移行シグネチャーを決定する工程は、母平均の差の検定、ウィルコクソンの順位和検定、 $t$  検定、ロジスティック回帰、及び一般化線形モデルのうちの 1 つを用いて、第 1 の複数の第 1 の単一細胞細胞成分の発現データセットと第 2 の複数の第 2 の単一細胞細胞成分の発現データセットとの間の複数の細胞成分にわたる細胞分量の差を決定する工程を含む。

## 【 0 0 1 9 】

ある実施形態において、細胞成分の発現の差の尺度は、母平均の差の検定、ウィルコクソンの順位和検定、 $t$  検定、ロジスティック回帰、及び一般化線形モデルのうちの 1 つを用いて、第 3 の複数の第 3 の単一細胞細胞成分の発現データセットと第 4 の複数の第 4 の単一細胞細胞成分の発現データセットとの間の細胞分量の差を定量する。第 3 の複数の第 3 の単一細胞細胞成分の発現データセット中の各それぞれの第 3 の単一細胞細胞成分の発現データセットが、複数の非摂動細胞の対応する単一細胞から得られ、第 4 の複数の第 4 の単一細胞細胞成分の発現データセット中の各それぞれの第 4 の単一細胞細胞成分の発現データセットが、摂動に曝された複数の摂動細胞中の第 4 の複数の細胞の対応する単一細胞から得られる。

40

## 【 0 0 2 0 】

ある実施形態において、単一細胞の移行シグネチャー及び摂動シグネチャーは、フィルタリングされて、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分

50



の数が減少される。ある実施形態において、単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングする工程は、閾値  $p$  値にしたがって、又は細胞成分の閾値数にしたがって、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させる工程を含む。

【0021】

ある実施形態において、それぞれの細胞成分についての対応する第2の重要度スコアを決定する工程は、それぞれの細胞成分についての重要度スコアを、複数の細胞成分中の各それぞれの細胞成分について、それぞれの細胞成分についての対応するマッチングスコアで置き換える工程を含む。マッチングスコアは、複数の細胞成分のために組み合わせられて、摂動についてのマッチングスコアが生成される。摂動が、それぞれの摂動についてのマッチングスコアに基づいて、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかの決定が行われる。

10

【0022】

ある実施形態において、対応するマッチングスコアは、別個の又は連続したスコアを含む。

【0023】

ある実施形態において、重要度スコアを置き換える工程は、それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞分量及びそれぞれの細胞成分についての摂動シグネチャーからの細胞分量が両方とも上方制御される場合、重要度スコアを第1のスコアで置き換える工程を含む。それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞分量が上方制御され、それぞれの細胞成分についての摂動シグネチャーからの細胞分量が下方制御される場合、重要度スコアは、第2のスコアで置き換えられる。さらに、それぞれの細胞成分についての摂動シグネチャーからの細胞分量が、有意に上方制御又は下方制御されない場合、重要度スコアは、第3のスコアで置き換えられる。

20

【0024】

ある実施形態において、重要度スコアを置き換える工程は、それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞分量及び細胞成分についての摂動シグネチャーからの細胞分量が両方とも、それらの同等物（例えば、それぞれ第1の細胞状態及び非摂動状態）と比較して下方制御される場合、重要度スコアを第1のスコアで置き換える工程を含む。それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞分量が、それらの同等物（例えば、それぞれ第1の細胞状態及び非摂動状態）と比較して下方制御され、細胞成分についての摂動シグネチャーからの細胞分量が上方制御される場合、重要度スコアは、第2のスコアで置き換えられる。さらに、細胞成分についての摂動シグネチャーからの細胞分量が、それらの同等物（例えば、それぞれ第1の細胞状態及び非摂動状態）と比較して有意に上方制御又は下方制御されない場合、重要度スコアは、第3のスコアで置き換えられる。

30

【0025】

ある実施形態において、複数の細胞成分は、複数の遺伝子を含む。

【0026】

ある実施形態において、第1の複数の第1の単一細胞細胞成分の発現データセット及び第2の複数の第2の単一細胞細胞成分の発現データセット中の各単一細胞細胞成分の発現データセットが、単一細胞リボ核酸 (RNA) シーケンシング (scRNA-seq)、scTag-seq、シーケンシングを用いたトランスポゼースでアクセス可能なクロマチン (transposase-accessible chromatin) の単一細胞アッセイ (scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seq、又はそれらの組合せ若しくは集合を含む方法を用いて生成される。

40

【0027】

ある実施形態において、本方法は、比較に基づいて、変化した細胞状態を促進するもの

50

として摂動を同定する工程をさらに含む。

【0028】

ある実施形態において、細胞の移行シグネチャー及び摂動シグネチャーは、異なるタイプの細胞成分を用いて生成される。ある実施形態において、細胞の移行シグネチャー及び摂動シグネチャーは、同じタイプの細胞成分を用いて生成される。

【0029】

ある実施形態において、電子的形態でアクセスする工程は、複数の摂動における各それぞれのシグネチャーについて行われ、それによって、複数の摂動シグネチャーを得る。さらに、比較する工程は、単一細胞の移行シグネチャー及び摂動シグネチャーを、複数の摂動シグネチャーにおける各それぞれのシグネチャーと比較し、それによって、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連する複数の摂動のサブセットを決定する。

10

【0030】

本開示の別の態様は、複数の単一細胞細胞成分の発現データセットにアクセスする工程を含む方法を提供する。各単一細胞細胞成分の発現データセットが、同じ「前駆」細胞型から移行された複数の細胞の1つの細胞から得られる。各データセットは、細胞成分のベクトル  $r_i$  を含む。細胞成分のベクトル  $r_i$  中の各エントリは、複数の細胞成分の1つに関連し、各エントリの値は、細胞についての細胞成分の量を表す。本方法は、データセットの次元削減を行って、行列  $M$  を生成する工程をさらに含む（例えば、図1の無次元削減成分ストア146-1のものなどの複数の無次元成分）。行列  $M$  は、第1の次元における行及び第2の次元における列を含む。各行は、複数の細胞の1つに対応する。行列  $M$  の値は、第1及び第2の次元空間中の点に位置する細胞成分の量から生成される値を含む。本方法は、クラスタリングを行って、クラスタのセット  $C_j$  を生成する工程をさらに含む。各クラスタは、行列  $M$  中の行のサブセットに対応する複数の点、及びそれらの対応する細胞を含む。本方法は、クラスタのセット  $C_j$  を用いて、細胞についての、異なって発現された細胞成分  $E_k$  のセットを決定する工程をさらに含む。

20

【0031】

特定の実施形態において、本方法は、点の相対的類似度の近似による行列  $M$  を用いて多様体学習を行って、行列  $N$  を生成する工程をさらに含む。行列  $N$  は、複数の行（行列  $M$  の行と同じ行）及び2つの列を含む。各行は、複数の細胞の1つに対応し、2つの列のそれぞれが、2次元空間の2つの次元の1つに対応する。行列  $N$  の値は、データセットに基づいて、各細胞の他の各細胞に対する相対的な細胞型を示す。

30

【0032】

データセットが得られる複数の細胞の複数の実施形態がある。特定の実施形態において、複数の細胞は、単一細胞細胞成分の発現データセットが得られるとき、様々な細胞型を有する不均一な細胞集団である。さらなる実施形態において、複数の細胞は、「前駆」細胞型を有する細胞の均一な集団であり、単一細胞細胞成分の発現データセットは、細胞が「前駆」細胞型から移行する際に複数の時点のそれぞれで得られ、複数のデータセットの異なるデータセットが、各独自の細胞及び時点の組合せのために収集されるようになっている。このような実施形態において、複数の時点は、少なくとも3つの時点を含み得る。さらなる実施形態において、複数の時点は、複数の細胞のかなりの割合が「前駆」細胞型から移行されていない「前駆」時点を含み得る。あるさらなる実施形態において、複数の時点は、複数の細胞のかなりの割合が「前駆」細胞型から移行された移行時点を含み得る。あるさらに他の実施形態において、複数の時点は、細胞のかなりの割合が「前駆」細胞型から少なくとも部分的に移行された少なくとも1つの中間時点を含み得る。

40

【0033】

複数の細胞成分はまた、様々であり得る。例えば、本明細書に開示される方法のある実施形態において、複数の細胞成分は、核酸、タンパク質、脂質、炭水化物、ヌクレオチド、及びそれらの任意の組合せからなる群から選択される。このような実施形態において、核酸は、DNA及びRNAからなる群から選択され得る。さらなる実施形態において、R

50

NAは、コーディング及びノンコーディングRNAからなる群から選択され得る。特定の実施形態において、複数の単一細胞細胞成分の発現データセットは、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスポーズでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seq及びそれらの任意の組合せ又は集合からなる群から選択される方法を用いて生成される。

#### 【0034】

次元削減は、データセットに対して行われて、複数の方法で(例えば、上述される行列Mの形態で)次元削減成分を生成し得る。特定の実施形態において、次元削減を行う工程は、単一細胞細胞成分の発現データセットに対する主成分分析(PCA)を行って、次元削減成分を生成する工程を含む。さらなる実施形態において、次元削減を行う工程は、データセットに対して拡散マップ及び/又はニューラルネットワークオートエンコーダを用いて達成されて、次元削減成分を生成し得る。

10

#### 【0035】

同様に、多様体学習は、(例えば、行列Mの形態で)次元削減成分を用いて行われて、複数の方法で行列Nなどの別のデータ形式を生成し得る。本開示のある実施形態において、多様体学習を行う工程は、行列Mにおけるデータの幾何学的配列(geometry)を推定して、行列Nを生成する工程を含み得る。このような実施形態において、多様体学習を行う工程は、局所線形埋め込み、局所線形アイソメトリックマッピング(ISOMAP)、t分布型確率的近傍埋め込み(t-SNE)、親和性に基づく軌道埋め込みのための熱拡散の可能性(Potential of Heat-Diffusion for Affinity Based Trajectory Embedding)(PHATE)、又は一様多様体近似及び投影(Uniform Manifold Approximation and Projection)(UMAP)を行う工程を含み得る。さらなる実施形態において、多様体学習を行う工程は、行列Mにおけるデータに基づいて力指向レイアウトを作成して、行列Nを生成する工程を含み得る。一実施形態において、力指向レイアウトは、Force Atlas 2アルゴリズムを用いて作成され得る。

20

#### 【0036】

クラスタリングはまた、複数の異なる方法を用いて行われ得る。特定の実施形態において、クラスタリングを行う工程は、各クラスター中の複数の点の組織化の予備知識を仮定しない。本明細書に開示されるさらなる実施形態において、クラスタリングを行う工程は、HDBSCAN及び/又はLouvainコミュニティ検出を行って、クラスターのセット $C_j$ を生成する工程を含む。さらなる実施形態において、クラスタリングを行う工程は、点に関連する単一細胞細胞成分の発現データセットが収集された時点に基づいて、各点をクラスター $C_j$ の1つに割り当てる工程を含む。ある実施形態において、クラスタリングを行う工程は、点がクラスターからどのくらい末端にあるかの測定に基づいて、点をクラスターに割り当てる拡散経路アルゴリズムを用いて複数の点を分析する工程を含む。

30

#### 【0037】

異なって発現された細胞成分 $E_k$ のセットを決定するために、複数の異なる方法が使用され得る。例えば、一実施形態において、各細胞成分について、クラスター $C_j$ の少なくとも1つについて、少なくとも1つのクラスター中の複数の点についての細胞成分の量は、少なくとも1つの他のクラスター中の複数の点についての細胞成分の量に対して比較され得る。次に、少なくとも1つのクラスター中の複数の点についての細胞成分の量が、少なくとも1つの他のクラスター中の複数の点についての細胞成分の量を超える閾値レベルであることに応答して、細胞成分は、異なって発現された細胞成分 $E_k$ のセットに加えられ得る。特定の実施形態において、少なくとも1つのクラスターは、クラスター $C_j$ の系統内(on-lineage)のものを含むことができ、系統内(on-lineage)クラスターは、所望の細胞型を有する複数の点を含む。さらなる実施形態において、少

40

50

なくとも1つの他のクラスターは、クラスター  $C_j$  の系統外 (off-line age) のものを含むことができ、系統外 (off-line age) クラスターは、所望されない細胞型を有する点を含む。

【0038】

さらなる実施形態において、異なって発現された細胞成分  $E_k$  のセットを決定するために、各細胞成分について、クラスターの少なくとも1つについて、少なくとも1つのクラスター中の複数の点についての細胞成分の量と、少なくとも1つの他のクラスター中の複数の点についての細胞成分の量との間の距離計量が、計算され得る。次に、距離計量が統計的に有意であることに応答して、細胞成分は、異なって発現された細胞成分  $E_k$  のセットに加えられ得る。

10

【0039】

特定の実施形態において、本明細書に記載される方法は、転写因子データベースに対して、異なって発現された細胞成分  $E_k$  のセットをスクリーニングして、異なって発現された転写因子のセットを同定する工程をさらに含み得る。異なって発現された転写因子のセットが同定されるこのような実施形態において、本方法は、異なって発現された細胞成分  $E_k$  のセットに対して経験的モード分解を行って、データセットの擬似時間表現を生成する工程と、擬似時間表現に基づいて、異なって発現された転写因子のセットを同定する工程とを含み得る。

【0040】

別の態様において、本開示は、複数の単一細胞細胞成分の発現データセットにアクセスする工程を含む方法を提供する。各データセットが、同じ「前駆」細胞型から移行された複数の細胞の1つの細胞から得られる。各データセットは、細胞成分のベクトル  $r_i$  を含む。細胞成分のベクトル  $r_i$  中の各エントリは、複数の細胞成分の1つに関連し、各エントリの値は、細胞についての細胞成分の量を表す。本方法は、 $kNN$  アルゴリズム及び単一細胞細胞成分の発現データセットを用いて  $kNN$  グラフを生成する工程と、クラスタリングを行って、クラスターのセット  $C_j$  を生成する工程と、クラスターのセット  $C_j$  を用いて、複数の細胞についての、異なって発現された細胞成分  $E_k$  のセットを決定する工程とをさらに含む。各クラスターは、複数の点を含み、各点は、複数の細胞の1つの細胞についての単一細胞細胞成分の発現データセットに対応する。ある実施形態において、異なって発現された細胞成分  $E_k$  のセットを決定する工程は、クラスター  $C_j$  中の複数の点間の距離計量を決定する工程を含む。

20

30

【0041】

別の態様において、本開示は、第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーにアクセスする工程を含む方法を提供する。本方法は、摂動に曝されていない非摂動細胞と摂動に曝された摂動細胞との間の細胞成分の発現の差の尺度を表す摂動シグネチャーにアクセスする工程をさらに含む。本方法は、単一細胞の移行シグネチャー及び摂動シグネチャーの比較に基づいて、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程をさらに含む。

【0042】

ある実施形態において、単一細胞の移行シグネチャーにアクセスする工程は、第1の複数の単一細胞細胞成分の発現データセット（各第1のデータセットが、第1の細胞状態における第1の複数の細胞の1つの細胞から得られる）、及び第2の複数の単一細胞細胞成分の発現データセット（各第2のデータセットが、変化した細胞状態における第2の複数の細胞の1つの細胞から得られる）に基づいて、単一細胞の移行シグネチャーを決定する工程を含む。第1及び第2の複数の単一細胞細胞成分の発現データセットの各データセットが、各細胞について、細胞成分のベクトル  $r_i$  を含むことができ、ベクトル中の各エントリが、複数の細胞成分の1つに関連し、各エントリの値が、細胞についての細胞成分の量を表す。ある実施形態において、第1の複数の単一細胞細胞成分の発現データセット及び第2の複数の単一細胞細胞成分の発現データセットに基づいて、単一細胞の移行シグネ

40

50

チャーを決定する工程は、母平均の差の検定、ウィルコクソンの順位和検定（マン・ホイットニーのU検定）、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて、第1及び第2の複数の単一細胞細胞成分の発現データセットの間の細胞成分量の差を決定する工程を含む。

【0043】

単一細胞の移行シグネチャーが、第1及び第2の複数の単一細胞細胞成分の発現データセットに基づいて、単一細胞の移行シグネチャーを決定する工程を含むこのような実施形態において、本方法は、第1及び第2の複数の単一細胞細胞成分の発現データセットを取得する工程をさらに含む。ここで、取得する工程は、第1及び第2の複数の単一細胞細胞成分の発現データセットの次元削減を行って、行列Mを生成する工程を含む。行列Mは、第1の次元における行及び第2の次元における列を含む。行列Mの各行は、複数の細胞の1つの細胞に対応する。行列Mの値は、第1及び第2の次元空間中の点に位置する細胞成分の量から生成される値を含む。第1及び第2の単一細胞細胞成分の発現データセットを取得する工程は、クラスタリングを行って、クラスターのセット $C_j$ を生成する工程をさらに含む。各クラスターは、行列M中の行のサブセットに対応する複数の点、及びそれらの対応する細胞を含む。第1及び第2の単一細胞細胞成分の発現データセットを取得する工程は、クラスターのセット $C_j$ の第1のクラスターから第1の複数の細胞を同定する工程と、クラスターのセット $C_j$ の第2のクラスターから第2の複数の細胞を同定する工程と、第1の複数の細胞から第1の複数の単一細胞細胞成分の発現データセットを取得する工程と、第2の複数の細胞から第2の複数の単一細胞細胞成分の発現データセットを取得する工程とをさらに含む。

10

20

【0044】

特定の実施形態において、第1及び第2の単一細胞細胞成分の発現データセットを取得する工程は、点の相対的類似度の近似による行列Mを用いて多様体学習を行って、行列Nを生成する工程をさらに含む。行列Nは、複数の行及び2つの列を含む。各行は、第1及び第2の複数の細胞の1つの細胞に対応し、列のそれぞれは、2次元空間中の1つ又は2つの次元に対応する。行列Nの値は、第1及び第2の複数の単一細胞細胞成分の発現データセットに基づいて、各細胞の他の各細胞に対する相対的な細胞状態を示す。

【0045】

特定の実施形態において、摂動シグネチャーにアクセスする工程は、摂動に曝されていない非摂動細胞の複数の非摂動単一細胞細胞成分の発現データセット、及び摂動に曝された摂動細胞の複数の摂動単一細胞細胞成分の発現データセットに基づいて、摂動シグネチャーを決定する工程を含む。非摂動細胞は、摂動細胞の摂動に曝されていない対照細胞であり得る。或いは、非摂動細胞は、摂動に曝された非関連摂動細胞の平均であり得る。ある実施形態において、複数の非摂動単一細胞細胞成分の発現データセット及び複数の摂動単一細胞細胞成分の発現データセットに基づいて、摂動シグネチャーを決定する工程は、母平均の差の検定、ウィルコクソンの順位和検定（マン・ホイットニーのU検定）、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて、複数の非摂動及び摂動単一細胞細胞成分の発現データセットの間の細胞成分量の差を決定する工程を含む。

30

40

【0046】

ある実施形態において、本方法は、転写因子である細胞成分を含む単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングする工程をさらに含む。さらなる実施形態において、本方法は、単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングして、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させる工程をさらに含む。具体的には、単一細胞の移行シグネチャー及び摂動シグネチャーは、閾値p値にしたがって又は細胞成分の閾値数にしたがって、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させるようにフィルタリングされ得る。

【0047】

50

本明細書に開示される方法のさらなる実施形態において、摂動シグネチャーは、複数の細胞成分を含むことができ、各細胞成分は、細胞成分の量の変化と、非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する重要度スコアに関連する。このような実施形態において、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程は、各細胞成分についての重要度スコアを、細胞成分についてのマッチングスコアで置き換える工程と、複数の細胞成分についてのマッチングスコアを組み合わせ、摂動についてのマッチングスコアを生成する工程と、摂動についてのマッチングスコアに基づいて、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とを含み得る。マッチングスコアは、別個のスコア又は連続したスコアを含み得る。重要度スコアを置き換える工程は、単一細胞の移行シグネチャーからの細胞成分量及びその細胞成分についての摂動シグネチャーからの細胞成分量が両方とも上方制御される場合、重要度スコアを第1のスコアで置き換える工程と、単一細胞の移行シグネチャーからの細胞成分量が上方制御され、その細胞成分についての摂動シグネチャーからの細胞成分量が下方制御される場合、重要度スコアを第2のスコアで置き換える工程と、その細胞成分についての摂動シグネチャーからの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアを第3のスコアで置き換える工程とを含み得る。或いは、重要度スコアを置き換える工程は、単一細胞の移行シグネチャーからの細胞成分量及びその細胞成分についての摂動シグネチャーからの細胞成分量が両方とも下方制御される場合、重要度スコアを第1のスコアで置き換える工程と、単一細胞の移行シグネチャーからの細胞成分量が下方制御され、その細胞成分についての摂動シグネチャーからの細胞成分量が上方制御される場合、重要度スコアを第2のスコアで置き換える工程と、その細胞成分についての摂動シグネチャーからの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアを第3のスコアで置き換える工程とを含み得る。

#### 【0048】

本明細書に開示される方法の代替的な実施形態において、マッチングスコアは、摂動シグネチャーの細胞成分に関連する重要度スコアを置き換えるのに使用されない。そうではなく、代替的な実施形態において、摂動シグネチャーは、複数の細胞成分を含むことができ、各細胞成分は、細胞成分の量の変化と、非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する重要度スコアに関連する。このような実施形態において、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程は、複数の細胞成分についての重要度スコアを単に組み合わせ、摂動についての重要度スコアを生成する工程と、摂動についての重要度スコアに基づいて、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とを含み得る。

#### 【0049】

ある実施形態において、摂動についてのマッチングスコアについての偽細胞成分発見率が、摂動への信頼度を決定するために推定される。このような実施形態において、偽細胞成分発見率は、複数の細胞成分の各細胞成分についての経験的な周辺発現頻度 (empirical marginal expression frequency) を計算し、複数の細胞成分についての経験的な周辺発現頻度を、それらの組合せにわたって合計して、独立に分布した発現を偶然呈するいくつかの細胞成分を同定する確率を生成し、確率に基づいて、摂動についてのマッチングスコアについての偽細胞成分発見率を推定することによって、推定される。

#### 【0050】

特定の実施形態において、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程は、摂動の共変量に依存する。例えば、ある実施形態において、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程は、摂動の共変量の閾値量が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連していることを決定する工程、及び決定に回答し

て、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連していることを決定する工程を含み得る。特定の実施形態において、摂動は、小分子への細胞の曝露を含み得る。摂動の共変量は、小分子の特定の用量、摂動細胞が小分子に曝される時間に対する、非摂動細胞と摂動細胞との間の細胞成分の発現の差が測定される時間、及び摂動細胞の細胞株を含み得る。

#### 【0051】

特定の実施形態において、細胞成分は、遺伝子を含み得る。単一細胞細胞成分の発現データセットは、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスポーズでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CyTOF/SCoP、E-MS/Absseq、miRNA-seq、CITE-seq、及びそれらの任意の組合せ又は集合からなる群から選択される方法を用いて生成され得る。

10

#### 【0052】

本明細書に開示される方法のある実施形態において、単一細胞の移行シグネチャー及び摂動シグネチャーの少なくとも1つが、データベースから得られる。摂動シグネチャーは、複数の摂動について複数の摂動シグネチャーを含むデータベースから得ることができる。このような実施形態において、データベース中の複数の摂動の各摂動について、摂動についての摂動シグネチャーは、データベースからアクセスされ、単一細胞の移行シグネチャー及び摂動シグネチャーの比較に基づいて、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうか決定される。

20

#### 【0053】

本明細書に開示される方法のさらなる実施形態において、本方法は、複数の摂動細胞から複数の摂動シグネチャーにアクセスする工程と、複数の摂動シグネチャーのそれぞれについて、単一細胞の移行シグネチャー及び摂動シグネチャーの比較に基づいて、摂動シグネチャーに関連する摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定することによって、変化した細胞状態を促進する摂動についてスクリーニングする工程とをさらに含み得る。摂動シグネチャーが、変化した細胞状態を促進する摂動についてスクリーニングされるこのような実施形態において、複数の摂動シグネチャーにアクセスする工程は、細胞を複数の摂動に曝して、複数の摂動細胞を生成する工程と、複数の摂動細胞から細胞分量を測定する工程とを含み得る。

30

#### 【0054】

本方法は、変化した細胞状態を促進する摂動を同定する工程をさらに含み得る。変化した細胞状態を促進する工程は、第1の細胞状態を含む細胞の集団中で、第1の細胞状態から変化した細胞状態への移行を促進する工程を含み得る。或いは、変化した細胞状態を促進する工程は、第1の細胞状態を含む細胞の集団中で、第1の状態、又は、任意に、変化した細胞状態以外の状態における細胞の数に対する、別の状態における細胞の数の比率を増加させる工程を含み得る。さらなる代替的な実施形態において、変化した細胞状態を促進する工程は、第1の細胞状態を含む細胞の集団中で、変化した細胞状態における細胞の絶対数を増加させる工程を含み得る。さらに他の代替的な実施形態において、変化した細胞状態を促進する工程は、第1の細胞状態を含む細胞の集団中で、第1の細胞状態又は、任意に、変化した細胞状態以外の状態における細胞の絶対数を減少させる工程を含み得る。

40

#### 【0055】

特定の実施形態において、細胞の移行シグネチャー及び摂動シグネチャーは、異なるタイプの細胞成分を用いて生成され得る。例えば、細胞の移行シグネチャーは、RNA発現(例えば、RNA転写物の数)に基づいて生成され得、摂動シグネチャーは、タンパク質発現(例えば、アミノ酸の数)に基づいて生成され得る。代替的な実施形態において、細胞の移行シグネチャー及び摂動シグネチャーは、同じタイプの細胞成分を用いて生成され得る。例えば、細胞の移行シグネチャー及び摂動シグネチャーは両方とも、RNA発現(例えば、RNA転写物の数)に基づいて生成され得る。

50

## 【0056】

別の態様において、本開示は、第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーにアクセスする工程を含む方法を提供する。本方法は、複数の摂動シグネチャーにアクセスする工程をさらに含み、各摂動シグネチャーが、摂動に曝されていない非摂動細胞と摂動に曝された摂動細胞との間の細胞成分の発現の差の尺度を表す摂動に関連する。本方法は、単一細胞の移行シグネチャー及び複数の摂動シグネチャーの比較に基づいて、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連する摂動のサブセットを決定する工程をさらに含む。

## 【0057】

特定の実施形態において、各摂動シグネチャーは、複数の細胞成分を含み、各細胞成分は、細胞成分の量の変化と、非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する重要度スコアに関連する。このような実施形態において、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連する摂動のサブセットを決定する工程は、各摂動シグネチャーについて、各細胞成分についての重要度スコアを、細胞成分についてのマッチングスコアで置き換える工程と、複数の細胞成分についてのマッチングスコアを組み合わせて、摂動についてのマッチングスコアを生成する工程とを含む。次に、本方法は、摂動を、それらのマッチングスコアにしたがってランク付けする工程と、摂動のランク付けされたリストに基づいて、摂動のサブセットを選択する工程とをさらに含む。

## 【0058】

別の態様において、本明細書に開示される本開示は、命令が符号化された非一時的なコンピュータ可読記録媒体を含むコンピュータプログラム製品を提供する。プロセッサによって実行されるとき、符号化された命令は、プロセッサに、本明細書に開示される方法の実施形態のいずれかを実行させる。さらに別の態様において、本明細書に開示される本発明は、命令が符号化された非一時的なコンピュータ可読記録媒体を含むシステムを提供する。プロセッサによって実行されるとき、符号化された命令は、プロセッサに、本明細書に開示される方法の実施形態のいずれかを実行させる。

## 【0059】

さらに別の態様において、本開示は、ニューロン及び/又は「前駆」細胞を促進するための方法を提供する。本方法は、線維芽細胞の開始集団を、線維芽細胞の開始集団の、「前駆」細胞及び/又はニューロンへの移行を促進する摂動シグネチャーを有する摂動に曝す工程を含む。このような実施形態において、摂動シグネチャーは、Brn2、Ascl1、Myt1、Zfp941、Taf5B、St18、Zkscan16、Camta1、及びArnt2の1つ以上の増大した活性、及び/又はAscl1、Atf3、Rorc、Scx、Satb1、Elf3、及びFosの1つ以上の低下した活性である。

## 【0060】

ニューロン及び/又は「前駆」細胞を促進するための方法の特定の実施形態において、ニューロン及び/又は「前駆」細胞は、ニューロン及び/又は「前駆」細胞の絶対数を増加させること、線維芽細胞の絶対数を減少させること、線維芽細胞の、ニューロン及び/又は「前駆」細胞への移行を促進すること、ニューロン又は「前駆」細胞の寿命を促進すること、線維芽細胞の寿命を減少させること、又は線維芽細胞に対するニューロン及び/又は「前駆」細胞の比率を増加させることのうちの1つ以上によって促進される。さらなる実施形態において、摂動は、フォルスコリン、PP1、PP2、及びトリコスタチンAを含まない。

## 【0061】

さらに別の態様において、本開示は、ニューロン及び/又は「前駆」細胞の量を増加させる方法を提供する。本方法は、線維芽細胞の集団を、線維芽細胞の集団の、ニューロンへの移行を促進する摂動シグネチャーを有する医薬組成物に曝す工程を含む。医薬組成物は、フォルスコリン、PP1、PP2、トリコスタチンA、BRD-K38615104、ゲルダナマイシン、マヌマイシンA、ミトキサントロン、クルクミン、アルボジブ、ポリノスタット、KI20227、又は上記の組合せ、例えば、上記の2、3、4、5つ

10

20

30

40

50



、若しくはそれ以上の組合せを含む。ある実施形態において、医薬組成物は、フォルスコリン、PP1、PP2、及びトリコスタチンAを含まない。

【0062】

さらに別の態様において、本開示は、ニューロン及び/又は「前駆」細胞を促進するための医薬組成物を提供する。医薬組成物は、フォルスコリン、PP1、PP2、トリコスタチンA、BRD-K38615104、ゲルダナマイシン、マヌマイシンA、ミトキサントロン、クルクミン、アルボシジブ、ポリノスタット、KI20227、又は上記の組合せからなる群から選択される摂動、及び薬学的に許容される賦形剤を含む。ある実施形態において、摂動は、フォルスコリン、PP1、PP2、及びトリコスタチンAを含まない。

