

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7169290号
(P7169290)

(45)発行日 令和4年11月10日(2022.11.10)

(24)登録日 令和4年11月1日(2022.11.1)

(51)国際特許分類

C 1 2 Q	1/6813(2018.01)	C 1 2 Q	1/6813	Z
C 1 2 Q	1/6851(2018.01)	C 1 2 Q	1/6851	Z
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z
C 1 2 N	15/09 (2006.01)	C 1 2 N	15/09	Z

F I

請求項の数 25 (全85頁)

(21)出願番号 特願2019-552517(P2019-552517)
 (86)(22)出願日 平成30年3月20日(2018.3.20)
 (65)公表番号 特表2020-513837(P2020-513837
 A)
 (43)公表日 令和2年5月21日(2020.5.21)
 (86)国際出願番号 PCT/US2018/023387
 (87)国際公開番号 WO2018/175458
 (87)国際公開日 平成30年9月27日(2018.9.27)
 審査請求日 令和3年3月22日(2021.3.22)
 (31)優先権主張番号 62/476,522
 (32)優先日 平成29年3月24日(2017.3.24)
 (33)優先権主張国・地域又は機関
 米国(US)

(73)特許権者 595117091
 ベクトン・ディキンソン・アンド・カン
 パニー
 BECTON, DICKINSON A
 ND COMPANY
 アメリカ合衆国 ニュー・ジャージー 0
 7417-1880 フランクリン・レ
 イクス ベクトン・ドライブ 1
 1 BECTON DRIVE, FRA
 NKLIN LAKES, NEW JE
 RSEY 07417-1880, UN
 ITED STATES OF AMER
 ICA
 (74)代理人 100094569
 弁理士 田中 伸一郎

最終頁に続く

(54)【発明の名称】 マルチプレットを決定するための合成マルチプレット

(57)【特許請求の範囲】

【請求項1】

マルチプレット発現プロファイルを同定するための方法であつて、
(a)複数の細胞を、複数のマイクロウェルを含む基材上に分配することにより、前記複数のマイクロウェル内に配置された複数の単一細胞を用意する工程；

(b)複数の確率バーコードを用いて、前記基材上に分配された前記複数の細胞における複数の標的に確率バーコードを付けて、前記複数の細胞の各細胞について複数の確率バーコード付き標的を生成する工程、ここで、前記複数の確率バーコードの各々は細胞標識及び分子標識を含み、前記複数の確率バーコードの少なくとも2つの確率バーコードの分子標識は異なる分子標識配列を含み、且つ前記複数の確率バーコードの少なくとも2つの確率バーコードは同一の細胞標識配列を有する細胞標識を含む；

(c)前記複数の確率バーコード付き標的のシーケンシングデータを取得する工程；

(d) (c)で取得された前記シーケンシングデータから前記複数の確率バーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程、ここで、前記複数の確率バーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルは、前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を含む；

(e) (d)で決定された前記複数の確率バーコードの前記細胞標識に関連付けられた前記複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程；及び

(f) 前記複数の確率バーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルを、(e)で作成された前記複数の合成マルチプレット発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程、を含む、方法。

【請求項2】

前記発現プロファイルが(f)におけるマルチプレットとして同定された場合には、前記発現プロファイルに関連付けられたシーケンシングデータを(c)で取得された前記シーケンシングデータから除去する工程を含む、請求項1に記載の方法。

【請求項3】

マルチプレット発現プロファイルを同定するための方法であつて、

(a) 複数の細胞の複数の発現プロファイルを受け取る工程、ここで、前記複数の発現プロファイルは、前記複数の細胞の各細胞について複数の標的の各標的の存在量(occurrence)を含み、前記複数の発現プロファイルの各発現プロファイルは、单一細胞由来の発現プロファイル又は2以上の細胞由来の発現プロファイルである；

(b) 前記複数の細胞の前記複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程；及び

(c) 前記複数の細胞の細胞に関連付けられた前記複数の発現プロファイルの発現プロファイルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程、を含む、方法。

10

20

【請求項4】

前記発現プロファイルが(c)においてマルチプレットとして同定された場合には、前記発現プロファイルを(a)で得られた前記複数の発現プロファイルから除去する工程を含む、請求項3に記載の方法。

【請求項5】

前記複数の細胞の前記複数の発現プロファイルから前記複数の合成マルチプレット発現プロファイルを作成する工程が、

前記複数の合成マルチプレット発現プロファイルの合成マルチプレット発現プロファイルについて、

(1) 前記複数の発現プロファイルの2以上の発現プロファイルを選択する工程と；

(2) (1)で選択された前記発現プロファイルを組合せて、前記合成マルチプレット発現プロファイルを作成する工程と、

を含む、請求項3又は4に記載の方法。

30

【請求項6】

(1) で選択された前記発現プロファイルを組合せて、前記合成マルチプレット発現プロファイルを作成する工程が、

前記複数の標的の各々について、選択された前記発現プロファイルにおける前記標的の存在量を組合せて、前記合成マルチプレット発現プロファイルにおける前記標的の存在量を作成する工程を含む、

請求項5に記載の方法。

40

【請求項7】

前記合成マルチプレット発現プロファイルにおける前記標的の前記存在量が、

(i) 選択された前記発現プロファイルにおける前記標的の前記存在量の合計であるか、或いは

(ii) 選択された前記発現プロファイルにおける前記標的の前記存在量の加重合計であるか、或いは

(iii) 選択された前記発現プロファイルにおける前記標的の前記存在量の平均であるか、或いは

(iv) 選択された前記発現プロファイルにおける前記標的の前記存在量の加重平均である、

50

請求項 6 に記載の方法。

【請求項 8】

前記複数の合成マルチプレット発現プロファイルの数が、(a) で得られた前記複数の発現プロファイルの約 10 パーセントである、請求項 3 ~ 7 のいずれか一項に記載の方法。

【請求項 9】

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 発現プロファイルマルチプレットの同定のための機械学習モデルを、(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び(a) で得られた前記複数の発現プロファイルの 1 つ以上の発現プロファイルから訓練する工程と；

(2) 前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、
を含む、請求項 3 ~ 8 のいずれか一項に記載の方法。

【請求項 10】

前記機械学習モデルを訓練する際に用いられる前記複数の発現プロファイルの前記 1 つ以上の発現プロファイルが、(a) で得られた発現プロファイルの約 10 パーセントを含む、請求項 9 に記載の方法。

10

【請求項 11】

前記機械学習モデルが、相關ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スペース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含む、請求項 9 又は 10 に記載の方法。

20

【請求項 12】

発現プロファイルマルチプレットの同定のための前記機械学習モデルを、(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び(a) で得られた前記複数の発現プロファイルの 1 つ以上の発現プロファイルから訓練する工程が、

30

(1) (b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程と；

(2) (a) で得られた前記複数の発現プロファイルの前記 1 つ以上の発現プロファイルを前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の発現プロファイルの 1 つ以上の投影された発現プロファイルを作成する工程と；

(3) 発現プロファイルマルチプレットの同定のための前記機械学習モデルを、(1) からの前記複数の合成マルチプレット発現プロファイルの前記投影された発現プロファイル及び(2) における前記複数の発現プロファイルの前記 1 つ以上の投影された発現プロファイルから訓練する工程と、

40

を含む、請求項 9 ~ 11 のいずれか一項に記載の方法。

【請求項 13】

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを投影して、前記複数の発現プロファイルの投影された発現プロファイルを作成する工程を含み、

ここで前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記複数の発現プロファイルの前記投影さ

50

れた発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む、

請求項 1 2 に記載の方法。

【請求項 1 4】

(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを、前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの前記投影された発現プロファイルを作成する工程が、
、 t 分布型確率的近傍埋め込み (t S N E) 法を用いる工程を含む、請求項 1 2 又は 1 3 に記載の方法。

【請求項 1 5】

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと前記複数の発現プロファイルの少なくとも 1 つの発現プロファイルとの間の第 1 の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと前記複数の合成マルチプレット発現プロファイルの少なくとも 1 つの合成マルチプレット発現プロファイルとの間の第 2 の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む、請求項 3 ~ 1 4 のいずれか一項に記載の方法。

【請求項 1 6】

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数の発現プロファイルを発現プロファイルの第 1 のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの第 2 のクラスターにクラスター化する工程と；

(3) 前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと発現プロファイルの前記第 1 のクラスターとの間の第 1 の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと合成マルチプレット発現プロファイルの第 2 のクラスターとの間の第 2 の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、請求項 3 ~ 1 4 のいずれか一項に記載の方法。

【請求項 1 7】

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数の発現プロファイルを発現プロファイルの第 1 のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの複数の第 2 のクラスターにクラスター化する工程と；

(3) 前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと発

10

20

30

40

50

現プロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと合成マルチプレット発現プロファイルの前記複数の第2のクラスターの1つ以上のクラスターとの間の第2の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、請求項3～14のいずれか一項に記載の方法。

【請求項18】

前記複数の細胞の前記複数の発現プロファイルを受け取る工程が、

(1) 複数のバーコードを用いて、前記複数の細胞における前記複数の標的にバーコードを付けて、前記複数の細胞の細胞について複数のバーコード付き標的を生成する工程、ここで、前記複数のバーコードの各々は細胞標識及び分子標識を含み、前記複数のバーコードの少なくとも2つのバーコードの分子標識は異なる分子標識配列を含み、且つ前記複数のバーコードの少なくとも2つのバーコードは同一の細胞標識配列を有する細胞標識を含む；

(2) 前記複数のバーコード付き標的のシーケンシングデータを取得する工程；及び

(3) (2)で取得された前記シーケンシングデータから前記複数のバーコードの細胞標識に関連付けられた前記複数の発現プロファイルを決定する工程、ここで、前記複数のバーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルは、前記シーケンシングデータ中の前記細胞標識及び前記複数の標的に関連付けられた識別可能な配列を有する分子標識の数を含む、を含む、請求項3～17のいずれか一項に記載の方法。

【請求項19】

前記シーケンシングデータから前記複数のバーコードの前記細胞標識に関連付けられた前記複数の発現プロファイルを決定する工程が、

前記複数のバーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルについて、前記シーケンシングデータ中の前記細胞標識及び前記複数の標的に各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程を含む、請求項18に記載の方法。

【請求項20】

前記シーケンシングデータ中の前記細胞標識及び前記複数の標的に各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程が、

前記複数の標的の1つ以上について、

(1) 前記シーケンシングデータ中の前記標的に関連付けられた識別可能な配列を有する分子標識の数をカウントする工程と；

(2) 前記シーケンシングデータ中の前記標的に関連付けられた識別可能な配列を有する分子標識の数に基づいて前記標的の数を推定する工程と、

を含む、請求項19に記載の方法。

【請求項21】

マルチプレットプロファイルを同定するための方法であって、

(a) 複数の細胞の複数のプロファイルを受け取る工程、ここで、前記複数のプロファイルの各プロファイルは、单一細胞由来のプロファイル又は2以上の細胞由来のプロファイルである；

(b) 前記複数の細胞の前記複数のプロファイルから複数の合成マルチプレットプロファイルを作成する工程；及び

(c) 前記複数の細胞に関連付けられた前記複数のプロファイルのプロファイルを、(b)で作成された前記複数の合成マルチプレットプロファイルのプロファイルに基づいてシングレット又はマルチプレットとして同定する工程、を含む、方法。

【請求項22】

10

20

30

40

50

前記複数の細胞の前記複数のプロファイルのプロファイルが、前記細胞のmRNA発現プロファイル、前記細胞のタンパク質発現プロファイル、前記細胞の突然変異プロファイル、前記細胞のメチル化プロファイル、又はそれらの任意の組合せを含む、請求項21に記載の方法。

【請求項23】

標的の数を決定するためのコンピュータシステムであって、
ハードウェア・プロセッサー；及び
命令(instructions)が保存されている非一時的なメモリーであって、前記ハードウェア
・プロセッサーによって実行されると、前記プロセッサーが請求項1～17、21又は2
2のいずれか一項に記載の方法を実施するように導く、前記非一時的なメモリー
を含む、前記コンピュータシステム。

10

【請求項24】

請求項1～17、21又は22のいずれか一項に記載の方法を実施するためのコードを
含む、コンピュータで読み込み可能なメディア。

【請求項25】

前記複数の発現プロファイルが、以下の工程：

前記複数の細胞を、複数のマイクロウェルを含む基材上に分配することにより、前記複
数のマイクロウェル中に配置された複数の単一細胞を用意する；

前記複数のマイクロウェル中の前記複数の細胞の各細胞における前記複数の標的を標識
する工程；及び

20

標識された複数の標的のシーケンシングデータを得る工程、
によって作成される、請求項3に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本願は、2017年8月24日に出願された米国仮特許出願第62/476,522号
明細書に対して、35U.S.C. § 119(e)の下での優先権を主張する。関連出願の
内容は、ここでその全体が参照により本明細書中に援用される。

【背景技術】

30

【0002】

本開示は、一般に、分子バーコーディングを用いる細胞の発現プロファイルの検出及び
分析の分野、より詳細にはマルチプレット発現プロファイルを同定する分野に関する。

【0003】

確率バーコーディングなどの方法及び技術は、細胞分析にとって有用である。例えば、
確率バーコーディングを用いて、細胞の生理的状態、例えば単細胞のタンパク質及び/又
は遺伝子発現プロファイルを解読し、例えば、逆転写、ポリメラーゼ連鎖反応(PCR)
増幅、及び次世代シーケンシング(NGS)を用いて、それらの状態を決定することができる。
しかし、検出される発現プロファイルは、2つ以上の異なるタイプの細胞に関連す
ることがあり、それは発現プロファイルの解釈を歪める可能性がある。

40

【発明の概要】

【課題を解決するための手段】

【0004】

本明細書に開示されるのは、マルチプレット発現プロファイルを同定するための方法で
ある。いくつかの実施形態では、本方法は、(a)複数のバーコード(例えば確率バーコ
ード)を用いて、複数の細胞における複数の標的にバーコードを付けて(例えば確率バ
ーコードを付けて)、複数の細胞の各細胞に対して複数のバーコード付き標的(例えば確率
バーコード付き標的)を生成する工程であって、複数のバーコードの各々が細胞標識及び
分子標識を含み、複数のバーコードの少なくとも2つのバーコードの分子標識が、異なる
分子標識配列を含み、また複数のバーコードの少なくとも2つのバーコードが、同一の細

50

胞標識配列を有する細胞標識を含む工程と；(b)複数のバーコード付き標的のシーケンシングデータを取得する工程と；(c)(b)で取得されたシーケンシングデータから複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程であつて、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルが、シーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を含む工程と；(d)(c)で決定された複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルから複数の合成マルチプレット発現プロファイル(synthetic multipplet expression profiles)を作成する工程と；(e)複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、(d)で作成された複数の合成ダブルット発現プロファイル(synthetic doublet expression profiles)の発現プロファイルに基づいてシングルレット(singlet)又はマルチプレットとして同定する工程と、を含む。

【0005】

いくつかの実施形態では、本方法は、発現プロファイルが(e)におけるマルチプレットとして同定される場合、発現プロファイルに関連付けられたシーケンシングデータを(b)で取得されたシーケンシングデータから除去する工程を含む。本方法は、発現プロファイルが(e)におけるマルチプレットとして同定される場合、該発現プロファイルを(c)で決定された複数の発現プロファイルから除去する工程を含みうる。複数のマルチプレットは、ダブルット、トリプレット、又はそれらの任意の組合せを含みうる。

【0006】

いくつかの実施形態では、複数の合成マルチプレット発現プロファイルを(c)で決定された複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルから作成する工程は、複数の合成マルチプレット発現プロファイルの合成マルチプレット発現プロファイルについて、(1)(c)で決定された複数の発現プロファイルの、第1の細胞標識配列に関連付けられた第1の発現プロファイルを選択する工程と；(2)(c)で決定された複数の発現プロファイルの、第2の細胞標識配列に関連付けられた第2の発現プロファイルを選択する工程と；(3)(1)で選択された第1の発現プロファイルと(2)で選択された第2の発現プロファイルとを組合せ、合成マルチプレット発現プロファイルを作成する工程と、を含み、ここで第1の細胞標識配列及び第2の細胞標識配列は、異なる細胞標識配列を含む。

【0007】

いくつかの実施形態では、(1)で選択された第1の発現プロファイルと(2)で選択された第2の発現プロファイルとを組合せ、合成マルチプレット発現プロファイルを作成する工程は、複数の標的の各々について、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数とを組合せ、合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数を作成する工程を含む。

【0008】

いくつかの実施形態では、合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との合計である。該合計は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との加重合計でありうる。

【0009】

いくつかの実施形態では、合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数は、第1の発現プロファイルにおける標的に関連付けら

10

20

30

40

50

れた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との平均である。該平均は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との加重平均でありうる。

【0010】

いくつかの実施形態では、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、(d)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、(1)発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(d)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び(c)で決定された複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程と；(2)複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。機械学習モデルを訓練する際に用いられる複数の発現プロファイルの1つ以上の発現プロファイルは、(c)で決定された複数の発現プロファイルの百分率を含みうる。該百分率は約10%でありうる。

【0011】

いくつかの実施形態では、該機械学習モデルは、分類モデルを含む。分類モデルは、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含みうる。機械学習モデルは、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク(Naïve Bayes network)、k近傍法(KNN)モデル、k平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含みうる。機械学習モデルは、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スパース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含みうる。

【0012】

いくつかの実施形態では、発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(d)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び(c)で決定された複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程は、(1)(d)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程と；(2)(c)で決定された複数の発現プロファイルの1つ以上の発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の発現プロファイルの1つ以上の投影された発現プロファイルを作成する工程と；(3)発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(1)からの複数の合成マルチプレット発現プロファイルの投影された発現プロファイル及び(2)における複数の発現プロファイルの1つ以上の投影された発現プロファイルから訓練する工程と、を含む。

【0013】

いくつかの実施形態では、本方法は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数のバーコードの発現プロファイルを投影して、複数のバーコードの投影された発現プロファイルを作成する工程を含み、ここで複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイル及び複数のバーコードの投影された発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工

10

20

30

40

50

程を含む。

【 0 0 1 4 】

いくつかの実施形態では、低次元空間は、二次元空間である。(d)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程は、t分布型確率的近傍埋め込み(t S N E)法を用いて、(d)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程を含みうる。

【 0 0 1 5 】

いくつかの実施形態では、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、作成された複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルと複数の発現プロファイルの少なくとも1つの発現プロファイルとの間の第1の距離、及び細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルと複数の合成マルチプレット発現プロファイルの少なくとも1つの合成マルチプレット発現プロファイルとの間の第2の距離に基づく発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む。

10

【 0 0 1 6 】

いくつかの実施形態では、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、作成された複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、(1)複数の発現プロファイルを発現プロファイルの第1のクラスターにクラスター化する工程と；(2)複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの第2のクラスターにクラスター化する工程と；(3)複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルと発現プロファイルの第1のクラスターとの間の第1の距離、及び細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルと合成マルチプレット発現プロファイルの第2のクラスターとの間の第2の距離に基づく発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

20

【 0 0 1 7 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のプロファイルをプロファイルの第1のクラスターにクラスター化する工程と；(2)複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの複数の第2のクラスターにクラスター化する工程と；(3)細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルとプロファイルの第1のクラスターとの間の第2の距離、及び細胞に関連付けられた複数のプロファイルのプロファイルと合成マルチプレットプロファイルの複数の第2のクラスターの1つ以上の第2のクラスターとの間の第2の距離に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

30

【 0 0 1 8 】

いくつかの実施形態では、複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の細胞における複数のバーコード付き標的を生成する工程は、複数の粒子の複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の各細胞における複数のバーコード付き標的を生成する工程であ

40

50

つて、複数の粒子の各々が複数のバーコードのサブセットを含み、ここでバーコードサブセットの各々が同一の細胞標識配列を含み且つ少なくとも 100 の異なる分子標識配列を有する工程を含む。

【 0019 】

いくつかの実施形態では、該粒子は、ビーズである。ビーズは、ストレプトアビジンビーズ、アガロースビーズ、磁気ビーズ、コンジュゲートビーズ、プロテイン A コンジュゲートビーズ、プロテイン G コンジュゲートビーズ、プロテイン A / G コンジュゲートビーズ、プロテイン L コンジュゲートビーズ、オリゴ T コンジュゲートビーズ、シリカビーズ、シリカ様ビーズ、抗ビオチンマイクロビーズ、抗蛍光色素マイクロビーズ、及びそれらの任意の組合せからなる群から選択されうる。該粒子は、ポリジメチルシロキサン (P D M S)、ポリスチレン、ガラス、ポリプロピレン、アガロース、ゼラチン、ヒドロゲル、常磁性体、セラミック、プラスチック、ガラス、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、セルロース、ナイロン、シリコーン、及びそれらの任意の組合せからなる群から選択される材料を含みうる。

10

【 0020 】

いくつかの実施形態では、複数の粒子の各々のバーコードは、少なくとも 1000 又は 10000 の異なる分子標識配列を有する分子標識を含む。バーコードの分子標識は、ランダム配列を含みうる。複数の粒子の各々は、少なくとも 10000 のバーコードを含みうる。

20

【 0021 】

いくつかの実施形態では、複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の各細胞における複数のバーコード付き標的を生成する工程は、(1) 標的のコピーをバーコードの標的結合領域と接触させる工程と；(2) 複数のバーコードを用いて複数の標的を逆転写して、複数の逆転写された標的を生成する工程と、を含む。

20

【 0022 】

いくつかの実施形態では、複数のバーコード付き標的のシーケンシングデータを取得する前に、バーコード付き標的を増幅し、複数の増幅されたバーコード付き標的を作成する。バーコード付き標的を増幅し、複数の増幅されたバーコード付き標的を作成する工程は、ポリメラーゼ連鎖反応 (P C R) によりバーコード付き標的を増幅する工程を含みうる。

30

【 0023 】

いくつかの実施形態では、(b) で取得されたシーケンシングデータから複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程は、複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルについて、(b) で取得されたシーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程を含む。(b) で取得されたシーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程は、複数の標的の 1 つ以上について、(1) (b) で取得されたシーケンシングデータ中の標的に関連付けられた識別可能な配列を有する分子標識の数をカウントする工程と；(2) (1) でカウントされたシーケンシングデータ中の標的に関連付けられた識別可能な配列を有する分子標識の数に基づいて標的の数を推定する工程と、を含みうる。

40

【 0024 】

本明細書では、マルチプレット発現プロファイルを同定するための方法が開示される。いくつかの実施形態では、本方法は、(a) 複数の細胞の複数の発現プロファイルを受ける工程であって、発現プロファイルが複数の細胞の各細胞における複数の標的の各標的の存在量 (occurrence) (又はコピー若しくは数) を含む工程と；(b) 複数の細胞の複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程と；(c) 複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシング

50

レット又はマルチプレットとして同定する工程と、を含む。

【0025】

いくつかの実施形態では、本方法は、発現プロファイルが(c)におけるマルチプレットとして同定される場合、該発現プロファイルを(a)で得られた複数の発現プロファイルから除去する工程を含む。複数のマルチプレットは、ダブレット、トリプレット、又はそれらの任意の組合せを含みうる。

【0026】

いくつかの実施形態では、複数の細胞の複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程は、複数の合成マルチプレット発現プロファイルの合成マルチプレット発現プロファイルについて、(1)複数の発現プロファイルのいくつか(a number of)の発現プロファイルを選択する工程と；(2)(1)で選択された発現プロファイルを組合せ、合成マルチプレット発現プロファイルを作成する工程と、を含みうる。(1)で選択された発現プロファイルを組合せ、合成マルチプレット発現プロファイルを作成する工程は、複数の標的の各々について、選択された発現プロファイルにおける標的の存在量を組合せ、合成マルチプレット発現プロファイルにおける標的の存在量を作成する工程を含みうる。

10

【0027】

いくつかの実施形態では、合成マルチプレット発現プロファイルにおける標的の存在量は、選択された発現プロファイルにおける標的の存在量の合計でありうる。該合計は、選択された発現プロファイルにおける標的の存在量の加重合計でありうる。いくつかの実施形態では、合成マルチプレット発現プロファイルにおける標的の存在量は、選択された発現プロファイルにおける標的の存在量の平均である。該平均は、選択された発現プロファイルにおける標的の存在量の加重平均でありうる。

20

【0028】

いくつかの実施形態では、複数の合成マルチプレット発現プロファイルの数は、およそ(a)で得られた複数の発現プロファイルの百分率である。該百分率は約10%でありうる。

【0029】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び該発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、(1)発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び(a)で得られた複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程と；(2)複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

30

【0030】

いくつかの実施形態では、機械学習モデルを訓練する際に用いられる複数の発現プロファイルの1つ以上の発現プロファイルは、(a)で得られた複数の発現プロファイルの百分率を含む。該百分率は約10%でありうる。

40

【0031】

いくつかの実施形態では、機械学習モデルは分類モデルを含む。分類モデルは、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの任意の組合せを含みうる。機械学習モデルは、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k近傍法(KNN)モデル、k平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含みうる。機械学習モデルは、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スパース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモ

50

ル、又はそれらの任意の組合せを含みうる。

【 0 0 3 2 】

いくつかの実施形態では、発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び(a)で得られた複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程は、(1) (b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程と；(2) (a)で得られた複数の発現プロファイルの1つ以上の発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の発現プロファイルの1つ以上の投影された発現プロファイルを作成する工程と；(3)発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(1)からの複数の合成マルチプレット発現プロファイルの投影された発現プロファイル及び(2)における複数の発現プロファイルの1つ以上の投影された発現プロファイルから訓練する工程と、を含む。

10

【 0 0 3 3 】

いくつかの実施形態では、本方法は、複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを投影して、複数の発現プロファイルの投影された発現プロファイルを作成する工程を含み、ここで複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて複数の発現プロファイルの投影された発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む。低次元空間は二次元空間でありうる。

20

【 0 0 3 4 】

いくつかの実施形態では、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程は、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程を含み、t分布型確率的近傍埋め込み(t S N E)法を用いる工程を含む。

30

【 0 0 3 5 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び該発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、細胞に関連付けられた複数の発現プロファイルの発現プロファイルと複数の発現プロファイルの少なくとも1つの発現プロファイルとの間の第1の距離、及び細胞に関連付けられた複数の発現プロファイルの発現プロファイルと複数の合成マルチプレット発現プロファイルの少なくとも1つの合成マルチプレット発現プロファイルとの間の第2の距離に基づく発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む。

40

【 0 0 3 6 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び該発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、(1)複数の発現プロファイルを発現プロファイルの第1のクラスターにクラスター化する工程と；(2)複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの第2のクラスターにクラスター化する工程と；(3)細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、細胞に

50

関連付けられた複数の発現プロファイルの発現プロファイルと発現プロファイルの第1のクラスターとの間の第1の距離、及び細胞に関連付けられた複数の発現プロファイルの発現プロファイルと合成マルチプレット発現プロファイルの第2のクラスターとの間の第2の距離に基づく発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

【0037】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、(b)で作成された複数の合成マルチプレット発現プロファイルの発現プロファイル及び該発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、(1)複数の発現プロファイルを発現プロファイルの第1のクラスターにクラスター化する工程と；(2)複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの複数の第2のクラスターにクラスター化する工程と；(3)細胞の細胞に関連付けられた複数の発現プロファイルの発現プロファイルを、細胞に関連付けられた複数の発現プロファイルの発現プロファイルと発現プロファイルの第1のクラスターとの間の第1の距離、及び細胞に関連付けられた複数の発現プロファイルの発現プロファイルと合成マルチプレット発現プロファイルの複数の第2のクラスターの1つ以上のクラスターとの間の第2の距離に基づく発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、含む。

【0038】

いくつかの実施形態では、複数の細胞の複数の発現プロファイルを受け取る工程は、(1)複数のバーコード(例えば確率バーコード)を用いて複数の細胞における複数の標的にバーコードを付けて(例えば確率バーコードを付けて)、複数の細胞の細胞において複数のバーコード付き標的(例えば確率バーコード付き標的)を生成する工程であって、複数のバーコードの各々が細胞標識及び分子標識を含み、複数のバーコードの少なくとも2つのバーコードの分子標識が異なる分子標識配列を含み、且つ複数のバーコードの少なくとも2つのバーコードが同一の細胞標識配列を有する細胞標識を含む工程と；(2)複数のバーコード付き標的のシーケンシングデータを取得する工程と；(3)(2)で取得されたシーケンシングデータから複数の確率バーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程であって、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルがシーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を含む工程と、を含む。

【0039】

いくつかの実施形態では、複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の細胞において複数のバーコード付き標的を生成する工程は、複数の粒子の複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の細胞において複数のバーコード付き標的を生成する工程であって、複数の粒子の各々が複数のバーコードのサブセットを含み、ここでバーコードのサブセットの各々が同一の細胞標識配列を含み且つ少なくとも100の異なる分子標識配列を有する工程を含む。

【0040】

いくつかの実施形態では、該粒子は、ビーズである。ビーズは、ストレプトアビジンビーズ、アガロースビーズ、磁気ビーズ、コンジュゲートビーズ、プロテインAコンジュゲートビーズ、プロテインGコンジュゲートビーズ、プロテインA/Gコンジュゲートビーズ、プロテインLコンジュゲートビーズ、オリゴTコンジュゲートビーズ、シリカビーズ、シリカ様ビーズ、抗ビオチンマイクロビーズ、抗蛍光色素マイクロビーズ、及びそれらの任意の組合せからなる群から選択されうる。該粒子は、ポリジメチルシロキサン(PDMS)、ポリスチレン、ガラス、ポリプロピレン、アガロース、ゼラチン、ヒドロゲル、常磁性体、セラミック、プラスチック、ガラス、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、セルロース、ナイロン、シリコーン、及びそれらの任

10

20

30

40

50

意の組合せからなる群から選択される材料を含みうる。

【0041】

いくつかの実施形態では、複数の粒子の各々のバーコードは、少なくとも1000又は10000の異なる分子標識配列を有する分子標識を含む。バーコードの分子標識は、ランダム配列を含みうる。複数の粒子の各々は、少なくとも10000のバーコードを含みうる。

【0042】

いくつかの実施形態では、複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の各細胞における複数のバーコード付き標的を生成する工程は、(1)標的のコピーをバーコードの標的結合領域と接触させる工程と；(2)複数のバーコードを用いて複数の標的を逆転写して、複数の逆転写された標的を生成する工程と、を含む。

10

【0043】

いくつかの実施形態では、本方法は、複数のバーコード付き標的のシーケンシングデータを取得する前に、バーコード付き標的を增幅し、複数の増幅されたバーコード付き標的を作成する工程を含む。バーコード付き標的を增幅し、複数の増幅されたバーコード付き標的を作成する工程は、ポリメラーゼ連鎖反応(PCR)によりバーコード付き標的を増幅する工程を含みうる。

【0044】

いくつかの実施形態では、複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルをシーケンシングデータから決定する工程は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルについて、シーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程を含む。シーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程は、複数の標的の1つ以上について、(1)シーケンシングデータ中の標的に関連付けられた識別可能な配列を有する分子標識の数をカウントする工程と；(2)シーケンシングデータ中の標的に関連付けられた識別可能な配列を有する分子標識の数に基づいて標的の数を推定する工程と、を含みうる。

20

【0045】

本明細書では、マルチプレットプロファイルを同定するための方法が開示される。いくつかの実施形態では、本方法は、(a)複数の細胞の複数のプロファイルを受け取る工程と；(b)複数の細胞の複数のプロファイルから複数の合成マルチプレットプロファイルを作成する工程と；(c)複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。複数の細胞の複数のプロファイルのプロファイルは、細胞のmRNA発現プロファイル、細胞のタンパク質発現プロファイル、細胞の突然変異プロファイル、細胞のメチル化プロファイル、又はそれらの任意の組合せを含みうる。

30

【0046】

いくつかの実施形態では、mRNA発現プロファイルは、複数の細胞の各細胞における複数の遺伝子の各遺伝子のmRNA分子の存在量を含みうる。各遺伝子の存在量は、遺伝子の絶対存在量、遺伝子の正規化された存在量、又はそれらの組合せを含みうる。遺伝子の正規化された存在量は、マッピングされたリード100万あたりの転写物1キロベース当たりのリード(Reads Per Kilobase of transcript per Million mapped reads)(RPKM)の単位又は閾値数(Ct)の単位で決定されうる。mRNA発現プロファイルは、シーケンシング、定量ポリメラーゼ連鎖反応(qPCR)、デジタルPCR、ハイブリダイゼーション、又はそれらの任意の組合せにより決定されうる。いくつかの実施形態では、細胞のタンパク質発現プロファイルは、複数の細胞の各細胞における複数の遺伝子の各遺伝子に対応するタンパク

40

50

質分子の存在量を含む。細胞の突然変異プロファイルは、細胞の複数のゲノム位置での細胞の突然変異プロファイルを含みうる。細胞のメチル化プロファイルは、細胞の複数のゲノム位置での細胞のメチル化プロファイルを含みうる。

【 0 0 4 7 】

いくつかの実施形態では、本方法は、該プロファイルが（c）におけるマルチプレットとして同定される場合、該プロファイルを（a）で得られた複数のプロファイルから除去する工程を含む。複数のマルチプレットは、ダブルレット、トリプレット、又はそれらの任意の組合せを含みうる。いくつかの実施形態では、複数の細胞の複数のプロファイルから複数の合成マルチプレットプロファイルを作成する工程は、複数の合成マルチプレットプロファイルの合成マルチプレットプロファイルについて、（1）複数のプロファイルの、いくつか（*a number of*）のプロファイルを選択する工程と；（2）（1）で選択されたプロファイルを組合せ、合成マルチプレットプロファイルを作成する工程と、を含む。

10

【 0 0 4 8 】

いくつかの実施形態では、（1）で選択されたプロファイルを組合せ、合成マルチプレットプロファイルを作成する工程は、複数の標的の各々について、選択されたプロファイルにおける対応する要素の値を組合せ、合成マルチプレットプロファイルにおける対応する要素の値を作成する工程を含む。合成マルチプレットプロファイルにおける対応する要素の値は、選択されたプロファイルにおける対応する要素の値の合計でありうる。該合計は、選択されたプロファイルにおける対応する要素の値の加重合計でありうる。合成マルチプレットプロファイルにおける標的の存在量は、選択されたプロファイルにおける対応する要素の値の平均でありうる。該平均は、選択されたプロファイルにおける対応する要素の値の加重平均でありうる。

20

【 0 0 4 9 】

いくつかの実施形態では、複数の合成マルチプレットプロファイルの数は、およそ（a）で得られた複数のプロファイルの百分率である。該百分率は、約10パーセントでありうる。

【 0 0 5 0 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、（b）で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、（1）プロファイルマルチプレットの同定を意図して機械学習モデルを、（b）で作成された複数の合成マルチプレットプロファイルのプロファイル及び（a）で得られた複数のプロファイルの1つ以上のプロファイルから訓練する工程と；（2）複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、機械学習モデルを用いてプロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

30

【 0 0 5 1 】

いくつかの実施形態では、機械学習モデルを訓練する際に用いられる複数のプロファイルの1つ以上のプロファイルは、（b）で得られた複数のプロファイルの百分率を含む。該百分率は、約10パーセントでありうる。機械学習モデルは、分類モデルを含みうる。

【 0 0 5 2 】

いくつかの実施形態では、分類モデルは、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含む。機械学習モデルは、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k近傍法（KNN）モデル、k平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含みうる。機械学習モデルは、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スパース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含みうる。

40

【 0 0 5 3 】

いくつかの実施形態では、プロファイルマルチプレットの同定を意図して機械学習モ

50

ルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び(a)で得られた複数のプロファイルの1つ以上のプロファイルから訓練する工程は、(1)(b)で作成された複数の合成マルチプレットプロファイルのプロファイルをプロファイル空間から低次元投影空間に投影して、複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程と；(2)(a)で得られた複数のプロファイルの1つ以上のプロファイルをプロファイル空間から低次元投影空間に投影して、複数のプロファイルの1つ以上の投影されたプロファイルを作成する工程と；(3)プロファイルマルチプレットの同定を意図して機械学習モデルを、(1)からの複数の合成マルチプレットプロファイルの投影されたプロファイル及び(1)における複数のプロファイルの1つ以上の投影されたプロファイルから訓練する工程と、を含む。

10

【0054】

いくつかの実施形態では、本方法は、複数の細胞の細胞に関連付けられた複数の複数のプロファイルのプロファイルを投影して、複数のプロファイルの投影されたプロファイルを作成する工程を含み、ここで複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、機械学習モデルを用いてプロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、機械学習モデルを用いて複数のプロファイルの投影されたプロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む。低次元空間は二次元空間でありうる。いくつかの実施形態では、(b)で作成された複数の合成マルチプレットプロファイルのプロファイルをプロファイル空間から低次元投影空間に投影して、複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程は、(b)で作成された複数の合成マルチプレットプロファイルのプロファイルをプロファイル空間から低次元投影空間に投影して、複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程を含み、t分布型確率的近傍埋め込み(t-SNE)法を用いる工程を含む。

20

【0055】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルと複数のプロファイルの少なくとも1つのプロファイルとの間の第1の距離、及び細胞に関連付けられた複数の発現プロファイルのプロファイルと複数の合成マルチプレット発現プロファイルの少なくとも1つの合成マルチプレット発現プロファイルとの間の第2の距離に基づいてシングレット又はマルチプレットとして同定する工程を含む。

30

【0056】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のプロファイルをプロファイルの第1のクラスターにクラスター化する工程と；(2)複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの第2のクラスターにクラスター化する工程と；(3)細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルとプロファイルの第1のクラスターとの間の第2の距離、及び細胞に関連付けられた複数のプロファイルのプロファイルと合成マルチプレットプロファイルの第2のクラスターとの間の第2の距離に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

40

【0057】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は

50

、複数のプロファイルをプロファイルの第1のクラスターにクラスター化する工程と；(2)複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの複数の第2のクラスターにクラスター化する工程と；(3)細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルとプロファイルの第1のクラスターとの間の第2の距離、及び細胞に関連付けられた複数のプロファイルのプロファイルと合成マルチプレットプロファイルの複数の第2のクラスターの1つ以上のクラスターとの間の第2の距離に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

【0058】

本明細書では、標的の数を決定するためのシステムが開示される。いくつかの実施形態では、該システムは、ハードウェア・プロセッサー；及び命令(*instructions*)が保存されている非一時的なメモリーであって、ハードウェア・プロセッサーによって実行されるとき、プロセッサーが上記のいずれかの方法の方法を実施するように導く、前記非一時的なメモリーを含む。本明細書では、標的の数を決定するためのコンピュータで読み込み可能なメディアが開示される。いくつかの実施形態では、コンピュータで読み込み可能なメディアは、上記の任意の方法を実施するためのコードを含む。

10

【図面の簡単な説明】

【0059】

【図1】非限定的な例示的確率バーコードを示す。

20

【図2】非限定的な例示的確率バーコーディング及びディジタルカウンティングを示す。

【図3】複数の標的からの確率バーコード標的のインデックス付きライブラリーを作製するための非限定的な例示的プロセスを示す概略図である。

【図4】図4は、合成ダブレット分析を分析パイプラインに組み込む非限定的な例示的ワークフローを示す。

【図5】図5は、合成ダブレット分析の非限定的な例示的方法を示すフローチャートである。

【図6】本開示の方法に使用するのに好適な非限定的な例示的機器を示す。

【図7】本開示の実施形態に関連して使用することができるコンピュータシステムの非限定的な例示的構造を示す。

【図8】本開示の方法に使用するのに好適な複数のコンピュータシステムを含むネットワークを示す非限定的な例示的構造を図示する。

30

【図9】本開示の方法に従う共有仮想アドレスメモリー空間を用いるマルチプロセッサー・コンピュータシステムの非限定的な例示的構造を示す。

【図10】図10Aは、ほぼ同数のJ u r k a t細胞及びR a m o s細胞を含むサンプル中の単細胞の発現プロファイルの非限定的な例示的t S N E投影プロットを示す。t S N E投影プロットは、J u r k a t細胞の発現プロファイルに対応するクラスター、R a m o s細胞の発現プロファイルに対応するクラスター、及び混合細胞型のダブルト発現プロファイルに対応するクラスターを含む。図10Bは、図10Aにおける発現プロファイルの非限定的な例示的t S N E投影プロットと、合成ダブルト発現プロファイルが2%であることを示す。合成ダブルト発現プロファイルに対応するクラスターは、図10Aにおけるダブルト発現プロファイルに対応するクラスターと重複する。

40

【図11】図11は、ほぼ同数のJ u r k a t細胞、K 5 6 2細胞、及びR a m o s細胞を含むサンプル中の単細胞の発現プロファイルの非限定的な例示的t S N E投影プロットを示す。合成ダブルト発現プロファイルに対応するクラスターは、観察された混合細胞型のダブルト発現プロファイルに対応するクラスターと重複する。

【図12】図12は、ヒトP B M Cサンプル中の単細胞から得られた発現プロファイルの非限定的な例示的t S N E投影プロットを示す。

【図13】図13は、12のサンプルからの単細胞の発現プロファイルの非限定的な例示的t S N E投影プロットである。

【図14】図14は、非限定的な例示的ユーザーインターフェースを用いる、合成ダブルト

50

ット発現プロファイルを作成、可視化、及び除去する非限定的な例示的ワークフローを示す。

【発明を実施するための形態】

【0060】

以下の詳細な説明では、その一部を成す添付の図面を参照にする。これら図面において、類似する符号は、文脈から他の解釈が要求されない限り、一般に、類似の構成要素を同一のものとみなす。詳細な説明、図面、及び特許請求の範囲に記載される例示的な実施形態は、限定的であることを意味しない。本明細書に提示される主題の精神又は範囲から逸脱することなく、他の実施形態を使用してもよく、また他の変更を実施してもよい。本明細書に概略的に記載され、図面に図示されるように、本開示の態様は、非常に多様な異なる構成で配置、代替、組合せ、分離、及び設計することができ、それらのすべては、本明細書において明示的に考慮され、本開示の一部を成すものとすることを理解されたい。

10

【0061】

本明細書で参照にされるすべての特許、公開特許出願、他の刊行物、並びに GenBank 及び他のデータベースからの配列は、関連技術に関してその全体を参照により組み込むものとする。

【0062】

少数の核酸又は標的、例えばメッセンジャーRNA (mRNA) 分子を定量化することは、例えば、さまざまな発生段階又はさまざまな環境条件下で細胞において発現される遺伝子を決定するために、臨床上重要である。しかし、特に、分子数が非常に小さい場合、核酸分子（例えば mRNA 分子）の絶対数を決定するのは極めて困難となりうる。サンプル中の分子の絶対数を決定するための一方法は、デジタルポリメラーゼ連鎖反応 (PCR) である。バーコード、例えば確率バーコードを、ユニーク分子標識 (ML、分子指標 (MI) とも称される) とともに用いて、分子の数をカウントすることができる。各細胞標識にとってユニークである分子標識を有するバーコードを用いて、各細胞内の分子数をカウントすることができる。バーコーディングにおける非限定的な例示的アッセイは、Precise (商標) アッセイ (Becton, Dickinson and Company (Franklin Lakes, NJ)) 又は Rhapsody (商標) アッセイ (Becton, Dickinson and Company (Franklin Lakes, NJ)) を含む。

20

【0063】

Precise (商標) アッセイは、RT 工程中に、サンプル中のすべてのポリ (A) - mRNA とハイブリダイズするために、ポリ (T) オリゴヌクレオチド上に多数、例えば 6561 ~ 65536 のユニーク分子標識を有する確率バーコードの非枯渇プールを用いることができる。分子標識に加えて、バーコードの細胞標識を用いて、Rhapsody (商標) プレートの各ウェル内の各単細胞を同定することができる。バーコードは、ユニバーサル PCR プライミング部位を含みうる。RT の最中に、標的遺伝子分子は、バーコードとランダムに反応する。各標的分子は、バーコードとハイブリダイズして、バーコード付き、例えば確率バーコード付きの相補的リボヌクレオチド酸 (cDNA) 分子を生成しうる）。標識した後、マイクロウェルプレートのマイクロウェルからのバーコード付き cDNA 分子を、PCR 増幅及びシーケンシングのために単一チューブ中にプールすることができる。未補正のシーケンシングデータを分析して、ユニーク分子標識を有するバーコードの数を取得しうる。

30

【0064】

本明細書では、細胞型を識別するために標的を同定するための方法が開示される。いくつかの実施形態では、本方法は、(a) 標的カウントのデータ構造を受け取る工程であって、標的カウントのデータ構造が複数の細胞の発現プロファイルを含み、ここで複数の細胞の発現プロファイルが複数の細胞の各細胞における複数の標的の各標的の数を含む工程と；(b) 複数の細胞の発現プロファイルを標的カウントのデータ構造及び複数の細胞の発現プロファイル間の距離に基づいて階層的にクラスター化し、複数の細胞の発現プロフ

40

50

イルを表すデンドログラムを作成する工程であって、デンドログラムが複数のノードを含み、ここで複数のノードがルートノード、複数のリーフノード、及び複数の非ルート、非リーフノードを含み、ここで複数のリーフノードの各リーフノードが複数の細胞の異なる細胞の発現プロファイルを表し、且フルートノードが複数の細胞の発現プロファイルを表す工程と；(c) デンドログラムのルートノードからデンドログラムの複数のリーフノードへとデンドログラムの複数のノードの各ノードを横切る間、(1)ノードの、ノードの子供ノードへの分離が有効であるか又は無効であるか(例えば、子供ノード間の差異が有意でない)を判定する工程と；(2)ノードの、ノードの子供ノードへの分離が無効である場合、マージングクラスター¹⁰セットにノードを加える工程と；(d) 反復的に、マージングクラスター¹⁰セットにおける各々の第1のノードについて、マージングクラスター¹⁰セットにおける第1のノードとマージングクラスター¹⁰セットにおける第1のノードに最も近い第2のノードとの間の距離がマージング距離閾値以内である場合、第1のノードを第2のノードとマージし、第1のノード及び第2のノードによって表される発現プロファイルを含むマージされたノードを作成する工程と；(e) マージングクラスター¹⁰セットにおける各ノードについて、ノードによって表される細胞の複数の標的の発現プロファイルに基づき、細胞型を識別するために標的を同定する工程と、を含む。

【0065】

細胞型を識別するため、標的を同定するためのシステムが本明細書で開示される。いくつかの実施形態では、該システムは、ハードウェア・プロセッサー；及び命令(instructions)が保存されている非一時的なメモリーであって、ハードウェア・プロセッサーによって実行されるとき、プロセッサーが本明細書で開示される方法のいずれかを実施するよう導く、前記非一時的なメモリーを含む。本明細書では、細胞型を識別するため、標的を同定するためのコンピュータで読み込み可能なメディアが開示される。いくつかの実施形態では、コンピュータで読み込み可能なメディアは、本明細書で開示される方法のいずれかを実施するためのコードを含む。²⁰

【0066】

定義

特に定義がない限り、本明細書で用いられる技術用語はすべて、本開示が属する分野の当業者により一般に理解されているものと同一の意味を有する。例えば、Singlet³⁰ on et al., Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, NY 1994); Sambrook et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, NY 1989)を参照されたい。本開示の目的のために、下記の用語を以下に定義する。

【0067】

本明細書で用いられる場合、「アダプター」という用語は、関連核酸の増幅又はシーケンシングを促進するための配列を意味しうる。関連核酸は、標的核酸を含みうる。関連核酸は、空間標識、標的標識、サンプル標識、指標標識、バーコード、確率バーコード、又は分子標識の1つ以上を含みうる。アダプターは、線状であってよい。アダプターは、事前にアデニル化されたアダプターであってよい。アダプターは、二本鎖又は一本鎖であつてよい。1つ以上のアダプターは、核酸の5'又は3'末端に配置することができる。アダプターが5'及び3'末端に既知の配列を含む場合、既知の配列は、同じ配列でも、異なる配列でもよい。ポリヌクレオチドの5'及び/又は3'末端に位置するアダプターは、表面上に固定された1つ以上のオリゴヌクレオチドにハイブリダイズする能力を有しうる。アダプターは、いくつかの実施形態では、ユニバーサル配列を含む。ユニバーサル配列は、2つ以上の核酸分子と共通のヌクレオチド配列の1領域であってよい。2つ以上の核酸分子は、異なる配列の領域を有しうる。従って、例えば、5'アダプターは、同一配列及び/又はユニバーサル核酸配列を含み、3'アダプターは、同一配列及び/又はユニバーサル配列を含みうる。複数の核酸分子の異なるメンバー中に存在しうるユニバーサル配列は、ユ⁴⁰

ニバーサル配列と相補的な單一ユニバーサルプライマーを用いて、複数の異なる配列の複製又は増幅を可能にし得る。同様に、核酸分子のコレクションの異なるメンバー中に存在しうる少なくとも1つ、2つ（例えば、ペア）若しくはそれ以上のユニバーサル配列は、ユニバーサル配列と相補的な少なくとも1つ、2つ（例えば、一対）若しくはそれ以上の單一ユニバーサルプライマーを用いて、複数の異なる配列の複製又は増幅を可能にし得る。従って、ユニバーサルプライマーは、こうしたユニバーサル配列とハイブリダイズすることができる配列を含む。標的核酸配列担持分子を修飾して、ユニバーサルアダプター（例えば、非標的核酸配列）を異なる標的核酸配列の一端又は両端に結合させることができる。標的核酸に結合した1つ以上のユニバーサルプライマーは、ユニバーサルプライマーのハイブリダイゼーションのための部位を提供することができる。標的核酸に結合した1つ以上のユニバーサルプライマーは、同じでも、互いに異なってもよい。

10

【0068】

本明細書で用いられる場合、「関連付けられる」又は「～に関連付けられる」という用語は、ある時点で2つ以上の種が共配置されているとして同定可能であることを意味しうる。関連付けは、2つ以上の種が類似の容器内にあることを意味しうる。関連付けは、インフォマティクス的関連付けでありうる。この場合、例えば、2つ以上の種に関するデジタル情報が記憶され、且つその情報を用いてこれらの種の1つ以上が共配置されたことを決定可能である。関連付けはまた、物理的関連付けでありうる。いくつかの実施形態では、2つ以上の関連付けられる種は、互いに又は共通の固体若しくは半固体の表面に「テザー連結」、「結合」、又は「固定」される。関連付けは、ビーズなどの固体又は半固体の支持体に標識を結合するための共有結合手段又は非共有結合手段を意味しうる。関連付けは、標的と標識との共有結合でありうる。

20

【0069】

本明細書で用いられる場合、「相補的」という用語は、2つのヌクレオチド間の精密なペアリングの能力を意味しうる。例えば、核酸の所与の位置のヌクレオチドが他の核酸のヌクレオチドと水素結合可能である場合、2つの核酸はその位置で互いに相補的であるとみなされる。2つの一本鎖核酸分子間の相補性は、ヌクレオチドの一部のみが結合する場合には「部分的」でありうるし、一本鎖分子間のすべてに相補性が存在する場合には完全でありうる。第1のヌクレオチド配列が第2のヌクレオチド配列に相補的である場合、第1のヌクレオチド配列は第2の配列の「相補体」であるといえる。第1のヌクレオチド配列が第2の配列の逆（すなわち、ヌクレオチドの順序が逆）の配列に相補的である場合、第1のヌクレオチド配列は第2の配列の「逆相補体」であるといえる。本明細書で用いられる場合、「相補体」、「相補的」、及び「逆相補体」という用語は、同義的に用いることが可能である。ある分子が他の分子にハイブリダイズしうる場合、それはハイブリダイズしている分子の相補体でありうることが、本開示から理解される。

30

【0070】

本明細書で用いられる場合、「ディジタルカウンティング」という用語は、サンプル中の標的分子の数を推定する方法を意味しうる。ディジタルカウンティングは、サンプル中の標的に関連付けられたユニーク標識の数を決定する工程を含みうる。この確率的方法は、分子をカウントする問題を、同一の分子の位置決定及び同定の問題から、所定の標識のセットの検出に関する一連のあり／なしのディジタル問題に変換する。

40

【0071】

本明細書で用いられる場合、「標識」という用語は、サンプル内の標的に関連付けられる核酸コードを意味しうる。標識は、例えば、核酸標識でありうる。標識は、全体又は一部が増幅可能な標識でありうる。標識は、全体又は一部がシーケンス可能標識でありうる。標識は、個別に同定可能な天然核酸の一部でありうる。標識は、既知の配列でありうる。標識は、核酸配列の接合（例えば、天然配列と非天然配列との接合）を含みうる。本明細書で用いられる場合、「標識」という用語は、「インデックス」、「タグ」、又は「標識タグ」という用語と同義的に用いうる。標識は、情報を伝達可能である。例えば、種々の実施形態では、標識は、サンプル同一性、サンプル源、細胞同一性、及び／又は標的を

50

決定するために使用可能である。

【0072】

本明細書で用いられる場合、「非枯渇リザーバー」という用語は、多種多様な標識から構成されたバーコードのプール（例えば確率バーコード）を意味しうる。非枯渇リザーバーは、非枯渇リザーバーが標的のプールに関連付けられる場合、各標的がユニークバーコード（例えば確率バーコード）に関連付けられる可能性が高くなるように、多数の異なるバーコードを含みうる。各標識標的分子のユニーク性は、ランダム選択の統計により決定可能であり、標識の多様性と比較してコレクション中の同一の標的分子のコピー数に依存する。得られる標識標的分子のセットのサイズは、バーコーディングプロセスの確率的性質により決定可能であり、次いで、検出されたバーコードの数の解析は、元のコレクション又はサンプル中に存在する標的分子の数の計算を可能にする。存在する標的分子のコピー数とユニークバーコードの数との比が低い場合、標識標的分子は極めてユニークである（すなわち、2つ以上の標的分子が1つの所与の標識で標識される確率は非常に低い）。

10

【0073】

本明細書で用いられる場合、「核酸」という用語は、ポリヌクレオチド配列又はその断片を意味する。核酸はヌクレオチドを含みうる。核酸は細胞に対して外因性又は内因性でありうる。核酸は細胞フリー環境中に存在しうる。核酸は遺伝子又はその断片でありうる。核酸はDNAでありうる。核酸はRNAでありうる。核酸は1つ以上のアナログ（例えば、修飾された骨格、糖又は核酸塩基）を含みうる。アナログのいくつかの例としては、限定されるものではないが、5'-プロモウラシル、ペプチド核酸、ゼノ核酸、モルホリノ体、ロックド核酸、グリコール核酸、トレオース核酸、ジデオキシヌクレオチド、コルジセピン、7'-デアザ-GTP、フルオロフオア（例えば、糖に結合されたローダミン又はフルオレセイン）、チオール含有ヌクレオチド、ビオチン結合ヌクレオチド、蛍光塩基アナログ、CpGアイランド、メチル-7'-ゲアノシン、メチル化ヌクレオチド、イノシン、チオウリジン、ブソイドウリジン、ジヒドロウリジン、キューオシン、及びワイオシンが挙げられる。「核酸」、「ポリヌクレオチド」、「標的ポリヌクレオチド」、及び「標的核酸」は、同義的に用いうる。

20

【0074】

核酸は、新しい又は向上した特徴（例えば、向上した安定性）を有する核酸を提供するために1つ以上の修飾（例えば、塩基修飾、骨格修飾）を含みうる。核酸は核酸アフィニティータグを含みうる。ヌクレオシドは塩基-糖の組合せでありうる。ヌクレオシドの塩基部分はヘテロ環塩基でありうる。かかるヘテロ環塩基の2つの最も一般的なクラスはプリン及びピリミジンである。ヌクレオシドは、ヌクレオシドの糖部分に共有結合されたリン酸基をさらに含むヌクレオシドでありうる。ペントフラノシリ糖を含むヌクレオシドでは、リン酸基は、糖の2'、3'、又は5'ヒドロキシリ部分に結合可能である。核酸を形成する際、リン酸基は、隣接ヌクレオシドを互いに共有結合して線状高分子化合物を形成可能である。ひいては、この線状高分子化合物のそれぞれの末端をさらに連結して環状化合物を形成可能である。しかしながら、線状化合物が一般に好適である。その他に、線状化合物は、内部ヌクレオチド塩基相補性を有しうるので、完全二本鎖又は部分二本鎖の化合物を生成するようにフォールディングしうる。核酸内では、リン酸基は、通常、核酸のヌクレオシド間骨格を形成するものとして参照可能である。結合又は骨格は、3'、5'ホスホジエステル結合でありうる。

30

【0075】

核酸は、修飾骨格及び/又は修飾ヌクレオシド間結合を含みうる。修飾骨格は、骨格中にリン原子を保持するもの及び骨格中にリン原子を有していないものを含みうる。リン原子を中に含有する好適な修飾核酸骨格は、例えば、ホスホロチオエート、キラルホスホロチオエート、ホスホロジチオエート、ホスホトリエステル、アミノアルキルホスホトリエステル、3'-アルキレンホスホネートや5'-アルキレンホスホネートなどのメチルや他のアルキルのホスホネート、キラルホスホネート、ホスフィネート、3'-アミノホスホルアミデートやアミノアルキルホスホルアミデートなどのホスホルアミデート、ホスホロジ

40

50

アミデート、チオノホスホルアミデート、チオノアルキルホスホネート、チオノアルキルホスホトリエステル、セレノホスフェート、及び通常 3' - 5' 結合、2' - 5' 結合アナログを有するボラノホスフェート、並びに 1 つ以上のヌクレオチド間結合が 3' - 3' 、 5' - 5' 、又は 2' - 2' 結合である逆極性を有するものを含みうる。

【 0076 】

核酸は、短鎖アルキル若しくはシクロアルキルのヌクレオシド間結合、混合ヘテロ原子及びアルキル若しくはシクロアルキルのヌクレオシド間結合、又は 1 つ以上の短鎖ヘテロ原子若しくはヘテロ環のヌクレオシド間結合により形成されるポリヌクレオチド骨格を含みうる。これらは、モルホリノ結合（ヌクレオシドの糖部分から部分的に形成される）、シロキサン骨格、スルフィド、スルホキシド、及びスルホン骨格、ホルムアセチル及びチオホルムアセチル骨格、メチレンホルムアセチル及びチオホルムアセチル骨格、リボアセチル骨格、アルケン含有骨格、スルファメート骨格、メチレンイミノ及びメチレンヒドロジノ骨格、スルホネート及びスルホンアミド骨格、アミド骨格を有するもの、並びに混合 N、O、S、及び C H₂ 構成部分を有する他のものを含みうる。

10

【 0077 】

核酸は核酸ミメティックを含みうる。「ミメティック」という用語は、フラノース環のみ又はフラノース環とヌクレオチド間結合の両方が非フラノース基で置き換えられているポリヌクレオチドを含むことを意図し得、フラノース環のみの置き換えは、糖サロゲートであるとして参照可能である。ヘテロ環塩基部分又は修飾ヘテロ環塩基部分は、適切な標的核酸とのハイブリダイゼーションのために保持可能である。かかる核酸の 1 つはペプチド核酸（PNA）でありうる。PNA では、ポリヌクレオチドの糖骨格は、アミド含有骨格特にアミノエチルグリシン骨格で置き換え可能である。ヌクレオチドは保持可能であり、且つ骨格のアミド部分のアザ窒素原子に直接的又は間接的に結合される。PNA 化合物中の骨格は、PNA にアミド含有骨格を与える 2 つ以上の結合されたアミノエチルグリシン単位を含みうる。ヘテロ環塩基部分は、骨格のアミド部分のアザ窒素原子に直接的又は間接的に結合可能である。

20

【 0078 】

核酸はモルホリノ骨格構造を含みうる。例えば、核酸は、リボース環の代わりに 6 員モルホリノ環を含みうる。これらの実施形態のいくつかでは、ホスホロジアミデート又は他の非ホスホジエステルのヌクレオシド間結合によりホスホジエステル結合を置き換え可能である。

30

【 0079 】

核酸は、モルホリノ環に結合されたヘテロ環塩基を有する結合されたモルホリノ単位（すなわちモルホリノ核酸）を含みうる。結合基は、モルホリノ核酸中のモルホリノモノマー単位を結合可能である。非イオン性モルホリノ系オリゴマー化合物は、細胞タンパク質とのより少ない望ましくない相互作用を有しうる。モルホリノ系ポリヌクレオチドは、核酸の非イオン性ミミックでありうる。モルホリノクラス内のさまざまな化合物は、異なる結合基を用いて連結可能である。ポリヌクレオチドミメティックのさらなるクラスは、シクロヘキセニル核酸（CeNA）として参照可能である。核酸分子中に通常存在するフラノース環は、シクロヘキセニル環で置き換え可能である。CeNA DMT 保護ホスホロアミダイトモノマーは、ホスホロアミダイト化学を用いたオリゴマー化合物合成のために調製及び使用が可能である。核酸鎖中への CeNA モノマーの取込みは、DNA / RNA ハイブリッドの安定性を増加可能である。CeNA オリゴアデニレートは、天然複合体に類似した安定性を有する核酸相補体との複合体を形成可能である。さらなる修飾は、2' - ヒドロキシル基が糖環の 4' 炭素原子に結合されて 2' - C, 4' - C - オキシメチレン結合を形成することにより二環式糖部分を形成するロックド核酸（LNA）を含みうる。結合は、2' 炭素原子と 4' 炭素原子とを架橋するメチレン（-CH₂-），基（式中、n は 1 又は 2 である）でありうる。LNA 及び LNA アナログは、相補的核酸との非常に高い二本鎖熱安定性（T_m = +3 ~ +10 °C）、3' - エキソヌクレアーゼ分解に対する安定性、及び良好な溶解性を示しうる。

40

50

【0080】

核酸はまた、核酸塩基（単に「塩基」ということが多い）の修飾又は置換を含みうる。本明細書で用いられる場合、「非修飾」又は「天然」の核酸塩基は、プリン塩基（例えば、アデニン（A）及びグアニン（G））、並びにピリミジン塩基（例えば、チミン（T）、シトシン（C）及びウラシル（U））を含みうる。修飾核酸塩基は、他の合成及び天然の核酸塩基、例えば、5-メチルシトシン（5-methyl-C）、5-ヒドロキシメチルシトシン、キサンチン、ヒポキサンチン、2-アミノアデニン、アデニン及びグアニンの6-メチル及び他のアルキル誘導体、アデニン及びグアニンの2-プロピル及び他のアルキル誘導体、2-チオウラシル、2-チオチミン及び2-チオシトシン、5-ハロウラシル及びシトシン、5-プロピニル（-C=C-CH₃）ウラシル及びシトシン、並びにピリミジン塩基の他のアルキニル誘導体、6-アゾウラシル、シトシン及びチミン、5-ウラシル（ブソイドウラシル）、4-チオウラシル、8-ハロ、8-アミノ、8-チオール、8-チオアルキル、8-ヒドロキシル、並びに他の8-置換アデニン及びグアニン、5-ハロ特に5-ブロモ、5-トリフルオロメチル及び他の5-置換ウラシル及びシトシン、7-メチルグアニン及び7-メチルアデニン、2-F-アデニン、2-アミノアデニン、8-アザグアニン及び8-アザアデニン、7-デアザグアニン及び7-デアザアデニン、並びに3-デアザグアニン及び3-デアアデニンを含みうる。修飾核酸塩基は、三環式ピリミジン、例えば、フェノキサジンシチジン（1H-ピリミド（5,4-b）（1,4）ベンゾオキサジン-2（3H）-オン）、フェノチアジンシチジン（1H-ピリミド（5,4-b）（1,4）ベンゾチアジン-2（3H）-オン）、置換フェノキサジンシチジン（例えば、9-（2-アミノエトキシ）-H-ピリミド（5,4-（b）（1,4）ベンゾオキサジン-2（3H）-オン）などのG-クランプ、フェノチアジンシチジン（1H-ピリミド（5,4-b）（1,4）ベンゾチアジン-2（3H）-オン）、置換フェノキサジンシチジン（例えば、9-（2-アミノエトキシ）-H-ピリミド（5,4-（b）（1,4）ベンゾオキサジン-2（3H）-オン）などのG-クランプ、カルバゾールシチジン（2H-ピリミド（4,5-b）インドール-2-オン）、ピリドインドールシチジン（H-ピリド（3',2':4,5）ピロロ[2,3-d]ピリミジン-2-オン）を含みうる。