10

【0063】

さらに別の態様において、本開示は、本明細書に開示される医薬組成物の1つを含む単位剤形を提供する。

【0064】

さらに別の態様において、本開示は、線維芽細胞の開始集団の、ニューロン及び/又は「前駆」細胞への移行を促進するための候補の摂動を同定する方法を提供する。本方法は、線維芽細胞の開始集団を摂動に曝す工程と、摂動についての摂動シグネチャーを同定する工程とを含む。摂動についての摂動シグネチャーは、1つ以上の細胞成分及び各細胞成分に関連する重要度スコアを含む。各細胞成分の重要度スコアは、線維芽細胞の集団の、摂動への曝露の後、細胞成分の発現の変化と、線維芽細胞の集団の、ニューロン及び/又は「前駆」細胞への細胞状態の変化との間の関連性を定量する。摂動シグネチャーは、Brn2、Ascl1、Myt1、Zfp941、Taf5B、St18、Zkscan16、Camta1、及びArnt2の1つ以上の増大した活性、及び/又はAscl1、Atf3、Rorc、Scx、Satb1、Elf3、及びFosの1つ以上の低下した活性を含む。本方法は、摂動シグネチャーに基づいて、線維芽細胞の集団の、ニューロン及び/又は「前駆」細胞への移行を促進するための候補の摂動として摂動を同定する工程をさらに含む。

20

【0065】

本開示のさらに別の態様は、非一時的なコンピュータ可読記録媒体を提供し、ここで、非一時的なコンピュータ可読記録媒体は、命令を記憶し、命令は、コンピュータシステムによって実行されるとき、コンピュータシステムに、本開示に記載される細胞を分析するための方法のいずれかを実行させる。

30

【0066】

本明細書に開示される実施形態は、添付の図面において、限定としてではなく、例として示されている。類似の参照番号は、図面を通して対応する部分を指す。

【図面の簡単な説明】

【0067】

【図1】本開示の一実施形態に係る、例示的なシステム及び計算装置のブロック図を示す。

【図2】本開示の様々な実施形態に係る、細胞を分析するためのプロセスのフローチャート及びシステムの特徴を提供し、ここで、破線のボックス中の要素は任意選択である。

40

【図3】本開示の一実施形態に係る、異なって発現された細胞成分のセットを決定するための細胞成分の発現の差のアッセイの第1の例についてのフローチャートである。

【図4A】本開示の一実施形態に係る、ある期間にわたって誘導された細胞状態の移行の軌道を辿るタイムラインを示す。

【図4B】本開示の一実施形態に係る、補足的な表1中例の行列Nについての力指向レイアウトアルゴリズムによって生成された多様体を示す。

【図5A】本開示の一実施形態に係る、図5Bの多様体を示す。

【図5B】本開示の一実施形態に係る、図4Bの多様体中の点として示される測定日のそれぞれにおける細胞のそれぞれの中のBAM転写因子のそれぞれの発現のレベルを示す。

50

【図6】本開示の一実施形態に係る、D A P I、M a p 2 抗体、及びT u j 1 抗体で染色された、A s c l 1 転写因子の発現が強制されたM E F 細胞の画像、D A P I、M a p 2 抗体、及びT u j 1 抗体で染色されたマウスニューロンの画像、並びにD A P I、M a p 2 抗体、及びT u j 1 抗体で染色された、A s c l 1 転写因子の発現が強制されていないM E F 細胞の画像を示す。

【図7A】本開示の一実施形態に係る、図4Bの多様体を、クラスタリングによって同定されるクラスターC<sub>j</sub>にグループ分けされた多様体中の点とともに示す。

【図7B】本開示の一実施形態に係る、M E F の、マウス筋細胞への移行ではなく、マウスニューロンへの移行に関連することが文献において知られている転写因子及び知られていない転写因子の両方を示す。

【図8A】本開示の一実施形態に係る、図4Aに関して説明されるM E F 細胞についての移行軌道のマップを示す。

【図8B】本開示の一実施形態に係る、細胞が図8Aの移行軌道のマップにおける第1の状態から第2の状態へと移行するように細胞中の遺伝子発現を変化させることによって細胞の移行軌道に影響を与える摂動を同定するための方法を示す。

【図9】本開示の一実施形態に係る、M E F の、マウス筋細胞への移行ではなく、マウスニューロンへの移行に関連する小分子摂動を示す。

【図10A】本開示の一実施形態に係る、各処理条件についてのニューロンの総数を示す棒グラフを提供し、ここで、ニューロンの総数が、陽性T u j 1 / M a p 2 シグナル及び神経形態に基づいて手動で計数され、各実験について、各処理条件のデータが、D M S O 処理されたウェル中のニューロンの数によって正規化された。

【図10B】本開示の一実施形態に係る、各処理条件についてのニューロンパーセントを示す棒グラフを提供する。

【発明を実施するための形態】

【0068】

ここで、実施形態への言及が詳細になされ、その例が、添付の図面に例示される。以下の詳細な説明において、多くの具体的な詳細が、本開示の十分な理解を提供するために記載される。しかしながら、本開示が、これらの具体的な詳細なしで実施され得ることが、当業者に明らかであろう。他の例では、周知の方法、手順、構成要素、回路、及びネットワークは、実施形態の態様を不必要に分かりにくくしないように、詳細に記載されていない。

【0069】

複数の例が、単一の例として本明細書に記載される構成要素、操作又は構造のために提供され得る。最後に、様々な構成要素、操作、及びデータストアの間の境界は、いくらか任意であり、特定の操作が、特定の例示的な形態に関連して示される。他の形態の機能が想定され、実施の範囲内に含まれ得る。一般に、例の形態における別個の構成要素として示される構造及び機能は、組み合わせられた構造又は構成要素として実現され得る。同様に、単一の構成要素として示される構造及び機能は、別個の構成要素として実現され得る。これらの及び他の変形、変更、追加、及び改良は、本実施の範囲内に含まれる。

【0070】

「第1」、「第2」などの用語は、本明細書において様々な要素を表すのに使用され得るが、これらの要素は、これらの用語によって限定されるべきではないことも理解されるであろう。これらの用語は、ある要素を別の要素と区別するために使用されるに過ぎない。例えば、第1のデータセットは、第2のデータセットと呼ばれ得、同様に、第2のデータセットは、本発明の範囲から逸脱せずに、第1のデータセットと呼ばれ得る。第1のデータセット及び第2のデータセットは両方ともデータセットであるが、それらは、同じデータセットではない。

【0071】

本明細書において使用される専門用語は、特定の実施を説明するためのものであるに過ぎず、特許請求の範囲を限定することは意図されていない。実施の説明及び添付の特許請

10

20

30

40

50

求の範囲において使用される際、単数形「a」、「an」、及び「the」は、文脈上明らかに別の意味に解釈すべき場合を除き、複数形も含むことが意図される。本明細書において使用される際の「及び/又は」という用語が、関連する列挙された項目の1つ以上のあらゆる可能な組合せを指し、包含することも理解されるであろう。「含む (comprises)」及び/又は「含む (comprising)」という用語は、本明細書において使用されるとき、記載される特徴、整数、工程、操作、要素、及び/又は構成要素の存在を特定するが、1つ以上の他の特徴、整数、工程、操作、要素、構成要素、及び/又はそれらの群の存在又は追加を除外しないことがさらに理解されるであろう。

【0072】

本明細書において使用される際、「場合」という用語は、文脈に応じて、記載される前提条件が真である「とき」又は真である「と」又は真であるという「決定に応じて」又は「決定にしたがって」又は「検出に応じて」を意味するものと解釈され得る。同様に、「(記載される前提条件が真であると)決定される場合」又は「(記載される前提条件が真である)場合」又は「(記載される前提条件が真である)とき」という語句は、文脈に応じて、記載される前提条件が真であると「決定すると」又は真であるという「決定に応じて」又は「決定にしたがって」又は真であると「検出すると」又は真であるという「検出に応じて」を意味するものと解釈され得る。

10

【0073】

さらに、参照番号が「第i (i<sup>th</sup>)」表記で示されるとき、参照番号は、包括的な構成要素、セット、又は実施形態を指す。例えば、「細胞成分i」と称される細胞成分は、複数の細胞成分中の第iの(i<sup>th</sup>)細胞成分を指す。

20

【0074】

上記の説明は、例示的な実施を実現するシステム、方法、技術、命令シーケンス、及び計算機械プログラム製品の例を含む。説明のために、多くの具体的な詳細が、本発明の主題の様々な実施の理解を提供するために記載される。しかしながら、本発明の主題の実施が、これらの具体的な詳細なしで行われ得ることが当業者には明らかであろう。一般に、周知の命令の例、プロトコル、構造及び技術は、詳細に示されていない。

【0075】

上記の説明は、説明のために、特定の実施に関して記載されている。しかしながら、以下の例示的な説明は、包括的であること又は開示される正確な形態に実施を限定することは意図されていない。多くの変更及び変形が、上記の教示を考慮して可能である。実施は、原理及びそれらの実際の適用を最良に説明し、それによって他の当業者が、実施及び様々な変更を伴う様々な実施を、考えられる特定の使用に合うように最良に利用することができるように選択及び記載される。

30

【0076】

明確にするために、本明細書に記載される実施の所定の特徴の全てが、示され、記載されているわけではない。いずれかのこのような実際の実施の発生の際、多くの実施に固有の決定が、使用事例及び事業に関連する制約の順守などの、設計者の具体的な目的を達成するためになされること、及びこれらの具体的な目的が、ある実施と別の実施とで及びある設計者と別の設計者とで異なることが理解されるであろう。さらに、このような設計努力は複雑で時間のかかるものであり得るが、それにもかかわらず本開示の利益を有する当業者の日常的な業務であることが理解されるであろう。

40

【0077】

本明細書の一部は、情報に関する操作のアルゴリズム及び記号表現に関して、本発明の実施形態を説明している。これらのアルゴリズム的記述及び表現は、研究の内容を他の当業者に効果的に伝えるために、データ処理の分野の当業者によって一般的に使用される。これらの操作は、機能的に、計算的に、又は論理的に記述されるが、コンピュータプログラム又は同等の電気回路、マイクロコードなどによって実施されることが理解される。

【0078】

本明細書において使用される文言は、主に、読みやすさ及び説明のために選択されてい

50

るが、本発明の主題の境界を明示するか又はそれを制限するように選択されていないことがある。したがって、本発明の範囲は、この詳細な説明によって限定されず、それに基づいて本出願を発行する特許請求の範囲によって限定されることが意図される。したがって、本発明の実施形態の開示内容は、例示的なものであることが意図され、本発明の範囲を限定するものではない。

【0079】

一般に、特許請求の範囲及び本明細書において使用される用語は、当業者によって理解される通常の意味を有するものと解釈されることが意図される。いくつかの用語が、さらなる明瞭性を提供するために以下に定義される。通常の意味と提供される定義との間に矛盾がある場合、提供される定義が使用されるものとする。

10

【0080】

本明細書に直接定義されていない用語は、本発明の技術分野内で理解されるような、それらに一般的に関連する意味を有することが理解されるべきである。いくつかの用語が、本発明の態様の組成物、デバイス、方法など、及びそれらの作製又は使用方法を説明する際に実行者にさらなる指針を提供するために本明細書に記載されている。同じことが、2つ以上の方法でも言えることが理解されるであろう。結果として、代替的な文言及び同義語が、本明細書に記載される用語のいずれか1つ以上に使用され得る。用語が本明細書において詳述又は説明されているか否かについては重要視されるべきではない。いくつかの同義語又は代替可能な方法、材料などが提供される。明記されない限り、1つ又はいくつかの同義語又は同等物の記載は、他の同義語又は同等物の使用を除外しない。用語の例を含む、例の使用は、例示のためのものに過ぎず、本明細書における本発明の態様の範囲及び意味を限定するものではない。

20

【0081】

本明細書において使用される際、細胞（例えば、細胞の摂動又は細胞摂動）に関して「摂動」という用語は、1つ以上の化合物による細胞の任意の処理を指す。これらの化合物は、「摂動因子（*perturbagen*）」と呼ばれ得る。ある実施形態において、摂動因子は、例えば、小分子、生物製剤、タンパク質、小分子と組み合わされたタンパク質、ADC、核酸、例えば*siRNA*若しくは干渉RNA、*cDNA*過剰発現野生型及び/又は突然変異型*shRNA*、*cDNA*過剰発現野生型及び/又は突然変異型ガイドRNA（例えば、*Cas9*システム若しくは他の遺伝子編集システム）、又は上記のいずれかの任意の組合せを含み得る。

30

【0082】

本明細書において使用される際、細胞（例えば、前駆細胞）に関して「前駆」という用語は、1つの細胞状態から少なくとも1つの他の細胞状態へと移行することが可能な任意の細胞を指す。

【0083】

本明細書において使用される際、細胞又は複数の細胞についての細胞成分の発現測定に関して「データセット」という用語は、ある文脈において、単一細胞から収集されるデータの高次元セット（例えば、単一細胞細胞成分の発現データセット）を指し得る。他の文脈において、「データセット」という用語は、単一細胞から収集されるデータの複数の高次元セット（例えば、複数の単一細胞細胞成分の発現データセット）を指し得、複数のデータの各セットは、複数の細胞の1つ細胞から収集される。

40

【0084】

本明細書において使用される際、「影響を与える」という用語は、細胞の移行の変化を指す。

【0085】

I. 例示的なシステムの実施形態

本開示のいくつかの態様の概説及び本開示において使用されるいくつかの定義が提供されてきたが、例示的なシステムの詳細が、図1とともに記載される。

【0086】

50

図1は、本開示のある実施形態に係るシステム100を示すブロック図を提供する。システム100は、摂動が細胞の移行に影響を与えるかどうかの予測を提供する。図1において、システム100は、計算装置として示される。当然ながら、コンピュータシステム100の他のトポロジーが可能である。例えば、ある実施形態において、システム100は、実際に、ネットワーク中で一緒にリンクされたいくつかのコンピュータシステムを構成し得るか、又はクラウドコンピューティング環境における仮想マシン若しくは容器であり得る。したがって、図1に示される例示的なトポロジーは、当業者に容易に理解されるように本開示の一実施形態の特徴を説明する役割を果たすに過ぎない。

#### 【0087】

図1を参照すると、ある実施形態において、コンピュータシステム100（例えば、計算装置）は、ネットワークインターフェース104を含む。ある実施形態において、ネットワークインターフェース104は、システム内のシステム100計算装置を、互いに、並びに任意選択の外部のシステム及びデバイスと、1つ以上の通信ネットワークを介して（例えば、ネットワーク通信モジュール118を介して）相互接続する。ある実施形態において、ネットワークインターフェース104は、任意に、インターネット、1つ以上のローカルエリアネットワーク（LAN）、1つ以上の広域ネットワーク（WAN）、他のタイプのネットワーク、又はこのようなネットワークの組合せを介して、ネットワーク通信モジュール118を介して通信を提供する。

#### 【0088】

ネットワークの例としては、ワールドワイドウェブ（WWW）、イントラネット及び/又は無線ネットワーク、例えば携帯電話ネットワーク、無線ローカルエリアネットワーク（LAN）及び/又はメトロポリタンエリアネットワーク（MAN）、及び無線通信による他のデバイスが挙げられる。無線通信は、任意に、移動通信用のグローバルシステム（GSM）、エンハンスドデータGSM環境（EDGE）、高速ダウンリンクパケットアクセス（HSDPA）、高速アップリンクパケットアクセス（HSUPA）、エボリューション、データオンリー（EV-DO）、HSPA、HSPA+、デュアルセルHSPA（DC-HSPDA）、ロングタームエボリューション（LTE）、近距離無線通信（NFC）、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、Wireless Fidelity（Wi-Fi）（例えば、IEEE 802.11a、IEEE 802.11ac、IEEE 802.11ax、IEEE 802.11b、IEEE 802.11g及び/又はIEEE 802.11n）、ボイスオーバーインターネットプロトコル（VoIP）、Wi-MAX、電子メール用のプロトコル（例えば、インターネットメッセージアクセスプロトコル（IMAP）及び/又はポストオフィスプロトコル（POP））、インスタントメッセージング（例えば、拡張メッセージング及びプレゼンスプロトコル（XMPP）、インスタントメッセージング及びプレゼンスレベライジングエクステンション用のセッションインイニシエーションプロトコル（SIMPLE）、インスタントメッセージング及びプレゼンスサービス（IMPS））、及び/又はショートメッセージサービス（SMS）を含む複数の通信規格、プロトコル及び技術のいずれか、又は本明細書の出願日の時点でまだ開発されていない通信プロトコルを含む任意の他の好適な通信プロトコルを使用する。

#### 【0089】

システム100は、ある実施形態において、1つ以上の処理装置（CPU）102（例えば、プロセッサ、処理コアなど）、1つ以上のネットワークインターフェース104、ユーザによる使用のための（任意に）ディスプレイ108及び入力システム110（例えば、入出力インターフェース、キーボード、マウスなど）を含むユーザインターフェース107、メモリ（例えば、非永続的メモリ111、永続的メモリ112）、及び上記の構成要素を相互接続するための1つ以上の通信バス114を含む。1つ以上の通信バス114は、任意に、システム構成要素間の通信を相互接続及び制御する回路（チップセットと呼ばれることもある）を含む。非永続的メモリ111が、典型的に、DRAM、SRA

10

20

30

40

50

M、DDR RAM、ROM、EEPROM、フラッシュメモリなどの高速ランダムアクセスメモリを含む一方、永続的メモリ112は、典型的に、CD-ROM、デジタル多用途ディスク(DVD)又は他の光学記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置又は他の磁気記憶デバイス、磁気ディスク記憶デバイス、光ディスク記憶デバイス、フラッシュメモリデバイス、又は他の不揮発性ソリッドステート記憶デバイスを含む。永続的メモリ112は、任意に、CPU102から遠隔に配置された1つ以上の記憶デバイスを含む。永続的メモリ112、及び非永続的メモリ112内の不揮発性メモリデバイスは、非一時的なコンピュータ可読記録媒体を含む。ある実施形態において、非永続的メモリ111或いは非一時的なコンピュータ可読記録媒体は、以下のプログラム、モジュール及びデータ構造、又はそのサブセットを、場合により、永続的メモリ112とともに保存する：

- ・様々な基本システムサービスを処理するため及びハードウェアに依存するタスクを実行するための手順を含む、オペレーティングシステム116(例えば、ANDROID、iOS、DARWIN、RTXC、LINUX、UNIX、OS X、WINDOWS、又はVxWorkなどの組み込みオペレーティングシステム)；

- ・システム100を、他のデバイス及び/又は通信ネットワーク104と接続するための任意選択のネットワーク通信モジュール(又は命令)118；

- ・複数のデータセット122を保存するデータセットストア120(各データセットは、1つ以上の識別子(例えば、サンプル識別子124及び/又は細胞/データセット識別子126)、関連する期間128、及び1つ以上の細胞成分132を含む細胞成分ベクトル130を含む)；並びに

- ・1つ以上の単一細胞の移行シグネチャー142及び1つ以上の摂動シグネチャー150を保存するシグネチャーストア140。

#### 【0090】

上述されるように、データセットストア120は、複数のデータセット120を含む。各データセットが、細胞の集団(例えば、それぞれのサンプル)の単一細胞の測定(例えば、図3の単一細胞の測定310)から得られる(例えば、収集される、通信されるなど)。各データセット122に関連するサンプル識別子(ID)124は、どのサンプルに細胞のデータセットが由来するかを示す。細胞/データセット識別子126は、どの細胞及び/又は複数のデータセット(例えば、データセットのサブセット)にデータセット122が関連するか及び/又は細胞の状態を示す。ある実施形態において、期間128は、データセット122の捕捉の期間(例えば、細胞が最初に培養される第1の期間 $t_0$ 、細胞発現の測定が行われる第2の期間 $t_1$ などの、細胞の増殖中の時間)に関連する。

#### 【0091】

さらに、ある実施形態において、各データセット120は、1つ以上の細胞成分132を含む細胞成分ベクトル130を含む。ある実施形態において、1つ以上の細胞成分132は、細胞の全ての細胞成分又は細胞のこれらの細胞成分のサブセットを含む。各細胞成分132は、測定(例えば、図3の単一細胞の測定310)に関連するデータの次元を表す。一般に、データセット122は、多量のデータを含む高い(例えば、3を超える、5を超える、10を超える、100を超えるなどの)次元を含む。さらに、ある実施形態において、各データセット122は、「前駆」細胞型から(例えば、第1の状態から変化した状態へと)移行された複数の細胞中の細胞から(例えば、サンプルから)得られる。

#### 【0092】

ある実施形態において、システムは、1つ以上の単一細胞の移行シグネチャー142及び1つ以上の摂動シグネチャー150を保存するシグネチャーストア140を含む。ある実施形態において、1つ以上の単一細胞の移行シグネチャー142は、1つ以上の所定のシグネチャー(例えば、トレーニングシグネチャー)を含む。ある実施形態において、1つ以上の単一細胞の移行シグネチャー142は、システム100によって決定され、及び/又は将来の使用のためにシステム内に保存される単一細胞の移行シグネチャーを含む。各単一細胞の移行シグネチャー142は、複数の細胞成分(例えば、図1の細胞成分13

10

20

30

40

50

2 - 1 - 1 ~ 1 3 2 - 1 - D) をさらに含む細胞成分の同定 1 4 4 を含む。さらに、単一細胞の移行シグネチャー 1 4 2 に関連する各細胞成分 1 3 2 は、対応する重要度スコア 1 3 4 を含む。ある実施形態において、次元削減が、データセット 1 2 2 に対して行われ (例えば、図 3 の次元削減 3 2 0)、これは、複数の次元削減成分 1 4 8 (例えば、図 1 の次元削減成分 1 4 8 - 1 - 1 から次元削減成分 1 4 8 - 1 - F) を生成する (例えば、図 1 の次元削減成分ストア 1 4 6 - 1 内に保存する及び / 又は図 3 の行列 M を生成する)。したがって、システム 1 0 0 は、ある実施形態において、次元削減 (例えば、図 3 の次元削減 3 2 0) を行って、複数の次元削減成分 1 4 8 を生成して (例えば、図 3 の行列 M を生成する)、データセット 1 2 2 の細胞成分 1 3 2 中に存在する潜在的パターンを保存する。ある実施形態において、この次元削減 (例えば、図 1 の次元削減成分削減成分 1 4 8 - 1 - 1 ~ 1 4 8 - 1 - F) の出力は、行列 (例えば、以下に示される行列 M) であり、これは、データセットの基本的な潜在構造も維持しながらデータセット 1 2 2 を圧縮形態で符号化する。

10

#### 【0093】

ある実施形態において、シグネチャー移行ストアは、多様体 1 4 9 を含む。ある実施形態において、この多様体 1 4 9 は、単一細胞の移行シグネチャー 1 4 2 の対応する次元削減成分 1 4 8 に関連する。この多様体 1 4 9 は、多様体に関連するデータセット 1 2 2 (例えば、単一細胞の移行シグネチャー 1 4 2 に関連するデータセット 1 2 2) の細胞成分ベクトル 1 3 0 を用いて多様体学習を行うことによって同定される。

20

#### 【0094】

シグネチャーストア 1 4 0 は、対応する摂動に関連する 1 つ以上の摂動シグネチャー 1 5 0 をさらに含む。各摂動シグネチャーは、複数の細胞成分 (例えば、図 1 の細胞成分 1 3 2 - 1 - 1 ~ 1 3 2 - 1 - H) を含む細胞成分の同定 1 5 2 を含む。ある実施形態において、細胞成分の同定 1 5 2 の細胞成分は、対応する単一細胞の移行シグネチャー 1 4 4 に関連する細胞成分の一部又は全てを含む (例えば、摂動シグネチャー 1 5 0 - 1 の細胞成分の同定 1 5 2 は、図 1 の単一細胞の移行シグネチャー 1 4 2 - 1 の細胞成分の同定 1 4 4 のサブセットを含む)。さらに、摂動シグネチャー 1 5 0 の各細胞成分は、対応する重要度スコア 1 3 4 を含む。

#### 【0095】

様々な実施形態において、上で同定される要素の 1 つ以上が、上述されたメモリデバイスの 1 つ以上に保存され、上述される機能を実行するための一連の命令に対応する。上で同定されるモジュール、データ、又はプログラム (例えば、一連の命令) は、別個のソフトウェアプログラム、手順、データセット、又はモジュールとして実装される必要はなく、したがって、これらのモジュール及びデータの様々なサブセットは、様々な実施において組み合わせられても又は他の形で再構成されてもよい。ある実施において、非永続的メモリ 1 1 1 は、任意に、上で同定されるモジュール及びデータ構造のサブセットを保存する。さらに、ある実施形態において、メモリは、上述されていないさらなるモジュール及びデータ構造を保存する。ある実施形態において、上で同定される要素の 1 つ以上は、システム 1 0 0 が必要に応じてこのようなデータの全て又は一部を読み出し得るようにシステム 1 0 0 によってアドレス可能な、システム 1 0 0 のもの以外のコンピュータシステムに保存される。

30

40

#### 【0096】

図 1 は、「システム 1 0 0」を示すが、この図は、本明細書に記載される実施の構造概略図としてより、コンピュータシステム中に存在し得る様々な特徴の機能記述として意図される。実際には、当業者によって認識されるように、別々に示される項目は、組み合わせられてもよく、いくつかの項目は、分離され得る。さらに、図 1 は、非永続的メモリ 1 1 1 中の特定のデータ及びモジュールを示すが、これらのデータ及びモジュールの一部又は全ては、その代わりに、永続的メモリ 1 1 2 又は 2 つ以上のメモリに保存され得る。例えば、ある実施形態において、少なくともデータセットストア 1 2 0 が、クラウドベースのインフラストラクチャの一部であり得る遠隔記憶デバイスに保存される。ある実施形態に

50

において、少なくともデータセットストア120が、クラウドベースのインフラストラクチャに保存される。ある実施形態において、データセットストア120及びシグネチャストア140はまた、遠隔記憶デバイスに保存され得る。

【0097】

本開示に係るシステムは、図1を参照して開示されているが、本開示に係る方法200が、ここで、図2を参照して詳述される。

【0098】

ブロック202. 図2のブロック202を参照すると、本方法は、単一細胞の移行シグネチャー（例えば、図1の単一細胞の移行シグネチャー142-1）に（例えば、電子的形態で）アクセスする工程を含む。単一細胞の移行シグネチャー142は、第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す。変化した細胞状態は、第1の細胞状態から変化した細胞状態への細胞の移行を経て生じる。単一細胞の移行シグネチャー142は、複数の細胞成分の同定（例えば、図1の細胞成分の同定144-1）を含む。複数の細胞成分中の各それぞれの細胞成分について（例えば、図1の細胞成分132-1-1~細胞成分132-1-D）、対応する第1の重要度スコア（例えば、重要度スコア134-1-1）は、それぞれの細胞成分の発現の変化と、第1の細胞状態と変化した細胞状態との間の細胞状態の変化との間の関連性を定量する。

10

【0099】

ある実施形態において、単一細胞の移行シグネチャーにアクセスする工程は、単一細胞の移行シグネチャー142を決定する工程を含む。この決定する工程は、第1の複数の第1の単一細胞細胞成分の発現データセット（例えば、データセット122-1、データセット122-2、及びデータセット122-3）、並びに第2の複数の第2の単一細胞細胞成分の発現データセット（例えば、データセット122-4、データセット122-5、及びデータセット122-6）に基づく。第1の複数の第1の単一細胞細胞成分の発現データセット中の各それぞれの第1の単一細胞細胞成分の発現データセット122が、第1の細胞状態の第1の複数の細胞の対応する単一細胞から得られる（例えば、図3の単一細胞の測定310）。さらに、第2の複数の第2の単一細胞細胞成分の発現データセット中の各それぞれの第2の単一細胞細胞成分の発現データセットが、変化した細胞状態の第2の複数の細胞の対応する単一細胞から得られる（例えば、図3の単一細胞の測定310）。

20

30

【0100】

ある実施形態において、単一細胞の移行シグネチャーを決定する工程は、複数の細胞成分132にわたる細胞分量の差を決定する工程を含む。この差は、第1の複数の第1の単一細胞細胞成分の発現データセットと第2の複数の第2の単一細胞細胞成分の発現データセットとの間の差である。ある実施形態において、この差は、母平均の差の検定、ウィルコクソンの順位和検定、t検定、ロジスティック回帰、又は一般化線形モデルのうちの1つを用いて決定される。

【0101】

ある実施形態において、第1の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセット122が、第1の複数の細胞成分ベクトル中の、対応する細胞成分ベクトル（例えば、図1のデータセット122-1の細胞成分ベクトル130-1）を含む。さらに、第2の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第2の複数の細胞成分ベクトル中の、対応する細胞成分ベクトル（例えば、データセット122-2の細胞成分ベクトル130-2）を含む。第1及び第2の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルが、複数の要素を含む。それぞれの細胞成分ベクトル130中の各それぞれの要素が、複数の細胞成分中の対応する細胞成分132に関連しており、第1及び第2の複数の単一細胞細胞成分の発現データセットのそれぞれのデータセットによって表される対応する単一細胞についての対応する細胞成分の量を表す対応する値（例えば、表2の細胞成分及び値）を含む。

40

【0102】

50



さらに、ある実施形態において、細胞成分 132 は、複数の遺伝子を含む。さらに、ある実施形態において、1つ以上のデータセット 122 が、単一細胞リボ核酸 (RNA) シーケンシング (scRNA-seq)、scTag-seq、シーケンシングを用いたトランスポーズでアクセス可能なクロマチンの単一細胞アッセイ (scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seq、及びそれらの任意の組合せ (例えば、表 1 の方法) を含む方法を用いて生成される。

#### 【0103】

ブロック 204 . ブロック 204 を参照すると、本方法は、摂動シグネチャー (例えば、図 1 の摂動シグネチャー 150 - 1) に (例えば、電子的形態で) アクセスする工程をさらに含む。摂動シグネチャー 150 は、複数の非摂動細胞と摂動に曝された複数の摂動細胞との間の細胞成分の発現の差の尺度を表す。摂動シグネチャー 150 は、複数の細胞成分の全て又は一部の同定 (例えば、図 1 の細胞成分の同定 152 - 1) を含む。複数の細胞成分の全て又は一部の中の各それぞれの細胞成分について (例えば、図 1 の細胞成分 132 - 3 - 1 ~ 細胞成分 132 - 3 - D)、複数の非摂動細胞と複数の摂動細胞との間のそれぞれの細胞成分の発現の変化と、複数の非摂動細胞と複数の摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第 2 の重要度スコア (例えば、図 1 の重要度スコア 134)。

#### 【0104】

ある実施形態において、方法 200 は、第 1 及び / 又は第 2 の複数の単一細胞細胞成分の発現データセット 122 に対して次元削減 (例えば、図 3 の次元削減 320) を行う工程を含む。この次元削減は、複数の次元削減成分 (例えば、図 1 の次元削減成分 148) を生成する。ある実施形態において、次元削減は、主成分アルゴリズム、ランダム射影アルゴリズム、独立成分分析アルゴリズム、又は特徴選択方法、因子分析アルゴリズム、サモンマッピング、曲線成分分析、確率的近傍埋め込み法 (SNE) アルゴリズム、IsoMap アルゴリズム、最大分散アンフォールディングアルゴリズム、局所線形埋め込みアルゴリズム、t-SNE アルゴリズム、非負値行列因子分解アルゴリズム、カーネル主成分分析アルゴリズム、グラフベースのカーネル主成分分析アルゴリズム、線形判別分析アルゴリズム、一般化判別分析アルゴリズム、一様多様体近似及び投影 (UMAP) アルゴリズム、LargeVis アルゴリズム、ラプラス固有写像 (Laplacian Eigenmap) アルゴリズム、又はフィッシャーの線形判別分析アルゴリズムである。例えば、Fodor, 2002, "A survey of dimension reduction techniques", Center for Applied Scientific Computing, Lawrence Livermore National, Technical Report UCL - ID - 148494; Cunningham, 2007, "Dimension Reduction", University College Dublin, Technical Report UCD - CSI - 2007 - 7, Zahorian et al., 2011, "Nonlinear Dimensionality Reduction Methods for Use with Automatic Speech Recognition", Speech Technologies. doi: 10.5772/16863. ISBN 978 - 953 - 307 - 996 - 7; 及び Lakshmi et al., 2016, "2016 IEEE 6th International Conference on Advanced Computing (IACC)", pp. 31 - 34. doi: 10.1109/IACC.2016.16, ISBN 978 - 1 - 4673 - 8286 - 1 (これらはそれぞれ、参照により本明細書に援用される) を参照されたい。したがって、ある実施形態において、次元削減は、主成分分析 (PCA) アルゴリズムであり、各それぞれの抽出された次元削減成分は、PCA によって得られるそれぞれの主成分を含む。このような実施形態において、複数の主成分中の主成分の数は、PCA アルゴリズムによって計算される主成分の閾値数に限定され得る。主成分の閾値数は

10

20

30

40

50

、例えば、5、10、20、50、100、1000、1500、又は任意の他の数であり得る。ある実施形態において、PCAアルゴリズムによって計算される各主成分は、PCAアルゴリズムによって固有値を割り当てられ、第1の複数の抽出された特徴の対応するサブセットが、最も高い固有値を割り当てられた主成分の閾値数に限定される。第1及び第2の複数の細胞成分ベクトル130中の各それぞれの細胞成分ベクトルについて、複数の次元削減成分が、それぞれの細胞成分ベクトルに適用されて、複数の次元削減成分中の各それぞれの次元削減成分についての次元削減成分値を含む対応する次元削減ベクトルを形成する（例えば、図3の行列Mを形成する）。これにより、対応する第1及び第2の複数の次元削減ベクトルが形成される。さらに、ある実施形態において、本方法は、クラスタリングを行って、クラスターのセットC<sub>j</sub>を生成する工程を含む（例えば、図3のクラスタリング340）。各クラスターは、第1及び第2の複数の次元削減ベクトルのサブセットに対応する複数の点を含む。クラスターのセットC<sub>j</sub>の第1のクラスターからの第1の複数の細胞及びクラスターのセットC<sub>j</sub>の第2のクラスターからの第2の複数の細胞の両方が同定される。

10

#### 【0105】

ある実施形態において、方法200は、対応する第1及び第2の複数の次元削減ベクトル130を用いて、多様体学習（例えば、図3の多様体学習330）を行う工程を含む。この多様体学習は、第1及び第2の複数の細胞における各細胞の他の各細胞に対する相対的な細胞状態を同定する（例えば、図3の行列Nを生成する）。多様体学習については、例えば、参照により本明細書に援用される、Wang et al., 2004, “Adaptive Manifold Learning”, Advances in Neural Information Processing Systems 17を参照されたい。

20

#### 【0106】

ある実施形態において、複数の非摂動細胞は、対照細胞（例えば、摂動に曝されていない細胞）である。さらに、ある実施形態において、非摂動細胞は、摂動に曝された非関連摂動細胞の平均である。

#### 【0107】

ある実施形態において、本方法は、単一細胞の移行シグネチャー及び/又は摂動シグネチャーをプルーニングする工程を含む。このプルーニングは、複数の細胞成分132を制限する（例えば、細胞成分を転写因子に制限する）。

30

#### 【0108】

ある実施形態において、細胞成分の発現の差の尺度（例えば、図3の、異なって発現された細胞成分350）は、第3の複数の第3の単一細胞細胞成分の発現データセットと第4の複数の第4の単一細胞細胞成分の発現データセットとの間の細胞成分量の差を定量する。同様に、ある実施形態において、この差は、母平均の差の検定、ウィルコクソンの順位和検定、t検定、ロジスティック回帰、又は一般化線形モデルのうちの1つを用いて決定される。さらに、第3の複数の第3の単一細胞細胞成分の発現データセット中の各それぞれの第3の単一細胞細胞成分の発現データセット122が、複数の非摂動細胞中の対応する単一細胞から得られる。さらに、第4の複数の第4の単一細胞細胞成分の発現データセット中の各それぞれの第4の単一細胞細胞成分の発現データセットが、摂動に曝された複数の摂動細胞中の第4の複数の細胞の対応する単一細胞から得られる。

40

#### 【0109】

ある実施形態において、それぞれの細胞成分についての対応する第2の重要度スコアを決定する工程は、複数の細胞成分中の各それぞれの細胞成分について、それぞれの細胞成分についての重要度スコアを、それぞれの細胞成分についての対応するマッチングスコアで置き換える工程を含む（例えば、細胞成分132-1-1に関連する重要度スコア134-1-1を、図1の重要度スコア134-d-Eで置き換える）。ある実施形態において、この置き換えにより、マッチングスコアが形成される。マッチングスコアは、複数の細胞成分について組み合わされて、摂動についてのマッチングスコアを生成する。したが

50

って、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連する（例えば、細胞の移行に影響する）かどうかは、それぞれの摂動についてのマッチングスコアに基づいて決定される。ある実施形態において、マッチングスコアは、別個のスコア又は連続したスコアを含む。

#### 【0110】

ある実施形態において、スコア134を置き換える工程は、それぞれの細胞成分についての単一細胞の移行シグネチャー142からの細胞成分量132及びそれぞれの細胞成分についての摂動シグネチャー150からの細胞成分量132が両方とも上方制御される場合、重要度スコアを第1のスコアで置き換える工程を含む。この置き換える工程は、それぞれの細胞成分についての単一細胞の移行シグネチャー142からの細胞成分量が上方制御され、それぞれの細胞成分についての摂動シグネチャー150からの細胞成分量が下方制御される場合、重要度スコア132を第2のスコアで置き換える工程をさらに含む。さらに、それぞれの細胞成分についての摂動シグネチャー150からの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアは、第3のスコアで置き換えられる。

10

#### 【0111】

ブロック206．ブロック206を参照すると、方法200は、単一細胞の移行シグネチャー142-1及び摂動シグネチャー150-1を比較する工程を含む。この比較により、摂動が細胞の移行に影響を与えるかどうかを決定する。

#### 【0112】

ある実施形態において、方法200は、単一細胞の移行シグネチャー142及び/又は摂動シグネチャー150をフィルタリングする工程を含む。このフィルタリングは、単一細胞の移行シグネチャー142及び摂動シグネチャー150に含まれる細胞成分132の数を減少させ、これは、方法200を行う（例えば、図3の後処理360を行う）のに必要なシグネチャーのデータサイズ及び時間を減少させるのに役立つ。

20

#### 【0113】

ある実施形態において、方法200は、比較206に基づいて（例えば、図3の後処理360に基づいて）変化した細胞状態を促進するものとして摂動を同定する工程を含む。ある実施形態において、単一細胞の移行シグネチャー142及び/又は摂動シグネチャー150は、異なるタイプの細胞成分を用いて生成される。同様に、ある実施形態において、単一細胞の移行シグネチャー142及び/又は摂動シグネチャー150は、同じタイプの細胞成分を用いて生成される。

30

#### 【0114】

##### II．単一細胞分析を行うために細胞をインビトロで培養する方法

細胞の運命の原因を同定するための本明細書に記載される技術を行う際、単一細胞から得られる細胞成分測定に関するデータセットを生成することが有用である。これらのデータセットを生成する（例えば、図3の単一細胞の測定310によって図1のデータセット122-1を生成する）ために、対象とする細胞の集団が、インビトロで培養される。対象とする1つ以上の細胞成分132の単一細胞の測定が、培養中に1つ以上の期間で行われて、データセット122を生成する。（例えば、図3の単一細胞の測定310）。ある実施形態において、対象とする細胞成分としては、DNA、修飾（例えば、メチル化）DNAを含む核酸、コーディング（例えば、mRNA）若しくはノンコーディングRNA（例えば、snRNA）を含むRNA、転写後修飾タンパク質（例えば、リン酸化、グリコシル化、ミリスチル化タンパク質など）を含むタンパク質、脂質、炭水化物、環状アデノシンーリン酸（cAMP）及び環状グアノシンーリン酸（cGMP）などの環状ヌクレオチドを含むヌクレオチド（例えば、アデノシン三リン酸（ATP）、アデノシン二リン酸（ADP）及びアデノシンーリン酸（AMP））、ニコチンアミドアデニンジヌクレオチド（NADP/NADPH）の酸化及び還元形態などの他の小分子細胞成分、及びそれらの任意の組合せが挙げられる。ある実施形態において、細胞成分測定は、RNAレベルなどの遺伝子発現測定を含む。

40

50

## 【0115】

いくつかの単一細胞細胞成分の発現測定技術のいずれか1つが、データセット122を収集するために使用され得る(例えば、表1の技術、図1の単一細胞の測定310の技術など)。例としては、限定はされないが、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスPOSEでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seqなどが挙げられる。細胞成分の発現測定は、測定される所望の細胞成分に基づいて選択され得る。例えば、scRNA-seq、scTag-seq、及びmiRNA-seqが、RNA発現を測定する。具体的には、scRNA-seqが、RNA転写物の発現を測定し、scTag-seqが、稀なmRNA種の検出を可能にし、miRNA-seqが、マイクロ-RNAの発現を測定する。CyTOF/SCoP及びE-MS/Abseqが、細胞におけるタンパク質発現を測定する。CITE-seqが、細胞における遺伝子発現及びタンパク質発現の両方を同時に測定し、scATAC-seqが、細胞におけるクロマチン構造を測定する。以下の表1は、上述される単一細胞細胞成分の発現測定技術のそれぞれを行うためのプロトコルの例へのリンクを提供する。

10

## 【0116】

## 【表 1】

表 1 - 測定プロトコルの例

技術	プロトコル
RNA-seq	Olsen <i>et al.</i> , (2018), "Introduction to Single-Cell RNA Sequencing," <i>Current protocols in molecular biology</i> 122(1), pg. 57.
Tag-seq	Rozenberg <i>et al.</i> , (2016), "Digital gene expression analysis with sample multiplexing and PCR duplicate detection: A straightforward protocol," <i>BioTechniques</i> , 61(1), pg. 26.
ATAC-seq	Buenrostro <i>et al.</i> , (2015), "ATAC-seq: a method for assaying chromatic accessibility genome-wide," <i>Current protocols in molecular biology</i> , 109(1), pg. 21.
miRNA-seq	Faridani <i>et al.</i> , (2016), "Single-cell sequencing of the small-RNA transcriptome," <i>Nature biotechnology</i> , 34(12), pg. 1264.
CyTOF/SCoPE-MS/Abseq	Bandura <i>et al.</i> , (2009), "Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry," <i>Analytic chemistry</i> , 81(16), pg. 6813.  Budnik <i>et al.</i> , (2018), "SCoPE-ME: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation," <i>Genome biology</i> , 19(1), pg. 161.  Shahi <i>et al.</i> , (2017), "Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding," <i>Scientific reports</i> , 7, pg. 44447.
CITE-seq	Stoeckius <i>et al.</i> , (2017), "Simultaneous epitope and transcriptome measurement in single cells," <i>Nature Methods</i> , 14(9), pg. 856.

10

20

30

40

50

## 【0117】

使用される細胞成分の発現測定技術は、細胞死をもたらす得る。或いは、細胞成分は、生細胞から抽出することによって、例えば、細胞を死滅させずに細胞質を抽出することによって測定され得る。この種の技術は、同じ細胞が複数の異なる時点で測定されるのを可能にする。

## 【0118】

同じ「前駆」細胞から生じる複数の異なる細胞型が集団中に存在するように、細胞集団が不均一である場合、単一細胞細胞成分の発現測定が、細胞が培養物中で成長するにつれて単一の時点で又は比較的少ない時点で行われ得る。細胞集団の不均一性の結果として、収集されたデータセット 122 は、移行の軌道に沿って様々なタイプの細胞を表す。

## 【0119】

単一の細胞型のみ又は比較的少ない細胞型、主に、対象とする「前駆」細胞が、集団中に存在するように、細胞集団が、実質的に均一である場合、単一細胞細胞成分の発現測定

は、細胞が移行するにつれて所定の期間にわたって複数回行われ得る。

【0120】

別個の単一細胞細胞成分の発現データセット122が、各細胞について、及び該当する場合、期間（例えば、図1の期間128）のそれぞれの時点で生成される。複数の異なる時点における細胞の集団からの単一細胞細胞成分の発現測定の実集は、同じ「前駆」細胞から生じる細胞型について時間とともに細胞発現の「擬似時間（pseudo-time）」表現として集合的に解釈され得る。擬似時間という用語は、2つの点で使用され、第1に、細胞状態の移行が、細胞間で必ずしも同じでないため、細胞の集団が、その「前駆」タイプの細胞の移行過程が時間とともに経る可能性の高い分布を提供し、第2に、複数の時点でのそれらの複数の細胞の発現の細胞成分の発現の測定が、異なる細胞の細胞成分の発現の測定がデータセットを生じさせる場合でさえ、可能な移行挙動を時間とともにシミュレートするという点である。意図的に簡略化した例として、細胞Xが、時点Aについてのデータセットを示し、細胞Yが、時点Bについてのデータセットを示す場合でさえ、これらの2つのデータセットは一緒に、時点Aと時点Bとの間の移行の擬似時間を表す。

10

【0121】

説明の便宜上、2つの異なる期間に「同じ」細胞について捕捉される2つのこのようなデータセット122（例えば、第1のデータセット122-1の第1の期間128-1、第2のデータセット122-2の第2の期間128-2など）（上で紹介されるように細胞を死滅させない技術が使用されると仮定する）は、本明細書において異なる「細胞」（及び対応する異なるデータセット）と呼ばれ、これは、実際には、このような細胞が、わずかに又は顕著に互いに移行されることが多く、ある場合には、様々な細胞成分の相対量から決定される際に全く異なる細胞型を有するためである。この文脈から見ると、異なる時点における単一細胞のこれらの2つの測定は、細胞自体が変化しているため、分析のために異なる細胞として解釈され得る。

20

【0122】

本明細書に記載される細胞（例えば、図1の細胞/データセット識別子126）/期間（例えば、図1の期間128）によるデータセットの分離は、説明の明確さのためのものであるため、実際に、これらのデータセットは、コンピュータメモリに保存され、1つ以上の集合データセットとして（例えば、全ての期間について細胞によって、全ての細胞及び期間について一度に）論理的に操作され得る。

30

【0123】

ある実施形態において、対象とする「前駆」細胞がそのベースライン状態から摂動されているデータセット122を収集することが有用である。これを行ういくつかの考えられる理由があり、例えば、1つ以上の細胞成分をロックアウトして（例えば、除去する、無効にするなど）、健康な細胞状態と病変した細胞状態との間の差を評価するためである。これらの実施形態において、プロセスは、細胞への所望の改変を導入するための工程も含み得る。例えば、1つ以上の摂動が、細胞に導入されてもよく、1つ以上の細胞成分をロックアウトするように設計された調整されたウイルスが導入されてもよく、CRISPRが、細胞成分を編集するのに使用されてもよいなどである。使用され得る技術の例としては、限定はされないが、RNA干渉（RNAi）、転写活性化因子様エフェクターヌクレアーゼ（TALEN）、又は亜鉛フィンガーヌクレアーゼ（ZFN）が挙げられる。

40

【0124】

どのように摂動が適用されるかに応じて、全ての細胞が同じように摂動されるわけではない。例えば、ウイルスが、特定の遺伝子をロックアウトするために導入される場合、そのウイルスは、集団中の全ての細胞に影響を与えるわけではない。より一般に、この特性は、単一の集団に対する多くの異なる摂動の影響を評価するために有利に使用され得る。例えば、多数の調整されたウイルスが導入されてもよく、これらはそれぞれ、異なる遺伝子をロックアウトさせることなどの異なる摂動を行う。ウイルスは、様々な細胞のいくつかのサブセットを様々に感染させて、対象とする遺伝子をロックアウトする。単一細胞シーケンシング、又は次に、別の技術が、どのウイルスがどの細胞に影響を与えたかを同定

50

するのに使用され得る。次に、得られた異なる単一細胞シーケンシングデータセットが、本明細書の他の箇所に記載される方法にしたがって遺伝子発現に対する遺伝子ノックアウトの影響を同定するために評価され得る。

【0125】

複数の異なる摂動の導入、CRISPRのバーコーディングなどの他のタイプの多摂動細胞改変が、同様に行われ得る。さらに、2つ以上のタイプの摂動が、分析される細胞の集団中に導入され得る。例えば、細胞は、異なって影響され得（例えば、異なるウイルスが導入される）、異なる摂動が、細胞の異なる部分集団中に導入され得る。

【0126】

さらに、細胞の集団の異なるサブセットが、多くの摂動を単に混合すること及びどの細胞がどの摂動によって影響されたかを事後解析することを超える様々な方法で摂動され得る。例えば、細胞の集団が、マルチウェルプレートの異なるウェル中に物理的に分割される場合、異なる摂動が、各ウェルに適用され得る。異なる細胞のために異なる摂動を達成する他の方法も可能である。

【0127】

以下に、方法が、単一細胞遺伝子発現測定を用いて例示される。本発明は、単一細胞から得られる他の細胞成分の測定を用いた類似の方法を包含するため、これは、限定ではなく、例示としてのものであることが理解されるべきである。本発明は、本開示に記載される方法を実施する個人又は組織によって行われる実験研究から直接得られる測定を用いた方法、並びに例えば、他の人によって行われ、及び第三者による刊行物に報告されたデータ、データベース、請負業者によって行われるアッセイ、又は開示される方法を実施するのに有用な好適な入力データの他の源を含む、任意の手段又は機構を介して利用可能な、実験研究の結果の報告から間接的に得られる測定を用いた方法を包含することがさらに理解されるべきである。

【0128】

本明細書に記載されるように、細胞中の遺伝子発現は、細胞をシーケンシングし、次に、シーケンシング中に同定された各遺伝子転写物の量を計数することによって測定され得る。ある実施形態において、シーケンシング及び定量される遺伝子転写物は、RNA、例えばmRNAを含み得る。代替的な実施形態において、シーケンシング及び定量される遺伝子転写物は、mRNAの下流の産物、例えば、転写因子などのタンパク質を含み得る。一般に、本明細書において使用される際、「遺伝子転写物」という用語は、遺伝子転写又は翻訳（翻訳後修飾を含む）の任意の下流の産物を示すのに使用され得、「遺伝子発現」は、一般に、遺伝子転写物の任意の尺度を指すのに使用され得る。

【0129】

本明細書の残りの部分は、遺伝子転写物及び遺伝子発現の分析に焦点を合わせているが、本明細書に記載される技術は全て、それらの細胞に関して単一細胞に基づいてデータを得る任意の技術に等しく適用可能である。例としては、単一細胞プロテオミクス（タンパク質発現）、クロマチン構造（クロマチン状態）、メチル化、又は他の定量可能な後成的影響が挙げられる。

【0130】

以下の説明は、複数の期間（例えば、図1の複数の期間128）にわたって単一細胞細胞成分の発現測定（例えば、図3の測定310）を行うために、細胞の集団をインビトロで培養するための概要の例を提供する。一般に、細胞をインビトロで培養するための方法が、当業者に公知である。当業者はまた、より長い又は短い期間にわたって増殖させるようにこのプロセスがどのように修正され得るかを理解し、さらなる又はより少ない単一細胞の測定工程を行うなどである。

【0131】

一実施形態において、第1の細胞状態の細胞を、変化した細胞状態の細胞になるまで培養するための方法が、以下の工程の1つ以上を含む：

0日目：第1の細胞状態のいくつかの細胞を、細胞の増殖に好適な培地中でプレート上

10

20

30

40

50

に解凍する。

1日目：第1の細胞状態のある数の細胞を、マルチウェルプレート中に播種する。該当する場合、細胞の細胞成分に影響を与えるためのさらなる工程を行う。例えば、対象とする細胞成分をロックアウトするために1つ以上のウイルスに同時に感染させる。

ウェル中の細胞について細胞成分の発現測定 of 反復  $t_1$  を行う。

1 + 1日目：何らかのさらなるプロセスが行われた場合、必要に応じて培地を交換する。

該当する場合、ウェル中の細胞について細胞成分の発現測定 of 反復  $t_1$  を行う。

1 + m日目：変化した細胞状態の細胞の増殖を補助するのに適切な培地に、培地を交換する。

該当する場合、ウェル中の細胞について細胞成分の発現測定 of 反復  $t_m$  を行う。

1 + n、o、p日目など：第1の細胞状態から変化した細胞状態へのさらなる細胞状態の移行を補助するために必要に応じて培地を交換する。該当する場合、第1の細胞状態から変化した細胞状態へのさらなる移行に影響を与えるためのさらなる工程を行う。例えば、細胞を変化した細胞状態にするための、対象とする摂動を加える。

該当する場合、ウェル中の細胞について、細胞成分の発現測定 of 反復  $t_n$ 、 $t_o$ 、 $t_p$  などを行う。

q日目：ウェル中の変化した細胞状態の細胞について細胞成分の発現測定 of 反復  $t_q$  を行う。

プレートを固定し、対象とする細胞成分 / タンパク質に適合した抗体で染色して、測定されるべきそれらを溶解 / 破壊させずに細胞をソート / 同定する。それはまた、細胞質の状況におけるほどの分解を伴って見られない表面マーカーを同定するのに使用され得る。各ウェルを走査することによる *Molecular Devices HCI IXM4* などの細胞イメージングシステムを用いてイメージングする。所望の変化した細胞状態にあるウェル当たりの細胞の数を定量する。

【0132】

表2は、1つ以上の時点における細胞の集団の単一細胞発現の測定（例えば、図3の単一細胞の測定310）から収集され得るデータの例を含む、いくつかのデータセット122の断片を示す。サンプルIDの列は、細胞のデータがどのサンプルに由来するかを示す（例えば、図1のサンプル識別子124-1）。実際には、集団中の細胞は、2つ以上のサンプル（例えば、第1のサンプル識別子124-1、第2のサンプル識別子124-2など）から取り出されてもよく、これらはそれぞれ、同じ又は異なる対象に由来していてもよい。細胞又はデータセットの列は、所与の行のデータがどの細胞又はデータセットに関連するかを示す（例えば、図1の細胞 / データセット識別子126-1）。データセット122は、或いは、データのベクトル  $r_i$ （例えば、図1の細胞成分ベクトル130-1）として表され得る。期間の列は、関連する場合、細胞の増殖中、その行のデータセットがいつ捕捉されたかを示す（例えば、図1の期間128-1）。

【0133】

表2の残りの列は、細胞の対象とする細胞成分（細胞成分132-1-1 ~ 132-1-B）に対応する。これは、細胞の全ての細胞成分、又は単にサブセットであり得る。各細胞成分132は、異なる列に関連する。データセットが、ベクトル  $r_i$  として表される場合、各細胞成分は、ベクトル中のエントリ  $i$  に対応する。ある実施形態において、各細胞の値は、単一細胞発現測定によって測定されるいくつかの細胞成分の（整数）数、又はそのいくつかの正規化（有理数）形態であり得る。

【0134】

10

20

30

40



## 【表 2】

表 2 - データセットの例

サンプル ID	細胞/データセット $r$	期間	細胞成分 1 0610005C13Rik	細胞成分 2 0610007C21Rik	細胞成分 3 0610007L01Rik
ID1	1_iN1_C01	$t_0$	0	0	0
ID1	1_iN1_C02	$t_0$	0	7.377047	0
ID1	1_iN1_C03	$t_1$	0	0	3.544937
ID2	1_iN1_C04	$t_1$	0	3.926875	0
ID2	1_iN1_C05	$t_2$	0	6.39981	6.946602
ID2	1_iN1_C07	$t_2$	0	7.995805	2.724768

10

## 【 0 1 3 5 】

III . 細胞成分の発現の差を決定するために単一細胞データセットを分析する方法

III . A . 概説

20

細胞状態の移行（すなわち、第 1 の細胞状態から変化した細胞状態への細胞の状態の移行）は、細胞における細胞成分 1 3 2 の発現の変化によって特徴付けられる。例えば、移行は、細胞における細胞成分の発現 1 3 2 の変化、ひいては細胞によって産生される細胞成分（例えば、mRNA、転写因子）の同定及び定量によって特徴付けられ得る。しかしながら、少なくとも現在、細胞状態の移行は、細胞内活性の複雑性のため、完全に決定性ではない。この複雑性を理解しようとして、本明細書は、細胞状態の移行の様々な段階での対象とする 1 つ以上の測定される細胞成分の可変の存在、非存在又は量に関連する可変の細胞成分の発現が、有意義な知識が抽出され得る高次元データセット（例えば、図 1 の細胞成分ベクトル 1 3 0）を提供するという理論の下で、細胞の集団の細胞中の細胞成分 1 3 2 を定量する単一細胞データセット 1 2 2 に統計的技術を適用する。本明細書において、データの次元は、データセット 1 2 2 に含まれる前駆細胞成分の測定に由来する。各細胞成分 1 3 2 は、次元を表し、集合的に、各細胞成分についての細胞成分測定データセット 1 2 2 が、「前駆」細胞の、異なる細胞型への移行に関する生物学的過程についての潜在情報を符号化する形状を有し得る。実際には、細胞成分 1 3 2 の数は、およそ数万であり得るため、本明細書に記載される計算は、知的に又は手動によって行うことが不可能ではないにしても実用的ではない。

30

## 【 0 1 3 6 】

一般に、これらの統計技術は、高次元データが、データセット中で符号化される潜在情報が何であれその形状を保ちながらより低次元の空間へと圧縮される方法として特徴付けられ得る（例えば、図 1 の細胞成分ベクトル 1 3 0 は、図 3 の行列  $M$  へと次元削減された 3 2 0）。低次元データは、細胞状態の移行の異なる段階間で異なって存在する細胞成分を同定するために評価される。方法への入力データが、細胞ごとの対象とする複数の細胞成分の単一細胞成分の発現データセット 1 2 2 である際、異なって発現された細胞成分のセットは、したがって、細胞の他の細胞成分と比べた存在、非存在又は量に関して、どの細胞成分が統計的に有意な過剰提示又は提示不足を有していたかを表す。いくつかの方法及び測定基準のいずれか 1 つを用いて、それらの細胞成分のうちのどれが、本明細書にしたがって「異なって発現される」とタグ付けされるように他の細胞成分に対して十分に「異なって」発現されるかを同定することができる。データセット 1 2 2 が得られる細胞の集団が、異なる型及び異なる移行段階の細胞を含む際、どの細胞成分が異なって存在するか（例えば、どの細胞成分が異なって発現されたか）を知ることは、どのような細胞

40

50

成分が、移行の過程又は他の移行において活性な細胞成分の発現に影響を与えるか又はそれに関連するかについての洞察を与える。

【 0 1 3 7 】

I I I . B . 使用事例

どの種類の方法が使用されるかにかかわらず、異なって発現された細胞成分の決定は、どのような結果が求められるかに応じて変化し得る。例えば、使用される方法が特定の細胞を系統内 ( o n - l i n e a g e ) 又は系統外 ( o f f - l i n e a g e ) であると同定する場合、どの細胞成分が異なって発現されるかの決定は、系統内 ( o n - l i n e a g e ) であると決定された細胞の細胞成分の発現レベルを、系統外 ( o f f - l i n e a g e ) であると決定された細胞の細胞成分と比較することによって行われ得る。それらの細胞成分の相対的発現は、どの細胞成分が、個別に又は組み合わせられて、1つの型又は別の型の細胞中で活性であることを示す。上記のように、この発現データは、異なって発現されるとフラグを付けられる細胞成分のサブセットを同定するのに使用され得る。次に、因果関係が、同定された細胞成分をインピットでロックアウトし、実験細胞集団の細胞の運命が、どの細胞成分が活性であるかの変化によって影響されるか否かを評価することによって決定され得る。

10

【 0 1 3 8 】

別の例として、使用される方法が、特定の細胞を系統内 ( o n - l i n e a g e ) であると同定し、他の細胞を系統内 ( o n - l i n e a g e ) 細胞型に向かう移行軌道に沿って「前駆」細胞又は中間細胞であると同定する場合、どの細胞成分が異なって発現されるかの決定は、系統内 ( o n - l i n e a g e ) であると決定された細胞の細胞成分の発現レベルを、系統内 ( o n - l i n e a g e ) 細胞の「前駆」及び/又は中間細胞であると決定された細胞の細胞成分と比較することによって行われ得る。前の段落にあるように、それらの細胞成分の相対的発現は、どの細胞成分が、個別に又は組み合わせられて、1つの型又は別の型の細胞中で活性であることを示し、同様に、この発現データは、異なって発現されるとフラグを付けられる細胞成分のサブセットを同定するのに使用され得る。また、上記のように、次に、因果関係が、同定された細胞成分をインピットでロックアウトし、実験細胞集団の細胞の運命が、どの細胞成分が活性であるかの変化によって影響されるか否かを評価することによって決定され得る。

20

【 0 1 3 9 】

別の例として、細胞の集団は、細胞の2つの部分集団、1つの健康な部分集団及び1つの不健康な部分集団を含み得る。細胞培養中、複数の異なる摂動が、不健康な部分集団中に導入され得る。本明細書に記載される方法とともに、その後の単一細胞発現測定によって、特に、健康な部分集団と比べて、摂動が不健康な部分集団中の細胞成分の細胞成分の発現の差にどのような影響を与えたかが決定され得る。例えば、1つ以上の摂動に曝された不健康な部分集団からの細胞のサブセットは、細胞の健康な部分集団と一致する細胞成分の発現を示すことがあり、これは、摂動が、細胞の不健康な部分集団に望ましい効果を与えたことを示す。

30

【 0 1 4 0 】

I I I . C . 低次元データを用いた、異なって発現された細胞成分の決定

図3は、一実施形態に係る、異なって発現された細胞成分132のセットを決定するための細胞成分の発現の差のアッセイの第1の例についてのフローチャートである。図3は、細胞成分の発現の差を用いて記載される一般的な場合の非限定的な例示的な実施形態を提供することに留意されたい。工程310において、単一細胞発現測定が、上記の第I節に記載されるように行われて、細胞の集団についての複数のデータセット122を生成する。上記のように、各細胞についての各データセット122が、細胞成分(例えば、図1の細胞成分132-1-1~132-1-B)のそれぞれの量を含む細胞成分のベクトル  $r_i$  (例えば、図1の細胞成分ベクトル130)として表され得る。単一細胞発現測定310から得られるデータセット122は、一般に、計算装置(例えば、図1のシステム100)の永続的メモリ(例えば、図1の永続的メモリ112)中にデジタル形式で保存