【0081】

本明細書で用いられる場合、「サンプル」という用語は、標的を含む組成物を意味しる。本開示の方法、デバイス、及びシステムによる分析に好適なサンプルとしては、細胞、組織、器官、又は生物が挙げられる。

【0082】

本明細書で用いられる場合、「サンプリングデバイス」又は「デバイス」という用語は、サンプルのセクションの採取及び/又は基材上へのセクションの配置を行いうるデバイスを意味しる。サンプルデバイスとは、例えば、蛍光活性化細胞選別（FACS）機、セルソーター機、生検針、生検デバイス、組織切片化デバイス、マイクロ流体デバイス、プレードグリッド、及び/又はミクロトームを意味しる。

【0083】

本明細書で用いられる場合、「固体担体」という用語は、複数のバーコード、例えば確率バーコードを結合しる離散した固体又は半固体の表面を意味しる。固体担体は、核酸を（例えば共有結合又は非共有結合で）固定しるプラスチック、セラミック、金属、又は高分子材料（例えばヒドロゲル）で構成された任意のタイプの中実、多孔性、又は中空のスフェア、ボール、ベアリング、シリンダー、又は他の類似の構成体を包含しる。固体担体は、球状（例えばマイクロスフェア）でありうるか又は非球状若しくは不規則形状、例えば、立方体形、直方体形、角錐形、円柱形、円錐形、扁球形、ディスク形などを有しる離散粒子を含みうる。アレイ状に離間して配置された複数の固体担体は、基材を含まないこもありうる。固体担体は、「ビーズ」という用語と同義的に用いうる。

【0084】

固体担体は「基材」を意味しる。基材は固体担体の1種でありうる。基材は、本開示

10

20

20

30

40

50

の方法を行いうる連続した固体又は半固体の表面を意味しうる。基材は、例えば、アレイ、カートリッジ、チップ、デバイス、及びスライドを意味しうる。

【0085】

本明細書で用いられる場合、「空間標識」という用語は、空間内の位置と関連させることができる標識を意味しうる。

【0086】

本明細書で用いられる場合、「バーコード」、例えば「確率バーコード」という用語は、標識を含むポリヌクレオチド配列を意味しうる。バーコードは、バーコーディング（例えば確率バーコーディング）に使用可能なポリヌクレオチド配列でありうる。バーコードは、サンプル中の標的を定量するために使用可能である。バーコードは、標識を標的に関連付けた後に起こりうるエラーの制御に使用可能である。例えば、バーコード、例えば確率バーコードは、增幅又はシーケンシングのエラーを評価するために使用可能である。標的に関連付けられたバーコードは、バーコード標的、例えば確率バーコード標的又は確率バーコードタグ標的と呼ぶことが可能である。

10

【0087】

本明細書で用いられる場合、「遺伝子特異的バーコード」、例えば「遺伝子特異的確率バーコード」という用語は、標識と、遺伝子特異的である標的結合領域とを含むポリヌクレオチド配列を意味しうる。バーコードは、バーコーディング（例えば確率バーコーディング）に使用可能であるポリヌクレオチド配列でありうる。バーコード（例えば確率バーコード）は、サンプル中の標的を定量するために使用可能である。バーコードは、標識を標的に関連付けた後に生じることがあるエラーに対して制御するために使用可能である。例えば、バーコード（例えば確率バーコード）は、增幅又はシーケンシングのエラーを評価するために使用可能である。

20

【0088】

本明細書で用いられる場合、「バーコーディング」、例えば「確率バーコーディング」という用語は、核酸のランダム標識化（例えばバーコーディング）を意味しうる。確率バーコーディングは、標識を標的に関連付けて、標的に関連付けられた標識を定量するため、再帰的ポアソンストラテジーを利用することができる。本明細書で用いられる場合、「バーコーディング」という用語は、「遺伝子特異的バーコーディング」、例えば「遺伝子特異的確率バーコーディング」と置き換え可能に用いることができる。

30

【0089】

本明細書で用いられる場合、「標的」という用語は、バーコード（例えば確率バーコード）に関連付け可能な組成物を意味しうる。本開示の方法、デバイス、及びシステムによる分析に好適な例示的な標的としては、オリゴヌクレオチド、DNA、RNA、mRNA、マイクロRNA、tRNAなどが挙げられる。標的は一本鎖又は二本鎖でありうる。いくつかの実施形態では、標的は、タンパク質、ペプチド、又はポリペプチドでありうる。いくつかの実施形態では、標的は脂質である。本明細書で用いられる場合、「標的」は、「種」と置き換え可能に用いることができる。

【0090】

本明細書で用いられる場合、「逆転写酵素」という用語は、逆転写酵素活性を有する（すなわち、RNA鑄型からのDNAの合成を触媒する）酵素のグループを意味しうる。一般的には、かかる酵素としては、限定されるものではないが、レトロウイルス逆転写酵素、レトロトランスポゾン逆転写酵素、レトロプラスミド逆転写酵素、レトロン逆転写酵素、細菌逆転写酵素、グループI I I型トロン由来逆転写酵素、及びそれらの突然変異体、変異体、又は誘導体が挙げられる。非レトロウイルス逆転写酵素としては、非LTRレトロトランスポゾン逆転写酵素、レトロプラスミド逆転写酵素、レトロン逆転写酵素、及びグループI I I型トロン逆転写酵素が挙げられる。グループI I I型トロン逆転写酵素の例としては、ラクトコッカス・ラクティス（Lactococcus lactis）L1、LtrB型トロン逆転写酵素、サーモシネコッカス（Thermosynechococcus）は、Tef4c型トロン逆転写酵素、又はジオバチルス・ステアロサ

40

50

ーモフィラス (*Geobacillus stearothermophilus*) Gs I - IIC イントロン逆転写酵素を伸長する。他のクラスの逆転写酵素としては、多くのクラスの非レトロウイルス逆転写酵素 (すなわち、レトロン、グループ IIE イントロン、及び特に多様性生成レトロエレメント) が挙げられる。

【 0 0 9 1 】

本明細書では、マルチプレット発現プロファイルを同定するための方法が開示される。いくつかの実施形態では、本方法は、(a) 複数のバーコードを用いて複数の細胞における複数の標的にバーコードを付けて、複数の細胞の細胞における複数のバーコード付き標的を作成する工程であって、複数のバーコードの各々が細胞標識及び分子標識を含み、複数のバーコードの少なくとも 2 つのバーコードの分子標識が異なる分子標識配列を含み、且つ複数のバーコードの少なくとも 2 つのバーコードが同一の細胞標識配列を有する細胞標識を含む工程と ; (b) 複数のバーコード付き標的のシーケンシングデータを取得する工程と ; (c) (b) で取得されたシーケンシングデータから複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程であって、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルがシーケンシングデータ中の細胞標識及び複数の標的に関連付けられた識別可能な配列を有する分子標識の数を含む工程と ; (d) (c) で決定された複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程と ; (e) 複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、(d) で作成された複数の合成ダブルレット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

10

【 0 0 9 2 】

バーコード

バーコーディング、例えば確率バーコーディングは、例えば、米国特許出願公開第 20150299784 号明細書、国際公開第 2015031691 号パンフレット、及び Fue et al., Proc Natl Acad Sci U.S.A. 2011 May 31; 108 (22) : 9026 - 31 (これら出版物の内容はその全体がここで援用される) に記載されている。いくつかの実施形態では、本明細書で開示されるバーコードは、標的を確率的に標識するために用いてもよいポリヌクレオチド配列 (例えば、バーコード、タグ) でありうる確率バーコードでありうる。バーコードは、確率バーコードの異なるバーコード配列の数と標識対象の標的のいずれかの存在量の数との比が、1 : 1、2 : 1、3 : 1、4 : 1、5 : 1、6 : 1、7 : 1、8 : 1、9 : 1、10 : 1、11 : 1、12 : 1、13 : 1、14 : 1、15 : 1、16 : 1、17 : 1、18 : 1、19 : 1、20 : 1、30 : 1、40 : 1、50 : 1、60 : 1、70 : 1、80 : 1、90 : 1、100 : 1、又はこれらの値のいずれか 2 つの間の数若しくは範囲、或いはそうした近似値でありうる場合、確率バーコードと呼ぶことができる。標的は、同一又はほぼ同一の配列を有する m RNA 分子を含む m RNA 種でありうる。バーコードは、確率バーコードの異なるバーコード配列の数と標識対象の標的のいずれかの存在量の数との比が、少なくとも、又は多くとも 1 : 1、2 : 1、3 : 1、4 : 1、5 : 1、6 : 1、7 : 1、8 : 1、9 : 1、10 : 1、11 : 1、12 : 1、13 : 1、14 : 1、15 : 1、16 : 1、17 : 1、18 : 1、19 : 1、20 : 1、30 : 1、40 : 1、50 : 1、60 : 1、70 : 1、80 : 1、90 : 1、若しくは 100 : 1 である場合、確率バーコードと呼ぶことができる。確率バーコードのバーコード配列は、分子標識と呼ぶことができる。

20

30

【 0 0 9 3 】

一つのバーコード、例えば確率バーコードは、1 つ以上の標識を含みうる。例示的な標識としては、ユニバーサル標識、細胞標識、バーコード配列 (例えば分子標識) 、サンプル標識、プレート標識、空間標識、及び / 又はプレ空間標識を挙げることができる。図 1 は、空間標識を有する例示的なバーコード 104 を示す。バーコード 104 は、バーコードを固体担体 105 に連結しうる 5 ' アミンを含んでよい。バーコードは、ユニバーサル標

40

50

識、次元標識、空間標識、細胞標識、及び／又は分子標識を含みうる。バーコード中のさまざまな標識（限定するものではないが、ユニバーサル標識、次元標識、空間標識、細胞標識、及び分子標識など）の順序は変動しうる。例えば、図1に示すように、ユニバーサル標識は、最も5'側の標識であってよく、分子標識は、最も3'側の標識であってもよい。空間標識、次元標識、及び細胞標識は、任意の順序であってよい。いくつかの実施形態では、ユニバーサル標識、空間標識、次元標識、細胞標識、及び分子標識は、任意の順序であってよい。バーコードは、標的結合領域を含みうる。標的結合領域は、サンプル中の標的（例えば、標的核酸、RNA、mRNA、DNA）と相互作用しうる。例えば、標的結合領域は、mRNAのポリ(A)テールと相互作用しうるオリゴ(dT)配列を含みうる。場合によっては、バーコードの標識（例えば、ユニバーサル標識、次元標識、空間標識、細胞標識、及びバーコード配列）は、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、若しくは20又はそれ以上のヌクレオチド分だけ離れてもよい。

【0094】

標識、例えば、細胞標識は、規定長さ、例えば、各々7ヌクレオチド（いくつかのハミングエラー訂正コードに使用されるビット数に相当する）の核酸部分配列の固有のセットを含んでもよく、これらは、エラー訂正能力を賦与するように設計することができる。エラー訂正部分配列のセットは、7つのヌクレオチド配列を含み、これらは、セット内の配列の任意のペア組合せが、規定の「遺伝子距離」（又はミスマッチ塩基の数）を呈示するように、設計することができ、例えば、3ヌクレオチドの遺伝子距離を呈示するように、1セットのエラー訂正部分配列を設計することができる。この場合、標識化標的核酸分子についてのシーケンシングデータのセット内のエラー訂正配列の見直しによって、增幅若しくはシーケンシングエラーを検出又は訂正することができる。いくつかの実施形態では、エラー訂正コードを作製するために用いられる核酸部分配列の長さは、例えば、約1、2、3、4、5、6、7、8、9、10、15、20、30、31、40、50ヌクレオチド長、又はこれらの値のいずれか2つの間の数若しくは範囲であってよい。いくつかの実施形態では、エラー訂正コードを作製するために、他の長さの核酸部分配列を使用することも可能である。

【0095】

バーコードは、標的結合領域を含みうる。標的結合領域は、サンプル中の標的と相互作用することができる。標的は、リボ核酸（RNA）、メッセンジャーRNA（mRNA）、microRNA、低分子干渉RNA（siRNA）、RNA分解産物、各々がポリ(A)テールを含有するRNA、及びそれらの任意の組合せであってもよいし、これらを含んでもよい。いくつかの実施形態では、複数の標的は、デオキシリボ核酸（DNA）を含みうる。

【0096】

いくつかの実施形態では、標的結合領域は、mRNAのポリ(A)テールと相互作用することができるオリゴ(dT)配列を含みうる。確率バーコードの標識（例えば、ユニバーサル標識、次元標識、空間標識、細胞標識、及びバーコード配列（例えば分子標識））の1つ以上は、確率バーコードの残りの標識の別の1つ又は2つからスペーサによって隔てることができる。スペーサは、例えば、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、若しくは20ヌクレオチド又はそれ以上であってよい。いくつかの実施形態では、確率バーコードの標識のいずれもスペーサによって隔てられない。

【0097】

ユニバーサル標識

バーコード（例えば確率バーコード）は1つ以上のユニバーサル標識又は配列を含みうる。いくつかの実施形態では、1つ以上のユニバーサル標識は、所与の固体担体に結合される確率バーコードのセット中のすべての確率バーコードで同一でありうる。いくつかの実施形態では、1つ以上のユニバーサル標識は、複数のビーズに結合されるすべての確率

10

20

30

40

50

バーコードで同一でありうる。いくつかの実施形態では、ユニバーサル標識は、シーケンシングプライマーにハイブリダイズ可能な核酸配列を含みうる。シーケンシングプライマーは、ユニバーサル標識を含む確率バーコードをシーケンスするために使用可能である。シーケンシングプライマー（例えば、ユニバーサルシーケンシングプライマー）は、高スループットシーケンシングプラットフォームに関連付けられるシーケンシングプライマーを含みうる。いくつかの実施形態では、ユニバーサル標識は、P C R プライマーにハイブリダイズ可能な核酸配列を含みうる。いくつかの実施形態では、ユニバーサル標識は、シーケンシングプライマー及びP C R プライマーにハイブリダイズ可能な核酸配列を含みうる。シーケンシングプライマー又はP C R プライマーにハイブリダイズ可能なユニバーサル標識の核酸配列は、プライマー結合部位として参照しうる。ユニバーサル標識は、確率バーコードの転写を開始するために使用しうる配列を含みうる。ユニバーサル標識は、確率バーコード又は確率バーコード内の領域の伸長のために、使用しうる配列を含みうる。ユニバーサル標識は、約1、2、3、4、5、10、15、20、25、30、35、40、45、50ヌクレオチド長、又はこれらの値のいずれか2つの間の数若しくは範囲であってよい。例えば、ユニバーサル標識は、少なくとも約10ヌクレオチドを含みうる。ユニバーサル標識は、少なくとも、又は多くとも、1、2、3、4、5、10、15、20、25、30、35、40、45、50ヌクレオチド長でありうる。いくつかの実施形態では、切断可能なリンカー又は修飾ヌクレオチドは、担体から確率バーコードを切断して除去することを可能にするユニバーサル標識配列の一部であってよい。

10

【0098】

次元標識

バーコード（例えば確率バーコード）は1つ以上の次元標識を含みうる。いくつかの実施形態では、次元標識は、確率標識化が行われた次元に関する情報を提供する核酸配列を含みうる。例えば、次元標識は、標的に確率バーコードが付された時点に関する情報を提供可能である。次元標識は、サンプルの確率バーコーディングの時点に関連付け可能である。次元標識は、確率標識化の時点で活性化可能である。異なる時点で異なる次元標識を活性化可能である。次元標識は、標的、標的のグループ、及び／又はサンプルに確率バーコードを付けた順序に関する情報を提供する。例えば、細胞集団は、細胞周期のG0期に確率バーコードを付けることが可能である。細胞は、細胞周期のG1期に確率バーコードで再びパルスすることが可能である。細胞は、細胞周期のS期に確率バーコードで再びパルスすることが可能であり、他の時期も同様である。各パルス時（例えば、細胞周期の各期）の確率バーコードは、異なる次元標識を含みうる。こうして、次元標識は、細胞周期のどの期に標的に標識したかに関する情報を提供する。次元標識は、多種多様な生物時間を精査することが可能である。例示的な生物時間としては、限定されるものではないが、細胞周期、転写（例えば転写開始）、及び転写物分解が挙げられうる。他の例として、薬剤治療及び／又は療法の前及び／又は後にサンプル（例えば、細胞、細胞集団）に確率標識を付けることが可能である。識別可能な標的のコピー数の変化は、薬剤及び／又は療法に対するサンプルの反応の指標でありうる。

20

【0099】

次元標識は、活性化可能であってよい。活性化可能な次元標識は、特定の時点で活性化可能でありうる。活性化可能な標識は、例えば、構成的に活性化することができる（例えば、オフに切り替わらない）。活性化可能な次元標識は、例えば、可逆的に活性化可能である（例えば、活性化可能な次元標識は、オン・オフの切替えが可能である）。例えば、次元標識は、少なくとも1、2、3、4、5、6、7、8、9、若しくは10回又はそれ以上可逆的に活性化可能でありうる。次元標識は、例えば、少なくとも1、2、3、4、5、6、7、8、9、若しくは10回又はそれ以上可逆的に活性化可能でありうる。いくつかの実施形態では、次元標識は、蛍光、光、化学的イベント（例えば、切断、他の分子のライゲーション、修飾（例えば、ペグ化、SUMO化、アセチル化、メチル化、脱アセチル化、脱メチル化）の付加、光化学的イベント（例えば、光ケージング）、及び非天然

30

40

50

ヌクレオチドの導入により活性化可能である。

【0100】

次元標識は、いくつかの実施形態では、所与の固体担体（例えばビーズ）に結合されるすべての確率バーコードで同一でありうるが、異なる固体担体（例えばビーズ）では異なりうる。いくつかの実施形態では、同一の固体担体上の確率バーコードの少なくとも60%、70%、80%、85%、90%、95%、97%、99%、又は100%は、同一の次元標識を含みうる。いくつかの実施形態では、同一の固体担体上の確率バーコードの少なくとも60%は、同一の次元標識を含みうる。いくつかの実施形態では、同一の固体担体上の確率バーコードの少なくとも95%は、同一の次元標識を含みうる。

【0101】

複数の固体担体（例えばビーズ）には、 10^6 程度又はそれ以上のユニーク次元標識配列が存在可能である。次元標識は、1、2、3、4、5、10、15、20、25、30、35、40、45、50、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。次元標識は、少なくとも、又は多くとも、1、2、3、4、5、10、15、20、25、30、35、40、45、50、100、200、若しくは300ヌクレオチド長でありうる。次元標識は、約5～約200ヌクレオチドを含みうる。次元標識は、約10～約150ヌクレオチドを含みうる。次元標識は、約20～約125ヌクレオチドを含みうる。

10

【0102】

空間標識

20

バーコード（例えば確率バーコード）は1つ以上の空間標識を含みうる。いくつかの実施形態では、空間標識は、確率バーコードに関連付けられる標的分子の空間配向に関する情報を提供する核酸配列を含みうる。空間標識は、サンプル中の座標に関連付け可能である。座標は固定座標でありうる。例えば、座標は基材を基準にして固定可能である。空間標識は二次元又は三次元のグリッドを基準にし得る。座標はランドマークを基準にして固定可能である。ランドマークは空間内で同定可能である。ランドマークはイメージング可能な構造体でありうる。ランドマークは生物学的構造体例えば解剖学的ランドマークでありうる。ランドマークは細胞ランドマーク（例えばオルガネラ）でありうる。ランドマークは、非天然ランドマーク、例えば、色コード、バーコード、磁性、蛍光、放射能、又はユニークなサイズ若しくは形状のような同定可能な識別子を有する構造体でありうる。空間標識は、物理的パーティション（例えば、ウェル、容器、又はドロップレット）に関連付け可能である。いくつかの実施形態では、空間内の1つ以上の位置にコードを付けるために複数の空間標識が一緒に使用される。

30

【0103】

空間標識は、所与の固体担体（例えばビーズ）に結合されるすべての確率バーコードで同一であってよいが、異なる固体担体（例えばビーズ）については異なっていてもよい。いくつかの実施形態では、同一の空間標識を含む、同一の固体担体上の確率バーコードのパーセンテージは、60%、70%、80%、85%、90%、95%、97%、99%、100%、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値でありうる。いくつかの実施形態では、同一の空間標識を含む、同一の固体担体上の確率バーコードのパーセンテージは、少なくとも、又は多くとも、60%、70%、80%、85%、90%、95%、97%、99%、若しくは100%でありうる。いくつかの実施形態では、同一の固体担体上の確率バーコードの少なくとも60%が、同一の空間標識を含んでよい。いくつかの実施形態では、同一の固体担体上の確率バーコードの少なくとも95%が、同一の空間標識を含んでよい。

40

【0104】

複数の固体担体（例えばビーズ）には、 10^6 程度又はそれ以上のユニーク空間標識配列が存在可能である。空間標識は、1、2、3、4、5、10、15、20、25、30、35、40、45、50、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。空間標識は、少なくとも、又は多くとも1

50

、 2、 3、 4、 5、 10、 15、 20、 25、 30、 35、 40、 45、 50、 100、 200、 若しくは 300 ヌクレオチド長でありうる。空間標識は、約 5 ~ 約 200 ヌクレオチドを含みうる。空間標識は、約 10 ~ 約 150 ヌクレオチドを含みうる。空間標識は、約 20 ~ 約 125 ヌクレオチドを含みうる。

【0105】

細胞標識

バーコード（例えば確率バーコード）は、1つ以上の細胞標識を含みうる。いくつかの実施形態では、細胞標識は、どの標的核酸がどの細胞に由来するかを決定するための情報を提供する核酸配列を含みうる。いくつかの実施形態では、細胞標識は、所与の固体担体（例えばビーズ）に結合されるすべての確率バーコードで同一であるが、異なる固体担体（例えばビーズ）については異なっている。いくつかの実施形態では、同一の細胞標識を含む、同一の固体担体上の確率バーコードのパーセンテージは、60%、70%、80%、85%、90%、95%、97%、99%、100%、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値でありうる。いくつかの実施形態では、同一の細胞標識を含む、同一の固体担体上の確率バーコードのパーセンテージは、最小で、又は最大で60%、70%、80%、85%、90%、95%、97%、99%、若しくは100%であってよい。例えば、同一の固体担体上の確率バーコードの少なくとも60%が、同一の細胞標識を含みうる。別の例として、同一の固体担体上の確率バーコードの少なくとも95%が、同一の細胞標識を含んでもよい。

10

【0106】

複数の固体担体（例えばビーズ）には、 10^6 程度又はそれ以上のユニーク細胞標識配列が存在可能である。細胞標識は、1、2、3、4、5、10、15、20、25、30、35、40、45、50、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。細胞標識は、少なくとも、又は多くとも、1、2、3、4、5、10、15、20、25、30、35、40、45、50、100、200、若しくは300 ヌクレオチド長でありうる。例えば、細胞標識は、約 5 ~ 約 200 ヌクレオチドを含みうる。別の例として、細胞標識は、約 10 ~ 約 150 ヌクレオチドを含みうる。さらに別の例として、細胞標識は、約 20 ~ 約 125 ヌクレオチドを含みうる。

20

【0107】

バーコード配列

バーコードは、1つ以上のバーコード配列を含みうる。いくつかの実施形態では、バーコード配列は、バーコードにハイブリダイズされた標的核酸種の特定のタイプを同定するための情報を提供する核酸配列を含みうる。バーコード配列は、バーコード（例えば標的結合領域）にハイブリダイズされた標的核酸種の特異的存在量に対するカウンター（例えば、粗い近似をもたらす）を提供する核酸配列を含みうる。

30

【0108】

いくつかの実施形態では、バーコード配列の多様なセットが所与の固体担体（例えばビーズ）に結合される。いくつかの実施形態では、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、 10^9 、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のユニーク分子標識配列が存在しうる。例えば、複数のバーコードは、識別可能な配列を有する約 6561 のバーコード配列を含みうる。別の例として、複数のバーコードは、識別可能な配列を有する約 65536 のバーコード配列を含みうる。いくつかの実施形態では、少なくとも、又は多くとも、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、若しくは 10^9 のユニークバーコード配列が存在しうる。ユニーク分子標識配列は、所与の固体担体（例えばビーズ）に結合されうる。

40

【0109】

バーコードは、1、2、3、4、5、10、15、20、25、30、35、40、45、50 ヌクレオチド長、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。バーコードは、少なくとも、又は多くとも

50

、 1、 2、 3、 4、 5、 10、 15、 20、 25、 30、 35、 40、 45、 50、 100、 200、 若しくは 300 ヌクレオチド長でありうる。

【0110】

分子標識

バーコード（例えば確率バーコード）は、1つ以上の分子標識を含みうる。いくつかの実施形態では、分子標識は、確率バーコードにハイブリダイズされた標的核酸種の特定のタイプを同定するための情報を提供する核酸配列を含みうる。分子標識は、確率バーコード（例えば標的結合領域）にハイブリダイズされた標的核酸種の特異的存在量に対するカウンターを提供する核酸配列を含みうる。

【0111】

いくつかの実施形態では、分子標識の多様なセットが所与の固体担体（例えばビーズ）に結合される。いくつかの実施形態では、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、 10^9 、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のユニーク分子標識配列が存在しうる。例えば、複数の確率バーコードは、識別可能な配列を有する約 6561 の分子標識を含みうる。別の例として、複数の確率バーコードは、識別可能な配列を有する約 65536 の分子標識を含みうる。いくつかの実施形態では、少なくとも、又は多くとも、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、若しくは 10^9 のユニーク分子標識配列が存在しうる。ユニーク分子標識配列は、所与の固体担体（例えばビーズ）に結合されうる。

【0112】

複数の確率バーコードを用いる確率バーコーディングについては、異なる分子標識配列の数と標的のいずれかの存在量の数との比は、1 : 1、2 : 1、3 : 1、4 : 1、5 : 1、6 : 1、7 : 1、8 : 1、9 : 1、10 : 1、11 : 1、12 : 1、13 : 1、14 : 1、15 : 1、16 : 1、17 : 1、18 : 1、19 : 1、20 : 1、30 : 1、40 : 1、50 : 1、60 : 1、70 : 1、80 : 1、90 : 1、100 : 1、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値でありうる。標的は、同一又はほぼ同一の配列を有するmRNA分子を含むmRNA種でありうる。いくつかの実施形態では、異なる分子標識配列の数と標的のいずれかの存在量の数との比は、少なくとも、又は多くとも、1 : 1、2 : 1、3 : 1、4 : 1、5 : 1、6 : 1、7 : 1、8 : 1、9 : 1、10 : 1、11 : 1、12 : 1、13 : 1、14 : 1、15 : 1、16 : 1、17 : 1、18 : 1、19 : 1、20 : 1、30 : 1、40 : 1、50 : 1、60 : 1、70 : 1、80 : 1、90 : 1、又は100 : 1である。

【0113】

分子標識は、1、2、3、4、5、10、15、20、25、30、35、40、45、50 ヌクレオチド長、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。分子標識は、少なくとも、又は多くとも、1、2、3、4、5、10、15、20、25、30、35、40、45、50、100、200、若しくは 300 ヌクレオチド長でありうる。

【0114】

標的結合領域

バーコード（例えば確率バーコード）は、1つ以上の標的結合領域を含みうる。いくつかの実施形態では、標的結合領域は、対象の標的とハイブリダイズすることができる。いくつかの実施形態では、標的結合領域は、標的（例えば、標的核酸、標的分子、例えば、分析される細胞核酸）、例えば、特定の遺伝子配列に特異的にハイブリダイズする核酸配列を含みうる。いくつかの実施形態では、標的結合領域は、特定の標的核酸の特定の位置に結合（例えばハイブリダイズ）しうる核酸配列を含みうる。いくつかの実施形態では、標的結合領域は、制限酵素部位オーバーハング（例えば、EcoRI 付着末端オーバーハング）への特異的なハイブリダイゼーションが可能な核酸配列を含みうる。次いで、確率バーコードは、制限部位オーバーハングに相補的な配列を含む任意の核酸分子にライゲートしうる。

10

20

30

40

50

【0115】

いくつかの実施形態では、標的結合領域は非特異的標的核酸配列を含みうる。非特異的標的核酸配列は、標的核酸の特定の配列に依存せずに複数の標的核酸に結合しうる配列を意味しうる。例えば、標的結合領域は、ランダムマルチマー配列を含みうるか又はmRNA分子のポリ(A)テールにハイブリダイズするオリゴ(dT)配列を含みうる。ランダムマルチマー配列は、例えば、ランダムダイマー、ランダムトリマー、ランダムクアトラマー、ランダムペントマー、ランダムヘキサマー、ランダムセプタマー、ランダムオクタマー、ランダムノナマー、ランダムデカマー、又は任意の長さのより高次のランダムマルチマーの配列でありうる。いくつかの実施形態では、標的結合領域は、所与のビーズに結合されたすべての確率バーコードで同一である。いくつかの実施形態では、所与のビーズに結合された複数の確率バーコードの標的結合領域は、2つ以上の異なる標的結合配列を含む。標的結合領域は、5、10、15、20、25、30、35、40、45、50、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。若しくはそれ以上又は概略で少なくともそうしたヌクレオチド長でありうる。標的結合領域は、多くとも約5、10、15、20、25、30、35、40、45、50ヌクレオチド長又はそれ以上でありうる。

10

【0116】

いくつかの実施形態では、標的結合領域は、ポリアデニル化末端を含むmRNAにハイブリダイズすることができるオリゴ(dT)を含みうる。標的結合領域は、遺伝子特異的でありうる。例えば、標的結合領域は、標的の特定の領域にハイブリダイズするように構成することができる。標的結合領域は、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。標的結合領域は、少なくとも、又は多くとも、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、若しくは30ヌクレオチド長でありうる。標的結合領域は、約5～30ヌクレオチド長であってもよい。確率バーコードが、遺伝子特異的標的結合領域を含む場合、この確率バーコードは、遺伝子特異的確率バーコードと呼ぶことができる。

20

【0117】

30

配向性

バーコード(例えば確率バーコード)は、確率バーコードの配向(例えばアライメント)のために使用することができる1つ以上の配向性を含みうる。確率バーコードは、等電点電気泳動用の部分を含みうる。異なる確率バーコードは、異なる等電点電気泳動点を含みうる。こうした確率バーコードをサンプルに導入した場合、サンプルは、確率バーコードを既知の形態にオリエントするために等電点電気泳動を行うことが可能である。こうして、オリエント性は、サンプルで確率バーコードの既知のマップを作成するために使用可能である。例示的なオリエント性としては、電気泳動移動度(例えば、確率バーコードのサイズに基づく)、等電点、スピン、伝導率、及び/又はセルファーセンブリーが挙げられる。例えば、セルファーセンブリーのオリエント性を含む確率バーコードは、活性化時に特定のオリエンテーションにセルファーセンブル可能である(例えば、核酸ナノ構造)。

40

【0118】

親和性

バーコード(例えば確率バーコード)は、1つ以上の親和性を含みうる。例えば、空間標識は、親和性を含みうる。親和性は、他のエンティティー(例えば細胞レセプター)との確率バーコードの結合を促進することができる化学的及び/又は生物学的部分を含みうる。例えば、親和性は、抗体、例えば、サンプル上の特定の部分(例えばレセプター)に特異的な抗体を含みうる。いくつかの実施形態では、抗体は、確率バーコードを特定の細胞型又は分子に誘導することができる。特定の細胞型若しくは分子及び/又はその近傍にある標的を確率標識化することができる。抗体は確率バーコードを特定の位置に誘導する

50

ことができる所以、いくつかの実施形態において、親和性は、空間標識のヌクレオチド配列に加え、空間情報も提供することができる。抗体は、治療用抗体、例えば、モノクローナル抗体又はポリクローナル抗体であってもよい。抗体は、ヒト化されていても、又はキメラであってもよい。抗体は、ネイキッド抗体又は融合抗体であってもよい。

【0119】

抗体は、全長（すなわち、天然に存在するか若しくは通常の免疫グロブリン遺伝子断片組換えプロセスにより形成される）免疫グロブリン分子（例えば IgG 抗体）又は免疫グロブリン分子の免疫活性（すなわち特異的結合）部分例えば抗体フラグメントでありうる。

【0120】

抗体フラグメントは、例えば、F(ab')2、F(ab')、Fab、Fv、sFvなどの抗体の一部でありうる。いくつかの実施形態において、抗体フラグメントは、全長抗体により認識される同一の抗原に結合可能である。抗体フラグメントは、抗体の可変領域からなる単離された断片、例えば、重鎖及び軽鎖の可変領域からなる「Fv」フラグメント並びに軽鎖及び重鎖の可変領域がペプチドリンカーにより接続された組換え一本鎖ポリペプチド分子（「scFv タンパク質」）を含みうる。例示的な抗体としては、限定されるものではないが、癌細胞に対する抗体、ウイルスに対する抗体、細胞表面レセプター（CD8、CD34、CD45）に結合する抗体、及び治療用抗体が挙げられうる。

10

【0121】

ユニバーサルアダプタープライマー

バーコードは、1つ以上のユニバーサルアダプタープライマーを含みうる。例えば、遺伝子特異的バーコード、例えば遺伝子特異的確率バーコードは、ユニバーサルアダプタープライマーを含みうる。ユニバーサルアダプタープライマーは、すべてのバーコードに対してユニバーサルであるヌクレオチド配列を意味しうる。ユニバーサルアダプタープライマーは、遺伝子特異的バーコードを構築するために使用することができる。ユニバーサルアダプタープライマーは、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値のヌクレオチド長でありうる。ユニバーサルアダプタープライマーは、少なくとも、又は多くとも、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、若しくは30ヌクレオチド長でありうる。ユニバーサルアダプタープライマーは、約5～30ヌクレオチド長であってもよい。

20

【0122】

リンカー

バーコードが2つ以上の標識タイプ（例えば、2つ以上の細胞標識又は2つ以上のバーコード配列、例えば1つの分子標識）を含む場合、標識は、リンカー標識配列で散在されてもよい。リンカー標識配列は、少なくとも約5、10、15、20、25、30、35、40、45、50又はそれ以上のヌクレオチド長でありうる。リンカー標識配列は、多くて約5、10、15、20、25、30、35、40、45、50又はそれ以上のヌクレオチド長でありうる。場合によっては、リンカー標識配列は、12ヌクレオチド長である。バーコードの合成を容易にするため、リンカー標識配列が使用可能である。リンカー標識は、エラー訂正（例えばハミング）コードを含みうる。

30

【0123】

固体担体

本明細書に開示されるバーコード（例えば確率バーコード）は、いくつかの実施形態において、固体担体と結合することができる。固体担体は、例えば、合成粒子であってよい。いくつかの実施形態では、固体担体上の複数の確率バーコード（例えば、第1の複数の確率バーコード）の分子標識（例えば、第1の分子標識）の一部又は全部が、少なくとも1ヌクレオチド異なる。同じ固体担体上の確率バーコードの細胞標識は、同じであってよい。異なる固体担体上の確率バーコードの細胞標識は、少なくとも1ヌクレオチド異なる。

40

50

りうる。例えば、第1の固体担体上の第1の複数の確率バーコードの第1の細胞標識は、同じ配列を有してよく、第2の固体担体上の第2の複数の確率バーコードの第2の細胞標識は、同じ配列を有してよい。第1の固体担体上の第1の複数の確率バーコードの第1の細胞標識と、第2の固体担体上の第2の複数の確率バーコードの第2の細胞標識とは、少なくとも1スクレオチド異なりうる。細胞標識は、例えば、約5～20スクレオチド長でありうる。分子標識は、例えば、約5～20スクレオチド長でありうる。合成粒子は、例えば、ビーズであってよい。

【0124】

ビーズは、例えば、シリカゲルビーズ、調節多孔性ガラスビーズ、磁気ビーズ、ダイナビーズ、セファデックス／セファロースビーズ、セルロースビーズ、ポリスチレンビーズ、又はそれらの任意の組合せであってよい。ビーズは、ポリジメチルシロキサン（PDM-S）、ポリスチレン、ガラス、ポリプロピレン、アガロース、ゼラチン、ヒドロゲル、常磁性材料、セラミック、プラスチック、ガラス、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、セルロース、ナイロン、シリコン、又はそれらの任意の組合せなどの材料を含みうる。

10

【0125】

いくつかの実施形態では、ビーズは、ポリマービーズ、例えば、変形性ビーズ又はゲルビーズであってよく、これらは、確率バーコードで官能化されている（例えば、10X Genomics（San Francisco, CA）からのゲルビーズなど）。いくつかの実施形態では、ゲルビーズは、ポリマーベースのゲルを含みうる。ゲルビーズは、例えば、1つ以上のポリマー前駆体を液滴中に封入することによって作製することができる。促進剤（例えば、テトラメチルエチレンジアミン（TEMED））にポリマー前駆体を曝露すると、ゲルビーズが作製されうる。

20

【0126】

いくつかの実施形態では、ポリマービーズは、例えば、所望の条件下で、溶解、溶融、又は分解しうる。所望の条件は、環境条件を含みうる。所望の条件は、制御された様式で、ポリマービーズの溶解、溶融、又は分解を引き起こしうる。ゲルビーズは、化学的刺激、物理的刺激、生物学的刺激、熱刺激、磁気刺激、電気刺激、光刺激、又はそれらの任意の組合せによって、溶解、溶融、又は分解しうる。

【0127】

30

例えば、オリゴスクレオチドバーコードなどの被検物質及び／若しくは試薬を、ゲルビーズの内側表面（例えば、オリゴスクレオチドバーコード及び／若しくはオリゴスクレオチドバーコードを作製するために用いられる材料の拡散を介して進入可能な内部）並びに／又はゲルビーズの外側表面、或いは本明細書に記載されるいづれか他のマイクロカプセルにカップリング／固定してもよい。カップリング／固定は、化学結合（例えば、共有結合、イオン結合）又は物理的現象（例えば、ファンデルワールス力、双極子-双極子相互作用など）の任意の形態を介するものであってよい。いくつかの実施形態では、ゲルビーズ又は本明細書に記載する任意の他のマイクロカプセルに対する試薬のカップリング／固定は、例えば、不安定部分（例えば、本明細書に記載の化学架橋剤をはじめとする、化学架橋剤）を介するなど、可逆性であってもよい。刺激を適用すると、不安定部分は、切断されて、固定された試薬が遊離されうる。いくつかの事例では、不安定部分は、ジスルフィド結合である。例えば、オリゴスクレオチドバーコードが、ジスルフィド結合を介してゲルビーズに固定されている場合、ジスルフィド結合を還元剤に曝露することにより、ジスルフィド結合を切断して、オリゴスクレオチドバーコードをビーズから遊離させることができる。不安定部分は、ゲルビーズ若しくはマイクロカプセルの一部として、試薬若しくは被検物質をゲルビーズ若しくはマイクロカプセルに連結する化学リンカーの一部として、及び／又は試薬若しくは被検物質の一部として含有させてもよい。

40

【0128】

いくつかの実施形態では、ゲルビーズは、限定するものではないが、以下のものをはじめとする、極めて多様なポリマーを含みうる：ポリマー、熱感受性ポリマー、感光性ポリ

50

マー、磁気ポリマー、pH感受性ポリマー、塩感受性ポリマー、化学的感受性ポリマー、高分子電解質、多糖、ペプチド、タンパク質、及び／又はプラスチック。ポリマーとしては、限定するものではないが、ポリ(N - イソプロピルアクリルアミド) (P N I P A A m)、ポリ(スルホン酸スチレン) (P S S)、ポリ(アリルアミン) (P A A m)、ポリ(アクリル酸) (P A A)、ポリ(エチレンイミン) (P E I)、ポリ(ジアリルジメチル - 塩化アンモニウム) (P D A D M A C)、ポリ(ピロール) (P P y)、ポリ(ビニルピロリドン) (P V P O N)、ポリ(ビニルピリジン) (P V P)、ポリ(メタクリル酸) (P M A A)、ポリ(メチルメタクリレート) (P M M A)、ポリスチレン (P S)、ポリ(テトラヒドロフラン) (P T H F)、ポリ(フタルアルデヒド) (P T H F)、ポリ(ヘキシルビオロゲン) (P H V)、ポリ(L - リシン) (P L L)、ポリ(L - アルギニン) (P A R G)、乳酸 - グリコール酸共重合体 (P L G A)などの材料が挙げられる。

【 0 1 2 9 】

多数の化学的刺激を用いて、ビーズの破壊又は分解をトリガーすることができる。これらの化学的变化の例として、限定するものではないが、ビーズ壁に対するpH媒介による变化、架橋の化学的切断を介したビーズ壁の崩壊、ビーズ壁の解重合トリガー、及びビーズ壁スイッチング反応が挙げられる。また、バルク変化を用いて、ビーズの破壊をトリガーしてもよい。

【 0 1 3 0 】

また、さまざまな刺激を介したマイクロカプセルに対するバルク又は物理的变化も、試薬を放出するようにカプセルを設計する上で多くの利点をもたらす。バルク又は物理的变化は、巨視的規模で起こり、その際、ビーズ破断は、刺激により誘導された機械物理的力の結果による。こうしたプロセスとしては、限定するものではないが、圧力誘導破断、ビーズ壁溶融、又はビーズ壁の多孔性変化が挙げられる。

【 0 1 3 1 】

生物学的刺激を用いて、ビーズの破壊又は分解をトリガーすることもできる。概して、生物学的トリガーは、化学的トリガーと類似しているが、多くの例では、生体分子、又は酵素、ペプチド、糖類、核酸などの生存系に一般的に存在する分子が使用される。例えば、ビーズは、特定のプロテアーゼによる切断に感受性のペプチド架橋を有するポリマーを含んでもよい。さらに具体的には、一例は、G F L G K ペプチド架橋を含むマイクロカプセルを含んでもよい。プロテアーゼカテプシンBなどの生物学的トリガーを加えると、シェルウェルのペプチド架橋が切断されて、ビーズの内容物が放出される。他の事例では、プロテアーゼを熱活性化してもよい。別の例では、ビーズは、セルロースを含有するシェル壁を含む。加水分解性酵素キトサンの添加は、セルロース結合の切断、シェル壁の解重合、及びその内部内容物の放出のための生物学的トリガーとして役立つ。

【 0 1 3 2 】

さらに、ビーズは、熱刺激の適用時にその内容物を放出するように誘導することもできる。温度の变化は、ビーズにさまざまな変化を引き起こし得る。熱の変化は、ビーズ壁が崩壊するように、ビーズの溶融を引き起こし得る。別の事例では、熱は、ビーズが破断又は破裂するように、ビーズの内部成分の内圧を高めうる。また別の事例では、熱は、ビーズを収縮した脱水状態に変形させうる。さらに、熱は、ビーズの壁内の熱感受性ポリマーに作用して、ビーズの破壊を引き起こしうる。

【 0 1 3 3 】

マイクロカプセルのビーズ壁に磁気ナノ粒子を含有させると、ビーズの破断トリガー、並びに多数のビーズの誘導を可能にし得る。本開示のデバイスは、いずれの目的で磁気ビーズを含んでもよい。一例では、高分子電解質含有ビーズに Fe_3O_4 ナノ粒子を組み込むと、振動磁界刺激の存在下で破断がトリガーされる。

【 0 1 3 4 】

ビーズはまた、電気刺激の結果として破壊又は分解することもできる。前のセクションに記載した磁気粒子と同様に、電気感受性ビーズも、ビーズの破断トリガー、並びに電界

10

20

30

40

50

下でのアライメント、導電性又はレドックス反応などの他の機能を可能にする。一例では、電気感受性材料を含有するビーズは、内部試薬の放出を制御することができるよう、電界下でアライメントされる。他の例では、電界は、ビーズ壁自体の内部でレドックス反応を誘導することもでき、これにより、多孔性が増加しうる。

【0135】

また、光刺激を用いて、ビーズを破壊することもできる。多数の光トリガーが考えられ、特定の範囲の波長の光子を吸収することができるナノ粒子及び発色団などのさまざまな分子を用いるシステムが挙げられる。例えば、金属酸化物コーティングをカプセルトリガーとして用いることができる。SiO₂でコーティングされた高分子電解質カプセルのUV照射は、ビーズ壁の崩壊を引き起こしうる。また別の例では、アゾベンゼン基などのフォトスイッチ材料をビーズ壁に組み込んでもよい。UV又は可視光線を適用すると、こうした化学物質は、光子の吸収時に、可逆的シス-トランス異性化を被る。この態様では、光子スイッチの組込みによって、光トリガー適用の際に、崩壊するか、又はより多孔性になりうるビーズ壁が得られる。

【0136】

例えば、図2に示す確率バーコードの非限定的な例200において、208でのマイクロウェルアレイの複数のマイクロウェルに、単一細胞などの細胞を導入した後、ビーズを212のマイクロウェルアレイの複数のマイクロウェルに導入することができる。各マイクロウェルは、1つのビーズを含みうる。ビーズは、複数の確率バーコードを含みうる。確率バーコードは、ビーズに結合した5'アミン領域を含みうる。確率バーコードは、ユニバーサル標識、分子標識、標的結合領域、又はそれらの任意の組合せを含んでもよい。

【0137】

本明細書に開示する確率バーコードは、固体担体（例えば、ビーズ）に関連（例えば、結合）させることができる。固体担体と結合した確率バーコードは、各々、ユニーク配列を有する少なくとも100又は1000の分子標識を含む群から選択される分子標識を含みうる。いくつかの実施形態では、固体担体と結合した異なる確率バーコードは、異なる配列の分子標識を含んでもよい。いくつかの実施形態では、固体担体と結合した、特定のパーセンテージの確率バーコードが、同じ細胞標識を含む。例えば、そのパーセンテージは、60%、70%、80%、85%、90%、95%、97%、99%、100%、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値でありうる。別の例として、パーセンテージは、少なくとも、又は多くとも60%、70%、80%、85%、90%、95%、97%、99%、若しくは100%でありうる。いくつかの実施形態では、固体担体と結合した確率バーコードは、同じ細胞標識を含みうる。異なる固体担体と結合した確率バーコードは、ユニーク配列を有する少なくとも100又は1000の細胞標識を含む群から選択される、異なる細胞標識を含んでもよい。

【0138】

いくつかの実施形態では、サンプル中の複数の標的に確率バーコードを付ける工程は、複数の確率バーコードと結合した複数の合成粒子を含む固体担体を用いて、実施することができる。いくつかの実施形態では、固体担体は、複数の確率バーコードと結合した複数の合成粒子を含みうる。さまざまな固体担体上の複数の確率バーコードの空間標識は、少なくとも1スクレオチド異なりうる。固体担体は、例えば、2次元又は3次元の複数の確率バーコードを含みうる。合成粒子は、ビーズであってよい。ビーズは、シリカゲルビーズ、調節多孔性ガラスビーズ、磁気ビーズ、ダイナビーズ、セファデックス/セファロースビーズ、セルロースビーズ、ポリスチレンビーズ、又はそれらの任意の組合せであってよい。固体担体は、ポリマー、マトリックス、ヒドロゲル、ニードルアレイデバイス、抗体、又はそれらの任意の組合せを含みうる。いくつかの実施形態では、固体担体は、浮動性であってよい。いくつかの実施形態では、固体担体は、半固体又は固体アレイに埋め込むことができる。確率バーコードは、固体担体と結合していなくてもよい。確率バーコードは、個別のスクレオチドであってもよい。確率バーコードは、基材と結合してもよい。

【0139】

10

20

30

40

50

本明細書で使用される場合、「テザー連結」、「結合」、及び「固定」という用語は、同義的に用いられて、確率バーコードを固体担体に結合するための共有結合又は非共有結合の手段を意味しうる。さまざまな異なるいずれの固体担体も、プレ合成された確率バーコードを結合するための、又は確率バーコードを *in situ* 固相合成するための固体担体として使用することができる。

【0140】

いくつかの実施形態では、固体担体はビーズである。ビーズは、核酸を（例えば共有結合又は非共有結合で）固定することができる、固体、多孔性、若しくは中空のスフェア、ボール、ペアリング、シリンダー、又は他の類似の構成体の1つ以上のタイプを包含しうる。ビーズは、例えば、プラスチック、セラミック、金属、若しくは高分子材料、又はそれらの任意の組合せから構成されうる。ビーズは、離散粒子であるか、又はそれを含んでもよく、離散粒子は、球状（例えばマイクロスフェア）であるか、又は非球状若しくは不規則形状、例えば、立方体形、直方体形、角錐形、円柱形、円錐形、扁球形、ディスク形などを有する。いくつかの実施形態では、ビーズは、非球状の形状でありうる。

10

【0141】

ビーズは、限定されるものではないが、常磁性材料（例えば、マグネシウム、モリブデン、リチウム、及びタンタル）、超常磁性材料（例えば、フェライト（Fe₃O₄；マグネタイト）ナノ粒子）、強磁性材料（例えば、鉄、ニッケル、コバルト、それらのいくつかの合金、及びいくつかの希土類金属化合物）、セラミック、プラスチック、ガラス、ポリスチレン、シリカ、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、アガロース、ヒドロゲル、ポリマー、セルロース、ナイロン、並びにそれらの任意の組合せなどのさまざまな材料を含みうる。いくつかの実施形態では、ビーズ（例えば、確率標識が結合されたビーズ）は、ヒドロゲルビーズである。いくつかの実施形態では、ビーズは、ヒドロゲルを含む。

20

【0142】

本明細書に開示するいくつかの実施形態は、1つ以上の粒子（例えば、ビーズ）を含む。粒子は各々、複数のオリゴヌクレオチド（例えば、確率バーコード）を含みうる。複数のオリゴヌクレオチドは各々、分子標識配列、細胞標識配列、及び標的結合領域（例えば、オリゴdT配列、遺伝子特異的配列、ランダム多量体、又はそれらの組合せ）を含みうる。複数のオリゴヌクレオチドの各々の細胞標識配列は、同じであってもよい。異なる粒子上のオリゴヌクレオチドの細胞標識配列は、異なる粒子上のオリゴヌクレオチドを同定できるように、相違してもよい。異なる細胞標識配列の数は、異なる実装において相違してもよい。いくつかの実施形態では、細胞標識配列の数は、10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10⁶、10⁷、10⁸、10⁹、又はこれらの値のいずれか2つの間の数若しくは範囲、又はそれ以上、或いはそうした近似値でありうる。いくつかの実施形態では、細胞標識配列の数は、少なくとも、又は多くとも10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、10⁶、10⁷、10⁸、若しくは10⁹でありうる。いくつかの実施形態では、複数の粒子の1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000以下、又はそれ以上が、同じ細胞配列のオリゴヌクレオチドを含む。いくつかの実施形態では、同じ細胞配列のオリゴヌクレオチドを含む複数の粒子は、多くとも0.1%、0.2%、0.3%、0.4%、0.5%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%又はそれ以上であってよい。いくつかの実施形態では、複数の粒子のいずれも同じ細

30

40

50

胞標識配列を含まない。

【0143】

各粒子の複数のオリゴヌクレオチドは、異なる分子標識配列を含みうる。いくつかの実施形態では、分子標識配列の数は、10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10⁶、10⁷、10⁸、10⁹、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値でありうる。分子標識配列の数は、少なくとも、又は多くとも10、100、200、300、400、500、600、700、800、900、100、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10⁶、10⁷、10⁸、若しくは10⁹でありうる。例えば、複数のオリゴヌクレオチドの少なくとも100は、異なる分子標識配列を含む。別の例として、単一粒子において、複数のオリゴヌクレオチドの少なくとも100、500、1000、5000、10000、15000、20000、50000、これらの値のいずれか2つの間の数若しくは範囲、又はそれ以上が、異なる分子標識配列を含む。いくつかの実施形態は、確率バーコードを含む複数の粒子を提供する。いくつかの実施形態では、標的の存在量（又はコピー若しくは数）と異なる分子標識配列の比は、少なくとも、1:1、1:2、1:3、1:4、1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:30、1:40、1:50、1:60、1:70、1:80、1:90、又はそれ以上でありうる。いくつかの実施形態では、複数のオリゴヌクレオチドの各々は、サンプル標識、ユニバーサル標識、又はその両方をさらに含む。粒子は、例えば、ナノ粒子又はミクロ粒子であってよい。

【0144】

ビーズのサイズは、変動しうる。例えば、ビーズの直径は、0.1マイクロメートル～50マイクロメートルの範囲であってよい。いくつかの実施形態では、ビーズの直径は、0.1、0.5、1、2、3、4、5、6、7、8、9、10、20、30、40、50マイクロメートル、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値でありうる。

【0145】

ビーズの直径は、基材のウェルの直径と関連させることができる。いくつかの実施形態では、ビーズの直径は、ウェルの直径よりも、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値だけ長い若しくは短い長さであってよい。ビーズの直径は、細胞（例えば、基材のウェルに閉じ込められた単一細胞）の直径に関連させることができる。いくつかの実施形態では、ビーズの直径は、細胞の直径よりも、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、200%、250%、300%、又はこれらの値のいずれか2つの間の数若しくは範囲、或いはそうした近似値だけ長い若しくは短い長さであってもよい。

【0146】

ビーズは、基材への埋込み及び／又は結合が可能である。ビーズは、ゲル、ヒドロゲル、ポリマー、及び／又はマトリックスへの埋込み及び／又は結合が可能である。基材（例えば、ゲル、マトリックス、スキヤフォールド、又はポリマー）内のビーズの空間位置は、位置アドレスとして機能可能なビーズ上の確率バーコードに存在する空間標識を用いて同定可能である。

【0147】

ビーズの例としては、限定されるものではないが、ストレプトアビジンビーズ、アガロースビーズ、磁気ビーズ、Dyna bead（登録商標）、MACS（登録商標）マイク

10

20

30

40

50

ロビーズ、抗体コンジュゲートビーズ（例えば、抗免疫グロブリンマイクロビーズ）、プロテインAコンジュゲートビーズ、プロテインGコンジュゲートビーズ、プロテインA/Gコンジュゲートビーズ、プロテインLコンジュゲートビーズ、オリゴ(dT)コンジュゲートビーズ、シリカビーズ、シリカ様ビーズ、抗ビオチンマイクロビーズ、抗蛍光色素マイクロビーズ、及びBcMag（商標）カルボキシル末端磁気ビーズが挙げられる。

【0148】

ビーズは、1つの蛍光光学チャネル又は複数の光学チャネルで蛍光を発するように量子ドット又は蛍光色素への関連付け（例えばそれらによる含浸）が可能である。ビーズは、常磁性又は強磁性にするために酸化鉄又は酸化クロムへの関連付けが可能である。ビーズは同定可能でありうる。例えば、ビーズは、カメラを用いてイメージング可能である。ビーズは、ビーズに関連付けられた検出可能なコードを有しうる。例えば、ビーズは、確率バーコードを含みうる。ビーズは、例えば、有機又は無機の溶液中での膨潤に起因してサイズ変化しうる。ビーズは疎水性でありうる。ビーズは親水性でありうる。ビーズは生体適合性でありうる。

10

【0149】

固体担体（例えばビーズ）は可視化可能である。固体担体は可視化タグ（例えば蛍光色素）を含みうる。固体担体（例えばビーズ）は識別子（例えば数）でエッチング可能である。識別子はビーズのイメージングにより可視化可能である。

【0150】

20

基材及びマイクロウェルアレイ

本明細書で使用される場合、基材はあるタイプの固体担体を意味しうる。基材は、本開示のバーコード（例えば確率バーコード）を含みうる固体担体を意味しうる。基材は、例えば、複数のマイクロウェルを含みうる。例えば、基材は、2つ以上のマイクロウェルを含むウェルアレイであってよい。いくつかの実施形態では、マイクロウェルは、規定の体積の小さい反応チャンバーを含みうる。いくつかの実施形態では、マイクロウェルは、1つ以上の細胞を閉じ込めることができる。いくつかの実施形態では、マイクロウェルは、1つの細胞のみを閉じ込めることができる。いくつかの実施形態では、マイクロウェルは、1つ以上の固体担体を閉じ込めることができる。いくつかの実施形態では、マイクロウェルは、1つの固体担体のみを閉じ込めることができる。いくつかの実施形態では、マイクロウェルは、単一細胞及び单一固体担体（例えば、ビーズ）を閉じ込める。

30

【0151】

確率バーコーディングの方法

本明細書においては、身体サンプル（例えば、組織、器官、腫瘍、細胞）における識別可能な位置の識別可能な標的の数を推定する方法を提供する。本方法は、サンプルと接近させて確率バーコードを配置する工程と、サンプルを溶解させる工程と、識別可能な標的を確率バーコードと関連させる工程と、標的を增幅する工程及び/又は標的をデジタルカウントする工程と、を含みうる。本方法は、さらに、確率バーコード上の空間標識から得られた情報を分析する工程及び/又は視覚化する工程をさらに含みうる。いくつかの実施形態では、一方法は、サンプル中の複数の標識を視覚化する工程を含む。サンプルのマップに複数の標的をマッピングする工程は、サンプルの二次元マップ又は三次元マップの作製を含みうる。二次元マップ又は三次元マップは、サンプル中の複数の標的に確率バーコードを付ける前又は後に作製することができる。サンプル中の複数の標的を視覚化する工程は、サンプルのマップに複数の標的をマッピングする工程を含みうる。サンプルのマップに複数の標的をマッピングする工程は、サンプルの二次元マップ又は三次元マップを作製するステップを含みうる。二次元マップ及び三次元マップは、サンプル中の複数の標的に確率バーコードを付ける前又は後に作製することができる。いくつかの実施形態では、二次元マップ及び三次元マップは、サンプルを溶解させる前又は後に作製することができる。二次元マップ又は三次元マップの作製前又は後にサンプルを溶解させる工程は、サンプルを加熱する工程と、サンプルを洗剤と接触させる工程と、サンプルのpHを変化させる工程、又はそれらの任意の組合せを含みうる。

40

50

【0152】

いくつかの実施形態では、複数の標的に確率バーコードを付ける工程は、複数の確率バーコードを複数の標的とハイブリダイズさせて、確率バーコード付き標的を作製する工程を含む。複数の標的に確率バーコードを付ける工程は、確率バーコード付き標的のインデックス付きライブラリーを作製する工程を含みうる。確率バーコード付き標的のインデックス付きライブラリーを作製する工程は、複数の確率バーコードを含む固体担体を用いて実施することができる。

【0153】

サンプルとバーコードの接触

本開示は、サンプル（例えば、細胞）を本開示の基材と接触させる方法を提供する。例えば、細胞、器官、又は組織薄片を含むサンプルをバーコード（例えば確率バーコード）と接触させることができる。例えば、重力流によって、細胞を接触させることができ、その場合、細胞は沈殿して単層を形成しうる。サンプルは、組織薄片であってよい。薄片を基材の上に配置することができる。サンプルは、一次元（例えば、平面表面を形成する）であってよい。サンプル（例えば、細胞）は、例えば、基材上に細胞を増殖させる／培養することによって、基材全体に広げることができる。

10

【0154】

確率バーコードが標的と近接して位置すると、標的は、確率バーコードとハイブリダイズしうる。識別可能な標的の各々が、本開示の識別可能な確率バーコードと結合し得るよう、確率バーコードを非枯渇的比率で接触させることができる。標的と確率バーコード同士の効率的な結合を確実にするために、標的を確率バーコードと架橋させることができる。

20

【0155】

細胞溶解

細胞及びバーコード（例えば確率バーコード）の分配後、細胞は標的分子を遊離するよう溶解可能である。細胞溶解は、さまざまな手段のいずれかにより、例えば、化学的若しくは生化学的手段により、浸透圧ショックにより、又は熱溶解、機械溶解、若しくは光学溶解により達成可能である。細胞は、界面活性剤（例えば、SDS、Li デシルスルフェート、Triton X-100、Tween-20、若しくはNP-40）、有機溶媒（例えば、メタノール若しくはアセトン）、又は消化酵素（例えば、プロテイナーゼK、ペプシン又はトリプシン）、或いはそれらの任意の組合せを含む細胞溶解緩衝液の添加により溶解可能である。標的と確率バーコードとの関連付けを向上させるために、例えば、温度の低下及び／又はライセートの粘度の増加により、標的分子の拡散速度を変化させることができる。

30

【0156】

いくつかの実施形態では、サンプルは濾紙を用いて溶解可能である。濾紙は濾紙の上を溶解緩衝液で浸漬可能である。濾紙は、サンプルの溶解及び基材へのサンプルの標的のハイブリダイゼーションを促進可能な加圧でサンプルに適用可能である。

【0157】

いくつかの実施形態では、溶解は、機械溶解、熱溶解、光学溶解、及び／又は化学溶解により行うことが可能である。化学溶解は、プロテイナーゼK、ペプシン、トリプシンなどの消化酵素の使用を含みうる。溶解は、基材への溶解緩衝液の添加により行うことが可能である。溶解緩衝液はトリスHClを含みうる。溶解緩衝液は、少なくとも約0.01、0.05、0.1、0.5、若しくは1M又はそれ以上のトリスHClを含みうる。溶解緩衝液は、多くとも約0.01、0.05、0.1、0.5、若しくは1M又はそれ以上のトリスHClを含みうる。溶解緩衝液は約0.1MトリスHClを含みうる。溶解緩衝液のpHは、少なくとも約1、2、3、4、5、6、7、8、9、若しくは10又はそれ以上でありうる。溶解緩衝液のpHは、多くとも約1、2、3、4、5、6、7、8、9、若しくは10又はそれ以上でありうる。いくつかの実施形態では、溶解緩衝液のpHは約7.5である。溶解緩衝液は塩（例えばLiCl）を含みうる。溶解緩衝液中の塩の