40

50

されるが、それらは、本明細書に記載される残りの工程を行うために、必要に応じて能動メモリ（例えば、図1の非永続的メモリ111）中にロードされ得る。一般に、図3のプロセスの残りの工程は、1つ以上の計算装置（例えば、図1のシステム100）によって行われる。計算装置の例が、図1に関して説明される。しかしながら、実際には、図3のプロセスは、本明細書に記載される工程の結果に基づいて行われるさらなるインビトロ試験又は臨床決定などの、コンピュータの外で行われ得るさらなる中間の又は後続の工程を含み得る。

#### 【0141】

##### III.C.1.次元削減

上で紹介されるように、細胞成分132のそれぞれが、データの異なる次元を表す際、データセット122は、全体で、高次元を有する。工程320において、次元削減が、計算装置（例えば、システム100）によって行われて、データセット122の細胞成分132の量中に存在する任意の潜在的パターンを保持しながらデータの次元を削減する。

10

#### 【0142】

次元削減工程320への入力は、一般に、上記の表2と同様に、個々の細胞の発現ベクトル（例えば、図1の細胞成分ベクトル130）を連結する行列である。次元削減320の出力は、データの基本的な潜在構造を維持しながら元のデータを圧縮形態で符号化する、本明細書において簡略化のために行列「M」と呼ばれる行列である。行列M中の各行は、細胞の特定の1つに関連する。行列M中の各列は、次元削減によって提供される削減された次元空間における次元の1つに関連する。各行・列のグループ分けでのエントリ中の値は、元の入力データセットに基づいて次元削減によって決定される。

20

#### 【0143】

ある実施形態において、これらの次元削減技術は、データのいくらかの不可逆圧縮をもたらすが、得られる出力行列Mは、計算的な記憶サイズがより小さいため、このプロセスの残りの工程において説明される他の下流の技術を用いて分析するのにより少ない計算処理能力を必要とし、これにより、現代の計算装置を用いて妥当な時間でこれらの工程の結果を得ることが計算的に実現可能になる。

#### 【0144】

様々な次元削減技術が使用され得る。例としては、限定はされないが、主成分分析（PCA）、非負値行列因子分解（NMF）、線形判別分析（LDA）、拡散マップ、又はオートエンコーダなどの（ニューラル）ネットワーク技術が挙げられる。

30

#### 【0145】

これらの段落において言及される技術のそれぞれは、変動の主な駆動因子を抽出し、元の入力データの次元を削減するために異なって動作するが、それぞれ、より低次元の空間において行列Mを出力する。

#### 【0146】

##### III.C.2.多様体学習

行列M中の削減された次元データ（例えば、次元削減成分ストア146）は、単一細胞発現データセット122から元の高次元データと比べて次元が有意に削減される。しかしながら、得られる行列Mは、非線形多様体（例えば、図1の多様体149）を埋め込む。工程330において、多様体学習技術は、多様体を抽出するために行列Mに適用される。多様体149自体が、擬似時間にわたって細胞の間で細胞成分の発現の差についての有用な情報を提供するだけでなく、それはまた、その情報を視覚化するのにも使用され得る。

40

#### 【0147】

多様体学習工程330への入力、次元削減工程320からの行列Mである。多様体学習330の出力は、本明細書において行列「N」又は多様体（例えば、図1の多様体149）と呼ばれる別の行列である。行列Nの構造は、行列Nの各行が、集団の元の細胞の1つに対応するようなものであり、本明細書においてこのプロセスの残りの工程のための「点」と呼ばれる。一実施形態において、行列Nは、多様体学習工程330が、使用される

50

特定の多様体学習アルゴリズムと関係なく出力するように構成される2つの次元に対応するX次元及びY次元と任意に呼ばれる2つの列を有する。X及びY次元は、多様体学習工程によって決定され、どの多様体アルゴリズムが使用されるかにしたがって、行列Mからのデータにどの次元が最良に適合するかにしたがって選択される。2つのこのような列を有する多様体は、図4Bに示されるように、視覚化に好都合である。他の実施形態において、多様体行列Nは、本明細書に導入される2次元形態を超えてさらなる次元を有する。

【0148】

行列Nの例が、以下の表3に示される。図4Bは、力指向レイアウトが次元削減工程に使用される実施形態において、以下の実施例1からのデータのプロットを示す。図4B中のプロットは、この及び同様の例の実験において、点がX/Y平面内の1つ以上の軌道に沿ってX/Y次元において分離するという点で、このプロセスにしたがって達成される結果の例示であり、ここで、一般に「前駆」細胞が、X/Y空間内の1つの一般的な領域において見られ、X/Y空間内の別の一般的な領域で中間細胞に向かって拡散し、X/Y空間内の1つ以上の異なる領域で終了し、これは、実際には、一般に、系統内(on-line age)又は系統外(off-line age)の移行された細胞であると確認される。一般に、同定される領域及び軌道の数は、「前駆」細胞のタイプ及び「前駆」細胞が移行することが知られている細胞のタイプに依存する。さらに、点の領域は、それらの中にいくらかの量の拡散を有することが多く、これは、移行のプロセスにおいて異なる進行段階の細胞を示唆する。

10

【0149】

20

【表3】

表3 - 出力行列N

点 (細胞)	X次元	Y次元
点1	$x_1$	$y_1$
点2	$x_2$	$y_2$
点3	$x_3$	$y_3$
点4	$x_4$	$y_4$
...	...	...
点 $p$	$x_p$	$y_5$

30

【0150】

様々な多様体学習技術が、行列Mに適用されて、行列Nを生成し得る。例としては、限定はされないが、力指向レイアウト(Fruchterman, T.M., & Reingold, E.M. (1991). Graph drawing by force directed placement. Software: Practice and experience, 21(11), 1129-1164) (例えば、Force Atlas 2)、t分布型確率的近傍埋め込み(t-SNE)、局所線形埋め込み(Roweis, S.T., & Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500), 2323-2326)、局所線形アイソメトリックマッピング(ISOMAP、Tenenbaum, J.B., De Silva, V., & Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500), 2319-2323)、カーネルPCA、グラフベースのカーネルPCA、親和性に基づく軌道埋め込みのための熱拡散の可能性(PHATE)、一般化判別分析(GDA)、一様多様体近似及び投影(UMAP)、又はカーネル判別分析が挙げられる。判別分析は、特に、いくらかの情報が各細胞の特定の細胞型に関して予め知られている場合に使用され

40

50

得る。力指向レイアウトは、細胞状態の移行のような基本的な生物学的過程から生じる基本的なデータの非線形形態様を符号化する新しいより低い次元を同定するそれらの能力のため、様々な特定の実施形態において有用である。力指向レイアウトは、データを最良に表す削減された次元を決定するための機構として物理学ベースのモデルを使用する。例として、力指向レイアウトは、物理学シミュレーションの形態を使用し、この実施形態において、セット中の各細胞/データセットが、「反発」力を割り当てられ、細胞の全体にわたって計算されるとき、これらの競合する「力」の下で一緒に「拡散する」データのセクターを同定するグローバルな「重力」が存在する。力指向レイアウトは、データの構造についての仮定をほとんど立てず、ノイズ除去手法を課さない。

【0151】

多様体学習330を行う工程は任意選択の工程であることに留意されたい。ある実施形態において、多様体学習は行われなない。

【0152】

### III.C.3. クラスタリング

工程340において、次元削減320によって提供される低次元空間中の点の位置におけるパターン（例えば、関連する複数の次元削減ベクトル146のサブセットに対応する）を同定するために、クラスタリングが行われて、 $j$  クラスターのセット  $C_j$  を生成する。これらのクラスターを用いて、同様の点（細胞/データセット）を統合して、低次元空間において互いに類似している点の群（例えば、第1のクラスター、第2のクラスターなど）についての統計的に関連する情報を引き出す。以下の表4は、ラスタリング340の出力であり得る点のクラスタリングの例を示す。

【0153】

【表4】

表4- クラスター割り当て

細胞/ 点	クラスター 割り当て $C_j$
点1	$C_1$
点2	$C_1$
点3	$C_1$
点4	$C_2$
点5	$C_2$
点6	$C_2$

【0154】

いくつかのクラスタリング技術のいずれか1つが使用され得、その例としては、限定はされないが、階層的クラスタリング、 $k$ -平均クラスタリング、及び密度に基づくクラスタリングが挙げられる。特定の一実施形態において、階層的密度に基づくクラスタリングアルゴリズムが使用される（HDBSCANと呼ばれる、Campello, R. J., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Transactions on Knowledge Discovery from Data (TKDD), 10(1), 5)。別の実施形態において、Louvainクラスタリングなどのコミュニティ検出に基づくクラスタリングアルゴリズムが使用される（Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfold

10

20

30

40

50

ing of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008)。

【0155】

クラスタリングのために、これらの技術は、行列Mのデータを使用して、クラスターを決定する。アルゴリズムと関係なく、一般に、行列Mの多次元空間内の互いに近い点が、同じクラスターに割り当てられる可能性が高く、互いにより離れた点は、同じクラスターに割り当てられる可能性が低い。図7Aは、点のそれぞれへの異なる視覚的マーカーで示されるクラスター割り当て1-10とともに、図4Bからのデータの例のプロットを提供する。クラスターの数は、操作者によって設定又は制限され得、及び/又は使用されるアルゴリズムに基づいて動的に決定され得る。

10

【0156】

III.C.4.細胞成分の発現の差の決定

次元削減320、任意選択の多様体学習330、及びクラスタリング340の工程は、一般に、集団の細胞、及びそれらの対応する単一細胞発現データセット122を、削減された次元空間内のクラスターへと組織化するように動作し、それによって、基本的な細胞成分ごとの発現測定データが、有意義な情報を抽出するために統合及び分析され得るようになっていく。ある実施形態において、この削減された次元空間は、本開示の方法を完成させるのに必要な時間及び/又は処理能力をさらに減少させる。

【0157】

クラスターから得られる情報の1つの項目は、どの細胞成分がどの他の細胞に対して集団において異なって発現されるかである。本明細書において、細胞成分のこのセットは、図3中で工程350において説明される、異なって発現された細胞成分 $E_k$ のセットと呼ばれる。異なって発現された細胞成分のセットを生成するためのいくつかの使用事例の例が、上記の第III.B.節に記載されている。

20

【0158】

異なって発現された細胞成分のセットを決定するためにクラスター $C_j$ 及びデータセット情報を使用するいくつかの方法がある。一実施形態において、所与の細胞成分(例えば、細胞成分A)が異なって発現されるかどうかの決定は、他のクラスター $C_m$ (ここで、 $m$ が1でない)の1つ以上の中の点によって、細胞成分Aの量に対する所与のクラスター $C_1$ 中の点(細胞)によって細胞成分Aの量を評価することによって決定される。正規化も使用され得る。例えば、全体として細胞中の細胞成分による発現のレベルは、細胞状態の移行の生物学と関係ない理由のため、細胞間で異なり得る。したがって、細胞成分量は、データセット中の各細胞についての細胞成分量の全体数に基づいて正規化され得る。

30

【0159】

上記の第III.B.節に記載されるように、細胞成分Aについてのどのクラスターの細胞成分量が所与のクラスター $C_1$ に対して比較されるかは、実施形態によって異なり得る。比較に使用される他のクラスターは、系統内(on-line age)細胞型に最も強く関連する、系統外(off-line age)細胞型に最も強く関連する、「前駆」細胞型に最も関連する、中間の細胞型に最も関連するなどのクラスターであり得る。比較はまた、2つ以上の他のクラスターに対してなされ得る。

40

【0160】

比較を所与として、細胞成分Aは、クラスターごとの総細胞成分量(同様に、クラスター中の全ての点について、若しくは平均などのいくつかの集合的な尺度(aggregate measure))、クラスターごとの正規化された細胞成分量、クラスターごとの中央値、平均、若しくは他の集合的な細胞成分量、他の細胞成分の細胞成分量に対する発現の比率などのいくつかの測定基準のいずれか1つにしたがって、異なって発現されると同定され得る。一実施形態において、細胞成分Aが異なって発現されることを確立するための基準は、閾値要件(threshold requirement)である。

【0161】

50

例えば、クラスター  $C_1$  中の細胞成分 A についての正規化された細胞分量は、少なくとも閾値だけ、1つ以上の他のクラスター  $C_m$  中の細胞成分 A についての正規化された細胞分量を超え得る。

【0162】

異なって発現された細胞成分の決定はまた、相対的なものであり得る。一実施形態において、複数の細胞成分/クラスターの組合せについての正規化された細胞分量、複数の細胞成分/クラスターの組合せについての距離計量、又は他の同様の測定基準が計算され得る。それらの測定基準は、ランク付け基準にしたがってランク付けされ得（例えば、クラスター中の最も多い正規化された細胞分量）、上位にランク付けされる細胞成分又は細胞成分/クラスターの組合せは、異なって発現された細胞成分であると決定され得る。

10

【0163】

一実施形態において、所与のクラスター中の所与の細胞成分についての細胞分量が、どの細胞成分が異なって発現されるかを同定するのに使用され得る。一実施形態において、これらの、異なって発現された細胞成分は、母平均の差の検定、ウィルコクソンの順位和検定（マン・ホイットニーのU検定）、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて同定される。

【0164】

当業者は、細胞成分/クラスターの組合せ当たりの細胞分量を含む他の測定基準も可能であることを理解するであろう。

【0165】

III.C.5. 後処理

異なって発現された細胞成分  $E_k$  のセットは、それ自体で有用な出力を表す。しかしながら、そのセットのサブセットを同定するために異なって発現された細胞成分のセットをさらに分析する（360）ことが有用であり得る。

20

【0166】

一実施形態において、異なって発現された細胞成分のセットは、転写因子データベース（例えば、図1のシグネチャストア140）に対してスクリーニングされて、セット中に存在する細胞成分に関連する転写因子のセットを同定する。例として、この情報は、ChIP-seqデータセットから得られる（どの転写因子が、細胞成分に位置合わせされるDNAのどの領域に結合するかについての情報）。

30

【0167】

特定の細胞について本明細書に記載されるデータセット122、例えば、元の入力データセット  $r$ （例えば、図1のデータセット122-1）又は異なって発現された細胞成分  $E_k$  のセット及び対応するデータセットは、いくつもの理由（例えば、技術的なノイズ、ドロップアウト、低い細胞分量など）で細胞分量を見落としていることがある。これらの及び任意のさらなる交絡因子を説明するために、単純モデルが、データセットに適合され得る。

【0168】

III.D. 細胞状態の移行に影響を与える摂動の予測

特定の細胞の移行を特徴付ける細胞成分の発現の差を、摂動への細胞の曝露によって引き起こされる細胞成分の発現の差にマッチさせることによって、特定の細胞状態の移行に影響を与える摂動が予測され得る。細胞の摂動は、1つ以上の化合物による細胞の任意の処理を含む。1つ以上の化合物は、例えば、小分子、生物製剤、タンパク質、小分子と組み合わされたタンパク質、ADC、siRNA若しくは干渉RNAなどの核酸、cDNA過剰発現野生型及び/又は突然変異型shRNA、cDNA過剰発現野生型及び/又は突然変異型ガイドRNA（例えば、Cas9システム又は他の細胞成分編集システム）、又は上記のいずれかの任意の組合せを含み得る。特定の細胞の移行について異なって発現された細胞成分は、摂動への細胞の曝露によって引き起こされる異なって発現された細胞成分と比較され得る。次に、特定の細胞の移行の細胞成分の発現の差にマッチする細胞成分の発現の差を引き起こす摂動は、特定の細胞の移行に影響を与えることが予測され得る。

40

50

## 【0169】

特定の細胞の移行を特徴付ける細胞成分の発現の差を、摂動への細胞の曝露によって引き起こされる細胞成分の発現の差にマッチさせることによって、特定の細胞の移行に影響を与える摂動を予測するために、第1に、特定の細胞の移行を特徴付ける最も異なって発現された細胞成分が同定される。ある実施形態において、これらの、異なって発現された細胞成分が、母平均の差の検定、ウィルコクソンの順位和検定（マン・ホイットニーのU検定）、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて同定される。代替的な実施形態において、任意の統計的方法が、特定の細胞の移行についての最も異なって発現された細胞成分を同定するのに使用され得る。細胞成分132の名称及び重要度スコア134の得られたランク付けされた表（又はリスト）は、「単一細胞の移行シグネチャー」とも呼ばれ得る（例えば、図1の単一細胞の移行シグネチャー142を含む）。各細胞成分132の重要度スコア134は、細胞成分の細胞成分の発現の変化と、元の細胞型と移行された細胞型との間の細胞型の変化との間の関連性を定量する。全体として、これらのスコア134は、元の細胞型（第1の細胞状態）と移行された細胞型（変化した細胞状態）との間の移行に関連する細胞成分の発現の差の全体的な尺度を形成する。

10

## 【0170】

同様に、摂動への細胞の曝露によって引き起こされる細胞成分の発現の差が、1つ以上の摂動について同定される。ある実施形態において、摂動への細胞の曝露によって引き起こされる細胞成分の発現の差を同定するために、摂動に曝された細胞における細胞成分の発現が、摂動に曝されていない対照細胞における細胞成分の発現又は関連していない摂動されたサンプルの平均と比較される（例えば、図3の後処理360）。ある実施形態において、この比較は、母平均の差の検定、ウィルコクソンの順位和検定（マン・ホイットニーのU検定）、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて行われる。代替的な実施形態において、任意の統計的方法が、比較を行うのに使用され得る。さらに他の代替的な実施形態において、摂動への細胞の曝露によって引き起こされる細胞成分の発現の差は、公知であり得、文献から同定され得る。細胞成分の名称及び重要度スコアの得られた同様のランク付けされた表（又はリスト）は、「摂動シグネチャー」と呼ばれる。

20

## 【0171】

ある実施形態において、技術的誤差、異なる実験アッセイ、並びに単一細胞の移行シグネチャー及び摂動シグネチャーの同定における他の変数による交絡を減少させるために、シグネチャーの一方又は両方が、特定の細胞成分の発現を促すことが知られているタンパク質である転写因子のみを含むようにフィルタリングされる。これらの転写因子は、例えば、文献から同定され得る。

30

## 【0172】

ある実施形態において、技術的誤差及び細胞の移行の曖昧さによる交絡をさらに減少させるために、シグネチャーの一方又は両方の最も異なって発現された細胞成分が、所与のp値及び/又は細胞成分の閾値数で切り捨てられる（truncated）（又はフィルタリング若しくはサブセット化される）。細胞の移行及び摂動曝露についての、異なって発現された細胞成分の得られた切り捨てられたセットは、順序付けられず、実施に応じて、10~25の細胞成分、又はそれ以上若しくはそれ以下を含み得る。

40

## 【0173】

シグネチャー（例えば、図1の単一細胞の移行シグネチャー142及び/又は摂動シグネチャー150）の一方又は両方の同定及び任意の処理の後、単一細胞の移行シグネチャー142についての、異なって発現された細胞成分は、摂動シグネチャー150についての、異なって発現された細胞成分と比較される。一実施形態において、比較を行うために、摂動の、異なって発現された細胞成分は、行列（例えば、図3の行列M、図1の細胞成分ベクトル130など）として表される。行列の各行は、単一の摂動に関連する。行列における各列は、異なって発現された細胞成分の1つに関連する。行列における各エントリ

50



は、特定の摂動について同定された異なって発現された細胞成分 1 3 2 についての重要度スコア 1 3 4 (例えば、p 値、t スコア) を含む。この行列は、単一細胞の移行シグネチャー 1 4 2 について同定された異なって発現された細胞成分のみを含むサブセットである。このフィルタリングは、前の段落に記載される方法を用いて (例えば、閾値 p 値、細胞成分の閾値数などによって) 達成され得る。

【 0 1 7 4 】

行列における各重要度スコア 1 3 4 は、個別のマッチングスコアで置き換えられる。各重要度スコアを、個別のマッチングスコアで置き換えるために、細胞の移行について有意に上方制御された細胞成分 1 3 2 及び細胞の移行について有意に下方制御された細胞成分が同定される。単一細胞の移行シグネチャー 1 4 2 によって同定された有意に上方制御された細胞成分のそれぞれについて、細胞成分がまた、その摂動についての摂動シグネチャー 1 5 0 について有意に上方制御される場合、その細胞成分 / 摂動の組合せについての行列における重要度スコアは、「 1 」の個別のマッチングスコアで置き換えられる。細胞成分が、単一細胞の移行シグネチャーと比べて摂動シグネチャーについて有意に下方制御される場合、その細胞成分 / 摂動の組合せについての行列における重要度スコアは、「 - 2 」の個別のマッチングスコアで置き換えられる。細胞成分が、摂動シグネチャーについて有意に上方制御又は下方制御されない場合、細胞成分 / 摂動の組合せについての行列における重要度スコアは、「 0 」の個別のマッチングスコアで置き換えられる。

10

【 0 1 7 5 】

逆に、単一細胞の移行シグネチャーにおいて同定された有意に下方制御された細胞成分のそれぞれについて、細胞成分がまた、摂動について有意に下方制御される場合、その細胞成分 / 摂動の組合せについての行列における重要度スコアは、「 - 1 」の個別のマッチングスコアで置き換えられる。細胞成分が、摂動について有意に上方制御される場合、その細胞成分 / 摂動の組合せについての行列における重要度スコアは、「 2 」の個別のマッチングスコアで置き換えられる。細胞成分が、摂動について有意に上方制御又は下方制御されない場合、その細胞成分 / 摂動の組合せについての行列における重要度スコアは、「 0 」の個別のマッチングスコアで置き換えられる。当業者は、これらの特定のスコア置き換えが、ある実施形態において他の数値で置き換えられ得ることを理解するであろう。

20

【 0 1 7 6 】

結果は、摂動の数によって示される行の数及び単一細胞の移行と異なる細胞成分によって示される列の数を有する行列並びに上述されるマッチングスコアを表すエントリである。

30

【 0 1 7 7 】

上述される個別のマッチングスコアによる、行列における重要度スコアの置き換えの後、行列の各行における個別のマッチングスコアが合計されて、各行についての合計されたマッチングスコアを生成する。次に、それぞれ摂動に対応する行列の行が、合計されたマッチングスコアが減少する順にランク付けされる。上位にランク付けされた行が、単一細胞の移行シグネチャーの同定された細胞の移行に関連する可能性が最も高い摂動に関連する。

【 0 1 7 8 】

ある実施形態において、行列における各行の合計されたマッチングスコアについて、偽細胞成分発見率の推定値が推定される。偽細胞成分発見率を推定するために、各細胞成分についての経験的な周辺発現頻度が計算され、経験的な周辺発現頻度が、それらの組合せにわたって各細胞成分について合計され、それにより、独立に分布した発現を仮定して、所与の数の細胞成分を偶然同定する確率を生成する (少なくとも、シグネチャーを生成するのに使用されるデータセットにおいて見られるのと同程度に稀であった発現がどの程度の可能性で観察されるか)。次に、その確率を用いて、偽細胞成分発見率を計算することができる。

40

【 0 1 7 9 】

特定の実施形態において、摂動の共変量が存在し得る。例えば、摂動が小分子である場

50

合、小分子の共変量は、小分子の特定の用量、小分子に曝された細胞が、小分子に曝された細胞の細胞成分を定量し、及び/又は(例えば、細胞株)を同定するために測定される時間を含み得る。ある実施形態において、その共変量の閾値量がまた、特定の細胞の移行に影響を与えることが予測される場合にのみ、摂動が特定の細胞の移行に影響を与えることが予測される。例えば、その共変量の少なくとも2つがまた、特定の細胞の移行に影響を与えることが予測される場合にのみ、摂動が、特定の細胞の移行に影響を与えることが予測され得る。

#### 【0180】

マッチングの代替的な方法が使用され得る。例えば、細胞成分は、ウェブインターフェイス(例えば、amp.pharm.mssm.edu/L1000CDS2/#/indexでのワールドワイドウェブにおける、L1000CDS2.Anultra-fastLINCSL1000CharacteristicDirectionSignatureSearchEngineなど)を用いてデータベースにマッチングされ得る。このマッチングの方法は、前の段落に記載されるマッチングの方法と同様に機能せず、後者は、はるかに高い感受性で結果を生じ、はるかに良好な規模であり、はるかに多いデータをカバーし(数万の代わりに数百万のサンプル)、有意な重複を考慮に入れ、有意な不一致を考慮に入れず、シグネチャーにおける意義のない情報を無視する。

10

#### 【0181】

特定の単一細胞状態の移行にマッチする摂動を発見することは、特定の単一細胞状態の移行についての高度に可変の細胞成分の発現及び摂動によって影響される高度に可変の細胞成分の発現のため困難であり得る。この問題を軽減するために、ある代替的な実施形態において、特定の軌道に沿って細胞状態の移行に影響を与える摂動のマッチング及びその後の同定は、トレーニングされたニューラルネットワークモデルによって行われ得る。

20

#### 【0182】

摂動が、特定の細胞状態の移行に影響を与える摂動である例が、第IV.E節において以下に示される上記の方法を用いて同定される。

#### 【0183】

III.E.摂動についての生物学的有用性を同定するための方法

ある実施形態において、開示される方法は、摂動についての生物学的有用性を同定するために使用される。これらの方法は、異なる状態又は表現型、例えば、病変及び正常表現型を有する細胞内に異なって存在することが示され得る任意の細胞成分(又は異なる細胞成分の組合せ)の測定を包含する。すなわち、細胞成分の存在、非存在、又は量は、細胞状態又は表現型に関連する。一実施形態において、本方法は、複数の細胞を摂動に曝す工程と;第1の細胞成分の発現の差のアッセイを行う工程であって、アッセイが、摂動への細胞の曝露の前及び後に複数の細胞から得られた第1の複数の単一細胞発現データセットにアクセスする工程を含み、データセットのそれぞれが、細胞成分のベクトル $r_i$ を含み、ベクトル中の各エントリが、複数の細胞成分の1つに関連し、各エントリの値が、細胞についての細胞成分の量を表す、工程と;統計技術を、第1の複数のデータセットに適用して、摂動への曝露に応答して異なって発現された細胞成分 $E_k$ のセットを生成する工程と;摂動への曝露に応答して異なって発現された細胞成分 $E_k$ のセットと、病変細胞表現型と正常細胞表現型との間に差に関連する異なって発現された細胞成分 $E_l$ のセットとの間の類似度を決定する工程とを含み、ここで、 $E_k$ と $E_l$ との間の有意な類似度が、病変細胞表現型と正常細胞表現型との間で細胞を移行させる摂動についての有用性を示す。

30

40

#### 【0184】

ある実施形態において、統計技術を適用する工程は、第1の複数のデータセット132に対して次元削減(例えば、図3の次元削減320)を行って、第1の行列Mを生成する工程であって、第1の行列Mが、第1の次元における行及び第2の次元における列を含み、行列Mの値が、第1及び第2の次元空間中のその点に位置する細胞成分の量から生成される値を含む、工程と;クラスタリングを行って、第1のクラスタのセット $C_j$ を生成

50

する工程であって、各クラスターが、第1の行列Mにおける行のサブセットに対応する複数の点及びそれらの対応する細胞応答状態を含む、工程と；第1のクラスターのセット $C_j$ を用いて、細胞についての摂動への曝露に反応して異なって発現された細胞成分 $E_k$ のセットを決定する工程とを含む。

【0185】

ある実施形態において、病変細胞表現型と正常細胞表現型との間の差に関連する異なって発現された細胞成分 $E_1$ のセットは、第2の細胞成分の発現の差アッセイを行うことによって決定され得、第2のアッセイは、正常細胞及び病変細胞などの異なる状態の複数の細胞から得られた第2の複数の単一細胞細胞成分の発現データセットにアクセスする工程であって、データセットのそれぞれが、細胞成分のベクトル $r_i$ を含み、ベクトル中の各エントリが、複数の細胞成分の1つに関連し、各エントリの値が、その細胞についてのその細胞成分の量を表す、工程と；統計技術を第2の複数のデータセットに適用する工程とを含む。

10

【0186】

ある実施形態において、統計技術を第2の複数のデータセットに適用する工程は、第2の複数のデータセットに対して次元削減を行って、第2の行列Mを生成する工程であって、第2の行列Mが、第1の次元における行及び第2の次元における列を含み、第2の行列Mの値が、第1及び第2の次元空間中のその点に位置する細胞成分の1つ以上の量から生成される値を含む、工程と；点の相対的類似度の近似による第2の行列Mを用いて多様体学習を行って、複数の行及び2つの列を含む第2の行列Nを生成する工程であって、各行が、細胞の1つに対応し、列のそれぞれが、2次元空間の2つの次元の1つに対応し、第2の行列Nの値が、データセットに基づいて、各細胞と他の各細胞との間の細胞表現型の相対的な差を示す、工程と；クラスタリングを行って、第2のクラスターのセット $C_j$ を生成する工程であって、各クラスターが、行列Nにおける行のサブセットに対応する複数の点及びそれらの対応する細胞応答状態を含む、工程と；第2のクラスターのセット $C_j$ を用いて、病変細胞表現型と正常細胞表現型との間の差を示す、細胞についての病変細胞表現型と正常細胞表現型との間の差に関連する異なって発現された細胞成分 $E_1$ のセットを決定する工程とを含む。

20

【0187】

ある実施形態において、摂動は、規制された臨床試験において得られた結果によって決定された許容されるヒト安全性プロファイルを有することが知られている。

30

【0188】

ある実施形態において、病変細胞表現型は、病変細胞と正常細胞との間の相違によって同定される。例えば、ある実施形態において、病変細胞表現型は、細胞の機能喪失、細胞の機能獲得、細胞の進行（例えば、分化状態への細胞の移行）、細胞の状態（例えば、分化状態へと移行する細胞の能力の欠如）、細胞の侵入（例えば、異常な位置における細胞の出現）、細胞の消失（例えば、細胞が通常存在する位置における細胞の非存在）、細胞の障害（例えば、細胞の中及び/又は周囲の構造的、形態的、及び/又は空間的变化）、細胞のネットワークの喪失（例えば、子孫細胞又は細胞の下流の細胞における通常の効果はなくす細胞の変化）、細胞のネットワークの獲得（例えば、細胞の下流の細胞の子孫細胞における新たな下流の効果を引き起こす細胞の変化）、過剰な細胞（例えば、細胞の過多）、細胞の不足（例えば、臨界閾値未満である細胞の密度、細胞内の細胞成分の比率及び/又は量の差、細胞内の移行の速度の差、又はそれらの任意の組合せによって同定され得る。