40

50

濃度は、少なくとも約 0.1、0.5、若しくは 1 M 又はそれ以上でありうる。溶解緩衝液中の塩の濃度は、多くとも約 0.1、0.5、若しくは 1 M 又はそれ以上でありうる。いくつかの実施形態では、溶解緩衝液中の塩の濃度は約 0.5 M である。溶解緩衝液は、界面活性剤（例えば、SDS、Li ドデシルスルフェート、トリトン X、トウイーン、NP-40）を含みうる。溶解緩衝液中の界面活性剤の濃度は、少なくとも約 0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、3%、4%、5%、6%、若しくは 7% 又はそれ以上でありうる。溶解緩衝液中の界面活性剤の濃度は、多くとも約 0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、1%、2%、3%、4%、5%、6%、若しくは 7% 又はそれ以上でありうる。いくつかの実施形態では、溶解緩衝液中の界面活性剤の濃度は約 1% Li ドデシルスルフェートである。本方法で溶解に使用される時間は、使用される界面活性剤の量に依存性しうる。いくつかの実施形態では、界面活性剤を多く使用するほど、溶解に必要な時間は短くなる。溶解緩衝液はキレート化剤（例えば、EDTA、EGTA）を含みうる。溶解緩衝液中のキレート化剤の濃度は、少なくとも約 1、5、10、15、20、25、若しくは 30 mM 又はそれ以上でありうる。溶解緩衝液中のキレート化剤の濃度は、多くとも約 1、5、10、15、20、25、若しくは 30 mM 又はそれ以上でありうる。いくつかの実施形態では、溶解緩衝液中のキレート化剤の濃度は約 10 mM である。溶解緩衝液は還元試薬（例えば、メルカプトエタノール、DTT）を含みうる。溶解緩衝液中の還元試薬の濃度は少なくとも約 1、5、10、15、20 mM 又はそれ以上でありうる。溶解緩衝液中の還元試薬の濃度は多くとも約 1、5、10、15、20 mM 又はそれ以上でありうる。いくつかの実施形態では、溶解緩衝液中の還元試薬の濃度は約 5 mM である。いくつかの実施形態では、溶解緩衝液は、約 0.1 M のトリス HCl、約 pH 7.5、約 0.5 M LiCl、約 1% リチウムドデシルスルフェート、約 10 mM EDTA、及び約 5 mM DTT を含みうる。

【0158】

溶解は、約 4、10、15、20、25、又は 30 の温度で行うことが可能である。溶解は、約 1、5、10、15、若しくは 20 分間又はそれ以上行うことが可能である。溶解細胞は、少なくとも約 100 000、200 000、300 000、400 000、500 000、600 000、若しくは 700 000 標的核酸分子又はそれ以上を含みうる。溶解細胞は、多くとも約 100 000、200 000、300 000、400 000、500 000、600 000、若しくは 700 000 標的核酸分子又はそれ以上を含みうる。

【0159】

標的核酸分子へのバーコードの結合

細胞の溶解及びそれからの核酸分子の放出の後、核酸分子は、共局在化された固体担体のバーコード（例えば確率バーコード）にランダムに関連付けすることができる。関連付けは、標的核酸分子の相補的部分への確率バーコードの標的認識領域のハイブリダイゼーションを含みうる（例えば、確率バーコードのオリゴ（dT）は、標的のポリ（A）テールと相互作用可能である）。ハイブリダイゼーションに使用されるアッセイ条件（例えば、緩衝液 pH、イオン強度、温度など）は、特定の安定なハイブリッドの形成を促進するように選択可能である。いくつかの実施形態では、溶解した細胞から放出された核酸分子は、基材上の複数のプローブに関連付けする（例えば、基板上のプローブとハイブリダイズする）ことができる。プローブが、オリゴ（dT）を含むとき、mRNA 分子は、プローブにハイブリダイズして、逆転写されうる。オリゴヌクレオチドのオリゴ（dT）部分は、cDNA 分子の第 1 鎮合成のためのプライマーとして作用しうる。例えば、図 2、216 に示す確率バーコードの非限定的な例において、mRNA 分子は、ビーズ上の確率バーコードをハイブリダイズすることができる。例えば、一本鎮ヌクレオチド断片は、確率バーコードの標的結合領域にハイブリダイズすることができる。

【0160】

10

20

30

40

50

結合は、確率バーコードの標的認識領域と標的核酸分子の一部とのライゲーションをさらに含みうる。例えば、標的結合領域は、制限部位オーバーハング（例えば、EcoRI付着末端オーバーハング）への特異的ハイブリダイゼーションが可能でありうる核酸配列を含みうる。アッセイ手順は、制限部位オーバーハングを生成するために制限酵素（例えばEcoRI）で標的核酸を処置する工程をさらに含みうる。次いで、確率バーコードは、制限部位オーバーハングに相補的な配列を含む任意の核酸分子にライゲートしうる。リガーゼ（例えばT4DNAリガーゼ）は2つの断片を連結するために使用しうる。

【0161】

例えば、図2、220に図示する確率バーコードの非限定的な例では、複数の細胞（又は複数のサンプル）からの標識標的（例えば、標的-バーコード分子）は、続いて、例えば、チューブ中にプールすることができる。例えば、確率バーコード及び／又は標的-バーコード分子が結合したビーズを回収することにより、標識標的をプールすることができる。

10

【0162】

結合した標的-バーコード分子の固体担体ベースのコレクションの回収は、磁気ビーズ及び外部印加磁界の使用により実現しうる。標的-バーコード分子をプールした後、すべてのさらなる処理を単一反応槽内で進行させることができる。さらなる処理は、例えば、逆転写反応、增幅反応、切断反応、解離反応、及び／又は核酸伸長反応を含みうる。さらなる処理反応は、マイクロウェル内で、すなわち、複数の細胞の標識標的核酸分子を最初にプールすることなく、実施することができる。

20

【0163】

逆転写

本開示は、（例えば、図2の224で）逆転写を用いて標的-バーコードコンジュゲート（例えば確率標的-バーコードコンジュゲート）を生成する方法を提供する。確率標的-バーコードコンジュゲートは、確率バーコードと標的核酸の全部又は一部の相補的配列と（すなわち、確率バーコード付きcDNA分子）を含みうる。関連付けられたRNA分子の逆転写は、逆転写酵素と共に逆転写プライマーを添加することによって起こりうる。逆転写プライマーは、オリゴ(dT)プライマー、ランダムヘキサヌクレオチドプライマー、又は標的特異的オリゴヌクレオチドプライマーでありうる。オリゴ(dT)プライマーは、12～18ヌクレオチド長、又は概ねそうしたヌクレオチド長であってよく、哺乳動物mRNAの3'末端の内因性ポリ(A)テールに結合することができる。ランダムヘキサヌクレオチドプライマーは、さまざまな相補的部位でmRNAと結合しうる。標的特異的オリゴヌクレオチドプライマーは、典型的には対象のmRNAを選択的にプライミングする。

30

【0164】

いくつかの実施形態では標識RNA分子の逆転写は、逆転写プライマーの添加によって起こりうる。いくつかの実施形態では、逆転写プライマーは、オリゴ(dT)プライマー、ランダムヘキサヌクレオチドプライマー、又は標的特異的オリゴヌクレオチドプライマーである。一般に、オリゴ(dT)プライマーは、12～18ヌクレオチド長であり、哺乳動物mRNAの3'末端の内因性ポリ(A)+テールに結合する。ランダムヘキサヌクレオチドプライマーは、さまざまな相補的部位でmRNAと結合しうる。標的特異的オリゴヌクレオチドプライマーは、典型的には対象のmRNAを選択的にプライミングする。

40

【0165】

逆転写は、繰返し行うことにより複数の標識cDNA分子を生成可能である。本明細書に開示される方法は、少なくとも約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、又は20回の逆転写反応を行う工程を含みうる。本方法は、少なくとも約25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、又は100回の逆転写反応を行う工程を含みうる。

【0166】

50

増幅

核酸増幅反応（例えば、図2の228で）は、標識標的核酸分子の複数のコピーを生成するために1回以上実施することができる。増幅は、複数の標的核酸配列が同時に増幅される、多重方式で実施してよい。増幅反応は、核酸分子にシーケンシングアダプターを付加するために使用することができる。増幅反応は、存在するのであれば、サンプル標識の少なくとも一部を増幅する工程を含みうる。増幅反応は、細胞及び／又は分子標識の少なくとも一部を増幅する工程を含みうる。増幅反応は、サンプルタグ、細胞標識、空間標識、分子標識、標的核酸、又はそれらの組合せの少なくとも一部を増幅する工程を含みうる。増幅反応は、複数の核酸の0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、100%、又はこれらの値のいずれか2つの間の範囲若しくは数を増幅する工程を含みうる。本方法は、サンプル標識、細胞標識、空間標識、及び／又は分子標識を含む標的-バーコード分子のcDNAコピーを1つ以上生成するために、cDNA合成反応を1回以上行う工程をさらに含みうる。

【0167】

いくつかの実施形態では、ポリメラーゼ連鎖反応（PCR）を用いて、増幅を実施することができる。本明細書で用いられる場合、PCRとは、DNAの相補鎖の同時プライマー伸長により特定のDNA配列のin vitro増幅を行う反応を意味しうる。本明細書で用いられる場合、PCRは、その反応の派生形、例えば、限定されるものではないが、RT-PCR、リアルタイムPCR、ネステッドPCR、定量PCR、多重PCR、デジタルPCR、及びアセンブリ-PCRを包含しうる。

【0168】

標識核酸の増幅は、非PCRベースの方法を含みうる。非PCRベースの方法の例としては、限定されるものではないが、多重置換増幅（MDA）、転写媒介増幅（TMA）、核酸配列ベースの増幅（NASBA）、鎖置換増幅（SDA）、リアルタイムSDA、ローリングサークル増幅、又はサークル-サークル増幅が挙げられる。他の非PCRベースの増幅方法としては、DNA若しくはRNA標的を増幅するためのDNA依存性RNAポリメラーゼ駆動RNA転写増幅又はRNA指向DNA合成及び転写の多重サイクル、リガーゼ連鎖反応（LCR）、及びQレプリカーゼ（Q）法、パリンドロームプローブの使用、鎖置換増幅、制限エンドヌクレアーゼを用いたオリゴヌクレオチド駆動増幅、プライマーが核酸配列にハイブリダイズされ且つ得られた二本鎖が伸長反応及び増幅の前に切断される増幅方法、5'エキソヌクレアーゼ活性の欠如した核酸ポリメラーゼを用いた鎖置換増幅、ローリングサークル増幅、及び分岐伸長増幅（RAM）が挙げられる。いくつかの実施形態では、増幅は、環化転写物を生成しうる。

【0169】

いくつかの実施形態では、本明細書に開示する方法は、標識アンプリコン（例えば確率標識アンプリコン）を生成するために標識核酸（例えば、標識RNA、標識DNA、標識cDNA）上でポリメラーゼ連鎖反応を実施する工程をさらに含む。標識アンプリコンは、二本鎖分子であってよい。二本鎖分子は、二本鎖RNA分子、二本鎖DNA分子、又はDNA分子にハイブリダイズされたRNA分子を含みうる。二本鎖分子の一方又は両方の鎖は、サンプル標識、空間標識、細胞標識、及び／又は分子標識を含みうる。確率標識アンプリコンは、一本鎖分子でありうる。一本鎖分子は、DNA、RNA、又はそれらの組合せを含みうる。本開示の核酸は、合成核酸又は改変核酸を含みうる。

【0170】

増幅は、1つ以上の非天然ヌクレオチドの使用を含みうる。非天然ヌクレオチドは、光不安定性又はトリガー性のヌクレオチドを含みうる。非天然ヌクレオチドの例としては、限定されるものではないが、ペプチド核酸（PNA）、モルホリノ核酸、及びロックド核酸（LNA）、さらにはグリコール核酸（GNA）及びトレオース核酸（TNA）が挙げられうる。非天然ヌクレオチドは、増幅反応の1サイクル以上に添加することができる。

10

20

30

40

50

非天然ヌクレオチドの添加は、増幅反応の特定のサイクル又は時点で産物を同定するために使用しうる。

【0171】

増幅反応を1回以上行う工程は、1つ以上のプライマーの使用を含みうる。1つ以上のプライマーは、例えば、少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、若しくは15ヌクレオチド又はそれ以上を含みうる。1つ以上のプライマーは、少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、若しくは15ヌクレオチド又はそれ以上を含みうる。1つ以上のプライマーは、12～15ヌクレオチド未満を含みうる。1つ以上のプライマーは、複数の確率標識標的の少なくとも一部にアニールしうる。1つ以上のプライマーは、複数の確率標識標的の3'末端又は5'末端にアニールしうる。1つ以上のプライマーは、複数の確率標識標的の内部領域にアニールしうる。内部領域は、複数の確率標識標的の3'末端から少なくとも約50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900、又は1000ヌクレオチドでありうる。1つ以上のプライマーは、プライマーの一一定パネルを含みうる。1つ以上のプライマーは、少なくとも1つ以上のカスタムプライマーを含みうる。1つ以上のプライマーは、少なくとも1つ以上の対照プライマーを含みうる。1つ以上のプライマーは、少なくとも1つ以上の遺伝子特異的プライマーを含みうる。

10

20

30

【0172】

1つ以上のプライマーは、ユニバーサルプライマーを含みうる。ユニバーサルプライマーは、ユニバーサルプライマー結合部位にアニールしうる。1つ以上のカスタムプライマーは、第1のサンプル標識、第2のサンプル標識、空間標識、細胞標識、分子標識、標的、又はそれらの任意の組合せにアニールしうる。1つ以上のプライマーは、ユニバーサルプライマー及びカスタムプライマーを含みうる。カスタムプライマーは、1つ以上の標的を増幅するように設計しうる。標的は、1つ以上のサンプル中の全核酸のサブセットを含みうる。標的は、1つ以上のサンプル中の全確率標識標的のサブセットを含みうる。1つ以上のプライマーは、少なくとも96カスタムプライマー又はそれ以上を含みうる。1つ以上のプライマーは、少なくとも960カスタムプライマー又はそれ以上を含みうる。1つ以上のプライマーは、少なくとも9600カスタムプライマー又はそれ以上を含みうる。1つ以上のカスタムプライマーは、2つ以上の異なる標識核酸にアニールしうる。2つ以上の異なる標識核酸は、1つ以上の遺伝子に相当しうる。

40

【0173】

任意の増幅スキームを本開示の方法で使用することができる。例えば、一スキームでは、第1ラウンドのPCRは、遺伝子特異的プライマー及びユニバーサルI11uminaシーケンシングプライマー1配列に対するプライマーを用いて、ビーズに結合された分子を増幅することができる。第2ラウンドのPCRは、I11uminaシーケンシングプライマー2配列がフランкиングするネステッド遺伝子特異的プライマーとユニバーサルI11uminaシーケンシングプライマー1配列に対するプライマーとを用いて第1のPCR産物を増幅可能である。第3ラウンドのPCRは、P5及びP7とサンプルインデックスを付加して、PCR産物をI11uminaシーケンシングライブラーにする。150bp×2シーケンシングを用いたシーケンシングは、リード1上の細胞標識及び分子標識、リード2上の遺伝子、並びにインデックス1リード上のサンプルインデックスを明らかにしうる。

40

【0174】

いくつかの実施形態では、核酸は、化学切断を用いて基材から除去可能である。例えば、核酸中に存在する化学基又は修飾塩基は、固体担体からのその除去を促進するために使用可能である。例えば、酵素は、基材から核酸を除去するために使用可能である。例えば

50

、核酸は、制限エンドヌクレアーゼ消化による基材からの除去が可能である。例えば、dUTP又はddUTPを含有する核酸のウラシル-d-グリコシラーゼ(UDG)処理は、基材から核酸を除去するために使用可能である。例えば、核酸は、ヌクレオチド切除を行う酵素、例えば、塩基除去修復酵素、例えば、脱プリン/脱ピリミジン(AP)エンドヌクレアーゼを用いて基材から除去可能である。いくつかの実施形態では、核酸は、光切断性基と光とを用いて基材から除去可能である。いくつかの実施形態では、切断性リンカーは、基材から核酸を除去するために使用可能である。例えば、切断性リンカーは、ビオチン/アビジン、ビオチン/ストレプトアビジン、ビオチン/ニュートラビジン、Ig-プロテインA、光不安定性リンカー、酸又は塩基不安定性リンカー基、又はアプタマーの少なくとも1つを含みうる。

10

【0175】

プローブが遺伝子特異的である場合、分子は、プローブにハイブリダイズし、逆転写及び/又は増幅が可能である。いくつかの実施形態では、核酸が合成された後(例えば、逆転写された後)、増幅が可能である。増幅は、複数の標的核酸配列が同時に増幅される条件で、多重方式で行いうる。増幅は、核酸にシーケンシングアダプターを付加しうる。

【0176】

いくつかの実施形態では、増幅は、例えばブリッジ増幅を用いて基材上に行うことが可能である。基材上でオリゴ(dT)プローブを用いてブリッジ増幅するのに適合していた末端を生成するために、cDNAにホモポリマーテールを付加することが可能である。ブリッジ増幅では、テンプレート核酸の3'末端に相補的なプライマーは、固体粒子に共有結合された各ペアの第1のプライマーでありうる。テンプレート核酸を含有するサンプルが粒子に接触して1回の熱サイクルが行われる場合、テンプレート分子は第1のプライマーにアニールし、且つ第1のプライマーはヌクレオチドの付加により順方向に伸長して、テンプレート分子とテンプレートに相補的な新たに形成されたDNA鎖とからなる二本鎖分子を形成する。次のサイクルの加熱工程では、二本鎖分子は変性されて、粒子からテンプレート分子を放出し、第1のプライマーを介して粒子に結合された相補的DNA鎖を残存させる。続くアニーリング・伸長工程のアニーリング段階では、相補鎖は、第1のプライマーから除去された位置の相補鎖のセグメントに相補的な第2のプライマーにハイブリダイズ可能である。このハイブリダイゼーションにより、相補鎖は、共有結合により第1のプライマーに且つハイブリダイゼーションにより第2のプライマーに固定されたブリッジを第1及び第2のプライマー間に形成可能である。伸長段階では、第2のプライマーは、同一の反応混合物中にヌクレオチドを添加することにより反対方向に伸長し、それによりブリッジを二本鎖ブリッジに変換可能である。次いで、次のサイクルが開始され、二本鎖ブリッジは変性されて、それぞれ第1及び第2のプライマーを介して粒子表面に結合された一方の末端と、それぞれ未結合の状態の他方の末端と、を有する2つの一本鎖核酸分子を与えることが可能である。この第2のサイクルのアニーリング・伸長工程では、各鎖は同一の粒子上のこれまで未使用であったさらなる相補的プライマーにハイブリダイズして新しい一本鎖ブリッジを形成可能である。この時点でハイブリダイズされる2つのこれまで未使用であったプライマーは伸長して2つの新しいブリッジを二本鎖ブリッジに変換可能である。

20

30

30

【0177】

増幅反応は、複数の核酸の少なくとも1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、又は100%を増幅する工程を含みうる。

40

【0178】

標識核酸の増幅は、PCRベースの方法又は非PCRベースの方法を含みうる。標識核酸の増幅は、標識核酸の指数関数的増幅を含みうる。標識核酸の増幅は、標識核酸の線形増幅を含みうる。増幅は、ポリメラーゼ連鎖反応(PCR)により行うことが可能である。PCRは、DNAの相補鎖の同時プライマー伸長により特定のDNA配列のin vi

50

tro 増幅を行う反応を意味しうる。PCRは、その反応の派生形、例えば、限定されるものではないが、RT-PCR、リアルタイムPCR、ネステッドPCR、定量PCR、多重PCR、ディジタルPCR、サプレッションPCR、セミサプレッショ PCR、及びアセンブリーPCRを包含しうる。

【0179】

いくつかの実施形態では、標識核酸の増幅は非PCRベースの方法を含む。非PCRベースの方法の例としては、限定されるものではないが、多重置換増幅(MDA)、転写媒介増幅(TMA)、核酸配列ベースの増幅(NASBA)、鎖置換増幅(SDA)、リアルタイムSDA、ローリングサークル増幅、又はサークル-サークル増幅が挙げられる。他の非PCRベースの増幅方法としては、DNA若しくはRNA標的を増幅するためのDNA依存性RNAポリメラーゼ駆動RNA転写増幅又はRNA指向DNA合成及び転写の多重サイクル、リガーゼ連鎖反応(LCR)、Qレプリカーゼ(Q)、パリンドロームプローブの使用、鎖置換増幅、制限エンドヌクレアーゼを用いたオリゴヌクレオチド駆動増幅、プライマーが核酸配列にハイブリダイズされ且つ得られた二本鎖が伸長反応及び増幅の前に切断される増幅方法、5'エキソヌクレアーゼ活性の欠如した核酸ポリメラーゼを用いた鎖置換増幅、ローリングサークル増幅、及び/又は分岐伸長増幅(RAM)が挙げられる。

【0180】

いくつかの実施形態では、本明細書に開示される方法は、増幅アンプリコン(例えば標的)上でネステッドポリメラーゼ連鎖反応を行う工程をさらに含む。アンプリコンは二本鎖分子でありうる。二本鎖分子は、二本鎖RNA分子、二本鎖DNA分子、又はDNA分子にハイブリダイズされたRNA分子を含みうる。二本鎖分子の一方又は両方の鎖は、サンプルタグ又は分子識別子標識を含みうる。代替的に、アンプリコンは一本鎖分子でありうる。一本鎖分子は、DNA、RNA、又はそれらの組合せを含みうる。本発明の核酸は、合成核酸又は改変核酸を含みうる。

【0181】

いくつかの実施形態では、本方法は、多数のアンプリコンを生成するために標識核酸を繰返し増幅する工程を含む。本明細書に開示される方法は、少なくとも約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、又は20回の増幅反応を行う工程を含みうる。代替的に、本方法は、少なくとも約25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、又は100回の増幅反応を行う工程を含む。

【0182】

増幅工程は、複数の核酸を含む1つ以上のサンプルに1つ以上の対照核酸を添加する工程をさらに含みうる。増幅工程は、複数の核酸に1つ以上の対照核酸を添加する工程をさらに含みうる。対照核酸は、対照標識を含みうる。

【0183】

増幅は、1つ以上の非天然ヌクレオチドの使用を含みうる。非天然ヌクレオチドは、光不安定性及び/又はトリガー性ヌクレオチドを含みうる。非天然ヌクレオチドの例としては、限定されるものではないが、ペプチド核酸(PNA)、モルホリノ核酸及びロックド核酸(LNA)、さらにはグリコール核酸(GNA)及びトレオース核酸(TNA)が挙げられる。非天然ヌクレオチドは、増幅反応の1サイクル以上に添加しうる。非天然ヌクレオチドの添加は、増幅反応の特定のサイクル又は時点で産物を同定するために使用しうる。

【0184】

増幅反応を1回以上行う工程は、1つ以上のプライマーの使用を含みうる。1つ以上のプライマーは1つ以上のオリゴヌクレオチドを含みうる。1つ以上のオリゴヌクレオチドは少なくとも約7~9ヌクレオチドを含みうる。1つ以上のオリゴヌクレオチドは12~15ヌクレオチド未満を含みうる。1つ以上のプライマーは、複数の標識核酸の少なくとも一部にアニールしうる。1つ以上のプライマーは、複数の標識核酸の3'末端及び/又は

10

20

30

30

40

50

5'末端にアニールしうる。1つ以上のプライマーは、複数の標識核酸の内部領域にアニールしうる。内部領域は、複数の標識核酸の3'末端から少なくとも約50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900、又は1000ヌクレオチドである。1つ以上のプライマーは、プライマーの一定パネルを含みうる。1つ以上のプライマーは、少なくとも1つ以上のカスタムプライマーを含みうる。1つ以上のプライマーは、少なくとも1つ以上の対照プライマーを含みうる。1つ以上のプライマーは、少なくとも1つ以上のハウスキーピング遺伝子プライマーを含みうる。1つ以上のプライマーは、ユニバーサルプライマーを含みうる。ユニバーサルプライマーは、ユニバーサルプライマー結合部位にアニールしうる。1つ以上のカスタムプライマーは、第1のサンプルタグ、第2のサンプルタグ、分子識別子標識、核酸、又はその産物にアニールしうる。1つ以上のプライマーは、ユニバーサルプライマー及びカスタムプライマーを含みうる。カスタムプライマー、1つ以上の標的核酸を増幅するように設計しうる。標的核酸は、1つ以上のサンプル中の全核酸のサブセットを含みうる。いくつかの実施形態では、プライマーには、本開示のアレイに結合されたプロープである。

【0185】

いくつかの実施形態では、サンプル中の複数の標的にバーコード（例えば確率バーコード）を付ける工程は、確率バーコード付き断片の指標インデックスライブラリーを作製する工程をさらに含む。異なる確率バーコードの分子標識は、互いに異なっていてもよい。確率バーコード付き標的の指標インデックスライブラリーを作製する工程は、サンプル中の複数の標的から複数の指標インデックスポリヌクレオチドを作製する工程を含む。例えば、第1の指標インデックス標的と第2の指標インデックス標的とを含む確率バーコード標的の指標インデックスライブラリーの場合、第1の指標インデックスポリヌクレオチドの標識領域は、第2の指標インデックスポリヌクレオチドの標識領域と、1、2、3、4、5、6、7、8、9、10、20、30、40、50ヌクレオチド異なって、概ね、少なくとも、若しくは多くともこうした値、又はこれらの値のいずれか2つの間の数若しくは範囲のヌクレオチド異なってもよい。いくつかの実施形態では、確率バーコード付き標的の指標インデックスライブラリーを作製する工程は、ポリ（T）領域及び標識領域などの複数のオリゴヌクレオチドと、複数の標識、例えば、mRNA分子を接触させる工程と；各々がcDNA領域及び標識領域を含む一本鎖標識cDNA分子を生成するために、逆転写酵素を用いて、第1鎖合成を実施する工程と、を含み、ここで、複数の標的は、異なる配列の少なくとも2つのmRNA分子を含み、複数のオリゴヌクレオチドは、異なる配列の少なくとも2つのオリゴヌクレオチドを含む。確率バーコード付き標的の指標インデックスライブラリーを作製する工程は、さらに、二本鎖標識cDNA分子を生成するために、一本鎖標識cDNA分子を増幅する工程と；標識アンプリコンを生成するために、二本鎖標識cDNA分子上でネステッドPCRを実施する工程と、を含む。いくつかの実施形態では、本方法は、アダプター-標識アンプリコンを作製する工程を含みうる。

【0186】

バーコーディング（例えば確率バーコーディング）は、個々の核酸（例えば、DNA又はRNA）分子を標識するために、核酸バーコード若しくはタグを使用しうる。いくつかの実施形態では、これは、DNAバーコード若しくはタグがmRNAから生成される際に、cDNA分子にこれらを付加する工程を含む。ネステッドPCRは、PCR増幅バイアスの最小化を実施することができる。アダプターは、例えば、次世代シーケンシング（NGS）を用いるシーケンシングのために付加することができる。シーケンシング結果を用いて、例えば、図2の232に位置する標的の1つ以上のコピーの細胞標識、分子標識、及びヌクレオチド断片の配列を決定することができる。

【0187】

10

20

30

40

50

図3は、確率バーコード付き標的、例えば、mRNAの指標インデックスライブライアを作製する非限定的な例示的プロセスを示す概略図である。ステップ1に示すように、逆転写プロセスは、ユニーク分子標識、細胞標識、及びユニバーサルPCR部位を含む各mRNA分子をコードすることができる。特に、分子識別子標識310のセットとmRNA分子302のポリ(A)テール領域308の確率論的ハイブリダイゼーションによって、mRNA分子302を逆転写して、cDNA領域306を含む標識cDNA分子304を生成することができる。分子識別子標識310の各々は、標的結合領域、例えば、ポリ(dT)領域312、標識領域314、及びユニバーサルPCR領域316を含みうる。

【0188】

いくつかの実施形態では、細胞標識は、3~20ヌクレオチドを含みうる。いくつかの実施形態では、分子標識は、3~20ヌクレオチドを含みうる。いくつかの実施形態では、複数の確率バーコードの各々は、1つ以上のユニバーサル標識及び細胞標識をさらに含み、ユニバーサル標識は、固体担体上の複数の確率バーコードについて同じであり、細胞標識は、固体担体上の複数の確率バーコードについて同じである。いくつかの実施形態では、ユニバーサル標識は、3~20ヌクレオチドを含みうる。いくつかの実施形態では、細胞標識は、3~20ヌクレオチドを含む。

【0189】

いくつかの実施形態では、標識領域314は、分子標識318及び細胞標識320を含みうる。いくつかの実施形態では、標識領域314は、1つ以上のユニバーサル標識、次元標識、及び細胞標識を含みうる。分子標識318は、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100ヌクレオチド長であっても、概ね、少なくとも、若しくは多くともそうしたヌクレオチド長であってもよいし、又はこれらの値のいずれかの間の数若しくは範囲のヌクレオチド長であってもよい。細胞標識320は、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100ヌクレオチド長であっても、概ね、少なくとも、若しくは多くともそうしたヌクレオチド長であってもよいし、又はこれらの値のいずれかの間の数若しくは範囲のヌクレオチド長であってもよい。ユニバーサル標識は、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100ヌクレオチド長であってもよいし、又はこれらの値のいずれかの間の数若しくは範囲のヌクレオチド長であってもよい。ユニバーサル標識は、固体担体上の複数の確率バーコードについて同じであってもよく、細胞標識は、固体担体上の複数の確率バーコードについて同じであってもよい。次元標識は、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100ヌクレオチド長であっても、概ね、少なくとも、若しくは多くともそうしたヌクレオチド長であってもよいし、又はこれらの値のいずれかの間の数若しくは範囲のヌクレオチド長であってもよい。

【0190】

いくつかの実施形態では、標識領域314は、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000の異なる標識を含むか、概ねそうした値の異なる標識を含むか、少なくとも、若しくは多くともそうしたヌクレオチド長であってもよいし、又はこれらの値のいずれかの間の数若しくは範囲のヌクレオチド長であってもよい。分子識別子標識310のセットは、10、20、40、50、70、80、90、10²、10³、10⁴、10⁵、10⁶、10⁷、10⁸、10⁹、10¹⁰、10¹¹、10¹²、10¹³、10¹⁴、10¹⁵、10²⁰の分子識別子標識310を含むか、概ねそうした値の識別子標識310を含むか、少なくとも、若しくは多くともそうした値の分子識別子標識310、又はこれらの値のいずれかの間の数若しくは範

10

20

30

40

50

囲の分子識別子標識 310 を含みうる。また、分子識別子標識 310 のセットは、例えば、各々、ユニーク標識領域 314 を含みうる。余剰の分子識別子標識 310 を除去するために、標識 cDNA 分子 304 を精製することができる。精製は、Ampure ビーズ精製を含みうる。

【0191】

工程 2 に示すように、工程 1 の逆転写プロセスからの産物を 1 チューブ中にプールし、第 1 PCR プライマーブール及び第 1 ユニバーサル PCR プライマーを用いて PCR 増幅することができる。プールする工程は、ユニーク標識領域 314 によって可能である。特に、ネステッド PCR 標識アンプリコン 322 を生成するために、標識 cDNA 分子 304 を増幅することができる。増幅は、多重 PCR 増幅を含みうる。増幅は、单一反応量で 96 多重プライマーを用いる多重 PCR 増幅を含みうる。いくつかの実施形態では、多重 PCR 増幅は单一反応量で 10、20、40、50、70、80、90、10²、10³、10⁴、10⁵、10⁶、10⁷、10⁸、10⁹、10¹⁰、10¹¹、10¹²、10¹³、10¹⁴、10¹⁵、10²⁰ の多重プライマーを使用するか、概ねそうした値の多重プライマー、少なくとも、若しくは多くともそうした値の多重プライマーを使用するか、又はこれらの値のいずれかの間の数若しくは範囲の多重プライマーを使用することができる。増幅は、特定の遺伝子を標的とするカスタムプライマー-326A～C の第 1 PCR プライマーブール 324 と、ユニバーサルプライマー-328 とを含みうる。カスタムプライマー-326 は、標識 cDNA 分子 304 の cDNA 部分 306' 内の 1 領域とハイブリダイズすることができる。ユニバーサルプライマー-328 は、標識 cDNA 分子 304 のユニバーサル PCR 領域 316 とハイブリダイズすることができる。