40

【0189】

ある実施形態において、病変細胞は、細胞株、生検サンプル細胞、及び初代培養細胞を含む。ある実施形態において、正常細胞は、初代培養細胞及び生検サンプル細胞を含む。ある実施形態において、細胞は、ヒト細胞である。

【0190】

ある実施形態において、本方法は、上記の方法を用いて同定される示される有用性に基

50

づいて、疾患を処置するのに有用な摂動を選択するのに使用される。ある実施形態において、本方法は、有効量の選択された摂動又は摂動誘導化合物から開発された製剤原料を対象に投与することによって、疾患に罹患している対象を治療する工程を含む。

#### 【0191】

##### 実施形態

実施形態1．複数の単一細胞細胞成分の発現データセットにアクセスする工程であって、各データセットが、同じ「前駆」細胞型から移行された複数の細胞の1つの細胞から得られ、各データセットが、細胞成分のベクトル  $r_i$  を含み、ベクトル中の各エントリが、複数の細胞成分の1つに関連し、各エントリの値が、細胞についての細胞成分の量を表す、工程と；データセットに対して次元削減を行って、行列  $M$  を生成する工程であって、行列  $M$  が、第1の次元における行及び第2の次元における列を含み、各行が、複数の細胞の1つの細胞に対応し、行列  $M$  の値が、第1及び第2の次元空間中のその点に位置する細胞成分の量から生成される値を含む、工程と；クラスタリングを行って、クラスタのセット  $C_j$  を生成する工程であって、各クラスタが、行列  $M$  中の行のサブセットに対応する複数の点及びそれらの対応する細胞を含む、工程と；クラスタのセット  $C_j$  を用いて、細胞についての、異なって発現された細胞成分  $E_k$  のセットを決定する工程とを含む方法。

10

#### 【0192】

実施形態2．点の相対的類似度の近似による行列  $M$  を用いて多様体学習を行って、複数の行及び2つの列を含む行列  $N$  を生成する工程をさらに含み、各行が、複数の細胞の1つに対応し、列のそれぞれが、2次元空間の2つの次元の1つに対応し、行列  $N$  の値が、データセットに基づいて、各細胞の他の各細胞に対する相対的な細胞型を示す、実施形態1に記載の方法。

20

#### 【0193】

実施形態3．単一細胞細胞成分の発現データセットが得られたとき、細胞が、様々な細胞型を有する不均一な細胞集団である、実施形態1～2のいずれか1つに記載の方法。

#### 【0194】

実施形態4．細胞が、「前駆」細胞型を有する細胞の実質的に均一な集団であり；単一細胞細胞成分の発現データセットが、細胞が「前駆」細胞型から移行された時点で、複数の点のそれぞれで得られ、複数のデータセットの異なるデータセットが、各細胞及び時点の組合せについて収集されるようになっている、実施形態1～2のいずれか1つに記載の方法。

30

#### 【0195】

実施形態5．複数の時点が、少なくとも3つの時点を含む、実施形態4に記載の方法。

#### 【0196】

実施形態6．複数の時点が、細胞のかなりの割合が「前駆」細胞型からまだ移行されていない「前駆」時点を含む、実施形態4～5のいずれか1つに記載の方法。

#### 【0197】

実施形態7．複数の時点が、細胞のかなりの割合が「前駆」細胞型から移行された移行時点を含む、実施形態4～6のいずれか1つに記載の方法。

40

#### 【0198】

実施形態8．複数の時点が、細胞のかなりの割合が「前駆」細胞型から少なくとも部分的に移行された少なくとも1つの中間時点を含む、実施形態4～7のいずれか1つに記載の方法。

#### 【0199】

実施形態9．複数の細胞成分が、核酸、タンパク質、脂質、炭水化物、ヌクレオチド、及びそれらの任意の組合せからなる群から選択される、実施形態1～8のいずれか1つに記載の方法。

#### 【0200】

実施形態10．核酸が、DNA及びRNAからなる群から選択される、実施形態9に記載

50

載の方法。

【0201】

実施形態11．RNAが、コーディング及びノンコーディングRNAからなる群から選択される、実施形態10に記載の方法。

【0202】

実施形態12．単一細胞細胞成分の発現データセットが、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスPOSEでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seq、及びそれらの任意の組合せ、並びに単一細胞細胞成分の発現データセットにおける活性化された経路を表す線形組合せなどの組合せを含む、上記のものの集合からなる群から選択される方法を用いて生成される、実施形態1～11のいずれか1つに記載の方法。

10

【0203】

実施形態13．次元削減を行う工程が、単一細胞細胞成分の発現データセットに対して主成分分析(PCA)を行って、行列Mを生成する工程を含む、実施形態1～12のいずれか1つに記載の方法。

【0204】

実施形態14．次元削減を行う工程が、単一細胞細胞成分の発現データセットに対して拡散マップを使用して、行列Mを生成する工程を含む、実施形態1～13のいずれか1つに記載の方法。

20

【0205】

実施形態15．次元削減を行う工程が、単一細胞細胞成分の発現データセットに対してニューラルネットワークオートエンコーダを使用して、行列Mを生成する工程を含む、実施形態1～14のいずれか1つに記載の方法。

【0206】

実施形態16．多様体学習を行う工程が、行列Mにおけるデータの幾何学的配列を推定して、行列Nを生成する工程を含む、実施形態2に記載の方法。

【0207】

実施形態17．多様体学習を行う工程が、局所線形埋め込みを行う工程を含む、実施形態16に記載の方法。

30

【0208】

実施形態18．多様体学習を行う工程が、局所線形アイソメトリックマッピング(ISOMAP)を行う工程を含む、実施形態16に記載の方法。

【0209】

実施形態19．多様体学習を行う工程が、t分布型確率的近傍埋め込み(t-SNE)を行う工程を含む、実施形態16に記載の方法。

【0210】

実施形態20．多様体学習を行う工程が、親和性に基づく軌道埋め込みのための熱拡散の可能性(PHATE)を行う工程を含む、実施形態16に記載の方法。

40

【0211】

実施形態21．多様体学習を行う工程が、一様多様体近似及び投影(UMAP)を行う工程を含む、実施形態16に記載の方法。

【0212】

実施形態22．多様体学習を行う工程が、力指向レイアウトを作成する工程を含む、実施形態16に記載の方法。

【0213】

実施形態23．力指向レイアウトが、Force Atlas 2アルゴリズムを用いて作成される、実施形態22に記載の方法。

【0214】

50

実施形態 24 . クラスタリングを行う工程が、各クラスター中の複数の点の組織化の予備知識を仮定しない、実施形態 1 ~ 23 のいずれか 1 つに記載の方法。

【0215】

実施形態 25 . クラスタリングを行う工程が、HDBSCANを行って、クラスターのセット  $C_j$  を生成する工程を含む、実施形態 1 ~ 24 のいずれか 1 つに記載の方法。

【0216】

実施形態 26 . クラスタリングを行う工程が、Louvainコミュニティ検出を行って、クラスターのセット  $C_j$  を生成する工程を含む、実施形態 1 ~ 25 のいずれか 1 つに記載の方法。

【0217】

実施形態 27 . クラスタリングを行う工程が、点に関連する単一細胞細胞成分の発現データセットがどの時点で収集されたかに基づいて、各点をクラスター  $C_j$  の 1 つに割り当てる工程を含む、実施形態 1 ~ 26 のいずれか 1 つに記載の方法。

【0218】

実施形態 28 . クラスタリングを行う工程が、点がクラスターからどのくらい末端にあるかの測定に基づいて、点をクラスターに割り当てる拡散経路アルゴリズムを用いて複数の点を分析する工程を含む、実施形態 1 ~ 27 のいずれか 1 つに記載の方法。

【0219】

実施形態 29 . 異なって発現された細胞成分  $E_k$  のセットを決定する工程が、各細胞成分について、クラスターの少なくとも 1 つについて、少なくとも 1 つのクラスター中の複数の点についての細胞成分の量を、少なくとも 1 つの他のクラスター中の複数の点についての細胞成分の量に対して比較する工程と；少なくとも 1 つのクラスター中の複数の点についての細胞成分の量が、少なくとも 1 つの他のクラスター中の複数の点についての細胞成分の量を超える閾値レベルであることに応答して、細胞成分を、異なって発現された細胞成分  $E_k$  のセットに加える工程とを含む、実施形態 1 ~ 28 のいずれか 1 つに記載の方法。

【0220】

実施形態 30 . 少なくとも 1 つのクラスターが、クラスター  $C_j$  の系統内 (on-lineage) のものを含み、系統内 (on-lineage) クラスターが、所望の細胞型を有すると同定可能な複数の点を含む、実施形態 29 に記載の方法。

【0221】

実施形態 31 . 少なくとも 1 つの他のクラスターが、クラスター  $C_j$  の系統外 (off-lineage) のものを含み、系統外 (off-lineage) クラスターが、所望されない細胞型を有すると同定可能な点を含む、実施形態 30 に記載の方法。

【0222】

実施形態 32 . 異なって発現された細胞成分  $E_k$  のセットを決定する工程が、各細胞成分について、クラスターの少なくとも 1 つについて、少なくとも 1 つのクラスター中の複数の点についての細胞成分の量と、少なくとも 1 つの他のクラスター中の複数の点についての細胞成分の量との間の距離計量を計算する工程と；距離計量が統計的に有意であることに応答して、細胞成分を、異なって発現された細胞成分  $E_k$  のセットに加える工程とを含む、実施形態 1 ~ 31 のいずれか 1 つに記載の方法。

【0223】

実施形態 33 . 異なって発現された細胞成分  $E_k$  のセットを、転写因子データベースに対してスクリーニングして、異なって発現された転写因子のセットを同定する工程をさらに含む、実施形態 1 ~ 32 のいずれか 1 つに記載の方法。

【0224】

実施形態 34 . 異なって発現された細胞成分  $E_k$  のセットに対して経験的モード分解を行って、データセットの擬似時間表現を生成する工程と；擬似時間表現に基づいて、異なって発現された転写因子のセットを同定する工程とをさらに含む、実施形態 33 に記載の方法。

10

20

30

40

50

## 【0225】

実施形態35．複数の単一細胞細胞成分の発現データセットにアクセスする工程であって、各データセットが、同じ「前駆」細胞型から移行された複数の細胞の1つの細胞から得られ、各データセットが、細胞成分のベクトル  $r_i$  を含み、ベクトル中の各エントリが、複数の細胞成分の1つに関連し、各エントリの値が、細胞についての細胞成分の量を表す、工程と；kNNアルゴリズムを用いて及び単一細胞細胞成分の発現データセットを用いて、kNNグラフを生成する工程と；クラスタリングを行って、クラスターのセット  $C_j$  を生成する工程であって、各クラスターが、複数の点を含み、各点が、複数の細胞の1つの細胞についての単一細胞細胞成分の発現データセットに対応する、工程と；クラスターのセット  $C_j$  を用いて、複数の細胞についての、異なって発現された細胞成分  $E_k$  のセットを決定する工程とを含む方法。

10

## 【0226】

実施形態36．異なって発現された細胞成分  $E_k$  のセットを決定する工程が、クラスター  $C_j$  中の複数の点間の距離計量を決定する工程を含む、実施形態35に記載の方法。

## 【0227】

実施形態37．第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーにアクセスする工程と；摂動に曝されていない非摂動細胞と摂動に曝された摂動細胞との間の細胞成分の発現の差の尺度を表す摂動シグネチャーにアクセスする工程と；単一細胞の移行シグネチャー及び摂動シグネチャーの比較に基づいて、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とを含む方法。

20

## 【0228】

実施形態38．単一細胞の移行シグネチャーにアクセスする工程が、第1の複数の単一細胞細胞成分の発現データセット（各第1のデータセットが、第1の細胞状態における第1の複数の細胞の1つの細胞から得られる）、及び第2の複数の単一細胞細胞成分の発現データセット（各第2のデータセットが、変化した細胞状態における第2の複数の細胞の1つの細胞から得られる）に基づいて、単一細胞の移行シグネチャーを決定する工程を含む、実施形態36に記載の方法。

## 【0229】

実施形態39．第1及び第2の複数の単一細胞細胞成分の発現データセットの各データセットが、細胞成分のベクトル  $r_i$  を含み、ベクトル中の各エントリが、複数の細胞成分の1つに関連し、各エントリの値が、細胞についての細胞成分の量を表す、実施形態38に記載の方法。

30

## 【0230】

実施形態40．第1及び第2の複数の単一細胞細胞成分の発現データセットを取得する工程をさらに含み、取得する工程が、第1及び第2の複数の単一細胞細胞成分の発現データセットに対して次元削減を行って、行列  $M$  を生成する工程であって、行列  $M$  が、第1の次元における行及び第2の次元における列を含み、各行が、複数の細胞の1つの細胞に対応し、行列  $M$  の値が、第1及び第2の次元空間中のその点に位置する細胞成分の量から生成される値を含む、工程と；クラスタリングを行って、クラスターのセット  $C_j$  を生成する工程であって、各クラスターが、行列  $M$  中の行のサブセットに対応する複数の点及びそれらの対応する細胞を含む、工程と；クラスターのセット  $C_j$  の第1のクラスターから第1の複数の細胞を同定する工程と；クラスターのセット  $C_j$  の第2のクラスターから第2の複数の細胞を同定する工程と；第1の複数の細胞から第1の複数の単一細胞細胞成分の発現データセットを取得する工程と；第2の複数の細胞から第2の複数の単一細胞細胞成分の発現データセットを取得する工程とを含む、実施形態38～39のいずれか1つに記載の方法。

40

## 【0231】

実施形態41．点の相対的類似度の近似による行列  $M$  を用いて多様体学習を行って、複数の行及び2つの列を含む行列  $N$  を生成する工程をさらに含み、各行が、第1及び第2の

50

複数の細胞の1つの細胞に対応し、列のそれぞれが、2次元空間の2つの次元の1つに対応し、行列Nの値が、第1及び第2の複数の単一細胞細胞成分の発現データセットに基づいて、各細胞の他の各細胞に対する相対的な細胞状態を示す、実施形態40に記載の方法。

【0232】

実施形態42．工程が、実施形態1～34のいずれか1つにしたがって行われる、実施形態40～41のいずれか1つに記載の方法。

【0233】

実施形態43．摂動シグネチャーにアクセスする工程が、摂動に曝されていない非摂動細胞の複数の非摂動単一細胞細胞成分の発現データセット及び摂動に曝された摂動細胞の複数の摂動単一細胞細胞成分の発現データセットに基づいて、摂動シグネチャーを決定する工程を含む、実施形態37～42のいずれか1つに記載の方法。

10

【0234】

実施形態44．非摂動細胞が、摂動細胞の摂動に曝されていない対照細胞であるか、又は非摂動細胞が、摂動に曝された非関連摂動細胞の平均である、実施形態37～43のいずれか1つに記載の方法。

【0235】

実施形態45．転写因子である細胞成分を含む単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングする工程をさらに含む、実施形態37～44のいずれか1つに記載の方法。

20

【0236】

実施形態46．第1の複数の単一細胞細胞成分の発現データセット及び第2の複数の単一細胞細胞成分の発現データセットに基づいて、単一細胞の移行シグネチャーを決定する工程が、母平均の差の検定、ウィルコクソンの順位和検定(マン・ホイットニーのU検定)、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて、第1及び第2の複数の単一細胞細胞成分の発現データセットの間の細胞分量の差を決定する工程を含む、実施形態38～42のいずれか1つに記載の方法。

【0237】

実施形態47．複数の非摂動及び摂動単一細胞細胞成分の発現データセットに基づいて、摂動シグネチャーを決定する工程が、母平均の差の検定、ウィルコクソンの順位和検定(マン・ホイットニーのU検定)、t検定、ロジスティック回帰、及び一般化線形モデルのうちの1つを用いて、複数の非摂動及び摂動単一細胞細胞成分の発現データセットの間の細胞分量の差を決定する工程を含む、実施形態43に記載の方法。

30

【0238】

実施形態48．単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングして、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させる工程をさらに含む、実施形態37～47のいずれか1つに記載の方法。

【0239】

実施形態49．単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングする工程が、閾値p値にしたがって、又は細胞成分の閾値数にしたがって、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させる工程を含む、実施形態48に記載の方法。

40

【0240】

実施形態50．摂動シグネチャーが、複数の細胞成分を含み、各細胞成分が、細胞成分の量の変化と、非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する重要度スコアに関連し、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程が、各細胞成分についての重要度スコアを、細胞成分についてのマッチングスコアで置き換える工程と；複数の細胞成分についてのマッチングスコアを組み合わせて、摂動についてのマッチングスコアを生成する工程と；摂動についてのマッチングスコアに基づいて、摂動が、第1の細胞状態と変化した細胞状態との

50



間の細胞の移行に関連しているかどうかを決定する工程とを含む、実施形態 37 ~ 49 のいずれか 1 つに記載の方法。

【0241】

実施形態 51 . マッチングスコアが、別個の又は連続したスコアを含む、実施形態 50 に記載の方法。

【0242】

実施形態 52 . 各重要度スコアを置き換える工程が、単一細胞の移行シグネチャーからの細胞成分量及びその細胞成分についての摂動シグネチャーからの細胞成分量が両方とも上方制御される場合、重要度スコアを第 1 のスコアで置き換える工程と；単一細胞の移行シグネチャーからの細胞成分量が上方制御され、その細胞成分についての摂動シグネチャーからの細胞成分量が下方制御される場合、重要度スコアを第 2 のスコアで置き換える工程と；その細胞成分についての摂動シグネチャーからの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアを第 3 のスコアで置き換える工程とを含む、実施形態 50 ~ 51 のいずれか 1 つに記載の方法。

10

【0243】

実施形態 53 . 重要度スコアを置き換える工程が、単一細胞の移行シグネチャーからの細胞成分量及びその細胞成分についての摂動シグネチャーからの細胞成分量が両方とも下方制御される場合、重要度スコアを第 1 のスコアで置き換える工程と；単一細胞の移行シグネチャーからの細胞成分量が下方制御され、その細胞成分についての摂動シグネチャーからの細胞成分量が上方制御される場合、重要度スコアを第 2 のスコアで置き換える工程と；その細胞成分についての摂動シグネチャーからの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアを第 3 のスコアで置き換える工程とを含む、実施形態 50 ~ 51 のいずれか 1 つに記載の方法。

20

【0244】

実施形態 54 . 摂動シグネチャーが、複数の細胞成分を含み、各細胞成分が、細胞成分の量の変化と、非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する重要度スコアに関連し、摂動が、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程が、複数の細胞成分についての重要度スコアを組み合わせて、摂動についての重要度スコアを生成する工程と；摂動についての重要度スコアに基づいて、摂動が、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とを含む、実施形態 37 ~ 49 のいずれか 1 つに記載の方法。

30

【0245】

実施形態 55 . 複数の細胞成分の各細胞成分についての経験的な周辺発現頻度を計算し；複数の細胞成分についての経験的な周辺発現頻度を、それらの組合せにわたって合計して、独立に分布した発現を偶然呈するいくつかの細胞成分を同定する確率を生成し；確率に基づいて、摂動についてのマッチングスコアについての偽細胞成分発見率を推定することによって、摂動についてのマッチングスコアについての偽細胞成分発見率を推定する工程をさらに含む、実施形態 50 ~ 53 のいずれか 1 つに記載の方法。

40

【0246】

実施形態 56 . 摂動が、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程が、摂動の共変量の閾値量が、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連していることを決定する工程と；決定に応答して、摂動が、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連していることを決定する工程とを含む、実施形態 37 ~ 55 に記載の方法。

【0247】

実施形態 57 . 摂動が、小分子への細胞の曝露を含み、摂動の 1 つ以上の共変量が、小分子の特定の用量、摂動細胞が小分子に曝される時間に対する、非摂動細胞と摂動細胞との間の細胞成分の発現の差が測定される時間、及び摂動細胞の細胞株を含む、実施形態 56 に記載の方法。

50

## 【0248】

実施形態58．細胞成分が、遺伝子を含む、実施形態37～57のいずれか1つに記載の方法。

## 【0249】

実施形態59．単一細胞細胞成分の発現データセットが、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスPOSEでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CyTOF/SCoP、E-MS/Abseq、miRNA-seq、CITE-seq、及びそれらの任意の組合せ又は集合からなる群から選択される方法を用いて生成される、実施形態37～58に記載の方法。

10

## 【0250】

実施形態60．単一細胞の移行シグネチャー及び摂動シグネチャーの少なくとも1つが、データベースから得られる、実施形態37～59に記載の方法。

## 【0251】

実施形態61．摂動シグネチャーが、複数の摂動についての複数の摂動シグネチャーを含むデータベースから得られ、方法が、データベース中の複数の摂動の各摂動について：データベースから摂動についての摂動シグネチャーにアクセスする工程と；単一細胞の移行シグネチャー及び摂動シグネチャーの比較に基づいて、摂動が、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とをさらに含む、実施形態60に記載の方法。

20

## 【0252】

実施形態62．複数の摂動細胞についての複数の摂動シグネチャーにアクセスする工程と；複数の摂動シグネチャーのそれぞれについて、決定工程を行い、それによって、変化した細胞状態を促進する摂動についてスクリーニングする工程とをさらに含む、実施形態37～61のいずれか1つに記載の方法。

## 【0253】

実施形態63．複数の摂動シグネチャーにアクセスする工程が、細胞を複数の摂動に曝して、複数の摂動細胞を生成する工程と；複数の摂動細胞から複数の細胞分量を測定する工程とを含む、実施形態62に記載の方法。

## 【0254】

実施形態64．変化した細胞状態を促進する摂動を同定する工程をさらに含む、実施形態37～63のいずれか1つに記載の方法。

30

## 【0255】

実施形態65．変化した細胞状態を促進する工程が、第1の細胞状態を含む細胞の集団中で、第1の細胞状態から変化した細胞状態への移行を促進する工程を含む、実施形態64に記載の方法。

## 【0256】

実施形態66．変化した細胞状態を促進する工程が、第1の細胞状態を含む細胞の集団中で、第1の状態、又は、任意に、変化した細胞状態以外の状態における細胞の数に対する、別の状態における細胞の数の比率を増加させる工程を含む、実施形態64に記載の方法。

40

## 【0257】

実施形態67．変化した細胞状態を促進する工程が、第1の細胞状態を含む細胞の集団中で、変化した細胞状態における細胞の絶対数を増加させる工程を含む、実施形態64に記載の方法。

## 【0258】

実施形態68．変化した細胞状態を促進する工程が、第1の細胞状態を含む細胞の集団中で、第1の細胞状態又は、任意に、変化した細胞状態以外の状態における細胞の絶対数を減少させる工程を含む、実施形態64に記載の方法。

## 【0259】

50

実施形態 69 . 細胞の移行シグネチャー及び摂動シグネチャーが、異なるタイプの細胞成分を用いて生成される、実施形態 37 ~ 68 のいずれか 1 つに記載の方法。

【0260】

実施形態 70 . 細胞の移行シグネチャー及び摂動シグネチャーが、同じタイプの細胞成分を用いて生成される、実施形態 37 ~ 68 のいずれか 1 つに記載の方法。

【0261】

実施形態 71 . 第 1 の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーにアクセスする工程と；複数の摂動シグネチャーにアクセスする工程であって、各摂動シグネチャーが、摂動に関連し、摂動に曝されていない非摂動細胞と摂動に曝された摂動細胞との間の細胞成分の発現の差の尺度を表す、工程と；単一細胞の移行シグネチャー及び複数の摂動シグネチャーの比較に基づいて、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連する摂動のサブセットを決定する工程とを含む方法。

10

【0262】

実施形態 72 . 各摂動シグネチャーが、複数の細胞成分を含み、各細胞成分が、細胞成分の量の変化と、非摂動細胞と摂動細胞との間の細胞状態の変化との間の関連性を定量する重要度スコアに関連し、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連する摂動のサブセットを決定する工程が、各摂動シグネチャーについて：各細胞成分についての重要度スコアを、細胞成分についてのマッチングスコアで置き換える工程と；複数の細胞成分についてのマッチングスコアを組み合わせ、摂動についてのマッチングスコアを生成する工程と；摂動を、それらのマッチングスコアにしたがってランク付けする工程と；摂動のランク付けされたリストに基づいて、摂動のサブセットを選択する工程とを含む、実施形態 71 に記載の方法。

20

【0263】

実施形態 73 . 命令が符号化された非一時的なコンピュータ可読記録媒体を含むコンピュータプログラム製品であって、命令が、プロセッサによって実行されるとき、プロセッサに、実施形態 37 ~ 72 のいずれか 1 つに記載の方法を実行させる、コンピュータプログラム製品。

【0264】

実施形態 74 . 命令が符号化された非一時的なコンピュータ可読記録媒体を含むシステムであって、命令が、プロセッサによって実行されるとき、プロセッサに、実施形態 37 ~ 72 のいずれか 1 つに記載の方法を実行させる、システム。

30

【0265】

実施形態 75 . ニューロン及び / 又は前駆細胞を促進するための方法であって、線維芽細胞を含む細胞の開始集団を、線維芽細胞を含む細胞の開始集団の、前駆細胞及び / 又はニューロンへの移行を促進する摂動シグネチャーを有する摂動に曝す工程を含み、ここで、摂動シグネチャーが、Brn2、Ascl1、Myt1、Zfp941、Taf5B、St18、Zksan16、Camta1、及びArnt2の1つ以上の増大した活性、及び / 又はAscl1、Atf3、Rorc、Scx、Satb1、Elf3、及びFosの1つ以上の低下した活性である、方法。

40

【0266】

実施形態 76 . ニューロン及び / 又は前駆細胞が、ニューロン及び / 又は前駆細胞の絶対数を増加させること、線維芽細胞の絶対数を減少させること、線維芽細胞の、ニューロン及び / 又は前駆細胞への移行を促進すること、ニューロン又は前駆細胞の寿命を促進すること、線維芽細胞の寿命を減少させること、又は線維芽細胞に対するニューロン及び / 又は前駆細胞の比率を増加させることのうちの1つ以上によって促進される、実施形態 75 に記載の方法。

【0267】

実施形態 77 . 摂動が、フォルスコリン、PP1、PP2、及びトリコスタチン A を含まない、実施形態 75 に記載の方法。

50

## 【0268】

実施形態78．ニューロン及び／又は前駆細胞の量を増加させる方法であって、線維芽細胞を含む細胞の集団を医薬組成物に曝す工程を含み、医薬組成物が、線維芽細胞を含む細胞の集団の、ニューロンへの移行を促進する摂動シグネチャーを有し、ここで、医薬組成物が、フォルスコリン、PP1、PP2、トリコスタチンA、BRD-K38615104、ゲルダナマイシン、マヌマイシンA、ミトキサントロン、クルクミン、アルボシジブ、ポリノスタット、KI20227、又は上記の組合せ、例えば、上記の2、3、4、5つ、若しくはそれ以上の組合せを含む、方法。

## 【0269】

実施形態79．医薬組成物が、フォルスコリン、PP1、PP2、及びトリコスタチンAを含まない、実施形態78に記載の方法。

10

## 【0270】

実施形態80．ニューロン及び／又は前駆細胞を促進するための医薬組成物であって、フォルスコリン、PP1、PP2、トリコスタチンA、BRD-K38615104、ゲルダナマイシン、マヌマイシンA、ミトキサントロン、クルクミン、アルボシジブ、ポリノスタット、KI20227、又は上記の組合せからなる群から選択される摂動、及び薬学的に許容される賦形剤を含む、医薬組成物。

## 【0271】

実施形態81．摂動が、フォルスコリン、PP1、PP2、及びトリコスタチンAを含まない、実施形態80に記載の医薬組成物。

20

## 【0272】

実施形態82．実施形態80又は81に記載の医薬組成物を含む単位剤形。

## 【0273】

実施形態83．線維芽細胞を含む細胞の開始集団の、ニューロン及び／又は前駆細胞への移行を促進するための候補の摂動を同定する方法であって、線維芽細胞を含む細胞の開始集団を摂動に曝す工程と；摂動についての摂動シグネチャーを同定する工程であって、摂動シグネチャーが、1つ以上の細胞成分及び各細胞成分に関連する重要度スコアを含み、各細胞成分の重要度スコアが、摂動への細胞の集団の曝露の後、細胞成分の発現の変化と、線維芽細胞からニューロン及び／又は前駆細胞への、細胞の集団の細胞状態の変化との間の関連性を定量する、工程と；摂動シグネチャーに基づいて、線維芽細胞を含む細胞の集団の、ニューロン及び／又は前駆細胞への移行を促進するための候補の摂動として摂動を同定する工程とを含み、ここで、摂動シグネチャーが、Brn2、Ascl1、Myt1、Zfp941、Taf5B、St18、Zkscan16、Camta1、及びArnt2の1つ以上の増大した活性、及び／又はAscl1、Atf3、Rorc、Scx、Satb1、Elf3、及びFosの1つ以上の低下した活性である、方法。

30

## 【実施例】

## 【0274】

IV．実施例0、1、2、及び3 - ニューロン及び筋細胞へと分化するマウス胚性線維芽細胞における因果関係の特定及び細胞の運命の制御

以下の実施例は、上記の第II及びIII節において紹介される方法を実証する。さらに詳細には、実施例は、細胞状態の移行の軌道に影響を与えることが知られている遺伝子及び／又は摂動を正確に同定する第II及びIII節の方法の能力を実証する。さらに、以下に説明される実施例は、細胞状態の移行の軌道を制御するのに使用され得る新規な生物学的洞察を生成する第II及びIII節の方法の能力を実証する。具体的には、実施例は、以前に知られていない、細胞状態の移行に影響を与える因子（例えば、遺伝子及び摂動）を同定する第II及びIII節の方法の能力を実証する。

40

## 【0275】

以下に説明される実施例は、第II及びIII節の方法を、公表されているデータ及びインビトロ実験データの組合せに適用して、細胞状態の移行の軌道に影響を与えるいくつかの公知の因子及び以前に知られていない因子（例えば、遺伝子及び摂動）を実証する。

50

公表されているデータ及びインビトロ実験データの組合せへの第 I I 及び I I I 節の方法のこの適用の結果が、図 4 B - 5 A 及び 7 A - 9 に示される。

【 0 2 7 6 】

これらの結果のいくつかはまた、インビトロ実験データのみを用いて実証された。このインビトロでの実証の結果が、図 6 に示される。インビトロ実験データは、以下の第 I V . A 節に記載されるプロトコルにしたがって、細胞を成長させ、測定することによって得られた。

【 0 2 7 7 】

I V . A . 実施例 0 : インビトロ細胞処理及びデータセット収集

この節は、上述されるインビトロ実験のためのプロトコルを説明する。このインビトロ実験からのデータを、公表されているデータとともにプールして、図 4 B - 5 A 及び 7 A - 9 を生成し、独立して使用して、図 6 を生成した。