【0192】

図 3 の工程 3 に示すように、工程 2 の PCR 増幅からの産物は、ネステッド PCR プライマーブール及び第 2 ユニバーサル PCR プライマーを用いて増幅することができる。ネステッド PCR は、PCR 増幅バイアスを最小限に抑えることができる。特に、ネステッド PCR 標識アンプリコン 322 は、ネステッド PCR によりさらに増幅することもできる。ネステッド PCR は、单一反応量でネステッド PCR プライマー-332a～c のネステッド PCR プライマーブール 330 と、第 2 ユニバーサル PCR プライマー-328' とを含む多重 PCR を含みうる。ネステッド PCR プライマーブール 328 は、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000 の異なるネステッド PCR プライマー-330 を含むか、概ねそうした値の異なるネステッド PCR プライマー-330 を含むか、少なくとも、若しくは多くともそうした値の異なるネステッド PCR プライマー-330、又はこれらの値のいずれかの間の数若しくは範囲の異なるネステッド PCR プライマー-330 を含みうる。ネステッド PCR プライマー-332 は、アダプター-334 を含有して、標識アンプリコン 322 の cDNA 部分 306' 内の 1 領域とハイブリダイズすることができる。ユニバーサルプライマー-328' は、アダプター-336 を含有して、標識アンプリコン 322 のユニバーサル PCR 領域 316 とハイブリダイズすることができる。このようにして、工程 3 は、アダプター-標識アンプリコン 338 を生成する。いくつかの実施形態では、ネステッド PCR プライマー-332 と第 2 ユニバーサル PCR プライマー-328' は、アダプター-334 及び 336 を含有しなくてもよい。それに代わり、アダプター-334 及び 336 は、アダプター-標識アンプリコン 338 を生成するために、ネステッド PCR の産物とライゲートすることができる。

【0193】

工程 4 に示すように、工程 3 からの PCR 産物は、ライブラー増幅プライマーを用いたシーケンシングのために PCR 增幅することができる。特に、アダプター-334 及び 336 を用いて、アダプター-標識アンプリコン 338 に対するアッセイをさらに 1 回以上実施することができる。アダプター-334 及び 336 は、プライマー-340 及び 342 とハイブリダイズすることができる。1 つ以上のプライマー-340 及び 342 は、PCR 增幅プライマーであってよい。1 つ以上のプライマー-340 及び 342 は、シーケンシングブ

10

20

30

40

50

ライマーであってよい。1つ以上のアダプター334及び336は、アダプター標識アンプリコン338のさらなる増幅のために使用することができる。1つ以上のアダプター334及び336は、アダプター標識アンプリコン338のシーケンシングのために使用することができる。プライマー342は、プレート指標インデックス344を含有することができ、これによって、分子識別子標識318の同じセットを用いて生成されたアンプリコンを、次世代シーケンシング(NGS)を用いた1回のシーケンシング反応でシーケンシングすることができる。

【0194】

合成マルチプレットを用いるマルチプレット発現プロファイルの同定

単細胞の発現プロファイルを決定するとき、2個の細胞は1個の細胞として同定されるかもしれない、2個の細胞の発現プロファイルは、(ダブレット発現プロファイルと称される)1個の細胞における発現プロファイルとして同定されるかもしれない。例えば、バーコーディング(例えば確率バーコーディング)を用いて2個の細胞の発現プロファイルを決定するとき、2個の細胞のmRNA分子は、同じ細胞標識を有するバーコードに関連付けられるかもしれない。別の例として、2個の細胞は、1個の粒子(例えばビーズ)に関連付けられるかもしれない。該粒子は、同じ細胞標識を有するバーコードを含みうる。細胞を溶解後、2個の細胞におけるmRNA分子は、該粒子のバーコード、ひいては同じ細胞標識に関連付けることができる。ダブレット発現プロファイルは、発現プロファイルの解釈を歪める可能性がある。本明細書には、合成マルチプレット発現プロファイルを用いてマルチプレット発現プロファイルを同定するためのシステム及び方法が開示される。

【0195】

図4は、合成ダブレット分析を分析パイプラインに組み込む、非限定的な例示的ワークフローを示す。既存の分析パイプラインにより、細胞の発現プロファイル、例えば数千個の細胞の発現プロファイルを作成することができる。分析パイプラインからの発現プロファイルは、合成ダブレットの発現プロファイル(合成ダブレット発現プロファイルとも称される)を作成するために使用可能である。合成ダブレット発現プロファイルの作成は、図5に関連して、以下に詳述される。つまり、合成ダブレット発現プロファイルを作成するため、2個の細胞の2つの発現プロファイルが組合せ可能である。

【0196】

いくつかの実施形態では、発現プロファイル及び合成ダブレット発現プロファイルは、任意選択的には、機械学習モデル(例えば教師あり機械学習モデル)を訓練するため、使用可能である。教師あり機械学習モデルは、1つ以上の発現プロファイルの各々をシングレット又はダブルットとして同定又は分類するため、使用可能である。教師あり機械学習モデルの出力は、1つ以上の発現プロファイルの各々がシングレット又はダブルットでの同定のいずれを有するかを示すアノテーションファイルでありうる。本明細書で開示される合成ダブレット分析方法は、ダブルット発現プロファイルの誤解を低減し、ダブルット発現プロファイルを生物学的発見及び洞察と混同するリスクを改善しうる。合成ダブルット分析方法は、完全に自動化又は半自動化されうる。いくつかの実施形態では、ユーザーは、作成された合成ダブルット発現プロファイルを用いて、例えば発現プロファイル及び合成ダブルット発現プロファイルの投影プロットを点検することにより、ダブルット発現プロファイルを手作業で同定することができる。

【0197】

図5は、合成ダブルット分析の非限定的な例示的方法500を示すフローチャートである。ブロック504で、複数の細胞における複数の標的では、任意選択的には、複数のバーコード(例えば確率バーコード)を用いてバーコードを付けて(例えば確率バーコードを付けて)、複数の細胞の細胞において複数のバーコード付き標的(例えば確率バーコード付き標的)を生成することができる。複数のバーコードの各々は、細胞標識及び分子標識を含みうる。複数のバーコードの少なくとも2つのバーコードの分子標識は、異なる分子標識配列を含みうる。複数のバーコードの少なくとも2つのバーコードは、同一の細胞標識配列を有する細胞標識を含みうる。ブロック508で、複数のバーコード付き標的の

10

20

30

40

50

シーケンシングデータが取得可能である。

【0198】

ブロック512で、任意選択的には、シーケンシングデータからの複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルが決定されうる。複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルは、シーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を含みうる。いくつかの実施形態では、シーケンシングデータから複数のバーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルについて、シーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程を含む。シーケンシングデータ中の細胞標識及び複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程は、複数の標的の1つ以上について、(1)シーケンシングデータ中の標的に関連付けられた識別可能な配列を有する分子標識の数をカウントする工程と；(2)(1)でカウントされたシーケンシングデータ中の標的に関連付けられた識別可能な配列を有する分子標識の数に基づいて標的の数を推定する工程と、を含みうる。

10

【0199】

いくつかの実施形態では、複数の標的にバーコードを付けて(例えば確率バーコードを付けて)、複数のバーコード付き標的(例えば確率バーコード付き標的)のシーケンシングデータを取得し、複数の発現プロファイルを決定する代わりに、複数の細胞の複数の発現プロファイルを受け取ることができる。複数の発現プロファイルは、複数の細胞の各細胞における複数の標的の各標的の存在量(又はコピー若しくは数)を含む。

20

【0200】

ブロック516で、複数の合成マルチプレット発現プロファイルが、決定された複数のバーコード(例えば確率バーコード)の細胞標識に関連付けられた複数の発現プロファイルから作成されうる。マルチプレットは、異なる実行において異なる可能性がある。いくつかの実施形態では、複数のマルチプレットは、ダブルレット、トリプレット、カルテット、クインテット、セクステット、セプテット、オクテット、ノネット、又はそれらの任意の組合せを含みうる。マルチプレットは、任意のnプレット(n-p1e1)でありうる。いくつかの実施形態では、nは、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、又はこれらの値のいずれか2つの間の範囲、或いはそうした近似値である。いくつかの実施形態では、nは、少なくとも、又は多くとも、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、若しくは20である。

30

【0201】

いくつかの実施形態では、ブロック516で決定された複数の確率バーコードの細胞標識に関連付けられた複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程は、複数の合成マルチプレット発現プロファイルの合成マルチプレット発現プロファイルについて、(1)複数の発現プロファイルの第1の発現プロファイルを選択する工程と；(2)複数の発現プロファイルの第2の発現プロファイルを選択する工程と；(3)(1)で選択された第1の発現プロファイルと(2)で選択された第2の発現プロファイルとを組合せ、合成マルチプレット発現プロファイルを作成する工程と、を含む。第1の発現プロファイルは、第1の細胞標識配列に関連付けられうる。第2の発現プロファイルは、第2の細胞標識配列に関連付けられうる。第1の細胞標識配列及び第2の細胞標識配列は、異なる細胞標識配列を含みうる。

40

【0202】

いくつかの実施形態では、(1)で選択された第1の発現プロファイルと(2)で選択された第2の発現プロファイルとを組合せ、合成マルチプレット発現プロファイルを作成する工程は、複数の標的の各々について、第1の発現プロファイルにおける標的に関連付

50

けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数とを組合せ、合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数を作成する工程を含む。

【0203】

いくつかの実施形態では、合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との合計である。該合計は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との加重合計でありうる。

10

【0204】

いくつかの実施形態では、合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との平均である。該平均は、第1の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数と第2の発現プロファイルにおける標的に関連付けられた識別可能な配列を有する分子標識の数との加重平均でありうる。

20

【0205】

ロック520で、複数のバーコード（例えば確率バーコード）の細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルは、ロック516で作成された複数の合成ダブルット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定されうる。複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、（1）発現プロファイルマルチプレットの同定を意図して機械学習モデルを、複数の合成マルチプレット発現プロファイルの発現プロファイル及び複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程と；（2）複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。機械学習モデルを訓練する際に用いられる複数の発現プロファイルの1つ以上の発現プロファイルは、複数の発現プロファイルの百分率を含みうる。

30

【0206】

機械学習モデルを訓練する際に用いられる複数の発現プロファイルの百分率は、異なる実行において異なる可能性があり、1%～50%の範囲、例えば約10パーセントでありうる。いくつかの実施形態では、該百分率は、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はこれらの値のいずれか2つの間の範囲、或いはそうした近似値でありうる。いくつかの実施形態では、該百分率は、少なくとも、又は多くとも、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はこれらの値のいずれか2つの間の範囲でありうる。

40

【0207】

いくつかの実施形態では、機械学習モデルは、分類モデルを含む。分類モデルは、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含

50

みうる。機械学習モデルは、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k 近傍法（KNN）モデル、k 平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含みうる。機械学習モデルは、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スパース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含みうる。

【0208】

いくつかの実施形態では、本方法は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数のバーコード（例えば確率バーコード）の発現プロファイルを投影して、複数のバーコードの投影された発現プロファイルを作成する工程を含む。複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のバーコードの細胞標識の細胞標識に関連付けられた複数の発現プロファイルの発現プロファイルを、機械学習モデルを用いて発現プロファイル及び複数のバーコードの投影された発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含みうる。

10

【0209】

いくつかの実施形態では、本方法は、任意選択的には、発現プロファイルがマルチプレットとして同定される場合、該発現プロファイルに関連付けられたシーケンシングデータをシーケンシングデータから除去する工程を含む。本方法は、発現プロファイルがマルチプレットとして同定される場合、該発現プロファイルを複数の発現プロファイルから除去する工程を含みうる。

20

【0210】

いくつかの実施形態では、発現プロファイルマルチプレットの同定を意図して機械学習モデルを、複数の合成マルチプレット発現プロファイルの発現プロファイル及び複数の発現プロファイルの 1 つ以上の発現プロファイルから訓練する工程は、（1）複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程と；（2）複数の発現プロファイルの 1 つ以上の発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の発現プロファイルの 1 つ以上の投影された発現プロファイルを作成する工程と；（3）発現プロファイルマルチプレットの同定を意図して機械学習モデルを、（1）からの複数の合成マルチプレット発現プロファイルの投影された発現プロファイル及び（2）における複数の発現プロファイルの 1 つ以上の投影された発現プロファイルから訓練する工程とを含む。

30

【0211】

いくつかの実施形態では、低次元空間は、二次元空間でありうる。複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程は、t 分布型確率的近傍埋め込み（t-SNE）法を用いて、複数の合成マルチプレット発現プロファイルの発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程を含みうる。

40

【0212】

合成マルチプレットを用いてのマルチプレットの同定

本明細書では、マルチプレットプロファイルを同定するための方法が開示される。いくつかの実施形態では、本方法は、（a）複数の細胞の複数のプロファイルを受け取る工程と；（b）複数の細胞の複数のプロファイルから複数の合成マルチプレットプロファイルを作成する工程と；（c）複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、（b）で作成された複数の合成マルチプレットプロファイルのプロファイル

50

に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。複数の細胞の複数のプロファイルのプロファイルは、細胞のmRNA発現プロファイル、細胞のタンパク質発現プロファイル、細胞の突然変異プロファイル、細胞のメチル化プロファイル、又はそれらの任意の組合せを含みうる。

【0213】

いくつかの実施形態では、mRNA発現プロファイルは、複数の細胞の各細胞における複数の遺伝子の各遺伝子のmRNA分子の存在量を含みうる。各遺伝子の存在量は、遺伝子の絶対存在量、遺伝子の正規化された存在量、又はそれらの組合せを含みうる。遺伝子の正規化された存在量は、マッピングされたリード100万あたりの転写物1キロベース当たりのリード(RPKM)の単位又は閾値数(Ct)の単位で決定されうる。mRNA発現プロファイルは、シーケンシング、定量ポリメラーゼ連鎖反応(qPCR)、デジタルPCR、ハイブリダイゼーション、又はそれらの任意の組合せにより決定されうる。いくつかの実施形態では、細胞のタンパク質発現プロファイルは、複数の細胞の各細胞における複数の遺伝子の各遺伝子に対応するタンパク質分子の存在量を含む。細胞の突然変異プロファイルは、細胞の複数のゲノム位置での細胞の突然変異プロファイルを含みうる。細胞のメチル化プロファイルは、細胞の複数のゲノム位置での細胞のメチル化プロファイルを含みうる。

10

【0214】

マルチプレットは、異なる実施において異なる可能性がある。いくつかの実施形態では、複数のマルチプレットは、ダブルット、トリプレット、カルテット、クインテット、セクステット、セプテット、オクテット、ノネット、又はそれらの任意の組合せを含みうる。マルチプレットは、任意のnプレットでありうる。いくつかの実施形態では、nは、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、又はこれらの値のいずれか2つの間の範囲、或いはそうした近似値である。いくつかの実施形態では、nは、少なくとも、又は多くとも、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、若しくは20である。

20

【0215】

いくつかの実施形態では、本方法は、プロファイルが(c)におけるマルチプレットとして同定される場合、該プロファイルを(a)で得られた複数のプロファイルから除去する工程を含みうる。複数のマルチプレットは、ダブルット、トリプレット、又はそれらの任意の組合せを含みうる。いくつかの実施形態では、複数の細胞の複数のプロファイルから複数の合成マルチプレットプロファイルを作成する工程は、複数の合成マルチプレットプロファイルの合成マルチプレットプロファイルについて、(1)複数のプロファイルの、プロファイルの数を選択する工程と；(2)(1)で選択されたプロファイルを組合せ、合成マルチプレットプロファイルを作成する工程と、を含む。

30

【0216】

いくつかの実施形態では、(1)で選択されたプロファイルを組合せ、合成マルチプレットプロファイルを作成する工程は、複数の標的の各々について、選択されたプロファイルにおける対応する要素の値を組合せ、合成マルチプレットプロファイルにおける対応する要素の値を作成する工程を含む。合成マルチプレットプロファイルにおける対応する要素の値は、選択されたプロファイルにおける対応する要素の値の合計でありうる。該合計は、選択されたプロファイルにおける対応する要素の値の加重合計でありうる。合成マルチプレットプロファイルにおける標的の存在量は、選択されたプロファイルにおける対応する要素の値の平均でありうる。該平均は、選択されたプロファイルにおける対応する要素の値の加重平均でありうる。

40

【0217】

いくつかの実施形態では、複数の合成マルチプレットプロファイルの数は、およそ(a)で得られた複数のプロファイルの百分率である。該百分率は、約10パーセントでありうる。いくつかの実施形態では、該百分率は、1、2、3、4、5、6、7、8、9、1

50

0、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はこれらの値のいずれか2つの間の範囲、或いはそうした近似値でありうる。いくつかの実施形態では、該百分率は、少なくとも、又は多くとも、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はこれらの値のいずれか2つの間の範囲でありうる。

【0218】

10

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、(1)プロファイルマルチプレットの同定を意図して機械学習モデルを(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び(a)で得られた複数のプロファイルの1つ以上のプロファイルから訓練する工程と；(2)複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、機械学習モデルを用いてプロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

【0219】

20

いくつかの実施形態では、機械学習モデルを訓練する際に用いられる複数のプロファイルの1つ以上のプロファイルは、(b)で得られた複数のプロファイルの百分率を含む。該百分率は、約10パーセントでありうる。いくつかの実施形態では、該百分率は、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はこれらの値のいずれか2つの間の範囲、或いはそうした近似値でありうる。いくつかの実施形態では、該百分率は、少なくとも、又は多くとも、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はこれらの値のいずれか2つの間の範囲でありうる。

【0220】

30

いくつかの実施形態では、機械学習モデルは、分類モデルを含みうる。いくつかの実施形態では、分類モデルは、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含む。機械学習モデルは、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k近傍法(KNN)モデル、k平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含みうる。機械学習モデルは、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スペース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含みうる。

【0221】

40

いくつかの実施形態では、プロファイルマルチプレットの同定を意図して機械学習モデルを、(b)で作成された複数の合成マルチプレットプロファイルのプロファイル及び(a)で得られた複数のプロファイルの1つ以上のプロファイルから訓練する工程は、(1)(b)で作成された複数の合成マルチプレットプロファイルのプロファイルをプロファイル空間から低次元投影空間に投影して、複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程と；(2)(a)で得られた複数のプロファイルの1つ以上のプロファイルをプロファイル空間から低次元投影空間に投影して、複数のプロフ

50

ファイルの 1 つ以上の投影されたプロファイルを作成する工程と ; (3) プロファイルマルチプレットの同定を意図して機械学習モデルを、(1) からの複数の合成マルチプレットプロファイルの投影されたプロファイル及び(1) における複数のプロファイルの 1 つ以上の投影されたプロファイルから訓練する工程と、を含む。

【 0 2 2 2 】

いくつかの実施形態では、本方法は、複数の細胞の細胞に関連付けられた複数の複数のプロファイルのプロファイルを投影して、複数のプロファイルの投影されたプロファイルを作成する工程を含み、ここで複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、機械学習モデルを用いてプロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、機械学習モデルを用いて複数のプロファイルの投影されたプロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む。低次元空間は、二次元空間でありうる。いくつかの実施形態では、(b) で作成された複数の合成マルチプレットプロファイルのプロファイルをプロファイル空間から低次元投影空間に投影して、複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程は、(b) で作成された複数の合成マルチプレットプロファイルのプロファイルをプロファイル空間から低次元投影空間に投影して、複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程を含み、 t 分布型確率的近傍埋め込み (t S N E) 法を用いる工程を含む。

10

【 0 2 2 3 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b) で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルと複数のプロファイルの少なくとも 1 つのプロファイルとの間の第 1 の距離、及び細胞に関連付けられた複数の発現プロファイルのプロファイルと複数の合成マルチプレット発現プロファイルの少なくとも 1 つの合成マルチプレット発現プロファイルとの間の第 2 の距離に基づいてシングレット又はマルチプレットとして同定する工程を含む。

20

【 0 2 2 4 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b) で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のプロファイルをプロファイルの第 1 のクラスターにクラスター化する工程と ; (2) 複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの第 2 のクラスターにクラスター化する工程と ; (3) 細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルとプロファイルの第 1 のクラスターとの間の第 1 の距離、及び細胞に関連付けられた複数のプロファイルのプロファイルと合成マルチプレットプロファイルの第 2 のクラスターとの間の第 2 の距離に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

30

【 0 2 2 5 】

いくつかの実施形態では、複数の細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、(b) で作成された複数の合成マルチプレットプロファイルのプロファイル及び該プロファイルに基づいてシングレット又はマルチプレットとして同定する工程は、複数のプロファイルをプロファイルの第 1 のクラスターにクラスター化する工程と ; (2) 複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの第 2 の複数のクラスターにクラスター化する工程と ; (3) 細胞の細胞に関連付けられた複数のプロファイルのプロファイルを、細胞に関連付けられた複数のプロファイルのプロファイルとプロファイルの第 1 のクラスターとの間の第 1 の距離、及び細胞に関連付けられた複数のプロファイルのプロファイルと合成マルチプレットプロファイルの第 2 のクラスターとの間の第 2 の距離に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

40

50

との間の第2の距離に基づいてシングレット又はマルチプレットとして同定する工程と、を含む。

【0226】

シーケンシング

いくつかの実施形態では、異なるバーコード付き標的（例えば確率バーコード付き標的）の数を推定する工程は、標識標的、空間標識、分子標識、サンプル標識、細胞標識、又はその任意の産物（例えば、標識アンプリコン、若しくは標識cDNA分子）の配列を決定する工程を含みうる。増幅された標的をシーケンシングに付すことができる。確率バーコード付き標的又はその任意の産物の配列を決定する工程は、サンプル標識の少なくとも一部、空間標識、細胞標識、分子標識、確率バーコード付き標的の少なくとも一部、その相補鎖、逆相補鎖、又はその任意の組合せの配列を決定するために、シーケンシング反応を実施する工程を含みうる。

【0227】

バーコード付き標的又は確率バーコード付き標的（例えば、増幅された核酸、標識核酸、標識核酸のcDNAコピーなど）の配列の決定は、さまざまなシーケンシング方法を用いて実施することができ、そうした方法として、限定するものではないが、ハイブリダイゼーションによるシーケンシング（SBH）、ライゲーションによるシーケンシング（SBL）、定量的インクリメンタル蛍光スクレオチド付加シーケンシング（QIFNAS）、段階的ライゲーション及び切断、蛍光共鳴エネルギー移動（FRET）、分子ビーコン、TaqManリポータプローブ消化、パイロシーケンシング、蛍光in situシーケンシング（FISH）、FISH EQビーズ、ワブル（wobble）シーケンシング、多重シーケンシング、重合コロニー（POLONY）シーケンシング；ナノグリッドローリングサークルシーケンシング（ROLONY）、対立遺伝子特異的オリゴライゲーションアッセイ（例えば、オリゴライゲーション（OLA）、ライゲートした線状プローブ及びローリングサークル増幅（RCA）読み出しを用いた単一テンプレート分子OLA、ライゲートした錠型（padlock）プローブ、又はライゲートした環状錠型プローブ及びローリングサークル増幅（RCA）を用いた単一テンプレート分子OLAなどが挙げられる。

【0228】

いくつかの実施形態では、バーコード付き標的（例えば確率バーコード標的）又はその任意の産物の配列を決定する工程は、ペアエンドシーケンシング、ナノポアシーケンシング、ハイスループットシーケンシング、ショットガンシーケンシング、ダイターミネータシーケンシング、マルチプルプライマーDNAシーケンシング、プライマーウォーキングを含み、サンガー（Sanger）ジデオキシシーケンシング、マクサム・ギルバート（Maxam Gilbert）シーケンシング、パイロシーケンシング、真の単一分子シーケンシング、又はそれらの任意の組合せを含む。或いは、確率バーコード付き標的又はその任意の産物の配列は、電子顕微鏡検査又は化学・感受性電界効果トランジスタ（chemFET）アレイにより決定することができる。

【0229】

Roche 454、 Illumina Solexa、 ABI - SOLiD、 ION Torrent、 Complete Genomics、 Pacific Biosciences、 Helicos、又はPolonator プラットホームといったプラットホームを用いた環状アレイシーケンシングなどのハイスループットシーケンシング方法を使用することができる。いくつかの実施形態では、シーケンシングは、MiSeqシーケンシングを含みうる。いくつかの実施形態では、シーケンシングは、HiSeqシーケンシングを含みうる。

【0230】

標識付き標的（例えば確率標識付き標的）は、生物のゲノムの遺伝子の約0.01%～生物のゲノムの遺伝子の約100%を占める核酸を含みうる。例えば、複数の多量体を含む標的相補領域を用いて、サンプル中の相補配列を含む遺伝子を捕捉することにより、生

10

20

30

40

50

物のゲノムの遺伝子の約 0 . 0 1 % ~ 生物のゲノムの遺伝子の約 1 0 0 % をシーケンシングすることができる。いくつかの実施形態では、確率バーコード付き標的は、生物のトランスクリプトームの転写物の約 0 . 0 1 % ~ 生物のトランスクリプトームの転写物の約 1 0 0 % を占める核酸を含む。例えば、ポリ (T) テールを含む標的相補的領域を用いて、サンプルから m R N A を捕捉することにより、生物のトランスクリプトームの転写物の約 0 . 5 0 1 % ~ 生物のトランスクリプトームの転写物の約 1 0 0 % をシーケンシングすることができる。

【 0 2 3 1 】

複数のバーコード (例えば確率バーコード) の空間標識及び分子標識の配列を決定する工程は、複数の確率バーコードの 0 . 0 0 0 0 1 % 、 0 . 0 0 0 1 % 、 0 . 0 0 1 % 、 0 . 0 1 % 、 0 . 1 % 、 1 % 、 2 % 、 3 % 、 4 % 、 5 % 、 6 % 、 7 % 、 8 % 、 9 % 、 1 0 % 、 2 0 % 、 3 0 % 、 4 0 % 、 5 0 % 、 6 0 % 、 7 0 % 、 8 0 % 、 9 0 % 、 9 9 % 、 1 0 0 % 、 又はこれらの値のいずれか 2 つの間の数若しくは範囲をシーケンシングする工程を含みうる。複数の確率バーコードの標識、例えば、サンプル標識、空間標識、及び分子標識の配列を決定する工程は、複数の確率バーコードの 1 、 1 0 、 2 0 、 3 0 、 4 0 、 5 0 、 6 0 、 7 0 、 8 0 、 9 0 、 1 0 0 、 1 0 3 、 1 0 4 、 1 0 5 、 1 0 6 、 1 0 7 、 1 0 8 、 1 0 9 、 1 0 1 0 、 1 0 1 1 、 1 0 1 2 、 1 0 1 3 、 1 0 1 4 、 1 0 1 5 、 1 0 1 6 、 1 0 1 7 、 1 0 1 8 、 1 0 1 9 、 1 0 2 0 、 又はこれらの値のいずれか 2 つの間の数若しくは範囲をシーケンシングする工程を含みうる。複数の確率バーコードの一部又は全部をシーケンシングする工程は、約 1 0 、 少なくとも、又は多くとも、 1 0 、 2 0 、 3 0 、 4 0 、 5 0 、 6 0 、 7 0 、 8 0 、 9 0 、 1 0 0 、 2 0 0 、 3 0 0 、 4 0 0 、 5 0 0 、 6 0 0 、 7 0 0 、 8 0 0 、 9 0 0 、 1 0 0 0 、 2 0 0 0 、 3 0 0 0 、 4 0 0 0 、 5 0 0 0 、 6 0 0 0 、 7 0 0 0 、 8 0 0 0 、 9 0 0 0 、 1 0 0 0 0 、 又はこれらの値のいずれか 2 つの間の数若しくは範囲のヌクレオチド又は塩基のリード長の配列を生成する工程を含みうる。

【 0 2 3 2 】

シーケンシング工程は、バーコード付き標的 (例えば確率バーコード付き標的) の少なくとも又は少なくとも約 1 0 、 2 0 、 3 0 、 4 0 、 5 0 、 6 0 、 7 0 、 8 0 、 9 0 、 1 0 0 以上 のヌクレオチド又は塩基対をシーケンシングする工程を含みうる。例えば、シーケンシング工程は、複数の確率バーコード付き標的に対するポリメラーゼ鎖反応 (P C R) 増幅を実施することにより、 5 0 、 7 5 、 若しくは 1 0 0 以上 のヌクレオチドのリード長を有するシーケンシングデータを生成する工程を含みうる。シーケンシング工程は、確率バーコード付き標的の少なくとも又は少なくとも約 2 0 0 、 3 0 0 、 4 0 0 、 5 0 0 、 6 0 0 、 7 0 0 、 8 0 0 、 9 0 0 、 1 , 0 0 0 以上 のヌクレオチド又は塩基対をシーケンシングする工程を含みうる。シーケンシング工程は、確率バーコード付き標的の少なくとも又は少なくとも約 1 , 5 0 0 、 2 0 0 0 、 3 0 0 0 、 4 0 0 0 、 5 0 0 0 、 6 0 0 0 、 7 0 0 0 、 8 0 0 0 、 9 0 0 0 、 若しくは 1 0 0 0 0 以上 のヌクレオチド又は塩基対をシーケンシングする工程を含みうる。

【 0 2 3 3 】

シーケンシング工程は、ラン当たり少なくとも約 2 0 0 、 3 0 0 、 4 0 0 、 5 0 0 、 6 0 0 、 7 0 0 、 8 0 0 、 9 0 0 、 1 , 0 0 0 以上 のシーケンシングリードを含みうる。いくつかの実施形態では、シーケンシング工程は、ラン当たり少なくとも又は少なくとも約 1 , 5 0 0 、 2 0 0 0 、 3 0 0 0 、 4 0 0 0 、 5 0 0 0 、 6 0 0 0 、 7 0 0 0 、 8 0 0 0 、 9 0 0 0 、 若しくは 1 0 0 0 0 以上 のシーケンシングリードを含みうる。シーケンシング工程は、ラン当たり約 1 , 6 0 0 , 0 0 0 , 0 0 0 以下 のシーケンシングリードを含みうる。シーケンシング工程は、ラン当たり約 2 0 0 , 0 0 0 , 0 0 0 以下 のリードを含みうる。

【 0 2 3 4 】

サンプル

いくつかの実施形態では、 1 つ以上のサンプル中に複数の標識が含有されうる。 1 サンプルは、 1 つ以上の細胞、又は 1 つ以上の細胞由来の核酸を含みうる。 1 サンプルは、单

10

20

30

40

50

一細胞、又は1細胞由来の核酸であってよい。1つ以上の細胞は、1つ以上の細胞型であってよい。1つ以上の細胞型の少なくとも1つは、脳細胞、心臓細胞、癌細胞、循環腫瘍細胞、器官細胞、上皮細胞、転移性細胞、良性細胞、一次細胞、循環細胞、又はそれらの任意の組合せである。

【0235】

本開示の方法に使用するためのサンプルは、1つ以上の細胞を含みうる。サンプルは、1つ以上の細胞を意味する。いくつかの実施形態では、複数の細胞は、1つ以上の細胞を含みうる。1つ以上の細胞型の少なくとも1つは、脳細胞、心臓細胞、癌細胞、循環腫瘍細胞、器官細胞、上皮細胞、転移性細胞、良性細胞、一次細胞、循環細胞、又はそれらの任意の組合せであってよい。いくつかの実施形態では、細胞は、癌組織、例えば、乳癌、肺癌、結腸癌、前立腺癌、卵巣癌、膀胱癌、脳癌、黒色腫及び非黒色腫皮膚癌などから切除された癌細胞である。いくつかの場合には、細胞は、癌に由来するが体液から採取される（例えば循環腫瘍細胞）。癌の非限定的な例としては、腺腫、腺癌、扁平上皮細胞癌、基底細胞癌、小細胞癌、大細胞未分化癌、軟骨肉腫、及び線維肉腫が挙げられる。サンプルは、組織、細胞単層、固定細胞、組織片、又はそれらの任意の組合せを含みうる。サンプルは、生体サンプル、臨床サンプル、環境サンプル、生体体液、組織、又は被検者からの細胞を含みうる。サンプルは、ヒト、哺乳動物、イヌ、ラット、マウス、魚類、ハエ、蠕虫、植物、真菌、細菌、ウイルス、脊椎動物、又は非脊椎動物から取得することができる。