【 0 2 7 8 】

この節は、第 I I 節に記載される一般化されたプロトコルを、ニューロン又は筋細胞へと分化するマウス胚性線維芽細胞 ( M E F ) を評価する特定の実施例に適用する。この特定の実施例において、ニューロンは、系統内 ( o n - l i n e a g e ) の細胞であり、筋細胞は、系統外 ( o f f - l i n e a g e ) の細胞であり、M E F は、「前駆」細胞であった。プロトコルは、遺伝子 A s c l 1 のレンチウイルス過剰発現及び摂動仲介を含むさらなる工程も含んでいた。

【 0 2 7 9 】

M E F 培地は、ダルベッコ改変イーグル培地 ( D M E M ) 中 1 0 % のウシ胎仔血清 ( F B S ) 、 1 x G l u t a m a x 、 1 x 非必須アミノ酸、P e n / s t r e p 、及び  $\beta$ -メルカプトエタノールであった。神経細胞用培地 ( n e u r o n a l m e d i a ) は、D M E M / F 1 2 、 N 2 、 B 2 7 、 1 x G l u t a m a x 、及びインスリン 2 5  $\mu$  g / m l であった。

【 0 2 8 0 】

後続のプロトコルが、以下に列挙される：

0 日目：1 0 0 万個の M E F 細胞を、M E F 培地中で 1 0 c m プレート中に解凍する。

1 日目：2 4 ウェルプレート中に 2 0 K / ウェルで播種する。

該当する場合、A s c l 1 ウイルスに同時にスピン感染させる ( s p i n i n f e c t ) ( 感染多重度 ( M O I ) 8 ) 。M E F 培地 ( 2 5 0  $\mu$  l / ウェル ) 及びポリブレン 8  $\mu$  g / m l の存在下で、3 2 で 1 時間にわたって 2 0 0 0 r p m で回転させる。単一細胞リボ核酸 ( R N A ) シーケンシング ( s c R N A - s e q ) を行って、各細胞についての d 2 データセットを得る。

2 日目：ウイルス実験のためにポリブレン ( M E F 培地 ) を洗い落とすために培地を交換する。

摂動実験のために、小分子 ( ジメチルスルホキシド ( D M S O ) 又はエタノール中で再懸濁された ) を加える。

3 日目：培地を神経細胞用培地に交換する。

摂動実験のために、分子 ( D M S O 又はエタノール中で再懸濁された ) を加える。

5 日目：半分培地交換 ( 該当する場合、小分子を加える )

8 日目：半分培地交換 ( 該当する場合、小分子を加える )

9 日目：半分培地交換 ( 該当する場合、小分子を加える )

1 1 日目：半分培地交換 ( 該当する場合、小分子を加える )

1 3 日目：半分培地交換 ( 該当する場合、小分子を加える )

1 5 日目：プレートを固定し、M a p 2 及び T u j 1 抗体で染色する。各ウェルを走査することによって、M o l e c u l a r D e v i c e s H C I I X M 4 又は他のハイコンテンツイメージング顕微鏡においてイメージングする。ウェルごとの M a p 2 / T u j 1 陽性ニューロンの数を定量する。

【 0 2 8 1 】

#### IV . B . 実施例 1、2、及び 3 : インビトロ細胞処理及びデータセット収集

図 4 A は、一実施形態に係る、ある期間にわたって誘導された細胞状態の移行の軌道を辿るタイムラインを示す。より具体的には、図 4 A は、0 日目から 22 日目までの 23 日の期間にわたる MEF の誘導された移行の軌道を辿るタイムラインを示す。

##### 【0282】

図 4 A に示されるように、23 日の期間の 0 日目に、MEF を入手した。代替的な実施形態において、任意の単一細胞の移行の軌道が、同様の方法にしたがって調べられ得る。例えば、代替的な実施例において、マウス胚性血球母細胞の移行の軌道が、同様の方法にしたがって調べられ得る。

##### 【0283】

23 日の期間の 0 日目に、集団 MEF の各 MEF に、適切な転写因子を形質導入した。図 4 A に示されるように、Ascl1 又は Brn2、Ascl1、及び Myt1l (まとめて BAM 転写因子として知られている) のみが、MEF において過剰発現された。具体的には、図 6 を生成する上記の第 IV . A 節のプロトコルを使用したインビトロ実験において、Ascl1 のみが、MEF において過剰発現された。逆に、図 4 B - 5 A 及び 7 A - 9 を生成するために第 I I 及び I I I 節の方法を適用した公表されているデータを生成するために、Brn2、Ascl1、及び Myt1l のそれぞれが、MEF において過剰発現された。

##### 【0284】

本明細書に開示される実施形態において、Ascl1 転写因子の発現を、レンチウイルス送達の後、Ascl1 の誘導性発現によって強制した。代替的な実施形態において、1 つ以上の転写因子の発現が、任意の代替的な手段によって強制され得る。例えば、代替的な実施形態において、1 つ以上の転写因子の発現が、トランスポゾン、mRNA 送達、又は別のタイプのウイルス送達によって強制され得る。

##### 【0285】

BAM 転写因子の 1 つ以上の強制発現が、強制された MEF の 1 つ以上に、マウス「前駆」細胞、マウスニューロン、及び / 又はマウス筋細胞へとより一般的に移行させることが知られている。具体的には、文献において公知であるように、Ascl1 プライミングが、MEF を、マウス「前駆」細胞へと移行するように誘導し、Ascl1 のみの発現が、マウス「前駆」細胞を、マウスニューロン及びマウス筋細胞へと移行するように誘導し、Brn2 及び Myt1l の発現が、マウス「前駆」細胞を、マウスニューロンへと移行するように誘導する。しかしながら、BAM 転写因子の 1 つ以上による細胞状態の移行のこの誘導は、100% の効率で起こるわけではない。具体的には、文献において公知であるように、BAM 転写因子は、20% の効率でマウスニューロンへの MEF の移行を誘導する。言い換えると、BAM 転写因子の 1 つ以上の発現にもかかわらず、一部の細胞は、予想どおりに移行しないことがある。ある実施形態において、この移行失敗は、再プログラミング失敗として知られている。

##### 【0286】

BAM 転写因子の 1 つ以上が強制的に発現されたマウス細胞を、23 日の期間にわたって監視した。より具体的には、Ascl1 の発現が強制されたマウス細胞については、集団中のマウス細胞の各単一マウス細胞についての単一細胞 RNA - シーケンシング (scRNA - seq) 測定が、23 日の期間中の 2、5、及び 22 日目に得られた。或いは、BAM 因子の全ての発現が強制されたマウス細胞については、集団中のマウス細胞の各単一マウス細胞についての scRNA - seq 測定が、23 日の期間中の 22 日目にのみ得られた。

##### 【0287】

代替的な実施形態において、RNA - シーケンシング測定が、任意の頻度で、任意の数の時点で取られ得る。より具体的には、細胞状態の移行軌道を正確に捕捉するために、RNA - シーケンシング測定が取られる時点は、理想的には、一般に、1 つ以上の移行軌道が分岐する時点に対応する。特定の日における単一細胞についての RNA - シーケンシ

10

20

30

40

50

グ測定は、その特定の日における単一細胞中での mRNA 発現の定量を含む。言い換えると、特定の日における単一細胞についての RNA - シーケンシング測定は、その特定の日における単一細胞中での各 mRNA 転写物の計数を含む。さらに、各 mRNA 転写物が、特定の遺伝子に関連するため、特定の日における単一細胞についての RNA - シーケンシング測定は、その特定の日における単一細胞中での遺伝子発現の定量を含む。しかしながら、実際には、細胞は、多くの場合、細胞状態の移行のそれらの状態が完全に均一であるわけではないため、所与の日における細胞状態の移行測定は、細胞状態の移行の様々な段階での細胞の分布を捕捉することが予測される。

#### 【0288】

Asc11 が MEF において過剰発現されたインビトロプロトコルを用いて、図 6 に示され、以下に詳細に記載される実証実験を行った。さらに、Asc11 が MEF において過剰発現されたインビトロプロトコルから得られた遺伝子発現測定を、BAM 因子の全てが過剰発現された MEF からの公表されている遺伝子発現測定とともにプールした。次に、インビトロ及び公表されているデータのこのプールを用いて、図 4B - 5A 及び 7A - 9 に示されるデータを生成した。上述されるように、これらの図は、細胞状態の移行に影響を与える遺伝子を正確に同定する第 I I 及び I I I 節の方法の能力を実証するだけでなく、細胞状態の移行の軌道、ひいては細胞の運命を制御するのに有用な新規な生物学的洞察を生成する第 I I 及び I I I 節に記載される方法の能力も実証する役割を果たす。

#### 【0289】

I V . C . 実施例 1 : 文献で同定された移行を実証する次元削減

上述されるように、Asc11 のみが過剰発現された MEF から 2、5、及び 22 日目に得られた遺伝子発現測定を、BAM 因子の全てが過剰発現された MEF から 22 日目に取られた公表されている遺伝子発現測定とともにプールした。第 I I 節において上述される方法を用いて、細胞における遺伝子発現を測定した日のそれぞれについて、細胞のそれぞれについての遺伝子発現測定を用いて、転写物のベクトル  $r_i$  のデータセットを生成した。転写物の各ベクトル  $r_i$  は、転写物のベクトル  $r_i$  に含まれる遺伝子発現測定が得られた特定の日における特定の細胞に関連していた。転写物のベクトル  $r_i$  における各転写物は、細胞のゲノム中の特定の遺伝子に関連しており、転写物のベクトル  $r_i$  における各エントリの値は、転写物のベクトル  $r_i$  に関連していた特定の日における転写物のシーケンシング深さ（転写物数）を表した。

#### 【0290】

第 I I I . C . 節に関して上述されるように、測定日のそれぞれにおいて細胞のそれぞれについて遺伝子発現測定を符号化したデータセットに対して、次元削減を行った。この実施例において、主成分分析（PCA）を用いて、次元削減を行い、次元削減された行列 M を生成した。

#### 【0291】

次に、多様体学習を、行列 M に対して行って、さらなる次元削減された行列 N を生成した。この実施例において、力指向レイアウトアルゴリズムを用いて、行列 N を生成した。行列 N が、補足的な表 1 に示される。行列 N はまた、図 4B に示される力指向レイアウト多様体としてプロットされる。図 4B の多様体中のプロットされたデータは、補足的な表 1 中の行列 N データに対応する。行列 N は、主に視覚化のために使用され、ある実施形態において、生成される必要がないことに留意されたい。言い換えると、ある実施形態において、多様体学習は、行列 M に対して行われない。

#### 【0292】

上述されるように、多様体中の各点は、行列 N の行のものに関連し、これは、遺伝子発現を細胞について測定した 4 日のうちの特定の日における細胞の特定の細胞に関連する。さらに、各点は、特定の日における特定の細胞について測定された遺伝子転写物数のデータセットに関連する。図 4B の多様体を解釈する際、多様体の行における次元  $x$  及び次元  $y$  の値が、その行に関連する日における細胞についての遺伝子転写物数に基づくため、多様体における点の位置決めは、他の点に対する、点に関連する日における細胞、及びひい

10

20

30

40

50

ては多様体中の、他の日における他の細胞についての遺伝子転写物数を反映している。結果として、多様体を視覚化することにより、23日の期間にわたる細胞についての様々な遺伝子の変化する遺伝子転写物数の観察を可能にする。

#### 【0293】

図4Bに示される多様体において、全ての点が、同一の色で同一の形状によって表される。したがって、図4Bの多様体において、点によって提供される識別可能な情報のみが、多様体におけるその位置(x, y)である。しかしながら、遺伝子ごとの転写物数及び遺伝子転写物数が各点について得られた特定の日は、図4Bにおいて識別可能でない。以下にさらに詳細に記載されるように、図5Aの多様体中の点の形状は、部分的に、遺伝子転写物数が各点について得られた日、及びひいては各細胞を示すように変化される。同様に、図5Bの多様体における点の陰影は、各点について、及びひいては各測定日における各細胞についての遺伝子ごとの遺伝子転写物数を示すように変化される。

10

#### 【0294】

図5Aは、一実施形態に係る、図4Bの多様体を示す。図5Aに示される多様体の実施形態において、多様体中の各点が、点に関連する細胞の転写因子発現が測定された日、及び細胞が移行過程にある定性的な段階で標識される。例えば、図5Aの多様体中の四角形状で標識された点は、点が、5日目に、早期誘導神経(iN)細胞として定性的に特徴付けられた細胞に関連することを示す。

#### 【0295】

多様体中の点のそれぞれを、点に関連する細胞についての遺伝子発現が測定された日及び細胞の移行の定性的段階で標識することによって、移行の軌道が同定され得る。例えば、移行の2つの異なる軌道が、図5A中の多様体の基礎をなす矢印によって示される。1つの同定された軌道は、MEF細胞の、マウスニューロンへの移行の軌道を描く。図5A中の別の同定された軌道は、MEF細胞の、マウス筋細胞への移行の軌道を描く。

20

#### 【0296】

移行の軌道に沿って異なる段階で点(例えば、細胞)の間の遺伝子発現の差を同定することによって、特定の軌道に沿った細胞の移行に寄与する遺伝子が同定され得る。しかしながら、おそらく、より重要なことには、移行の2つ以上の軌道が分岐する接続において点(例えば、細胞)の間の遺伝子発現の差を同定することによって、この移行軌道の分岐に寄与する遺伝子が同定され得る。次に、これらの同定された遺伝子は、移行の特定の軌道及び/又は段階に関連することが予測され得る。例えば、遺伝子Aの増加したレベルの発現が、5日目の早期筋細胞として標識された細胞と比べて、5日目の早期iN細胞として標識された細胞において同定される場合、MEFからマウス筋細胞への移行の軌道ではなく、遺伝子Aの発現は、MEFからマウスニューロンへの移行の軌道に関連すると仮説が立てられ得る。

30

#### 【0297】

上述されるように、図5Aは、細胞の移行過程の定性的時点、及び細胞の移行過程の定性的段階の両方に基づいて、移行の軌道を確認する。しかしながら、図5Aは、点(例えば、異なる時点での細胞)について遺伝子ごとの遺伝子発現のレベルを示さない。したがって、図5Aに示される情報に基づいて、どの遺伝子が移行のどの軌道に関連するかを予測することが可能でない。しかしながら、上述されるように、図5Bの多様体中の点の陰影は、各点について遺伝子ごとの相対的な遺伝子転写物数を示すように変化される。遺伝子ごとの点(例えば、異なる時点での細胞)についての遺伝子発現のこの表示に基づいて、どの遺伝子が移行のどの軌道に関連するかの予測が行われ得る。

40

#### 【0298】

図5Bは、一実施形態に係る、図4Bの多様体中の点として示される測定日(Asc11について2、5、及び22日目並びにBrn2及びMyt11について22日目)のそれぞれにおける細胞のそれぞれの中の3つのBAM転写因子のそれぞれの発現のレベルを示す。具体的には、図5Bは、図4Bの多様体の3つの異なる形態を示す。図5Bに示される多様体の第1の形態は、多様体の点のそれぞれについてのAsc11転写因子の発現

50



のレベルを示し、図 5 B に示される多様体の第 2 の形態は、多様体の点のそれぞれについての *B r n 2* 転写因子の発現のレベルを示し、図 5 B に示される多様体の第 3 の形態は、多様体の点のそれぞれについての *M y t 1 1* 転写因子の発現のレベルを示す。

【 0 2 9 9 】

図 5 B において、多様体中の点（例えば、ある時点での細胞）についての転写因子の発現のレベルが、転写因子についての 1 0 0 万のマッピングされたリード数当たりの転写物のキロベース当たりのフラグメント（*F P K M*）の *l o g* として測定される。*l o g*（*F P K M*）の比較的低い値は、比較的低いレベルの転写因子発現を示す。一方、*l o g*（*F P K M*）の比較的高い値は、比較的高いレベルの転写因子発現を示す。図 5 B の多様体において、点についての転写因子の比較的低いレベルの発現（例えば、*l o g*（*F P K M*）の比較的低い値）が、点に比較的暗い陰影を付けることによって示される。対照的に、点についての転写因子の比較的高いレベルの発現（例えば、*l o g*（*F P K M*）の比較的高い値）が、点に比較的明るい陰影を付けることによって示される。

10

【 0 3 0 0 】

図 5 A に描かれる移行の軌道を、遺伝子ごとの *B A M* 転写因子の発現レベルを示す図 5 B の多様体と比較することによって、特定の移行軌道に沿って細胞の進行に影響を与える転写因子を同定した。

【 0 3 0 1 】

まず、*A s c 1 1* 転写因子の発現を示す図 5 B の多様体を参照すると、2 3 日の期間中の 0 日目に、マウス細胞に、単に *A s c 1 1* 又は *B A M* を形質導入した。したがって、0 日目の細胞は、検出可能なレベルで *A s c 1 1* を発現しなかった。*A s c 1 1* を発現しなかったこれらの 0 日目の細胞は *M E F* であった。次に、2 3 日の期間の 2 日目に、*A s c 1 1* が、2 日目の細胞に関連する点の比較的暗い陰影によって示されるように、比較的低いレベルで発現された。*A s c 1 1* を発現したこれらの 2 日目の細胞は、図 5 A に示される移行の軌道に沿って進行し始めた。具体的には、2 日目の細胞の一部が、前駆マウス細胞になり、2 日目の細胞の一部が、*M E F* からニューロンへの移行軌道における中間細胞になり、2 日目の細胞の一部が、*M E F* から筋細胞への移行軌道における誘導された細胞になった。同様に、2 3 日の期間の 5 日目に、5 日目の細胞における *A s c 1 1* の発現は、5 日目の細胞に関連する点の比較的明るい陰影によって示されるように、2 日目の細胞と比べて増加された。*A s c 1 1* の増加した発現を有していたこれらの 5 日目の細胞は、図 5 A に示される移行の軌道に沿ってさらに進行した。具体的には、*M E F* からニューロンへの移行軌道における 5 日目の細胞は、中間及び早期 *i N* 細胞になった一方、*M E F* から筋細胞への移行軌道における 5 日目の細胞は、早期筋細胞になった。最後に、2 3 日の期間の 2 2 日目に、2 2 日目の細胞における *A s c 1 1* の発現が、5 日目の細胞と比べて増加したか又は同じままであった。*A s c 1 1* を発現したこれらの 2 2 日目の細胞は、図 5 A に示される移行の軌道に沿ってさらに進行した。具体的には、*M E F* からニューロンへの移行軌道における 2 2 日目の細胞が、成熟したマウスニューロンになった一方、*M E F* から筋細胞への移行軌道における 2 2 日目の細胞は、成熟したマウス筋細胞になった。2 2 日目に残っているマウス前駆細胞はなかった。

20

30

【 0 3 0 2 】

*A s c 1 1* 発現の誘導の後の *M E F* 細胞状態の移行のこれらの観察は、文献において公知である傾向にしたがうものである。具体的には、上に簡単に説明されるように、*A s c 1 1* プライミングが、*M E F* を、マウス前駆細胞へと移行するように誘導し、*A s c 1 1* のみの発現が、マウス前駆細胞を、マウスニューロン及びマウス筋細胞へと移行するように誘導する。図 5 B の *A s c 1 1* 多様体に関して上に説明されるように、0 日目の *M E F* における *A s c 1 1* の強制発現の後、*M E F* は、マウス前駆細胞、マウス筋細胞、及びマウスニューロンのいずれか 1 つへと移行した。

40

【 0 3 0 3 】

次に、*B r n 2* 転写因子の発現を示す図 5 B の多様体を参照すると、2 3 日の期間中の 0 日目に、*M E F* に、*B A M* 因子を形質導入した。*B r n 2* 発現が、2 3 日の期間中の 2

50

2日目のみに測定された。図5Bに見られるように、23日の期間の22日目に、22日目のマウスニューロンは、Brn2を強く発現した。したがって、Brn2の発現が、MEFからマウスニューロンへの移行軌道に沿ってMEF細胞の進行に関連していることが推定され得る。

【0304】

Brn2発現の誘導の後のMEF細胞状態の移行のこの観察は、文献において公知である傾向にしたがうものである。具体的には、上に簡単に説明されるように、Brn2発現は、マウス前駆細胞を、マウスニューロンへと移行するように誘導する。図5BのBrn2多様体に関して上述されるように、Brn2を発現したMEFは、マウスニューロンへと移行した。

10

【0305】

最後に、Myt11転写因子の発現を示す図5Bの多様体を参照すると、23日の期間中の0日目に、MEFに、BAM因子を形質導入した。Myt11発現が、23日の期間中の22日目のみに測定された。23日の期間の22日目に、22日目のマウスニューロンは、Myt11を強く発現した。したがって、Brn2転写因子と同様に、Myt11の発現が、MEFからマウスニューロンへの移行軌道に沿ってMEF細胞の進行に関連していることが推定され得る。

【0306】

Myt11発現の誘導の後のMEF細胞状態の移行のこの観察は、文献において公知である傾向にしたがうものである。具体的には、上に簡単に説明されるように、Myt11発現は、マウス前駆細胞を、マウスニューロンへと移行するように誘導する。図5BのMyt11多様体に関して上述されるように、Myt11を発現したMEFは、マウスニューロンへと移行した。

20

【0307】

したがって、第II及びIII節の方法を用いて図5B中のAsc11、Brn2、Myt11多様体を生成することによって得られるこれらの観察は、文献において実証される観察と一致する。Asc11による、Brn2による、Myt11による移行の観察のこの一貫性は、細胞状態の移行に影響を与える遺伝子を正確に同定する第II及びIII節の方法の能力を実証するのに役立つ。

【0308】

細胞状態の移行に影響を与える遺伝子を正確に同定する第II及びIII節の方法の能力をさらに実証するために、図5A及び5Bの多様体に基づいてなされた上記の観察を確認するために、インビトロ実験を行った。具体的には、Asc11発現が、MEFを、マウス「前駆」細胞、マウスニューロン、及び/又はマウス筋細胞へと移行するように誘導するという上記の観察を確認するために、インビトロ実験を行った。

30

【0309】

インビトロ実験を、第IV.A節において上に説明されるプロトコルにしたがって行った。上述されるように、このプロトコルにおいて、Asc11のみの発現を、MEFにおいて強制した。23日の期間の0日目、23日の期間の15日目のMEFにおけるAsc11転写因子の強制発現の後、マウス細胞を、DAPI、Map2抗体、及びTuj1抗体で染色した。DAPIは、DNA中のアデニン-チミンリッチ領域を染色することが知られている。したがって、DAPIは、細胞核を染色する。Map2抗体及びTuj1抗体は、神経系細胞を染色することが知られている。したがって、マウス細胞を、DAPI、Map2抗体、及びTuj1抗体で染色することによって、全マウス細胞の量に対するマウスニューロンの量が同定され得、ひいては、MEFの移行に対するAsc11の過剰発現の影響が決定され得る。Asc11転写因子の発現を強制したマウス細胞のこのセットは、本明細書において、インビトロ実験における実験群と呼ばれる。

40

【0310】

インビトロ実験における陽性対照群として、マウスニューロンのみを含むマウス細胞のサンプルも、DAPI、Map2抗体、及びTuj1抗体で染色した。陰性対照群として

50

、A s c l 1 発現を強制していない M E F 細胞のサンプルも、D A P I、M a p 2 抗体、及び T u j 1 抗体で染色した。

【0311】

実験群、陽性対照群、及び陰性対照群を、D A P I、M a p 2 抗体、及び T u j 1 抗体で染色した後、各染色剤で染色された各群を、M o l e c u l a r D e v i c e s H C I I X M 4 においてイメージングした。得られた画像が、図6に示される。図6は、一実施形態に係る、D A P I、M a p 2 抗体、及び T u j 1 抗体で染色された、A s c l 1 の発現が強制された M E F 細胞の画像、D A P I、M a p 2 抗体、及び T u j 1 抗体で染色されたマウスニューロンの画像、並びに D A P I、M a p 2 抗体、及び T u j 1 抗体で染色された、A s c l 1 の発現が強制されていない M E F 細胞の画像を示す。

10

【0312】

まず、陰性対照群の画像を参照すると、図6に示されるように、A s c l 1 の発現が強制されていない D A P I で染色された M E F 細胞の核が視認可能であるが、A s c l 1 の発現が強制されていない M E F 細胞の M a p 2 及び T u j 1 染色を示す画像中にニューロンはほとんどないし全くない。言い換えると、サンプル中に存在する多くのマウス細胞（特に M E F ）があった一方、ニューロンは存在しなかった。これは、A s c l 1 の発現がこのサンプルの M E F 細胞内で強制されず、したがって、ニューロンへの M E F 細胞の移行が誘導されなかったため、予測される結果である。

【0313】

次に、陽性対照群の画像を参照すると、図6に示されるように、D A P I で染色されたマウスニューロンの核が視認可能であるが、これらの同じマウスニューロンはまた、マウスニューロンの M a p 2 及び T u j 1 染色を示す画像中でも視認可能である。言い換えると、陽性対照サンプル中の細胞の全てが、ニューロンとして正確に同定された。

20

【0314】

最後に、実験群の画像を参照すると、図6に示されるように、A s c l 1 の発現を強制した D A P I で染色された M E F 細胞の核が視認可能である。さらに、これらの D A P I で染色された細胞の一部がまた、M a p 2 及び T u j 1 で染色され、これは、これらの選択された細胞がマウスニューロンであったことを示す。したがって、A s c l 1 の強制発現が、M E F からマウスニューロンへの移行の誘導に関連していることが示唆され得る。

30

【0315】

図5A及び5Bに関して上述されるインシリコ実験において観察されるように、図6のインビトロ実験は、M E F 細胞における A s c l 1 の強制発現が、M E F 細胞の、マウスニューロンへの移行をもたらし得ることを確認する。図5A及び5B中でなされる観察のこの確認は、細胞状態の移行に影響を与える遺伝子を正確に同定する第 I I 及び I I I 節の方法の能力をさらに実証する。

【0316】

I V . D . 実施例 2 : クラスタリング

第 I I I . C . 節において上述されるように、次元削減による行列 M の生成の後、クラスタリングを行って、行列 M 中のデータをグループ分けして、クラスタのセット  $C_j$  を生成する。クラスタのセット  $C_j$  中の各クラスタは、点のセットを含む。

40

【0317】

図7Aは、一実施形態に係る、図4Bの多様体を、クラスタリングによって同定されるクラスタ  $C_j$  にグループ分けされた多様体中の点とともに示す。図7Aの実施形態において、クラスタリングは、L o u v a i n コミュニティ検出、具体的には G e n L o u v a i n コミュニティ検出を用いて行った。図7Aに見られるように、クラスタリングは、多様体中の点の10の独自のクラスタ  $C_j$  を同定した。

【0318】

一般に、クラスタリングは、点に関連する値の閾値類似性、例えば、多様体の削減された次元空間におけるそれらの位置、それらの関連する遺伝子転写物数などに基づいて、多様体中の点を所与のクラスタに割り当てる。特に、図7Aの多様体について、クラスタ

50

リングは、多様体中の点間の閾値類似性に基づいて、点を所与のクラスターに割り当てる。例えば、群 8 に含まれる図 7 A の多様体中の点は全て、マウスニューロン、又はマウスニューロンと遺伝的に類似した他の細胞に関連する可能性が高い。同様に、群 9 に含まれる図 7 A の多様体中の点は全て、マウス筋細胞、又はマウス筋細胞と遺伝的に類似した他の細胞に関連する可能性が高い。

#### 【0319】

上述されるように、細胞状態の移行を誘導することが文献において知られている遺伝子の正確な同定を可能にすることに加えて、第 I I 及び I I I 節の方法はまた、文献において知られていない、細胞状態の移行に影響を与える因子（例えば、遺伝子及び摂動）の同定も可能にする。図 7 B は、一実施形態に係る、M E F の、マウス筋細胞への移行ではなく、マウスニューロンへの移行に関連することが文献において知られている転写因子及び知られていない転写因子の両方を示す。特に、図 7 B は、マウス「前駆」細胞において低発現される場合、マウス「前駆」細胞の、マウス筋細胞への移行の阻害に関連する転写因子、及びマウス「前駆」細胞において過剰発現される場合、マウス「前駆」細胞の、マウスニューロンへの移行に関連する転写因子を示す。マウス「前駆」細胞において、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連する転写因子を低発現させることによって、及びマウス「前駆」細胞において、マウス「前駆」細胞からマウスニューロンへの移行の誘導に関連する転写因子を過剰発現させることによって、マウス「前駆」細胞が、マウス筋細胞ではなく、マウスニューロンへと移行するように誘導され得る。

#### 【0320】

第 1 の細胞状態から別の特定の細胞状態への移行、又は第 1 の細胞状態から任意の他の細胞状態への移行に関連する転写因子を同定するために、クラスターが使用され得る。具体的には、第 1 の細胞状態に関連するクラスター中の点に関連する遺伝子転写物数を同定し、別の特定の細胞状態、又は第 1 の細胞状態以外の任意の細胞状態に関連する別のクラスター中の点に関連する遺伝子転写物数と比較する。クラスター間の遺伝子転写物数のこの比較は、母平均の差の検定、ウィルコクソンの順位和検定、t 検定、ロジスティック回帰、及び一般化線形モデルなどの任意の示差発現試験を用いて行われ得る。

#### 【0321】

例として、M E F からマウスニューロンへの移行に関連する転写因子を同定するために、図 7 A に関して説明されたクラスターを使用した。まず、マウス「前駆」細胞において過剰発現される場合、マウス「前駆」細胞からマウスニューロンへの移行に関連する転写因子を同定するために、マウスニューロンに関連する図 7 A のクラスター（例えば、図 7 A のクラスター 8 ）に含まれる点に関連する遺伝子転写物数を同定し、マウスニューロンに関連していない図 7 A の別のクラスターに含まれる点に関連する遺伝子転写物数と比較した。図 7 B の実施形態において、この比較は、ウィルコクソンの順位和検定を用いて行われた。しかしながら、代替的な実施形態において、比較は、統計分析の任意の他の方法を用いて行われ得る。この比較に基づいて、マウスニューロンに関連する図 7 A のクラスター中の点に関連する細胞において過剰発現された遺伝子は、マウス「前駆」細胞からマウスニューロンへの移行に関連していることが予測された。これらの遺伝子の転写及び翻訳から生じる転写因子は、マウス「前駆」細胞において過剰発現される場合、マウス「前駆」細胞からマウスニューロンへの移行に関連する図 7 B 中の転写因子として同定された。

#### 【0322】

同様に、マウス「前駆」細胞において低発現される場合、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連する転写因子を同定するために、マウス筋細胞に関連する図 7 A のクラスター（例えば、図 7 A のクラスター 9 ）に含まれる点に関連する遺伝子転写物数を同定し、マウス筋細胞に関連していない図 7 A の別のクラスターに含まれる点に関連する遺伝子転写物数と比較した。上述されるように、図 7 B の実施形態において、この比較は、ウィルコクソンの順位和検定を用いて行われた。しかしながら、代替的な実施形態において、比較は、統計分析の任意の他の方法を用いて行われ得る。この比較に基づ

いて、マウス筋細胞に関連する図7Aのクラスター中の点に関連する細胞において低発現された遺伝子は、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連していることが予測された。これらの遺伝子の転写及び翻訳から生じる転写因子は、マウス「前駆」細胞において低発現される場合、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連する図7B中の転写因子として同定された。

#### 【0323】

図7Bに見られるように、マウス「前駆」細胞において過剰発現される場合、マウス「前駆」細胞からマウスニューロンへの移行に関連する転写因子としては、Zfp941、Brn2、Myt1l、Taf5b、St18、Zkscan16、Camta1、及びArnt2が挙げられる。マウス「前駆」細胞において低発現される場合、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連する転写因子としては、Atf3、Rorc、Scx、Satb1、Elf3、及びFosが挙げられる。実施例1に関して上に詳細に説明されるように、Brn2及びMyt1l転写因子は、マウス「前駆」細胞からマウスニューロンへの移行の誘導に関連していることが文献において知られている。しかしながら、図7Bに示される残りの転写因子は、MEFの、マウス筋細胞への移行ではなく、マウスニューロンへの移行に関連することが文献において知られていない。したがって、上記の第II節及び第III節の方法を使用することによって、細胞を、移行の特定の軌道を進むように誘導することが文献において知られている及び知られていない遺伝子及び/又は転写因子の両方が同定され得る。次に、これらの同定された転写因子を用いて、細胞状態の移行、ひいては細胞の運命を制御することができる。

10

20

#### 【0324】

##### IV.E.実施例3：摂動に誘導される移行

第III.D及びIII.E節において説明されるように、細胞状態の移行に影響を与える遺伝子及び転写因子の同定を可能にすることに加えて、第II及びIII節の方法はまた、細胞状態の移行に影響を与える、小分子などの摂動の同定も可能にする。まず、細胞を、移行の特定の軌道を進むように誘導する摂動を同定するために、移行の可能な軌道が同定される。

#### 【0325】

図8Aは、一実施形態に係る、図4Aに関して説明されるMEF細胞についての移行軌道のマップを示す。移行軌道のこのマップを構築するために、図4Bの多様体を使用した。具体的には、同様の遺伝子転写物数に関連する多様体中の点を、複数の状態(図8A中で丸として表される)にグループ分けした。状態間に位置する可変の遺伝子転写物数の点を用いて、状態間の移行経路(図8A中で線として表される)を同定した。図8Aに示される移行軌道のマップは、細胞における遺伝子発現を変化させ、それによって細胞を、移行軌道のマップにおいてある状態から別の状態へと進行させることによって、細胞の移行軌道に影響を与える摂動を同定するのに使用され得る。ある実施形態において、図8Aに示される移行軌道のマップを生成するために、標準的なマーカー遺伝子のセットによる細胞の分類が使用され得る。このような実施形態において、同じ細胞型として同定される細胞は、移行軌道のマップにおいて同じ移行軌道に沿った細胞であると予測される。代替的な実施形態において、図8Aに示される移行軌道のマップを生成するために、図4Bの多様体のブランチが同定され、移行軌道のマップにおける異なる移行軌道を規定することが予測される。

30

40

#### 【0326】

図8Bは、一実施形態に係る、細胞が図8Aの移行軌道のマップにおける第1の状態から第2の状態へと移行するように細胞における遺伝子発現を変化させることによって細胞の移行軌道に影響を与える摂動を同定するための第III.D節に記載される方法の例を示す。具体的には、細胞に曝されるとき、細胞が第1の状態から第2の状態へと移行するように、細胞に遺伝子発現を変化させる摂動を同定するために、図8Bの方法は、細胞が第1の状態から第2の状態へと移行した後の細胞における遺伝子発現の変化を、摂動へのビヒクル細胞の曝露の後のビヒクル細胞における遺伝子発現の変化と比較する。細胞が

50

第1の状態から第2の状態へと移行した後の遺伝子発現の変化が、摂動へのビヒクル細胞の曝露の後のビヒクル細胞における遺伝子発現の変化と一致する（例えば、同等であるか又は類似している）場合、摂動は、細胞における遺伝子発現を変化させることによって、摂動に曝された細胞の、第1の状態から第2の状態への移行を誘導することが予測される。このように、摂動は、細胞状態の移行の特定の軌道に関連していることが予測される。

#### 【0327】

具体的には、図8Bに示される例を参照すると、図8Bは、6つの異なる遺伝子、遺伝子1～6についての、状態1の細胞、状態2の細胞、ビヒクル細胞、及び小分子摂動に曝されたビヒクル細胞の遺伝子発現のレベルを示す。所与の遺伝子についての遺伝子発現のレベルは、陰影によって示される。水玉模様の陰影は、検出不可能な遺伝子発現を示す一方、網目模様の陰影は、検出可能な遺伝子発現を示す。言い換えると、図8Bの実施形態において、遺伝子発現が、検出可能な遺伝子発現又は検出不可能な遺伝子発現のバイナリベースで測定される。しかしながら、代替的な実施形態において、遺伝子発現のレベルは、バイナリベースで測定されず、より定量ベースで測定される。

10

#### 【0328】

各細胞中の各遺伝子についての遺伝子発現のレベルを調べることに目を向けると、状態1の細胞については、遺伝子1～3の発現が検出不可能であったが、遺伝子4～6の発現は検出可能であった。対照的に、状態2の細胞については、遺伝子4～6の発現が検出不可能であったが、遺伝子1～3の発現は検出可能であった。ビヒクル細胞については、遺伝子1～3の発現が検出不可能であったが、遺伝子4～6の発現は検出可能であった。対照的に、摂動に曝されたビヒクル細胞については、遺伝子4～6の発現が検出不可能であったが、遺伝子1～3の発現は検出可能であった。

20

#### 【0329】

次に、各遺伝子について、状態1の細胞における遺伝子の発現のレベルを、状態2の細胞における遺伝子の発現のレベルと比較して、状態1から状態2への細胞の移行の後の遺伝子の発現のレベルの変化を決定した。遺伝子1～3に関連する暗色の網目模様の陰影によって示されるように、遺伝子1～3の発現は、状態1から状態2への細胞の移行の後に増加した。一方、遺伝子4～6に関連する暗色の水玉模様の陰影によって示されるように、遺伝子4～6の発現は、状態1から状態2への細胞の移行の後に低下した。

30

#### 【0330】

同様に、各遺伝子について、ビヒクル細胞における遺伝子の発現のレベルを、摂動に曝されたビヒクル細胞における遺伝子の発現のレベルと比較して、摂動へのビヒクル細胞の曝露の後の遺伝子の発現のレベルの変化を決定した。遺伝子1～3に関連する暗色の網目模様の陰影によって示されるように、遺伝子1～3の発現は、摂動へのビヒクル細胞の曝露の後に増加した。一方、遺伝子4～6に関連する暗色の水玉模様の陰影によって示されるように、遺伝子4～6の発現は、摂動へのビヒクル細胞の曝露の後に低下した。

#### 【0331】

最後に、状態1から状態2への細胞の移行の後の細胞における遺伝子発現の変化を、摂動へのビヒクル細胞の曝露の後のビヒクル細胞における遺伝子発現の変化と比較した。移行した細胞における遺伝子発現の変化を、ビヒクル細胞における遺伝子発現の変化と比較するために、任意の示差発現試験が使用され得る。例えば、母平均の差の検定、ウィルコクソンの順位和検定、t検定、ロジスティック回帰、及び一般化線形モデル比較アルゴリズムのいずれか1つが使用され得る。

40

#### 【0332】

図8Bに示されるように、遺伝子1～3の発現は、状態1から状態2へと移行した細胞、及び摂動に曝されたビヒクル細胞の両方において増加した。さらに、遺伝子4～6の発現は、状態1から状態2へと移行した細胞、及び摂動に曝されたビヒクル細胞の両方において低下した。状態1から状態2へと移行した細胞及び摂動に曝されたビヒクル細胞における遺伝子発現の変化のこの類似性に基づいて、摂動への状態1の細胞の曝露が、細胞に

50

おける遺伝子発現を変化させることによって、状態1の細胞の、状態2への移行を誘導し得ることが予測され得る。したがって、摂動を用いて、状態1から状態2への細胞の移行を制御することができる。

#### 【0333】

図8Bに関して上述される方法は、一般的状態1から一般的状態2への細胞の移行の誘導に関連する摂動の同定に関する。したがって、図8に関して上述される方法は、図8Aの移行軌道のマップにおいて任意の他の状態に関連付けられた任意の状態からの細胞の移行の誘導に関連する摂動を同定するのに使用され得る。しかしながら、図8Aの移行軌道のマップにおける一般的状態を参照するのではなく、図9は、図8Aの移行軌道のマップにおける特定の状態を同定し、次に、細胞が、マウス筋細胞ではなくマウスニューロンになるような、図9におけるある同定された状態から別の同定された状態への細胞の移行の誘導又は阻害に関連する特定の摂動を同定する。具体的には、図9は、MEF状態、マウス「前駆」細胞状態、マウス筋細胞状態、及びマウスニューロン状態を同定し、次に、細胞が、マウス筋細胞ではなく、マウスニューロンになるような、これらの状態の1つから別の状態への細胞の誘導又は阻害に関連する特定の摂動を同定する。

10

#### 【0334】

図9は、一実施形態に係る、MEFの、マウス筋細胞への移行ではなく、マウスニューロンへの移行に関連する小分子摂動を示す。特に、図9は、MEFに曝されたとき、MEFの、マウス「前駆」細胞への移行に関連する小分子摂動のセット、マウス「前駆」細胞に曝されたとき、マウス「前駆」細胞の、マウス筋細胞への移行の阻害に関連する小分子摂動のセット、及びマウス「前駆」細胞に曝されたとき、マウス「前駆」細胞の、マウスニューロンへの移行に関連する小分子摂動を示す。MEFからマウスニューロンへの移行の誘導に関連する摂動にMEFを曝すことによって、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連する摂動にMEFを曝すことによって、及びマウス「前駆」細胞からマウスニューロンへの移行の誘導に関連する摂動にMEFを曝すことによって、MEFは、マウス筋細胞ではなく、マウスニューロンへと移行するように誘導され得る。

20

#### 【0335】

図9に示される小分子摂動のそれぞれは、図8Bに関して上述される方法を実施することによって同定された。例えば、MEFからマウス「前駆」細胞への移行に関連しているものとして小分子摂動BRD-K38615104を同定するために、MEFからマウス「前駆」細胞への移行の後のMEFにおける遺伝子発現の変化が、BRD-K38615104へのビヒクル細胞の曝露の後のビヒクル細胞における遺伝子発現の変化に一致する(例えば、同等であるか又は類似している)ことが、図8Bの方法を用いて決定された。したがって、BRD-K38615104は、MEFにおける遺伝子発現を変化させることによって、MEFからマウス「前駆」細胞への移行を誘導することが予測された。同様に、マウス「前駆」細胞からマウス筋細胞の移行の阻害に関連しているものとして小分子摂動ダサチニブを同定するために、マウス「前駆」細胞からマウス筋細胞への移行の後のマウス「前駆」細胞における遺伝子発現の変化が、ダサチニブへのビヒクル細胞の曝露の後のビヒクル細胞における遺伝子発現の変化の逆数であることが、図8Bの方法を用いて決定された。したがって、ダサチニブは、マウス「前駆」細胞からマウス筋細胞への移行を阻害することが予測された。

30

40

#### 【0336】

図9に見られるように、MEFに曝される時、MEFからマウス「前駆」細胞への移行に関連する小分子摂動としては、BRD-K38615104、ゲルダナマイシン、マヌマイシンA、ミトキサントロン、クルクミン、及びトリコスタチンAが挙げられる。マウス「前駆」細胞に曝される時、マウス「前駆」細胞からマウスニューロンへの移行に関連する小分子摂動としては、アルボシジブ、ポリノスタット、KI20227、フォルスコリン、PP1、及びPP2が挙げられる。マウス「前駆」細胞に曝される時、マウス「前駆」細胞からマウス筋細胞への移行の阻害に関連する小分子摂動としては、アルボシジブ、ゲルダナマイシン、キナクリン、CGP-60474、及びダサチニブが挙げら

50

れる。

【0337】

図9において同定される小分子摂動の2つである、アルボシジブ及びゲルダナマイシンは、2つの異なる状態のマウス細胞の移行を誘導及び/又は阻害することによって、マウス細胞からマウスニューロンへの移行の誘導に関連する。具体的には、図9に示されるように、アルボシジブは、マウス「前駆」細胞からマウスニューロンへの移行の誘導、及びマウス「前駆」細胞からマウス筋細胞への移行の阻害の両方に関連する。同様に、ゲルダナマイシンは、MEFからマウス「前駆」細胞への移行の誘導、及びマウス「前駆」細胞からマウス筋細胞への移行の阻害の両方に関連する。したがって、MEFを、アルボシジブ及びゲルダナマイシンの両方に曝すことによって、MEFは、マウスニューロンへと移行することが予測され得る。

10

【0338】

図9において同定される小分子摂動のいくつかは、移行の示される軌道に関連していることが文献において知られている。具体的には、フォルスコリン、PP1、及びPP2は、マウス「前駆」細胞からマウスニューロンへの移行の誘導に関連していることが文献において知られている。同様に、トリコスタチンAは、MEFからマウス「前駆」細胞への移行の誘導に関連していることが文献において知られている。図8Bの方法及び文献において公知の情報によってなされる予測のこの一貫性は、細胞状態の移行に影響を与える摂動を正確に同定する図8Bの方法の能力を実証する。

20

【0339】

細胞状態の移行に影響を与えることが文献において知られている摂動を正確に同定することに加えて、図8Bの方法はまた、細胞状態の移行に影響を与えることが文献において知られていない摂動も同定することが可能である。具体的には、図9に示される残りの小分子摂動は、MEFの、マウス筋細胞への移行ではなく、マウスニューロンへの移行に関連することが文献において知られていない。したがって、図8Bに関して上述される方法を使用することによって、細胞を、移行の特定の軌道を進めるように誘導することが文献において知られている摂動及び知られていない摂動の両方が同定され得る。次に、これらの同定された摂動を用いて、細胞状態の移行、ひいては細胞の運命を制御することができる。

30

【0340】

V. 実施例4

この実施例の実験は、ニューロン及び/又は前駆細胞を促進するための方法を実証した。本明細書に記載される実験において、線維芽細胞(すなわち、初代マウス線維芽細胞)の開始集団を、Asc11過剰発現レンチウイルスを含む組成物に曝した。48時間後、化合物(例えば、フォルスコリン、グレサチニブ、PD-0325901)、又はビヒクル(すなわち、DMSO又はエタノール)を、組成物に加えた。ニューロンの総数を、陽性Tuj1/Map2シグナル及び神経形態に基づいて手で計数した。各実験について、各処理条件についてのニューロンの総数を、その実験に対する、DMSOで処理されたウェル中のニューロンの数によって正規化した。図10A及び図10Bに示されるように、線維芽細胞の開始集団から発生したニューロンの存在が、これらの実験において検出された。ニューロンの総数及びニューロンのパーセンテージの両方の倍率変化は、組成物に加えらるる化合物に応じて、増加するか、減少するか、又は同じままであった。これらの実験は、本発明の方法が、線維芽細胞を含む細胞の開始集団からニューロン及び/又は前駆細胞を促進するのに有用であることを実証した。

40

【0341】

細胞培養及び化合物処理

2回目の継代の初代マウス胚性線維芽細胞(MEF)を、DMEM中10%のFBS、1xGlutamax、1xMEMの非必須アミノ酸、1mMのピルビン酸ナトリウム、0.05U/mlのpen/strep、及び55µMのβ-メルカプトエタノールを含むMEF培養培地中で、20,000~45,000個/ウェル(ロットに応じて)で、

50



24ウェルプレート上で平板培養した。培養物中で24時間後、スピんフェクション (spinfection) (90分間にわたって32で、2000rpmでプレートを回転させる) によって、8 $\mu$ g/mlのポリブレンを含有するMEF培養培地中で、MEFを、Asc11過剰発現レンチウイルスに感染させた。レンチウイルス生成については以下を参照されたい。48時間後、培地を、DMEM/F12、1%のN2、2%のB27 1:50、1 $\times$ Glutamax、25 $\mu$ g/mlのインスリン、化合物又はビヒクル (DMSO又はエタノール) を含有する0.05U/mlのpen/strepを含む神経細胞用培地に交換した。化合物及びそれらの濃度を、以下から選択した: BI-2536 (200nM)、シロスタゾール (1000nM)、ダブラフェニブ (2500nM)、エストラジオールシピオナート (2000nM)、EX-527 (5000nM)、フェドラチニブ (1000nM)、フォレチニブ (200nM)、フォルスコリン (5000nM)、グレサチニブ (2500nM)、インジルピン-3-オキシム (2000nM)、KI20227 (250nM)、KU 0060648 (200nM)、m-3M3 FBS (1000nM)、マヌマイシン (800nM)、PD-0325901 (5000nM)、PHA-665752 (1000nM)、キナクリン (200nM)、ロットレリン (1000nM)、セルメチニブ (100nM)、トログリタゾン (5000nM)、及びベムラフェニブ (5000nM)。半分培地交換を、補充される化合物とともに2~3日ごとに行った。

10

#### 【0342】

##### 免疫蛍光染色

Asc11感染の12日後の時点で、細胞を、4%のバラホルムアルデヒドで固定し、透過処理し (0.2%のTriton X100)、5%の血清 (ロバ、子牛、ヤギ血清混合物) 中でブロックし、4で一晚、又は室温で2時間にわたってウサギ抗Tuj1 (1:1000) 及びマウス抗Map2 (1:500) 抗体で染色した後、二次抗体及びDAPI染色を行った。

20

#### 【0343】

##### イメージング及び分析

イメージングを、Molecular Devices ImageXpress Microにおいて行い; ウェルごとに36の画像を、10倍対物レンズで撮影した。ニューロンの総数を、陽性Tuj1/Map2シグナル及び神経形態に基づいて手動で計数した。各実験について、各処理条件についてのニューロンの総数を、その実験について、DMSOで処理されたウェル中のニューロンの数によって正規化した。

30

#### 【0344】

##### レンチウイルス生成

パッケージングプラスミド (Packaging plasmid) (Systems Bio、LV510A-1) 又は同様のもの、及びAsc11過剰発現プラスミド (Origeneレンチウイルス発現ベクトルへとクローニングされたAsc11 cDNA cat# PS100064) とともに、Mirus TransIT Lenti Transfection Reagent (Mirus、MIR 6603) によって293T細胞をトランスフェクトすることによって、レンチウイルスをパッケージングし、16,500RPMで1.5時間にわたってBeckman Coulter超遠心機において濃縮した。48時間の時点でウサギ抗Asc11 (1:200; Abcam、ab74065-100UG) 免疫蛍光染色によって判定した際に、90%以上の細胞のレンチウイルス感染を有する実験のみを追跡した。

40

#### 【0345】

##### V. 実施例5

実施形態1. 摂動が細胞の移行に影響を与えるかどうかを予測するための方法であって、方法が、メモリ及び1つ以上のプロセッサを含むコンピュータシステムにおいて: 第1の細胞状態と変化した細胞状態との間の細胞成分の発現の差の尺度を表す単一細胞の移行シグネチャーに電子的形態でアクセスする工程であって、変化した細胞状態は、第1の細

50

胞状態から変化した細胞状態への細胞の移行を経て生じ、単一細胞の移行シグネチャーは、複数の細胞成分の同定及び、複数の細胞成分中の各それぞれの細胞成分について、それぞれの細胞成分の発現の変化と、第1の細胞状態と変化した細胞状態との間の細胞状態の変化との間の関連性を定量する対応する第1の重要度スコアを含む、工程と；複数の非摂動細胞と摂動に曝された複数の摂動細胞との間の細胞成分の発現の差の尺度を表す摂動シグネチャーに電子的形態でアクセスする工程であって、摂動シグネチャーは、複数の細胞成分の全て又は一部と同定及び、複数の細胞成分の全て又は一部の中の各それぞれの細胞成分について、(i)複数の非摂動細胞と複数の摂動細胞との間のそれぞれの細胞成分の発現の変化と、(ii)複数の非摂動細胞と複数の摂動細胞との間の細胞状態の変化との間の関連性を定量する対応する第2の重要度スコアを含む、工程と；単一細胞の移行シグネチャー及び摂動シグネチャーを比較し、それによって、摂動が細胞の移行に影響を与えるかどうかを決定する工程とを含む、方法。

10

## 【0346】

実施形態2．単一細胞の移行シグネチャーにアクセスする工程が、(i)第1の複数の第1の単一細胞細胞成分の発現データセット、及び(ii)第2の複数の第2の単一細胞細胞成分の発現データセットに基づいて、単一細胞の移行シグネチャーを決定する工程を含み、ここで：第1の複数の第1の単一細胞細胞成分の発現データセット中の各それぞれの第1の単一細胞細胞成分の発現データセットが、第1の細胞状態の第1の複数の細胞の対応する単一細胞から得られ、第2の複数の第2の単一細胞細胞成分の発現データセット中の各それぞれの第2の単一細胞細胞成分の発現データセットが、変化した細胞状態の第2の複数の細胞の対応する単一細胞から得られる、実施形態1に記載の方法。

20

## 【0347】

実施形態3．第1の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第1の複数の細胞成分ベクトル中の、対応する細胞成分ベクトルを含み、第2の複数の単一細胞細胞成分の発現データセットの各それぞれのデータセットが、第2の複数の細胞成分ベクトル中の、対応する細胞成分ベクトルを含み、第1及び第2の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルが、複数の要素を含み、それぞれの細胞成分ベクトル中の各それぞれの要素が、複数の細胞成分中の対応する細胞成分に関連しており、第1及び第2の複数の単一細胞細胞成分の発現データセットのそれぞれのデータセットによって表される対応する単一細胞についての対応する細胞成分の量を表す対応する値を含む、実施形態2に記載の方法。

30

## 【0348】

実施形態4．第1及び/又は第2の複数の単一細胞細胞成分の発現データセットに対して次元削減を行って、複数の次元削減成分を生成する工程と；第1及び第2の複数の細胞成分ベクトル中の各それぞれの細胞成分ベクトルについて、複数の次元削減成分を、それぞれの細胞成分ベクトルに適用して、複数の次元削減成分中の各それぞれの次元削減成分についての次元削減成分値を含む対応する次元削減ベクトルを形成し、それによって、対応する第1及び第2の複数の次元削減ベクトルを形成する工程と；クラスタリングを行って、クラスタのセット $C_j$ を生成する工程であって、各クラスタが、第1及び第2の複数の次元削減ベクトルのサブセットに対応する複数の点を含む、工程と；クラスタのセット $C_j$ の第1のクラスタから第1の複数の細胞を同定する工程と；クラスタのセット $C_j$ の第2のクラスタから第2の複数の細胞を同定する工程とをさらに含み、方法が、任意に、対応する第1及び第2の複数の次元削減ベクトルを用いて多様体学習を行って、第1及び第2の複数の細胞における各細胞の他の各細胞に対する相対的な細胞状態を同定する工程をさらに含む、実施形態3に記載の方法。

40

## 【0349】

実施形態5．複数の非摂動細胞が、摂動に曝されていない対照細胞であり、又は非摂動細胞が、摂動に曝された非関連摂動細胞の平均である、実施形態1～4のいずれか1つに記載の方法。

## 【0350】

50

実施形態 6 . 単一細胞の移行シグネチャー及び摂動シグネチャーをブルーニングして、複数の細胞成分を、任意に RNA レベルで測定される転写因子に制限する工程をさらに含む、実施形態 1 ~ 5 のいずれか 1 つに記載の方法。

【 0 3 5 1 】

実施形態 7 . 単一細胞の移行シグネチャーを決定する工程が、母平均の差の検定、ウィルコクソンの順位和検定、t 検定、ロジスティック回帰、及び一般化線形モデルのうちの 1 つを用いて、( i ) 第 1 の複数の第 1 の単一細胞細胞成分の発現データセットと、第 2 の複数の第 2 の単一細胞細胞成分の発現データセットとの間の複数の細胞成分にわたる細胞成分量の差を決定する工程を含む、実施形態 2 に記載の方法。

【 0 3 5 2 】

実施形態 8 . 細胞成分の発現の差の尺度が、ウィルコクソンの順位和検定、t 検定、ロジスティック回帰、及び一般化線形モデルのうちの 1 つを用いて、( i ) 第 3 の複数の第 3 の単一細胞細胞成分の発現データセットと、( i i ) 第 4 の複数の第 4 の単一細胞細胞成分の発現データセットとの間の細胞成分量の差を定量し、ここで：第 3 の複数の第 3 の単一細胞細胞成分の発現データセット中の各それぞれの第 3 の単一細胞細胞成分の発現データセットが、複数の非摂動細胞中の対応する単一細胞から得られ、第 4 の複数の第 4 の単一細胞細胞成分の発現データセット中の各それぞれの第 4 の単一細胞細胞成分の発現データセットが、摂動に曝された複数の摂動細胞中の第 4 の複数の細胞の対応する単一細胞から得られる、実施形態 1 に記載の方法。

【 0 3 5 3 】

実施形態 9 . 単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングして、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させる工程をさらに含み、任意に、単一細胞の移行シグネチャー及び摂動シグネチャーをフィルタリングする工程が、閾値 p 値にしたがって、又は細胞成分の閾値数にしたがって、単一細胞の移行シグネチャー及び摂動シグネチャーに含まれる細胞成分の数を減少させる工程を含む、実施形態 1 ~ 8 のいずれか 1 つに記載の方法。

【 0 3 5 4 】

実施形態 1 0 . それぞれの細胞成分についての対応する第 2 の重要度スコアを決定する工程が、それぞれの細胞成分についての重要度スコアを、複数の細胞成分中の各それぞれの細胞成分について、それぞれの細胞成分についての対応するマッチングスコアで置き換える工程と；複数の細胞成分についてのマッチングスコアを組み合わせ、摂動についてのマッチングスコアを生成する工程と；摂動が、それぞれの摂動についてのマッチングスコアに基づいて、第 1 の細胞状態と変化した細胞状態との間の細胞の移行に関連しているかどうかを決定する工程とを含み、任意に、対応するマッチングスコアが、別個の又は連続したスコアを含む、実施形態 1 ~ 9 のいずれか 1 つに記載の方法。

【 0 3 5 5 】

実施形態 1 1 . 重要度スコアを置き換える工程が、それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞成分量及びそれぞれの細胞成分についての摂動シグネチャーからの細胞成分量が両方とも上方制御される場合、重要度スコアを第 1 のスコアで置き換える工程と；それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞成分量が上方制御され、それぞれの細胞成分についての摂動シグネチャーからの細胞成分量が下方制御される場合、重要度スコアを第 2 のスコアで置き換える工程と；それぞれの細胞成分についての摂動シグネチャーからの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアを第 3 のスコアで置き換える工程とを含む、実施形態 1 0 に記載の方法。

【 0 3 5 6 】

実施形態 1 2 . 重要度スコアを置き換える工程が、それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞成分量及び細胞成分についての摂動シグネチャーからの細胞成分量が両方とも下方制御される場合、重要度スコアを第 1 のスコアで置き換える工程と；それぞれの細胞成分についての単一細胞の移行シグネチャーからの細胞成分量が

10

20

30

40

50

下方制御され、細胞成分についての摂動シグネチャーからの細胞成分量が上方制御される場合、重要度スコアを第2のスコアで置き換える工程と；細胞成分についての摂動シグネチャーからの細胞成分量が、有意に上方制御又は下方制御されない場合、重要度スコアを第3のスコアで置き換える工程とを含む、実施形態10に記載の方法。

【0357】

実施形態13．複数の細胞成分が、任意にRNAレベルで測定される複数の遺伝子を含む、実施形態1～12のいずれか1つに記載の方法。

【0358】

実施形態14．第1の複数の第1の単一細胞細胞成分の発現データセット及び第2の複数の第2の単一細胞細胞成分の発現データセット中の各単一細胞細胞成分の発現データセットが、単一細胞リボ核酸(RNA)シーケンシング(scRNA-seq)、scTag-seq、シーケンシングを用いたトランスPOSEでアクセス可能なクロマチンの単一細胞アッセイ(scATAC-seq)、CytOF/SCOPE、EMSAbsseq、miRNA-seq、CITE-seq、及びそれらの任意の組合せ、並びに単一細胞細胞成分の発現データセットにおける活性化された経路を表す線形組合せなどの組合せを含む、上記のものの集合からなる群から選択される方法を用いて生成される、実施形態2に記載の方法。

10

【0359】

実施形態15．比較に基づいて、変化した細胞状態を促進するものとして摂動を同定する工程、又は比較に基づいて、変化した細胞状態を阻害するものとして摂動を同定する工程をさらに含む、実施形態1～14のいずれか1つに記載の方法。

20

【0360】

実施形態16．細胞の移行シグネチャー及び摂動シグネチャーが、異なるタイプの細胞成分を用いて生成される、実施形態1～15のいずれか1つに記載の方法。

【0361】

実施形態17．細胞の移行シグネチャー及び摂動シグネチャーが、同じタイプの細胞成分を用いて生成される、実施形態1～16のいずれか1つに記載の方法。

【0362】

実施形態18．電子的形態でアクセスする工程が、複数の摂動における各それぞれのシグネチャーについて行われ、それによって、複数の摂動シグネチャーを得て、比較する工程が、前記単一細胞の移行シグネチャー及び前記摂動シグネチャーを、複数の摂動シグネチャーにおける各それぞれのシグネチャーと比較し、それによって、第1の細胞状態と変化した細胞状態との間の細胞の移行に関連する複数の摂動のサブセットを決定する、実施形態1～17のいずれか1つに記載の方法。

30

【0363】

実施形態19．1つ以上のプロセッサ及びメモリを含むコンピュータシステムであって、メモリが、実施形態1～18のいずれか1つに記載の方法を実行するための命令を保存する、コンピュータシステム。

【0364】

実施形態20．摂動が細胞の移行に影響を与えるかどうかを予測するための、コンピュータによって実行可能な1つ以上のコンピュータプログラムを保存する非一時的なコンピュータ可読媒体であって、コンピュータが、1つ以上のプロセッサ及びメモリを含み、1つ以上のコンピュータプログラムが、実施形態1～18のいずれか1つに記載の方法を実行するための、コンピュータが実行可能な命令を集合的に符号化する、非一時的なコンピュータ可読媒体。

40

【0365】

引用される参考文献及び代替的な実施形態

本明細書に引用される参考文献は、各個々の刊行物又は特許若しくは特許出願が、あらゆる目的のために全体が参照により援用されることが具体的に且つ個別に示されているのと同程度に、あらゆる目的のために全体が参照により本明細書に援用される。