10

【0236】

いくつかの実施形態では、細胞は、ウイルスに感染していてウイルスオリゴヌクレオチドを含有する細胞である。いくつかの実施形態では、ウイルス感染は、一本鎖（+鎖又は「センス」）DNAウイルス（例えば、パルボウイルス）、又は二本鎖RNAウイルス（例えば、レトロウイルス）などのウイルスにより引き起こされうる。いくつかの実施形態では、細胞は、細菌である。これらは、グラム陽性又はグラム陰性菌のいずれかを含みうる。いくつかの実施形態では、細胞は、真菌である。いくつかの実施形態では、細胞は、原生動物又はその他の寄生体である。

20

【0237】

本明細書で使用されるとき、「細胞」という用語は、1つ以上の細胞を意味しうる。いくつかの実施形態では、細胞は、正常細胞、例えば、さまざまな発生段階のヒト細胞、又はさまざまな器官若しくは組織型に由来するヒト細胞である。いくつかの実施形態では、非ヒト細胞、例えば、他のタイプの哺乳動物細胞（例えば、マウス、ラット、ブタ、イヌ、ウシ、又はウマ）である。いくつかの実施形態では、細胞は、他のタイプの動物又は植物細胞である。他の実施形態では、細胞は、任意の原核細胞又は真核細胞でありうる。

30

【0238】

本明細書で使用されるとき、細胞は、細胞をビーズに関連付ける前にソートされる。例えば、細胞は、蛍光活性化細胞ソーティング又は磁気活性化細胞ソーティング、又はより一般的にはフローサイトメトリーによりソートすることができる。細胞はサイズ別に濾過することができる。いくつかの実施形態では、リテンテートは、ビーズに関連付けられる細胞を含有する。いくつかの実施形態では、フロースルーは、ビーズに関連付けられる細胞を含有する。

40

【0239】

サンプルは、複数の細胞を意味しうる。サンプルは、細胞の単層を意味しうる。サンプルは、薄い切片（例えば、組織薄片）を意味しうる。サンプルは、一次元のアレイに配置することができる細胞の固体又は半固体コレクションを意味しうる。

【0240】

データ解析及び表示ソフトウェア

データ解析及び標的の空間分解能の可視化

本開示は、バーコーディング（例えば確率バーコーディング）及び空間標識を使ってデジタルカウンティングを用いて標的の数及び位置を推定する方法を提供する。本開示の方法から得られるデータはマップ上に可視化可能である。サンプルの標的の数及び位置の

50

マップは、本明細書に記載の方法を用いて生成された情報を用いて構築可能である。マップは、標的の物理的位置を決定するために使用可能である。マップは、複数の標的の位置を同定するために使用可能である。複数の標的は標的の同一種でありうるか、又は複数の標的は複数の異なる標的でありうる。例えば、脳のマップを構築して複数の標的のディジタルカウント及び位置を示すことが可能である。

【0241】

マップは、単一のサンプルのデータから生成可能である。マップは、複数のサンプルのデータを用いて構築可能であり、それにより組合せマップを生成可能である。マップは、何十、何百、及びノ又は何千ものサンプルのデータで構築可能である。複数のサンプルから構成されるマップは、複数のサンプルに共通する領域に関連付けられる標的のディジタルカウントの分布を示すことが可能である。例えば、レプリケートアッセイは同一のマップ上に表示可能である。少なくとも1、2、3、4、5、6、7、8、9、若しくは10レプリケート又はそれ以上を同一のマップ上に表示（例えばオーバーレイ）しうる。多くとも1、2、3、4、5、6、7、8、9、若しくは10レプリケート又はそれ以上を同一のマップ上に表示（例えばオーバーレイ）しうる。標的の空間分布及び数は、さまざまな統計量により表すことが可能である。

10

【0242】

複数のサンプルからのデータを合わせることにより、合わせたマップの位置的解像度を高めることができる。複数のサンプルの配向は、共通のランドマークにより登録することができ、そこで、サンプル全域に及ぶ個別の位置測定値は、少なくとも部分的に非連続的である。具体的な例は、ミクロトームを用いて、1つの軸上でサンプルを切断してから、別の軸に沿って第2のサンプルを切断するものである。合わせたデータベースは、標的のディジタルカウントを伴う三次元の空間位置を付与するであろう。前述のアプローチを多重化することにより、ディジタルカウント統計学の高解像度三次元マップが可能になるであろう。

20

【0243】

機器システムのいくつかの実施形態では、システムは、単一細胞バーコーディングアッセイ（例えば確率バーコーディングアッセイ）を行うことにより生成されたシーケンスデータセットのデータ解析を提供するためのコードを含むコンピュータ可読媒体を含むであろう。データ解析ソフトウェアにより提供しうるデータ解析機能の例としては、限定されるものではないが、（i）アッセイの実施時に生成された確率バーコードライブライアリをシーケンスすることにより提供されるサンプル標識、細胞標識、空間標識、分子標識、及び標的シーケンスデータのデコーディング/デマルチプレクシングのためのアルゴリズム、（ii）リード数/遺伝子/細胞及びユニーク転写物分子数/遺伝子/細胞を決定するためのアルゴリズム、（iii）例えば、遺伝子発現データにより細胞をクラスター化するための又は転写物分子数/遺伝子/細胞などの決定の信頼区間を予測するためのシーケンスデータの統計解析、（iv）例えば、主成分分析、階層的クラスタリング、k平均値クラスタリング、自己組織化マップ、神経回路網などを用いて、希少細胞のサブ集団を同定するためのアルゴリズム、（v）遺伝子配列データを既知の参照配列にアライメントするための及び突然変異、多型体マーカー、及びスプライス変異体を検出するための配列アライメント機能、並びに（vi）増幅又はシーケンシングエラーを補償するための分子標識の自動クラスタリングが挙げられる。いくつかの実施形態では、データ解析の全部又は一部を行うために市販のソフトウェアを使用しうる。例えば、全細胞コレクションで各細胞に存在する1遺伝子以上のコピー数の表を編集するためにSeven Bridges（<https://www.sbggenomics.com/>）ソフトウェアを使用しうる。いくつかの実施形態では、データ解析ソフトウェアは、有用なグラフ形式のシーケンシング結果、例えば、細胞集団の各細胞に存在する1遺伝子以上のコピー数を示すヒートマップを出力するためのオプションを含みうる。いくつかの実施形態では、データ解析ソフトウェアは、例えば、細胞集団の各細胞に存在する1遺伝子以上のコピー数と、あるタイプの細胞、あるタイプの希少細胞型、又は特異的疾患若しくは病態を有する被験体に由

30

40

50

来る細胞と、を相關付けることにより、シーケンシング結果から生物学的意味を抽出するためのアルゴリズムをさらに含みうる。ある実施形態では、データ解析ソフトウェアは、異なる生物学的サンプル全体にわたり細胞集団を比較するためのアルゴリズムをさらに含みうる。

【0244】

いくつかの実施形態では、データ分析の機能性のすべては、単一のソフトウェアパッケージ内に含めることができる。いくつかの実施形態では、データ解析能力の完全セットは、ソフトウェアパッケージ式を含みうる。いくつかの実施形態では、データ分析ソフトウェアは、ユーザーがアッセイ装置システムに依存せずに利用可能となるスタンドアロンパッケージでありうる。いくつかの実施形態では、ソフトウェアは、ウェブベースでありえ、ユーザーによるデータの共有を可能にし得る。

10

【0245】

いくつかの実施形態では、データ解析機能性のすべてを単一のソフトウェアパッケージ内にパッケージすることができる。いくつかの実施形態では、データ解析能力の完全セットは、一式のソフトウェアパッケージを含みうる。いくつかの実施形態では、データ解析ソフトウェアは、アッセイ機器システムとは独立に、ユーザーが利用可能なスタンドアロンパッケージであってよい。いくつかの実施形態では、ソフトウェアは、ウェブベースでありうるとともに、ユーザーによるデータの共有が可能になりうる。

【0246】

システムプロセッサー及びネットワーク

20

一般的には、本開示の機器システム方法にての使用に適したコンピュータ又はプロセッサーは、図6に示すように、固定媒体612を有するサーバー609に任意選択的に接続可能な媒体611又はネットワークポート605から命令を読み取り可能な論理装置としてさらに理解しうる。システム600は、図6に示すように、CPU601、ディスクドライブ603、キーボード615やマウス616などのオプションの入力デバイス、及びオプションのモニター607を含みうる。データ通信は、ローカル位置又はリモート位置のサーバーに対して指定の通信媒体を介して達成可能である。通信媒体は、データを送受信する任意の手段を含みうる。例えば、通信媒体は、ネットワーク接続、無線接続、又はインターネット接続でありうる。かかる接続は、World Wide Webによる通信を提供可能である。本開示に関するデータは、図6に示すように、かかるネットワーク又は接続を介してあるパーティー622による受信又は閲覧のために伝送可能である。

30

【0247】

図7が示すコンピュータシステム700の第1のアーキテクチャー例の例示的な実施形態は、本開示の実施形態例との関連で使用可能である。図7が示すように、コンピュータシステム例は、処理命令用のプロセッサー702を含みうる。プロセッサーの例としては、限定されるものではないが、Intel Xeon(商標)プロセッサー、AMD Opteron(商標)プロセッサー、Samsung 32ビットRISC ARM 1176JZ(F)-S v1.0(商標)プロセッサー、ARM Cortex-A8 Samsung S5PC100(商標)プロセッサー、ARM Cortex-A8 Apple A4(商標)プロセッサー、Marvell PXA 930(商標)プロセッサー、又は機能的に等価なプロセッサーが挙げられる。実行のマルチスレッドは並列処理に使用可能である。いくつかの実施形態では、クラスター接続の單一コンピュータシステムであるか、又は複数のコンピュータ、携帯電話、若しくは個人用携帯情報端末デバイスを含むネットワーク接続の分散システムであるかにかかわらず、複数のプロセッサー又は複数のコアを備えたプロセッサーも使用可能である。

40

【0248】

図7が示すように、高速キャッシュ704は、プロセッサー702が最近使用した又は頻繁に使用する命令又はデータに対する高速メモリーを提供するために、プロセッサー702に接続又は導入することが可能である。プロセッサー702は、プロセッサーバス708によりノースブリッジ706に接続可能である。ノースブリッジ706は、メモリー

50

バス 712 によりランダムアクセスメモリー (RAM) 710 に接続され、プロセッサー 702 により RAM 710 へのアクセスを管理する。ノースブリッジ 706 はまた、チップセットバス 716 によりサウスブリッジ 714 に接続可能である。サウスブリッジ 714 は、ひいては、周辺機器用バス 718 に接続される。周辺機器用バスは、例えば、PCI I、PCI-X、PCI Express、又は他の周辺機器用バスでありうる。ノースブリッジ及びサウスブリッジはプロセッサーチップセットと呼ばれることが多く、プロセッサーと RAM と周辺機器用バス 718 上の周辺機器要素との間のデータ転送を管理する。いくつかの代替アーキテクチャーでは、ノースブリッジの機能、個別のノースブリッジチップを使用する代わりにプロセッサー中に組込み可能である。

【0249】

10

いくつかの実施形態では、システム 700 は、周辺機器用バス 718 に結合されたアクセラレータカード 722 を含みうる。アクセラレーターは、ある特定の処理を加速するためにフィールドプログラマブルゲートアレイ (FPGA) 又は他のハードウェアを含みうる。例えば、アクセラレーターは、アダプティブデータリストラクチャリングのために、又は拡張セット処理で使用される代数式を評価するために、使用可能である。

【0250】

ソフトウェア及びデータは、外部記憶装置 724 に記憶され、プロセッサーによる使用のために RAM 710 又はキャッシュ 704 にロード可能である。システム 700 は、管理システムリソース用のオペレーティングシステムを含む。オペレーティングシステムの例は、限定されるものではないが、Linux、Windows (商標)、Mac OS (商標)、BlackBerry OS (商標)、iOS (商標)、及び他の機能的に等価なオペレーティングシステム、さらには本発明の実施形態例に従ってデータ記憶及び最適化を管理するためのオペレーティングシステムの上で動作するアプリケーションソフトを含む。

20

【0251】

この例では、システム 700 はまた、ネットワークインターフェースカード (NIC) 720 及び 721 を含み、ネットワーク接続記憶装置 (NAS) などの外部記憶装置及び分散並列処理に使用可能な他のコンピュータシステムへのネットワークインターフェースを提供する周辺機器用バスに接続される。

【0252】

30

図 8 は、本開示の方法での使用に好適な、複数のコンピュータシステム 802a、及び 802b、複数の携帯電話及び個人用携帯情報端末 802c、並びにネットワーク接続記憶装置 (NAS) 804a、及び 804b を含むネットワーク 800 の例示的な図を示す。実施形態例では、システム 812a、812b、及び 812c は、データ記憶を管理し、ネットワーク接続記憶装置 (NAS) に記憶されたデータに対するデータアクセスを最適化することができる。データに数学モデルを使用することができ、分散並列処理コンピュータシステム 812a、及び 812b、並びに携帯電話及び個人用携帯情報端末システム 812c を用いて評価することができる。コンピュータシステム 812a、及び 812b、並びに携帯電話及び個人用携帯情報端末システム 812c はまた、ネットワーク接続記憶装置 (NAS) 814a 及び 814b に記憶されたデータのアダプティブデータリストラクチャリングのために並列処理を提供可能である。図 8 は、一例を示すに過ぎず、多種多様な他のコンピュータアーキテクチャー及びシステムが、本発明の種々の実施形態に関連して使用することができる。例えば、ブレードサーバーを用いて、並列処理を提供することができる。プロセッサーブレードは、並列処理を提供するためにバックプレーンを介して接続可能である。記憶装置はまた、バックプレーンに接続してもよいし、又は個別ネットワークインターフェースを介してネットワーク接続記憶装置 (NAS) として存在してもよい。

40

【0253】

いくつかの実施形態例では、プロセッサーは、個別メモリー空間を保持可能であるとともに、ネットワークインターフェースを介してバックプレーンに又は他のプロセッサーに

50

による並列処理のために他のコネクターにデータを伝送可能である。他の実施形態では、プロセッサーの一部又は全部は、共有仮想アドレスメモリー空間を使用可能である。

【0254】

図9に示すマルチプロセッサー・コンピュータ・システム900の例示的なブロック図は、実施形態例に従って共有仮想アドレスメモリー空間を使用する。システムは、共有メモリーサブシステム904にアクセス可能な複数のプロセッサー902a-fを含む。システムは、メモリーサブシステム904中で複数のプログラマブルハードウェアメモリーアルゴリズムプロセッサー(MAP)906a-fを組込む。各MAP906a-fは、メモリー908a-fと1つ以上のフィールドプログラマブルゲートアレイ(FPGA)910a-fとを含みうる。MAPは、設定可能な機能ユニットを提供し、特定のアルゴリズム又はアルゴリズムの一部は、それぞれのプロセッサーと緊密に連携して処理するためにFPGA910a-fに提供可能である。例えば、MAPは、データモデルに関する代数式を評価するために及び実施形態例でアダプティブデータリストラクチャリングを行うために使用可能である。この例では、各MAPは、こうした目的のためにすべてのプロセッサーによりグローバルにアクセス可能である。一構成では、各MAPは、関連付けられたメモリー908a-fにアクセスするためにダイレクトメモリーアクセス(DMA)を使用可能であり、それにより、それぞれのマイクロプロセッサー902a-fに依存せずに且つ非同期的に課題を実行可能になる。この構成では、MAPは、アルゴリズムのパイプライン実行及び並行実行のために他のMAPに結果を直接供給可能である。

10

【0255】

以上のコンピュータアーキテクチャー及びシステムは、単なる例にすぎず、一般的プロセッサー、共プロセッサー、FPGA、及び他のプログラマブルロジックデバイス、システムオンチップ(SOC)、特定用途向け集積回路(ASIC)、及び他の処理素子及び論理素子の任意の組合せを使用するシステムを含めて、多種多様な他のコンピュータ、携帯電話、及び個人用携帯情報端末のアーキテクチャー及びシステムを実施形態例との関連で使用可能である。いくつかの実施形態では、コンピュータシステムの全部又は一部は、ソフトウェア又はハードウェアで実現可能である。任意のさまざまなデータ記憶媒体は、ランダムアクセスメモリー、ハードドライブ、フラッシュメモリー、テープドライブ、ディスクアレイ、ネットワーク接続記憶装置(NAS)、並びに他のローカル又は分散データ記憶デバイス及びシステムを含めて、実施形態例との関連で、使用可能である。

20

【0256】

実施形態例では、本開示のコンピュータサブシステムは、以上の又は他のコンピュータアーキテクチャー及びシステムのいずれかで実行されるソフトウェアモジュールを用いて実現可能である。他の実施形態では、システムの機能は、ファームウェア、プログラマブルロジックデバイス、例えば、フィールドプログラマブルゲートアレイ(FPGA)、システムオンチップ(SOL)、特定用途向け集積回路(ASIC)、又は他の処理素子及び論理素子で、部分的に又は完全に実現可能である。例えば、セットプロセッサー及びオプティマイザーは、アクセラレーターカードなどのハードウェアアクセラレーターカードを用いてハードウェアアクセラレーションで実現可能である。

30

【実施例】

【0257】

以上論述した実施形態のいくつかの態様を以下の実施例でさらに詳しく開示するが、これらの実施例は、本開示の範囲を何ら制限することを意図しない。

40

【0258】

実施例1

1:1の混合細胞型の合成ダブレット発現プロファイル

本実施例は、サンプル中の1:1の混合細胞型の発現プロファイルを、合成ダブルト発現プロファイルを用いてダブルト発現プロファイルとして決定することについて記載する。

【0259】

50

単一の Jurkat 細胞及びRamos 細胞の発現プロファイルを決定した。サンプルは、ほぼ同数の Jurkat 細胞及びRamos 細胞を含有した（1：1 の混合細胞型とも称する）。発現プロファイルは、発現プロファイル空間内の 3 つのクラスター： Jurkat 細胞の発現プロファイルに対応するクラスター、Ramos 細胞の発現プロファイルに対応するクラスター、及び混合細胞型のダブレット発現プロファイルに対応するクラスターに含まれた。各ダブレット発現プロファイルは、単細胞の発現プロファイルを決定する中で 2 個の細胞が 1 個の細胞として同定されていることに起因していることがある。これら 3 つのクラスターの可視化について図 10 A に示す。図 10 A は、ほぼ同数の Jurkat 細胞及びRamos 細胞を含むサンプル中の単細胞の発現プロファイルの非限定的な例示的 t S N E 投影プロットを示す。図 10 A 中の t S N E 投影プロットは、 Jurkat 細胞の発現プロファイルに対応するクラスター、Ramos 細胞の発現プロファイルに対応するクラスター、及び混合細胞型のダブレット発現プロファイルに対応するクラスターを含む。

【0260】

2 個の細胞の観察された発現プロファイルを組合せることにより、合成ダブレット発現プロファイルを作成した。細胞のペアを複数回、無作為にサンプリングすることにより、複数の合成ダブレット発現プロファイルを作成した。複数の合成ダブレット発現プロファイルの可視化について図 10 B に示す。図 10 B は、図 10 A における発現プロファイルの非限定的な例示的 t S N E 投影プロット及び合成ダブレット発現プロファイルが 2 % であることを示す。いくつかの合成ダブレット発現プロファイルが、混合細胞型（ Jurkat 細胞及びRamos 細胞）を表した。図 10 B に示すように、これらの合成ダブレット発現プロファイルに対応するクラスターは、図 10 A におけるダブレット発現プロファイルに対応するクラスターと重複する。いくつかの合成ダブレット発現プロファイルは、単細胞型（ Jurkat 細胞又はRamos 細胞）を表した。 Jurkat 細胞型の 2 個の細胞の発現プロファイルから作成した合成ダブレット発現プロファイルは、図 10 における Jurkat 細胞の発現プロファイルの様相を呈した。同様に、Ramos 細胞型の 2 個の細胞の発現プロファイルから作成した合成ダブレット発現プロファイルは、図 10 におけるRamos 細胞の発現プロファイルの様相を呈した。

【0261】

まとめると、これらのデータは、1：1 の混合細胞型の細胞のダブレット発現プロファイルが、発現プロファイル空間内でのそれらの近接性が理由で合成ダブレット発現プロファイルを用いて同定されたことを示す。

【0262】

実施例 2

1：1：1 の混合細胞型の合成ダブレット発現プロファイル

本実施例は、サンプル中の 1：1：1 の混合細胞型の発現プロファイルを、合成ダブレット発現プロファイルを用いてダブレット発現プロファイルとして決定することについて記載する。

【0263】

単一の Jurkat 細胞及びRamos 細胞の発現プロファイルを決定した。サンプルは、ほぼ同数の Jurkat 細胞、K562 細胞、及びRamos 細胞を含有した（1：1 の混合細胞型とも称する）。発現プロファイルは、発現プロファイル空間内の 6 つのクラスター： Jurkat 細胞の発現プロファイルに対応するクラスター、K562 細胞の発現プロファイルに対応するクラスター、Ramos 細胞の発現プロファイルに対応するクラスター、及び混合細胞型のダブレット発現プロファイルに対応する 3 つのクラスターに含まれた。3 つのクラスターは、 Jurkat 細胞と K562 細胞、 Jurkat 細胞とRamos 細胞、及び K562 細胞とRamos 細胞のダブレット発現プロファイルに対応する。これら 3 つのクラスターの可視化について図 11 に示す。図 11 は、ほぼ同数の Jurkat 細胞、K562 細胞及びRamos 細胞を含むサンプル中の単細胞の発現プロファイルの非限定的な例示的 t S N E 投影プロットを示す。

10

20

30

40

50

【0264】

2個の細胞の観察された発現プロファイルを合計することにより、合成ダブレット発現プロファイルを作成した。細胞のペアを複数回、無作為にサンプリングすることにより、複数の合成ダブレット発現プロファイルを作成した。合成ダブレット発現プロファイルには、3つのタイプ：単一のJ u r k a t 細胞と単一のK 5 6 2 細胞の発現プロファイルから作成した発現プロファイル、単一のJ u r k a t 細胞と単一のR a m o s 細胞の発現プロファイルから作成した発現プロファイル、及び単一のK 5 6 2 細胞と単一のR a m o s 細胞の発現プロファイルから作成した発現プロファイルが存在した。複数の合成ダブレット発現プロファイルの可視化についても図11に示す。図11において、3つのタイプの合成ダブレット発現プロファイルに対応する3つのクラスターは、観察された混合細胞型のダブレット発現プロファイルに対応するクラスターと重複する。

10

【0265】

まとめると、これらのデータは、1：1：1の混合細胞型の細胞のダブレット発現プロファイルが、発現プロファイル空間内でのそれらの近接性が理由で合成ダブレット発現プロファイルを用いて同定されたことを示す。

【0266】

実施例3

ヒトP B M Cサンプル中の細胞の合成ダブレット発現プロファイル

本実施例は、ヒト末梢血単核球（P B M C）サンプル中の細胞の発現プロファイルを、合成ダブレット発現プロファイルを用いてダブルレット発現プロファイルとして決定することについて記載する。

20

【0267】

ヒトP B M Cサンプル中の単細胞の発現プロファイルを決定した。発現プロファイルは、異なるタイプのダブルレット発現プロファイルに対応するクラスターを含む、発現プロファイル空間内の複数のクラスターに含まれた。ダブルレット発現プロファイルのあるタイプに対応するクラスター及び合成ダブルレット発現プロファイルに対応するクラスターは、発現プロファイル空間内で近接していた。低次元t S N E 投影空間内でのこれらのクラスターの可視化について図12に示す。図12は、ヒトP B M Cサンプルから取得された発現プロファイルの非限定的な例示的t S N E 投影プロットを示す。特定のサンプル、例えばヒトP B M Cサンプル中の細胞の多様性が理由で、ダブルレット発現プロファイルのあるタイプに対応するクラスターと合成ダブルレット発現プロファイルに対応するクラスターとの近接性は、図12中の発現プロファイルのt S N E 投影プロットにおいて容易に可視化されない場合がある。クラスター間のさらなる近接性は、異なる低次元投影空間内で可視化されてもよい。

30

【0268】

まとめると、これらのデータは、多様な細胞のダブルレット発現プロファイルが、高次元発現プロファイル空間内又は適切な低次元投影空間内で合成ダブルレット発現プロファイルを用いて同定されてもよいことを示す。

【0269】

実施例4

マウス細胞の12のサンプルの合成ダブルレット発現プロファイル

本実施例は、12のマウス細胞サンプルからの細胞の発現プロファイルを、合成ダブルレット発現プロファイルを用いてシングレット発現プロファイル及びマルチプレット発現プロファイルとして決定することについて記載する。

40

【0270】

マウス2匹から6つの組織（骨髄、脂肪（性腺白色脂肪組織（g W A T））、結腸、肝臓、肺、及び脾臓）を単離した。蛍光標示式細胞分取器（F A C S）を用いて、単離した組織からC D 4 5 + 単細胞を単離し、選別し、12のサンプルを作成した。各マウスからの6つの組織のC D 4 5 + 単細胞を、R N Aシーケンシング用のB D（商標）S i n g l e - C e l l M u l t i p l e x i n g K i tを用いて6つの異なる核酸サンプル標識

50

で標識し、Rhapsody（商標）カートリッジに負荷した。Rhapsody（商標）磁気ビーズを用いて、細胞からのmRNA分子を捕捉し、Rhapsody（商標）Immune Response Panel - Mouse（マウス）を用いて、バーコード化し、増幅した。細胞の発現プロファイルを、合成マルチプレット発現プロファイルを用いてシングレット発現プロファイル及びマルチプレット発現プロファイルとして決定及び同定した。図13は、マウス2匹からの6つの組織の12のサンプルからのCD45+単細胞の発現プロファイルの非限定的な例示的t-SNE投影プロットである。図13は、マルチプレット発現プロファイルを複数のクラスターとして示す。合成マルチプレット発現プロファイルを用いて、マルチプレット発現プロファイルを同定し、ここで各々は、2つ以上の細胞型又はサブタイプの細胞の発現プロファイルを含んだ。2つの異なる組織からのCD45+細胞のマルチプレットを、細胞から取得された配列データにおけるサンプル標識配列の存在に基づいて同定した。合成マルチプレット発現プロファイルを用いてマルチプレットを同定する性能は、核酸サンプル標識を用いてサンプルをインデックス化し、マルチプレットを同定する性能と同等であった。さらに、追加的な発現プロファイルを、合成マルチプレット発現プロファイルを用いてマルチプレットとして同定した。サンプル標識でなく、合成マルチプレット発現プロファイルを用いてマルチプレットとして同定した発現プロファイルは、本実施例中の1つのサンプル標識配列を用いて標識できた、1つの組織からの2つ以上の細胞型又はサブタイプの2つ以上の細胞の発現プロファイルを含んでもよい。

【0271】

まとめると、これらのデータは、異なる組織からの細胞、又は異なる細胞型若しくはサブタイプのマルチプレット発現プロファイルが、合成マルチプレット発現プロファイルを用いて同定可能であることを示す。

【0272】

実施例5

合成ダブレット発現プロファイルを作成、可視化、及び除去する

本実施例は、合成ダブルット発現プロファイルを手作業で作成、可視化、及び除去するワークフローを説明する。

【0273】

以下のステップを含む、合成ダブルット発現プロファイルを作成、可視化、及び除去するためのワークフローを実施した。ワークフローを図14に関連して例示し、それは、例えばBD（商標）（Franklin Lake, NJ）Data Viewの非限定的な例示的ユーザーインターフェース1404を用いての合成ダブルット発現プロファイルを作成、可視化、及び除去する非限定的な例示的ワークフローを示す。

ステップ1. 発現プロファイルを有するデータファイルをロードする

ステップ2. 合成ダブルット発現プロファイルを有する合成ダブルットファイルを、例えば「合成ダブルット」ボタン1408を選択することにより作成する。合成ダブルット発現プロファイルを作成するため、ロードされたデータファイル中の発現プロファイルのペアを無作為にサンプリングし、各ペアにおける2つの発現プロファイルを合わせる。合成ダブルット発現プロファイルを合成ダブルットファイルとして保存する

ステップ3. クリアし、データファイル及び合成ダブルットファイルをリロードする

ステップ4. 2つの異なる色として、又はヒートマップで色付けして発現プロファイル及び合成ダブルット発現プロファイルを示す、1つ以上の投影プロット1412、1416を生成する

ステップ5. (データファイルから又は合成ダブルットファイルからの) 発現プロファイルの起源に基づいてプロットをアノテートする

ステップ6. 「アノテーションを追加する」及び「ドロー」オプションを用いて、多数の合成ダブルット発現プロファイルを有するクラスターの弧を描く。これらをダブルットとしてアノテートする

ステップ7. 「アノテーションを追加する」及び「既存のアノテーションを組合せる」

10

20

30

40

50

オプションを用いて、発現プロファイル及び「ダブレット」アノテーションの起源に基づく4つのグループを伴う新しいアノテーションを作成する。4つのグループは、「ダブレット」としてアノテートされていないデータファイルからの発現プロファイル；「ダブレット」としてアノテートされたデータファイルからの発現プロファイル；「ダブレット」としてアノテートされていない合成ダブレットデータファイルからの発現プロファイル；及び「ダブレット」としてアノテートされた合成ダブレットデータファイルからの発現プロファイルに対応する

ステップ8. この新しいアノテーションをハイライトする。例えば「フィルター発現プロファイル」ボタン1420を選択することにより、データファイルからの発現プロファイルにフィルターをかける。オプションを選択し、ここでハイライトされたアノテーションに基づいて細胞にフィルターをかける。3つのグループを選択し、「ダブレット」としてアノテートされていないデータファイルからの発現プロファイル；「ダブレット」としてアノテートされたデータファイルからの発現プロファイル；及び「ダブレット」としてアノテートされた合成ダブレットデータファイルからの発現プロファイルを除去する

ステップ9. 残りの発現プロファイルを新しいデータファイルとして保存する

【0274】

まとめると、これらのデータは、合成ダブレット発現プロファイルを、1つ以上のtSNE投影プロットを用いて手作業で作成、可視化、及び除去することを示す。

【0275】

以上に記載の実施形態の少なくともいくつかでは、実施形態で使用される1つ以上のエレメントは、他の実施形態で互換的に使用可能である。ただし、かかる交換が技術的に実現可能である場合に限る。特許請求された主題の範囲から逸脱することなく、以上に記載の方法及び構造に種々の他の省略、追加、及び変更を行いうることは、当業者であれば分かるであろう。かかる変更及び変化はすべて、添付の特許請求の範囲に規定される主題の範囲内に含まれることが意図される。

【0276】

本明細書に記載の実質的に任意の複数形及び／又は単数形の用語の使用に関連して、文脈上及び／又は適用上適切であれば、当業者は複数形から単数形へ及び／又は単数形から複数形への変換が可能である。明確にするために種々の単数形／複数形の入替えを本明細書に明示的に記述しうる。本明細書及び添付の特許請求の範囲で用いられる場合、特に文脈上明確に規定されていない限り、単数形の「a」、「a n」、及び「t h e」には、複数の参照語が含まれる。本明細書での「o r (又は)」の意味はいずれも、特に明記されていない限り、「a n d / o r (及び／又は)」を包含することが意図される。