50

## 【 0 3 6 6 】

本発明は、非一時的なコンピュータ可読記録媒体に埋め込まれたコンピュータプログラム機構を含むコンピュータプログラム製品として実装され得る。例えば、コンピュータプログラム製品は、図 1 又は 2 の任意の組合せで示されるプログラムモジュールを含み得る。これらのプログラムモジュールは、CD-ROM、DVD、磁気ディスク記憶装置製品、又は任意の他の非一時的なコンピュータ可読データ又はプログラム記憶装置製品に保存され得る。

## 【 0 3 6 7 】

本発明の多くの変更及び変形が、当業者に明らかであるように、本発明の趣旨及び範囲から逸脱せずに行われ得る。本明細書に記載される特定の実施形態は、例として提供されるに過ぎない。実施形態は、本発明の原理及びその実際の用途を最良に説明し、それによって他の当業者が、本発明及び様々な変更を有する様々な実施形態を、考えられる特定の用途に合うように最良に利用することができるように選択及び記載された。本発明は、このような特許請求の範囲が権利を有する均等物の全範囲とともに、添付の特許請求の範囲のみによって限定されるものである。

10

## 【 0 3 6 8 】

【表 5】

## 補足的な表1

	x	y			
0	292.9590886	705.1857209	54	355.1414568	642.7888097
1	374.6541132	595.9985647	55	376.677819	229.8185684
2	-19.38276064	716.5409291	56	265.8882156	570.9775465
3	644.688266	-111.5344681	57	471.3562122	337.6768758
4	322.2346012	395.5017566	58	137.9160196	798.8754115
5	152.7499645	735.6219439	59	400.8680379	504.8839273
6	494.5462458	257.0674805	60	198.5849661	710.7693083
7	432.4564718	515.1958575	61	606.6521059	-509.7572535
8	140.8424799	774.5022913	62	225.4256559	689.4398313
9	464.4809307	283.6841628	63	469.9972441	143.4721685
10	291.8623356	574.2620559	64	232.1374399	757.0158223
11	97.07154344	743.7991899	65	226.3789216	643.0198475
12	43.69841021	767.4559498	66	738.477066	41.45517208
13	320.7317393	738.7282443	67	586.5634942	147.1840513
14	247.9474996	291.1524828	68	716.7266637	-359.8871297
15	200.4509778	742.0161885	69	316.7344784	591.7914456
16	433.3434906	371.8609426	70	140.801892	543.732719
17	170.6087748	755.2888557	71	355.2827082	354.411774
18	602.1577532	257.8131808	72	280.7942953	796.1915258
19	445.9835374	455.4914589	73	284.963372	82.61134257
20	177.9019201	813.1096354	74	332.6970419	518.5559573
21	808.5955637	-208.5263147	75	411.8110794	266.7794701
22	89.82707295	816.1114036	76	382.6336111	626.7278101
23	479.7784168	-112.8197427	77	298.1778232	776.7635117
24	860.715019	-197.2301183	78	555.4030268	228.9955405
25	492.8841929	218.5930701	79	432.7647179	241.3787013
26	339.8283861	690.392759	80	384.2042425	421.0252928
27	149.4278669	680.6605704	81	492.3343172	12.03358116
28	223.6446878	803.5071625	82	267.8191562	-138.3291145
29	417.1707573	417.2615233	83	847.7556161	-174.1663311
30	99.03864487	683.332534	84	541.5396082	-415.1513959
31	382.721431	555.22895	85	520.5799179	21.95926784
32	162.0561419	644.127785	86	537.2485129	-213.1905582
33	200.267714	774.22094	87	636.4348944	-466.9994817
34	848.4594185	-72.34808992	88	636.3489348	19.06806914
35	268.2163485	695.7546768	89	393.5393329	-268.9946479
36	252.1473308	655.8261974	90	785.9406294	-140.0325714
37	126.4134128	412.1389212	91	477.1344647	-494.0936324
38	271.3307885	511.8307189	92	733.2929278	-300.6038544
39	353.2309599	282.9238861	93	724.5250405	-87.92260621
40	263.2929281	220.5472408	94	808.551751	-81.29303912
41	-236.3670523	541.1518713	95	363.2923532	-402.2831256
42	465.4341182	423.6497983	96	781.3238933	-302.7763969
43	679.7364573	-463.9095819	97	561.6407264	-270.8026998
44	583.5582967	-344.174618	98	725.4134987	-137.3051546
45	162.9296925	838.8955218	99	569.2010139	35.84068973
46	303.56886	681.5465862	100	593.0072608	-183.803066
47	604.5129157	100.1546147	101	462.6139553	68.24138424
48	117.0693135	723.5327479	102	620.9697368	51.12408879
49	521.7201039	297.3321782	103	422.8574274	-320.4548981
50	300.60139	480.2801613	104	786.0866476	-91.11583893
51	110.8288737	851.5641072	105	402.7391764	-439.7250854
52	241.7052693	582.640065	106	398.9005319	-228.5642935
53	274.1473831	751.9519303	107	509.951701	-394.6145652

10

20

30

40

【表 6】

108	384.8061749	-84.17493609	164	-590.1911886	-134.3217373
109	569.0737847	-230.8710451	165	-781.0024313	-155.0530685
110	377.3159906	-40.62880392	166	-370.273676	23.14103967
111	226.152504	-235.2579455	167	-870.5453051	-79.50528178
112	404.039674	-373.1420998	168	-737.6398733	-18.17552605
113	537.6256187	-446.2795533	169	-437.4358812	31.86828718
114	663.4627561	-192.6176209	170	-789.5181555	-125.0321147
115	474.2668537	-69.75627991	171	-748.6858151	-214.8080234
116	747.3457141	-5.647140309	172	-652.7952349	41.20341279
117	544.8569144	-78.16983422	173	-787.0300957	-39.10950006
118	296.3980478	-224.8809905	174	-382.781317	-122.031103
119	260.5447659	-394.8730357	175	-459.2922017	-100.2497458
120	215.8139222	-309.3801446	176	-1013.648221	213.2949868
121	526.7281561	-350.864597	177	-376.19131	-89.46049067
122	716.4765281	-200.3204533	178	-377.4088089	-23.03812659
123	347.395669	-138.8038114	179	-900.9061151	52.23467137
124	489.5767434	-163.0308792	180	-789.0565976	-89.43258183
125	706.1063659	-410.8218199	181	-883.7875002	-179.2004471
126	661.5195229	-247.7761725	182	-815.7958619	-89.1302342
127	629.0244112	-250.884033	183	-992.4329143	156.6352187
128	861.5078812	-264.8657422	184	-845.3559874	-213.7163084
129	509.1152599	-456.619506	185	-833.0558495	-177.3140009
130	440.5776014	-62.64558011	186	-892.4398673	-149.1835258
131	823.9910195	-284.7041831	187	-638.4570606	-66.08309919
132	528.3097523	-476.6018224	188	-733.1453101	-138.8679106
133	528.6998461	-11.32332005	189	-824.9813994	-49.40482982
134	640.6338998	-308.5704094	190	-442.8329584	-48.22282782
135	748.0650823	-349.8444491	191	-303.2900893	-62.08204323
136	245.6327051	-4.682363994	192	-378.4226756	-64.85523963
137	645.2431043	-141.4374096	193	-1014.110255	263.2637815
138	430.9649022	-108.2216251	194	-703.3821521	-79.63206877
139	547.316555	-180.841155	195	-1061.556178	298.1987987
140	671.3783221	-325.557468	196	-578.1017679	209.2722555
141	417.5524522	-145.6894216	197	-397.9077377	-46.56708336
142	699.8571739	-265.0475523	198	-935.97811	204.5283192
143	585.9935863	-428.0759414	199	-328.0118588	-69.18380624
144	749.329451	-322.3169203	200	-251.796289	66.94312521
145	636.7559818	-399.3802132	201	-412.1046852	-19.12629274
146	835.4546022	-315.6254608	202	-1060.224854	266.6457265
147	425.3919014	-25.27147061	203	-732.9925382	153.1166422
148	523.9193173	-113.5700454	204	-525.3661035	253.8142789
149	254.3915387	-286.1509508	205	200.3885141	29.61427703
150	743.0553842	-168.3968679	206	-313.5452179	-19.60141447
151	336.6267218	-250.6428123	207	-285.0414633	107.1246899
152	482.6828643	-139.7346036	208	-1473.634035	799.9533324
153	-494.7330005	-56.03030691	209	-1011.94183	385.2788015
154	-752.0305594	-173.9089457	210	-452.3230446	-1203.930107
155	-673.7690983	-113.730056	211	810.68978	-253.5587957
156	-859.5870798	-22.06943053	212	-1492.073178	970.8142665
157	-923.5296174	103.529707	213	-1460.767934	914.9541654
158	-369.6618882	-43.77239245	214	-191.2004832	-91.74360449
159	-849.6874488	-110.1884344	215	-1636.051075	930.0650453
160	-892.9356778	113.8219709	216	-1566.812784	1012.514253
161	-744.3633801	-94.61202087	217	-1504.908289	890.6777928
162	-818.2681693	-125.5071708	218	-1519.317885	1005.714622
163	-763.9751888	115.4465471	219	-1533.625269	1068.262969

10

20

30

40

【 0 3 7 0 】

【表 7】

220	-1636.103988	973.4266963	276	-71.49536896	642.9115696
221	-399.1888452	-1252.609284	277	-182.3503122	905.1465444
222	-1566.957146	975.8168791	278	-312.3035442	-685.7343163
223	-1609.858896	961.3718502	279	-330.1756081	815.2471162
224	-1601.661403	984.9681923	280	-396.2282268	-662.7181382
225	-1610.797154	1044.435169	281	-158.7947144	737.1809729
226	-1434.321223	837.4344285	282	366.0232079	-186.9663003
227	-1618.61781	1086.246642	283	-60.56980165	786.5366894
228	-497.898372	-1049.764737	284	-217.3511357	765.5545062
229	-1596.147271	936.1974239	285	-597.2363354	-794.6747245
230	-1616.720328	1014.948265	286	-198.7047088	-336.2320429
231	-1567.814229	1052.955665	287	-478.8798362	-1206.701925
232	-204.2291143	-12.17503384	288	75.32541167	-302.6097324
233	-1670.893476	982.7077687	289	-437.9894975	-1123.247693
234	-1562.771641	836.903796	290	-172.8620362	955.2024567
235	-1274.276718	610.9868859	291	448.8157384	-240.7631613
236	-1535.915783	1006.58881	292	-244.4714526	800.7282973
237	-1650.682415	1019.465694	293	353.6505551	-326.6336007
238	-1524.182997	977.7089269	294	-135.301662	763.8703634
239	-524.7566037	-1102.132075	295	-310.5588962	852.8546552
240	-1609.013256	909.5991955	296	-696.8566331	-487.1825639
241	-1508.015199	1030.311782	297	-421.0922692	748.0684235
242	-915.0402933	309.261403	298	-255.9695809	-343.902937
243	-1645.011514	1051.053554	299	-235.2427064	936.5821591
244	-1627.985088	896.7519543	300	-620.5149865	-886.9828334
245	54.60783988	46.11744393	301	-129.328065	926.5450404
246	-1559.708568	938.2312785	302	302.0644251	-352.2447281
247	-1442.149664	758.3161505	303	-342.6855001	773.704035
248	-320.5928299	-106.3511041	304	-566.0436414	-478.6265257
249	-157.0861516	-83.09609891	305	-244.2965795	-112.6708507
250	-428.1811291	-1007.72501	306	-449.643089	-1100.563662
251	-272.5163626	-87.54651401	307	-464.6885558	-1074.707812
252	-146.999199	866.083403	308	-68.66972972	578.3858961
253	-613.0961278	-532.8535664	309	307.0357686	-193.4856106
254	-385.2549262	803.8327949	310	230.2016127	-355.8154448
255	473.33345	-257.4672531	311	385.5731293	-334.5244516
256	-278.418256	896.3042008	312	478.6284	-363.9641181
257	-257.897376	964.4410063	313	-695.6406796	-761.847312
258	-101.6588352	-114.7229664	314	-605.723566	-251.8766458
259	-250.2214407	621.8342482	315	-529.7214129	-1130.534451
260	-289.3527446	-795.6425939	316	-267.7197429	-118.81822
261	-189.8727007	831.4207487	317	-500.1458532	-1157.421321
262	-318.8133504	775.1673847	318	-470.8628684	-981.2061023
263	-663.9420342	-760.7854227	319	-371.0940203	-913.1472336
264	-300.2568764	938.5857675	320	-230.9171642	878.0894167
265	422.0440188	-277.1384651	321	-599.3762515	-937.5683057
266	-376.5780954	-1067.366599	322	-73.7071824	828.3052437
267	-480.572652	-1171.594662	323	31.79960939	-322.1999788
268	-419.7416236	-843.7273631	324	-15.67298881	582.9978845
269	-430.40124	-1068.741207	325	-251.0465086	761.1669009
270	-352.0698541	-1092.336735	326	186.7232748	-492.1827404
271	-299.5467262	-911.3166677	327	-191.3677344	933.293115
272	-147.635184	819.0818859	328	133.6371696	-405.2718901
273	284.6024545	-308.3650509	329	-39.56106229	484.6992835
274	-89.95073517	701.9526346	330	-107.4926555	630.3918785
275	594.6679208	-289.7241624	331	-357.2339947	832.4105136

10

20

30

40

【 0 3 7 1 】



【表 8】

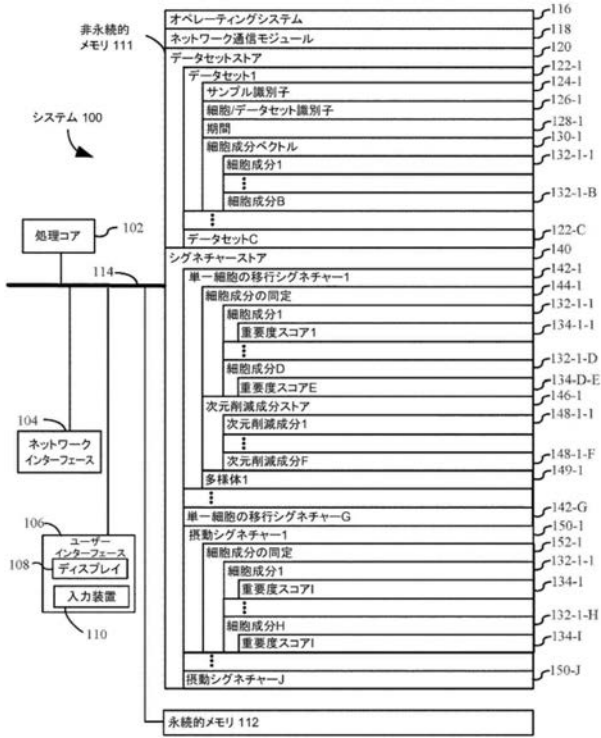
332	-374.6902631	-1186.054859	369	-80.97859781	-774.2828386
333	-399.9303104	-1151.054458	370	-44.01461995	-344.6215693
334	-501.8839635	-997.8236834	371	-414.4883378	-963.9064518
335	-415.2502713	-1213.56788	372	-374.1797009	-949.2800878
336	-97.78630012	823.0603775	373	-482.8997867	-1095.714719
337	452.7220978	-356.303574	374	-428.385547	-1198.038919
338	-187.5032343	866.0517318	375	-467.6739632	-1130.539902
339	189.7833559	-340.9935829	376	-385.8849809	-731.4034683
340	-404.6932064	855.929578	377	61.15450278	-426.4883923
341	746.3533845	-266.2656985	378	-676.8001687	-838.1180892
342	-149.2445028	652.3696661	379	-673.9297813	23.91043665
343	384.1799451	-417.0021581	380	-547.2029152	-692.958461
344	-254.7295489	856.0585652	381	185.096998	-126.6624536
345	515.0188104	-304.3697333	382	-21.76754275	-526.1392038
346	-119.9326629	898.4543687	383	-566.3527916	-1041.877279
347	-646.7570031	-850.3130306	384	343.3211308	-349.3066494
348	-227.7166574	812.5845214	385	-471.4022291	-713.8246475
349	-667.41269	-813.4695855	386	14.25231433	-593.0560185
350	-350.4325873	869.5823574	387	-671.8363046	-872.4771127
351	-629.6469551	-979.3159524	388	-626.571634	-765.2101338
352	-346.2799403	-1122.895243	389	455.1603625	-216.8347232
353	-510.7372588	-1193.85305	390	-628.6889684	-919.5718853
354	-385.3067915	-986.307804	391	450.7754866	-280.4444406
355	-309.4066867	-149.4820066	392	155.6661198	-254.2648132
356	-446.2571647	-691.8036739	393	-485.114941	-685.7250585
357	-322.2046006	-781.4684253	394	215.2075216	-381.2364804
358	-57.50674612	871.6000215	395	281.0292909	-366.7052951
359	-591.8808177	-553.2850621	396	-665.5292398	-639.6013519
360	-299.8092449	715.0119834	397	-592.8960195	-753.6753068
361	-651.5471611	-789.7689615	398	344.6874105	-31.30720444
362	-316.4290629	885.4767089	399	-701.470463	-825.0957739
363	-676.0616298	-539.7217594	400	-718.2334019	-854.3294898
364	-210.2976814	904.6664887	401	-367.9842936	-451.9427781
365	282.9910976	13.23776322	402	172.2477735	-462.6789888
366	-240.2963947	899.6751261	403	120.619482	-487.4043328
367	-630.0944712	-804.5742854	404	502.8416095	-199.0993372
368	-273.9414322	931.8500003			

10

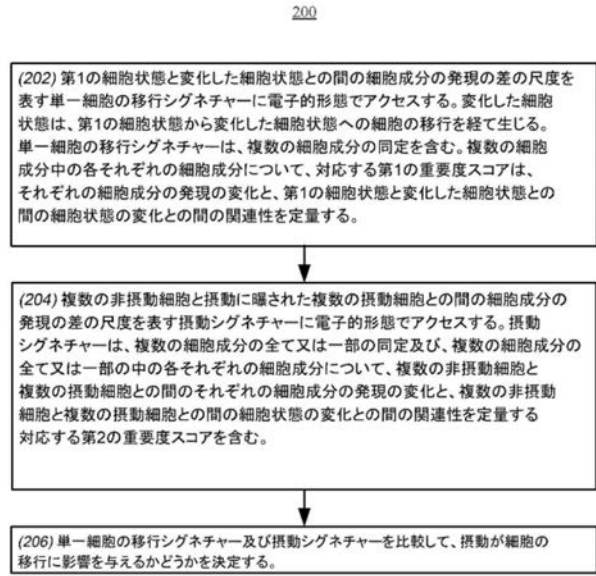
20

30

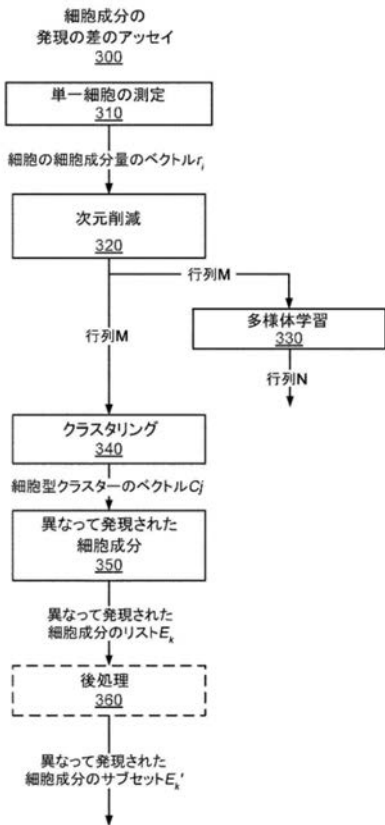
【 図 1 】



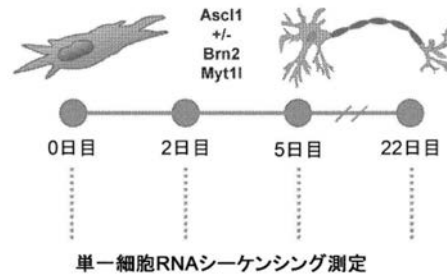
【 図 2 】



【 図 3 】



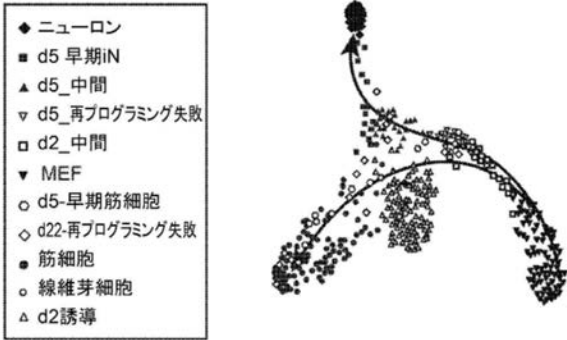
【 図 4 A 】



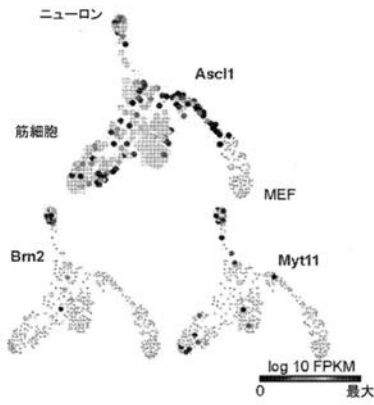
【 図 4 B 】



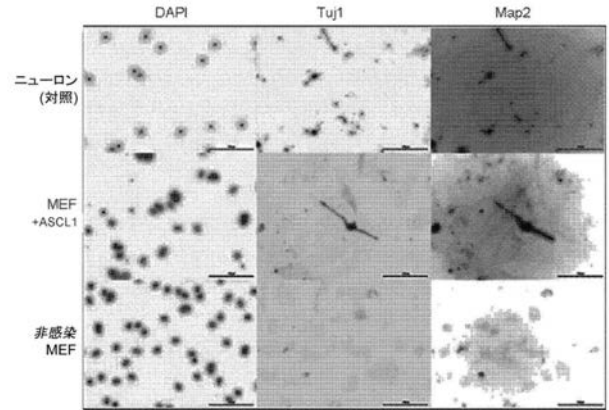
【 図 5 A 】



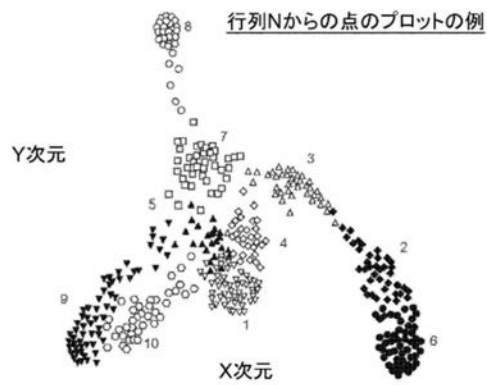
【 図 5 B 】



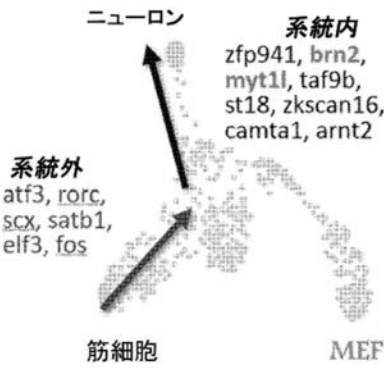
【 図 6 】



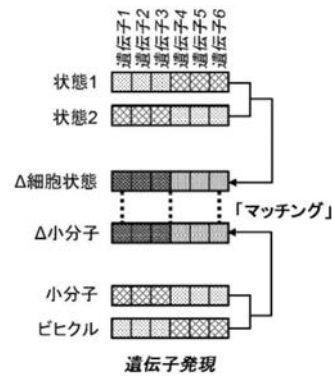
【 図 7 A 】



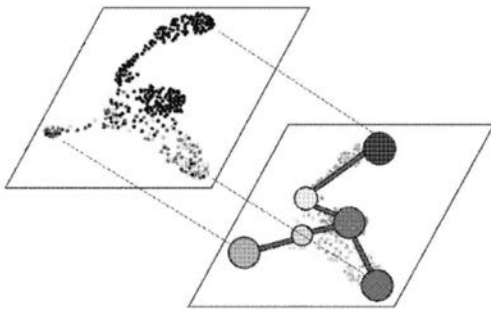
【 図 7 B 】



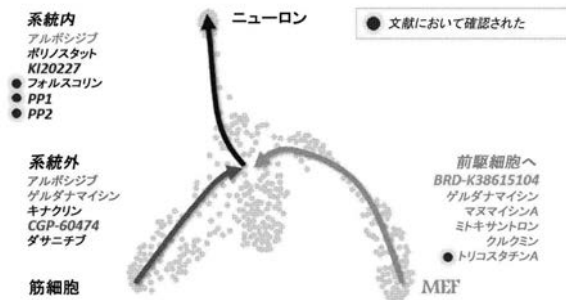
【 図 8 B 】



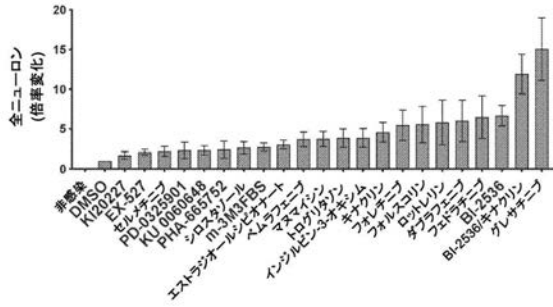
【 図 8 A 】



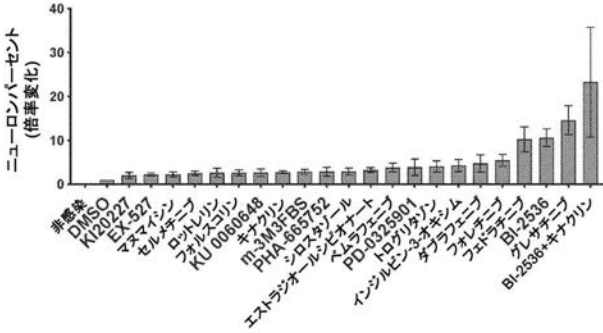
【 図 9 】



【 図 1 0 A 】



【 図 1 0 B 】



## 【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US2019/041976
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC(8) - C12N 9/22; C12N 15/1093; C12N 15/11; C12N 15/113; C12Q 1/68; C12Q 1/6886 (2019.01) CPC - C12N 9/22; C12N 15/1093; C12N 15/11; C12N 15/113; C12N 2310/14; C12N 2310/20; C12N 2800/80; C12Q 1/68; C12Q 1/6886; C12Q 2600/106; C12Q 2600/136; G16B 5/00 (2019.08)		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) See Search History document		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched USPC - 435/6.1; 435/6.11; 435/320.1; 506/2; 506/9; 506/14; 705/2; 705/3 (keyword delimited)		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) See Search History document		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2017/075294 A1 (THE BROAD INSTITUTE INC. et al) 04 May 2017 (04.05.2017) entire document	1-11, 13-20
Y		12
Y	LAMB et al. "The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease," Science, 29 September 2006 (29.09.2006), Vol. 313, Pgs. 1929-1935. entire document	12
A	BARGAJE et al. "Cell Population Structure Prior to Bifurcation Predicts Efficiency of directed Differentiation in Human Induced Pluripotent Cells," Proc. Natl. Acad. Sci. USA, 06 February 2017 (06.02.2017), Vol. 114, Pgs. 9, 2271-2276. entire document	1-20
A	CARRO et al. "The Transcriptional Network for Mesenchymal Transformation of Brain Tumors," Nature, 21 January 2010 (21.01.2010), Vol. 463, Pgs. 318-325. entire document	1-20
A	US 2014/0206546 A1 (CELLECTA, INC.) 24 July 2014 (24.07.2014) entire document	1-20
A	US 2008/0033658 A1 (DALTON et al) 07 February 2008 (07.02.2008) entire document	1-20
A	GEORGE et al. "Survival Outcomes in Cancer Patients Predicted by a Partial EMT Gene Expression Scoring Metric," Cancer Research, 25 September 2017 (25.09.2017), Vol. 77, No. 22, Pgs. 6415-6428. entire document	1-20
A	KAHVEJIAN et al. "What Would You Do if You Could Sequence Everything?," Nat Biotechnol, 31 October 2008 (31.10.2008), Vol. 26, No. 10, Pgs. 1125-1133. entire document	1-20
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 20 September 2019		Date of mailing of the international search report 15 OCT 2019
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300		Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2019/041976

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2016/0312302 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 27 October 2016 (27.10.2016) entire document	1-20
A	- WELCH et al. "MATCHER: Manifold Alignment Reveals Correspondence Between Single Cell Transcriptome and Epigenome Dynamics," Genome Biology, 24 July 2017 (24.07.2017), Vol. 18, Pgs. 1-19. entire document	1-20

## フロントページの続き

(31)優先権主張番号 62/698,701

(32)優先日 平成30年7月16日(2018.7.16)

(33)優先権主張国・地域又は機関  
米国(US)

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

(特許庁注：以下のものは登録商標)

1. BLUETOOTH
2. ANDROID
3. Linux
4. UNIX
5. WINDOWS
6. TRITON

(72)発明者 カーヴジアン、アヴァック  
アメリカ合衆国 2142 マサチューセッツ州 ケンブリッジ、ケンブリッジ パークウェイ、  
55 スイート 800イー フラッグシップ パイオニアリング イノベーションズ プイアイ  
、エルエルシー 内(72)発明者 プルージス、ニコラス マッカートニー  
アメリカ合衆国 2142 マサチューセッツ州 ケンブリッジ、ケンブリッジ パークウェイ、  
55 スイート 800イー フラッグシップ パイオニアリング イノベーションズ プイアイ  
、エルエルシー 内(72)発明者 レチン、マイケル レイモンド  
アメリカ合衆国 2142 マサチューセッツ州 ケンブリッジ、ケンブリッジ パークウェイ、  
55 スイート 800イー フラッグシップ パイオニアリング イノベーションズ プイアイ  
、エルエルシー 内(72)発明者 ウルフ、ファビアン アレクサンダー  
アメリカ合衆国 2142 マサチューセッツ州 ケンブリッジ、ケンブリッジ パークウェイ、  
55 スイート 800イー フラッグシップ パイオニアリング イノベーションズ プイアイ  
、エルエルシー 内(72)発明者 ホセイニ、プアヤ  
アメリカ合衆国 2142 マサチューセッツ州 ケンブリッジ、ケンブリッジ パークウェイ、  
55 スイート 800イー フラッグシップ パイオニアリング イノベーションズ プイアイ  
、エルエルシー 内

Fターム(参考) 4B029 AA07 BB11 CC02 FA02 FA12

4B063 QA01 QA18 QA20 QQ08 QQ52 QR08 QR55 QR62 QS25 QS34

QX02 QX10