【0277】

一般的には、本明細書特に添付の特許請求の範囲（例えば添付の特許請求の範囲の本文）で用いられる用語は「オープン」用語であることが一般に意図されることは当業者であれば理解されよう（例えば、「i n c l u d i n g (~を含む)」という用語は「~を含むがこれらに限定されるものではない」と解釈すべきであり、「h a v i n g (~を有する)」という用語は「少なくとも~を有する」と解釈すべきであり、「i n c l u d e s (~を含む)」という用語は「~を含むがこれらに限定されるものではない」と解釈すべきであるなど）。さらに、導入クレームレシテーションの特定数が意図される場合、かかる意図は請求項で明示的にリサイトされ、かかるレシテーションの不在下ではかかる意図は存在しないことは当業者であれば理解されよう。例えば、理解の一助として、以下の添付の特許請求の範囲は、クレームレシテーションを導入するために導入語句「a t l e a s t o n e (少なくとも1つ)」及び「o n e o r m o r e (1つ以上)」の使用を含みうる。しかしながら、かかる語句が用いられたとしても、不定冠詞「a」又は「a n」によるクレームレシテーションの導入が、かかる導入クレームレシテーションを含む任意の特定の請求項を、一方のかかるレシテーションを含む実施形態のみに限定することを意味するものと解釈すべきでない。たとえ同一の請求項が導入語句「o n e o r m o r e (1つ以上)」又は「a t l e a s t o n e (少なくとも1つ)」と不定冠詞例え

10

20

30

40

50

ば「a」又は「a n」とを含む場合でさえも、そのように解釈すべきでない（例えば、「a」及び／又は「a n」は「a t l e a s t o n e（少なくとも1つ）」又は「o n e o r m o r e（1つ以上）」を意味するものと解釈すべきである）。定冠詞を用いてクレームレシテーションを導入する場合にも、同じことが当てはまる。そのほかに、たとえ特定数の導入クレームレシテーションが明示的にリサイトされたとしても、かかるレシテーションは少なくともリサイトされた数を意味すると解釈すべきであることは当業者であれば分かるであろう（例えば、「2つのレシテーション」という他の修飾語を含まないベアのレシテーションは、少なくとも2つのレシテーション又は2つ以上レシテーションを意味する）。さらに、「A、B、及びCの少なくとも1つ」に類似した条件が用いられる場合、一般的には、かかる構成は当業者がその条件を理解する意味であることが意図される（例えば、「A、B、及びCの少なくとも1つを有する系」は、限定されるものではないが、A単独、B単独、C単独、AとBの両方、AとCの両方、BとCの両方、及び／又はAとBとCの全部などを有する系を含であろう）。「A、B、又はCの少なくとも1つなど」に類似した条件が用いられる場合、一般的には、かかる構成は当業者がその条件を理解する意味であることが意図される（例えば、「A、B、又はCの少なくとも1つを有する系」は、限定されるものではないが、A単独、B単独、C単独、AとBの両方、AとCの両方、BとCの両方、及び／又はAとBとCの全部などを有する系を含であろう）。さらに、2つ以上の代替用語を表す実質上任意の選言的な語及び／又は語句は、明細書、請求項、又は図面にかかわらず、用語の1つ、用語のいずれか、又は用語の両方を含む可能性が企図されると理解すべきであることは当業者であれば理解されよう。例えば、「A又はB」という語句は「A」又は「B」又は「A及びB」の可能性を含むものと理解されよう。

【0278】

その他に、本開示の特徴又は態様がマーカッシュグループにより記述される場合、それにより、本開示は、マーカッシュグループの任意の個別のメンバー又はメンバーのサブグループにより記述されることは当業者であれば分かるであろう。

【0279】

当業者であれば理解されるであろうが、あらゆる目的で、例えば、明細書の提供に関して、本明細書に開示された範囲はすべて、あらゆる可能なサブ範囲及びそのサブ範囲の組合せをも包含する。いずれの列挙された範囲も、十分に記述されたものとして且つその範囲が少なくとも2等分、3等分、4等分、5等分、10等分などされうるものとして容易に認識可能である。例えば、限定されるものではないが、本明細書で考察した各範囲は、下3分の1、中3分の1、上3分の1に容易に分解可能である。同様に、当業者であれば理解されるであろうが、「～まで」、「少なくとも～」、「～超」、「～未満」などの表現はすべて、リサイトされた数を含み、以上で考察したように後続的にサブ範囲に分解可能な範囲を意味する。最終的に、当業者であれば理解されるであろうが、範囲は各個別のメンバーを含む。従って、例えば、1～3個の物品を有するグループは、1、2、又は3個の物品を有するグループを意味する。同様に、1～5個の物品を有するグループは、1、2、3、4、又は5個の物品を有するグループを意味し、他も同様である。

【0280】

種々の態様及び実施形態を本明細書に開示してきたが、他の態様及び実施形態は当業者には自明であろう。本明細書に開示される種々の態様及び実施形態は、例示を目的としたものであり、限定を意図したものではなく、真の範囲及び趣旨は、以下の特許請求の範囲により示される。

本発明のまた別の態様は、以下のとおりであってもよい。

〔1〕マルチプレット発現プロファイルを同定するための方法であって、

（a）複数の確率バーコードを用いて、複数の細胞における複数の標的に確率バーコードを付けて、前記複数の細胞の各細胞において複数の確率バーコード付き標的を生成する工程であって、前記複数の確率バーコードの各々が細胞標識及び分子標識を含み、ここで前記複数の確率バーコードの少なくとも2つの確率バーコードの分子標識が異なる分子標

10

20

30

40

50

識配列を含み、且つ前記複数の確率バーコードの少なくとも 2 つの確率バーコードが同一の細胞標識配列を有する細胞標識を含む工程と；

(b) 前記複数の確率バーコード付き標的のシーケンシングデータを取得する工程と；

(c) (b) で取得された前記シーケンシングデータから前記複数の確率バーコードの細胞標識に関連付けられた複数の発現プロファイルを決定する工程であって、前記複数の確率バーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルが、前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を含む工程と；

(d) (c) で決定された前記複数の確率バーコードの前記細胞標識に関連付けられた前記複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程と；

(e) 前記複数の確率バーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルを、(d) で作成された複数の合成ダブレット発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む、方法。

[2] 前記発現プロファイルが(e) におけるマルチプレットとして同定される場合、前記発現プロファイルに関連付けられたシーケンシングデータを(b) で取得された前記シーケンシングデータから除去する工程を含む、前記[1] に記載の方法。

[3] 前記発現プロファイルが(e) におけるマルチプレットとして同定される場合、前記発現プロファイルを(c) で決定された前記複数の発現プロファイルから除去する工程を含む、前記[1] ~ [2] のいずれか一項に記載の方法。

[4] 前記複数のマルチプレットがダブルレット、トリプレット、又はそれらの任意の組合せを含む、前記[1] ~ [3] のいずれか一項に記載の方法。

[5] (c) で決定された前記複数の確率バーコードの前記細胞標識に関連付けられた前記複数の発現プロファイルから前記複数の合成マルチプレット発現プロファイルを作成する工程が、

前記複数の合成マルチプレット発現プロファイルの合成マルチプレット発現プロファイルについて、

(1) (c) で決定された前記複数の発現プロファイルの第 1 の発現プロファイルであって、第 1 の細胞標識配列に関連付けられた第 1 の発現プロファイルを選択する工程と；

(2) (c) で決定された前記複数の発現プロファイルの第 2 の発現プロファイルであって、第 2 の細胞標識配列に関連付けられた第 2 の発現プロファイルを選択する工程と；

(3) (1) で選択された前記第 1 の発現プロファイルと(2) で選択された前記第 2 の発現プロファイルとを組合せ、前記合成マルチプレット発現プロファイルを作成する工程と、

を含み、ここで前記第 1 の細胞標識配列及び前記第 2 の細胞標識配列が異なる細胞標識配列を含む、前記[1] ~ [4] のいずれか一項に記載の方法。

[6] (1) で選択された前記第 1 の発現プロファイルと(2) で選択された前記第 2 の発現プロファイルとを組合せ、前記合成マルチプレット発現プロファイルを作成する工程が、

前記複数の標的の各々について、前記第 1 の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の数と前記第 2 の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の数とを組合せ、前記合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の数を作成する工程を含む、前記[5] に記載の方法。

[7] 前記合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の前記数が、前記第 1 の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数と前記第 2 の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数との合計である、前記[6] に記載の方法。

10

20

30

40

50

〔8〕前記合計が、前記第1の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数と前記第2の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数との加重合計である、前記〔7〕に記載の方法。

〔9〕前記合成マルチプレット発現プロファイルに関連付けられた識別可能な配列を有する分子標識の前記数が、前記第1の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数と前記第2の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数との平均である、前記〔6〕に記載の方法。

〔10〕前記平均が、前記第1の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数と前記第2の発現プロファイルにおける前記標的に関連付けられた識別可能な配列を有する分子標識の前記数との加重平均である、前記〔9〕に記載の方法。 10

〔11〕前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(d)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び(c)で決定された前記複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程と； 20

(2) 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、を含む、前記〔1〕～〔10〕のいずれか一項に記載の方法。

〔12〕前記機械学習モデルを訓練する際に用いられる前記複数の発現プロファイルの前記1つ以上の発現プロファイルが、(c)で決定された前記複数の発現プロファイルの百分率を含む、前記〔11〕に記載の方法。

〔13〕前記百分率が約10パーセントである、前記〔12〕に記載の方法。

〔14〕前記機械学習モデルが分類モデルを含む、前記〔11〕～〔13〕のいずれか一項に記載の方法。 30

〔15〕前記分類モデルが、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含む、前記〔14〕に記載の方法。

〔16〕前記機械学習モデルが、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k近傍法(KNN)モデル、k平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含む、前記〔11〕～〔15〕のいずれか一項に記載の方法。

〔17〕前記機械学習モデルが、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スパース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含む、前記〔11〕～〔16〕のいずれか一項に記載の方法。 40

〔18〕発現プロファイルマルチプレットの同定を意図して前記機械学習モデルを、(d)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び(c)で決定された前記複数の発現プロファイルの前記1つ以上の発現プロファイルから訓練する工程が、

(1) (d)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程と；

(2) (c)で決定された前記複数の発現プロファイルの前記1つ以上の発現プロファイルを前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の発現プロファイルを作成する工程と； 50

ロファイルの 1 つ以上の投影された発現プロファイルを作成する工程と；

(3) 発現プロファイルマルチプレットの同定を意図して前記機械学習モデルを、(1)からの前記複数の合成マルチプレット発現プロファイルの前記投影された発現プロファイル及び(2)における前記複数の発現プロファイルの前記 1 つ以上の投影された発現プロファイルから訓練する工程と、
を含む、前記〔 11 〕～〔 17 〕のいずれか一項に記載の方法。

〔 19 〕 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の確率バーコードの前記発現プロファイルを投影して、前記複数の確率バーコードの投影された発現プロファイルを作成する工程を含み、

ここで前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイルに基づいてシングルレット又はマルチプレットとして同定する工程が、

前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイル及び前記複数の確率バーコードの前記投影された発現プロファイルに基づいてシングルレット又はマルチプレットとして同定する工程を含む、
前記〔 18 〕に記載の方法。

〔 20 〕 前記低次元空間が二次元空間である、前記〔 18 〕～〔 19 〕のいずれか一項に記載の方法。

〔 21 〕 (d) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程が、
、 t 分布型確率的近傍埋め込み (t S N E) 法を用いて、(d) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程を含む、前記〔 18 〕～〔 20 〕のいずれか一項に記載の方法。

〔 22 〕 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングルレット又はマルチプレットとして同定する工程が、

前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと前記複数の発現プロファイルの少なくとも 1 つの発現プロファイルとの間の第 1 の距離、及び

前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと前記複数の合成マルチプレット発現プロファイルの少なくとも 1 つの合成マルチプレット発現プロファイルとの間の第 2 の距離

に基づく前記発現プロファイルに基づいてシングルレット又はマルチプレットとして同定する工程

を含む、前記〔 1 〕～〔 21 〕のいずれか一項に記載の方法。

〔 23 〕 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングルレット又はマルチプレットとして同定する工程が、

(1) 前記複数の発現プロファイルを発現プロファイルの第 1 のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの第 2 のクラスターにクラスター化する工程と；

10

20

30

40

50

(3) 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと発現プロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと合成マルチプレット発現プロファイルの第2のクラスターとの間の第2の距離に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔1〕～〔21〕のいずれか一項に記載の方法。

〔24〕 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数の発現プロファイルを発現プロファイルの第1のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの複数の第2のクラスターにクラスター化する工程と；

(3) 前記複数の確率バーコードの前記細胞標識の前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと発現プロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞標識に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと合成マルチプレット発現プロファイルの前記複数の第2のクラスターの1つ以上のクラスターとの間の第2の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔1〕～〔21〕のいずれか一項に記載の方法。

〔25〕 前記複数の確率バーコードを用いて、前記複数の細胞における前記複数の標的に確率バーコードを付けて、前記複数の細胞の各細胞において前記複数の確率バーコード付き標的を生成する工程が、

複数の粒子の前記複数の確率バーコードを用いて、前記複数の細胞における前記複数の標的に確率バーコードを付けて、前記複数の細胞の各細胞において前記複数の確率バーコード付き標的を生成する工程を含み、

ここで前記複数の粒子の各々が前記複数の確率バーコードのサブセットを含み、ここで確率バーコードの前記サブセットの各々が同一の細胞標識配列を含み且つ少なくとも100の異なる分子標識配列を有する、

前記〔1〕～〔24〕のいずれか一項に記載の方法。

〔26〕 前記粒子がビーズである、前記〔25〕に記載の方法。

〔27〕 前記ビーズが、ストレプトアビジンビーズ、アガロースビーズ、磁気ビーズ、コンジュゲートビーズ、プロテインAコンジュゲートビーズ、プロテインGコンジュゲートビーズ、プロテインA/Gコンジュゲートビーズ、プロテインLコンジュゲートビーズ、オリゴTコンジュゲートビーズ、シリカビーズ、シリカ様ビーズ、抗ビオチンマイクロビーズ、抗蛍光色素マイクロビーズ、及びそれらの任意の組合せからなる群から選択される、前記〔26〕に記載の方法。

〔28〕 前記粒子が、ポリジメチルシリコサン(PDMS)、ポリスチレン、ガラス、ポリプロピレン、アガロース、ゼラチン、ヒドロゲル、常磁性体、セラミック、プラスチック、ガラス、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、セルロース、ナイロン、シリコーン、及びそれらの任意の組合せからなる群から選択される材料を含む、前記〔25〕～〔27〕のいずれか一項に記載の方法。

〔29〕 前記複数の粒子の各々の前記確率バーコードが、少なくとも1000の異なる分

10

20

30

40

50

子標識配列を有する分子標識を含む、前記〔25〕～〔28〕のいずれか一項に記載の方法。

〔30〕前記複数の粒子の各々の前記確率バーコードが、少なくとも10000の異なる分子標識配列を有する分子標識を含む、前記〔25〕～〔29〕のいずれか一項に記載の方法。

〔31〕前記確率バーコードの前記分子標識が、ランダム配列を含む、前記〔25〕～〔30〕のいずれか一項に記載の方法。

〔32〕前記複数の粒子の各々が、少なくとも10000の確率バーコードを含む、前記〔25〕～〔31〕のいずれか一項に記載の方法。

〔33〕前記複数の確率バーコードを用いて、前記複数の細胞における前記複数の標的に確率バーコードを付けて、前記複数の細胞の各細胞において前記複数の確率バーコード付き標的を生成する工程が、
10

(1) 前記標的のコピーを前記確率バーコードの標的結合領域と接触させる工程と；

(2) 前記複数の確率バーコードを用いて前記複数の標的を逆転写して、複数の逆転写された標的を生成する工程と、

を含む、前記〔1〕～〔32〕のいずれか一項に記載の方法。

〔34〕前記複数の確率バーコード付き標的の前記シーケンシングデータを取得する前に、前記確率バーコード付き標的を増幅し、複数の増幅された確率バーコード付き標的を作成する工程を含む、前記〔1〕～〔33〕のいずれか一項に記載の方法。

〔35〕前記確率バーコード付き標的を増幅し、前記複数の増幅された確率バーコード付き標的を作成する工程が、前記確率バーコード付き標的をポリメラーゼ連鎖反応（PCR）により増幅する工程を含む、前記〔34〕に記載の方法。
20

〔36〕前記複数の確率バーコードの前記細胞標識に関連付けられた前記複数の発現プロファイルを（b）で取得された前記シーケンシングデータから決定する工程が、

前記複数の確率バーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルについて、（b）で取得された前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程を含む、前記〔1〕～〔35〕のいずれか一項に記載の方法。

〔37〕（b）で取得された前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の前記数を決定する工程が、
30

前記複数の標的の1つ以上について、

(1) (b)で取得された前記シーケンシングデータ中の前記標的に関連付けられた識別可能な配列を有する分子標識の前記数をカウントする工程と；

(2) (1)でカウントされた前記シーケンシングデータ中の前記標的に関連付けられた識別可能な配列を有する分子標識の前記数に基づいて前記標的の前記数を推定する工程と、

を含む、前記〔36〕に記載の方法。

〔38〕マルチプレット発現プロファイルを同定するための方法であって、
40

(a) 複数の細胞の複数の発現プロファイルを受け取る工程であって、前記複数の発現プロファイルが前記複数の細胞の各細胞における複数の標的の各標的の存在量を含む工程と；

(b) 前記複数の細胞の前記複数の発現プロファイルから複数の合成マルチプレット発現プロファイルを作成する工程と；

(c) 前記複数の細胞に関連付けられた前記複数の発現プロファイルの発現プロファイルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの発現プロファイルに基づいてシングルレット又はマルチプレットとして同定する工程と、
を含む、方法。

〔39〕前記発現プロファイルが(c)におけるマルチプレットとして同定される場合、
50

前記発現プロファイルを(a)で得られた前記複数の発現プロファイルから除去する工程を含む、前記〔38〕に記載の方法。

〔40〕前記複数のマルチプレットが、ダブルット、トリプレット、又はそれらの任意の組合せを含む、前記〔38〕～〔39〕のいずれか一項に記載の方法。

〔41〕前記複数の細胞の前記複数の発現プロファイルから前記複数の合成マルチプレット発現プロファイルを作成する工程が、

前記複数の合成マルチプレット発現プロファイルの合成マルチプレット発現プロファイルについて、

(1) 前記複数の発現プロファイルのいくつか(a number of)の発現プロファイルを選択する工程と；

(2) (1)で選択された前記発現プロファイルを組合せて、前記合成マルチプレット発現プロファイルを作成する工程と、

を含む、前記〔38〕～〔40〕のいずれか一項に記載の方法。

〔42〕(1)で選択された前記発現プロファイルを組合せて、前記合成マルチプレット発現プロファイルを作成する工程が、

前記複数の標的の各々について、選択された前記発現プロファイルにおける前記標的の存在量を組合せて、前記合成マルチプレット発現プロファイルにおける前記標的の存在量を作成する工程を含む、

前記〔41〕に記載の方法。

〔43〕前記合成マルチプレット発現プロファイルにおける前記標的の前記存在量が、選択された前記発現プロファイルにおける前記標的の前記存在量の合計である、前記〔42〕に記載の方法。

〔44〕前記合計が、選択された前記発現プロファイルにおける前記標的の前記存在量の加重合計である、前記〔43〕に記載の方法。

〔45〕前記合成マルチプレット発現プロファイルにおける前記標的の前記存在量が、選択された前記発現プロファイルにおける前記標的の前記存在量の平均である、前記〔42〕に記載の方法。

〔46〕前記平均が、選択された前記発現プロファイルにおける前記標的の前記存在量の加重平均である、前記〔45〕に記載の方法。

〔47〕前記複数の合成マルチプレット発現プロファイルの前記数が、およそ(a)で得られた前記複数の発現プロファイルの百分率である、前記〔38〕～〔46〕のいずれか一項に記載の方法。

〔48〕前記百分率が約10パーセントである、前記〔47〕に記載の方法。

〔49〕前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 発現プロファイルマルチプレットの同定を意図して機械学習モデルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び(a)で得られた前記複数の発現プロファイルの1つ以上の発現プロファイルから訓練する工程と；

(2) 前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔38〕～〔48〕のいずれか一項に記載の方法。

〔50〕前記機械学習モデルを訓練する際に用いられる前記複数の発現プロファイルの前記1つ以上の発現プロファイルが、(a)で得られた前記複数の発現プロファイルの百分率を含む、前記〔49〕に記載の方法。

〔51〕前記百分率が約10パーセントである、前記〔50〕に記載の方法。

〔52〕前記機械学習モデルが分類モデルを含む、前記〔49〕～〔51〕のいずれか一

10

20

30

40

50

項に記載の方法。

[5 3] 前記分類モデルが、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含む、前記〔 5 2 〕に記載の方法。

[5 4] 前記機械学習モデルが、ニューラルネットワーク、線形回帰モデル、ロジスティック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k 近傍法 (K N N) モデル、k 平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含む、前記〔 4 9 〕～〔 5 3 〕のいずれか一項に記載の方法。

[5 5] 前記機械学習モデルが、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スパース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含む、前記〔 4 9 〕～〔 5 4 〕のいずれか一項に記載の方法。

[5 6] 発現プロファイルマルチプレットの同定を意図して前記機械学習モデルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び(a)で得られた前記複数の発現プロファイルの 1 つ以上の発現プロファイルから訓練する工程が、

(1) (b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを発現プロファイル空間から低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの投影された発現プロファイルを作成する工程と；

(2) (a) で得られた前記複数の発現プロファイルの前記 1 つ以上の発現プロファイルを前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の発現プロファイルの 1 つ以上の投影された発現プロファイルを作成する工程と；

(3) 発現プロファイルマルチプレットの同定を意図して前記機械学習モデルを、(1)からの前記複数の合成マルチプレット発現プロファイルの前記投影された発現プロファイル及び(2)における前記複数の発現プロファイルの前記 1 つ以上の投影された発現プロファイルから訓練する工程と、

を含む、前記〔 4 9 〕～〔 5 5 〕のいずれか一項に記載の方法。

[5 7] 前記複数の細胞の前記細胞に関連付けられた前記複数の前記複数の発現プロファイルの前記発現プロファイルを投影して、前記複数の発現プロファイルの投影された発現プロファイルを作成する工程を含み、

ここで前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、前記機械学習モデルを用いて前記複数の発現プロファイルの前記投影された発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む、

前記〔 5 6 〕に記載の方法。

[5 8] 前記低次元空間が二次元空間である、前記〔 5 6 〕～〔 5 7 〕のいずれか一項に記載の方法。

[5 9] (b) で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを、前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの前記投影された発現プロファイルを作成する工程が、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイルを前記発現プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレット発現プロファイルの前記投影された発現プロファイルを作成する工程を含み、t 分布型確率的近傍埋め込み (t S N E) 法を用いる工程を含む、前記〔 5 6 〕～〔 5 8 〕のいずれか一項に記載の方法。

[6 0] 前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び前記発現プロファイルに基づいてシングレット又はマルチ

10

20

30

40

50

プレットとして同定する工程が、

前記細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと前記複数の発現プロファイルの少なくとも1つの発現プロファイルとの間の第1の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと前記複数の合成マルチプレット発現プロファイルの少なくとも1つの合成マルチプレット発現プロファイルとの間の第2の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む、前記〔38〕～〔59〕のいずれか一項に記載の方法。

〔61〕前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数の発現プロファイルを発現プロファイルの第1のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの第2のクラスターにクラスター化する工程と；

(3) 前記細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと発現プロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと合成マルチプレット発現プロファイルの第2のクラスターとの間の第2の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔38〕～〔59〕のいずれか一項に記載の方法。

〔62〕前記複数の細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、(b)で作成された前記複数の合成マルチプレット発現プロファイルの前記発現プロファイル及び前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数の発現プロファイルを発現プロファイルの第1のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレット発現プロファイルを合成マルチプレット発現プロファイルの複数の第2のクラスターにクラスター化する工程と；

(3) 前記細胞の前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルを、

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと発現プロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記発現プロファイルと合成マルチプレット発現プロファイルの前記複数の第2のクラスターの1つ以上のクラスターとの間の第2の距離

に基づく前記発現プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔38〕～〔59〕のいずれか一項に記載の方法。

〔63〕前記複数の細胞の前記複数の発現プロファイルを受け取る工程が、

(1) 複数のバーコードを用いて、前記複数の細胞における前記複数の標的にバーコードを付けて、前記複数の細胞の細胞において複数のバーコード付き標的を生成する工程であって、前記複数のバーコードの各々が細胞標識及び分子標識を含み、前記複数のバーコードの少なくとも2つのバーコードの分子標識が異なる分子標識配列を含み、且つ前記複

10

20

30

40

50

数のバーコードの少なくとも 2 つのバーコードが同一の細胞標識配列を有する細胞標識を含む工程と

(2) 前記複数のバーコード付き標的のシーケンシングデータを取得する工程と；

(3) (2) で取得された前記シーケンシングデータから前記複数のバーコードの細胞標識に関連付けられた前記複数の発現プロファイルを決定する工程であって、前記複数のバーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルが、前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を含む工程と、

を含む、前記〔38〕～〔62〕のいずれか一項に記載の方法。

〔64〕 前記複数のバーコードを用いて、前記複数の細胞における前記複数の標的にバーコードを付けて、前記複数の細胞の前記細胞における前記複数のバーコード付き標的を生成する工程が、

複数の粒子の前記複数のバーコードを用いて、前記複数の細胞における前記複数の標的にバーコードを付けて、前記複数の細胞の前記細胞における前記複数のバーコード付き標的を生成する工程を含み、

ここで前記複数の粒子の各々が前記複数のバーコードのサブセットを含み、ここでバーコードの前記サブセットの各々が同一の細胞標識配列を含み且つ少なくとも 100 の異なる分子標識配列を有する、

前記〔63〕に記載の方法。

〔65〕 前記粒子がビーズである、前記〔64〕に記載の方法。

〔66〕 前記ビーズが、ストレプトアビジンビーズ、アガロースビーズ、磁気ビーズ、コンジュゲートビーズ、プロテイン A コンジュゲートビーズ、プロテイン G コンジュゲートビーズ、プロテイン A / G コンジュゲートビーズ、プロテイン L コンジュゲートビーズ、オリゴ T コンジュゲートビーズ、シリカビーズ、シリカ様ビーズ、抗ビオチンマイクロビーズ、抗蛍光色素マイクロビーズ、及びそれらの任意の組合せからなる群から選択される、前記〔65〕に記載の方法。

〔67〕 前記粒子が、ポリジメチルシロキサン (P D M S)、ポリスチレン、ガラス、ポリプロピレン、アガロース、ゼラチン、ヒドロゲル、常磁性体、セラミック、プラスチック、ガラス、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、セルロース、ナイロン、シリコーン、及びそれらの任意の組合せからなる群から選択される材料を含む、前記〔63〕～〔67〕のいずれか一項に記載の方法。

〔68〕 前記複数の粒子の各々の前記バーコードが、少なくとも 1000 の異なる分子標識配列を有する分子標識を含む、前記〔63〕～〔67〕のいずれか一項に記載の方法。

〔69〕 前記複数の粒子の各々の前記バーコードが、少なくとも 10000 の異なる分子標識配列を有する分子標識を含む、前記〔63〕～〔68〕のいずれか一項に記載の方法。

〔70〕 前記バーコードの前記分子標識が、ランダム配列を含む、前記〔63〕～〔69〕のいずれか一項に記載の方法。

〔71〕 前記複数の粒子の各々が、少なくとも 10000 のバーコードを含む、前記〔63〕～〔70〕のいずれか一項に記載の方法。

〔72〕 前記複数のバーコードを用いて、前記複数の細胞における前記複数の標的にバーコードを付けて、前記複数の細胞の各細胞における前記複数のバーコード付き標的を生成する工程が、

(1) 前記標的のコピーを前記バーコードの標的結合領域と接触させる工程と；

(2) 前記複数のバーコードを用いて前記複数の標的を逆転写して、複数の逆転写された標的を生成する工程と、

を含む、前記〔63〕～〔71〕のいずれか一項に記載の方法。

〔73〕 前記複数のバーコード付き標的の前記シーケンシングデータを取得する前に、前記バーコード付き標的を増幅し、複数の増幅されたバーコード付き標的を作成する工程を含む、前記〔63〕～〔72〕のいずれか一項に記載の方法。

〔74〕 前記バーコード付き標的を増幅し、前記複数の増幅されたバーコード付き標的を

10

20

30

40

50

作成する工程が、ポリメラーゼ連鎖反応（P C R）により前記バーコード付き標的を増幅する工程を含む、前記〔73〕に記載の方法。

〔75〕前記シーケンシングデータから前記複数のバーコードの前記細胞標識に関連付けられた前記複数の発現プロファイルを決定する工程が、

前記複数のバーコードの前記細胞標識の細胞標識に関連付けられた前記複数の発現プロファイルの発現プロファイルについて、前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の数を決定する工程を含む、前記〔63〕～〔74〕のいずれか一項に記載の方法。

〔76〕前記シーケンシングデータ中の前記細胞標識及び前記複数の標的の各標的に関連付けられた識別可能な配列を有する分子標識の前記数を決定する工程が、

前記複数の標的の1つ以上について、

(1) 前記シーケンシングデータ中の前記標的に関連付けられた識別可能な配列を有する分子標識の前記数をカウントする工程と；

(2) 前記シーケンシングデータ中の前記標的に関連付けられた識別可能な配列を有する分子標識の前記数に基づいて前記標的の前記数を推定する工程と、
を含む、前記〔75〕に記載の方法。

〔77〕マルチプレットプロファイルを同定するための方法であって、

(a) 複数の細胞の複数のプロファイルを受け取る工程と；

(b) 前記複数の細胞の前記複数のプロファイルから複数の合成マルチプレットプロファイルを作成する工程と；

(c) 前記複数の細胞の細胞に関連付けられた前記複数のプロファイルのプロファイルを、(b)で作成された前記複数の合成マルチプレットプロファイルのプロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、
を含む、方法。

〔78〕前記複数の細胞の前記複数のプロファイルのプロファイルが、前記細胞のm R N A発現プロファイル、前記細胞のタンパク質発現プロファイル、前記細胞の突然変異プロファイル、前記細胞のメチル化プロファイル、又はそれらの任意の組合せを含む、前記〔77〕に記載の方法。

〔79〕前記m R N A発現プロファイルが、前記複数の細胞の各細胞における複数の遺伝子の各遺伝子のm R N A分子の存在量を含む、前記〔78〕に記載の方法。

〔80〕各遺伝子の前記存在量が、前記遺伝子の絶対存在量、前記遺伝子の正規化された存在量、又はそれらの組合せを含む、前記〔79〕に記載の方法。

〔81〕前記遺伝子の前記正規化された存在量が、マッピングされたリード100万あたりの転写物1キロベース当たりのリード（R P K M）の単位又は閾値数（C t）の単位で決定される、前記〔80〕に記載の方法。

〔82〕前記m R N A発現プロファイルが、シーケンシング、定量ポリメラーゼ連鎖反応（q P C R）、デジタルP C R、ハイブリダイゼーション、又はそれらの任意の組合せにより決定される、前記〔78〕～〔81〕のいずれか一項に記載の方法。

〔83〕前記細胞の前記タンパク質発現プロファイルが、前記複数の細胞の各細胞における複数の遺伝子の各遺伝子に対応するタンパク質分子の存在量を含む、前記〔78〕～〔82〕のいずれか一項に記載の方法。

〔84〕前記細胞の前記突然変異プロファイルが、前記細胞の複数のゲノム位置での前記細胞の突然変異プロファイルを含む、前記〔78〕～〔83〕のいずれか一項に記載の方法。

〔85〕前記細胞の前記メチル化プロファイルが、前記細胞の複数のゲノム位置での前記細胞のメチル化プロファイルを含む、前記〔78〕～〔84〕のいずれか一項に記載の方法。

〔86〕前記プロファイルが(c)におけるマルチプレットとして同定される場合、前記プロファイルを(a)で得られた前記複数のプロファイルから除去する工程を含む、前記〔85〕に記載の方法。

10

20

30

40

50

[8 7] 前記複数のマルチプレットが、ダブルット、トリプレット、又はそれらの任意の組合せを含む、前記〔 7 7 〕～〔 8 6 〕のいずれか一項に記載の方法。

[8 8] 前記複数の細胞の前記複数のプロファイルから前記複数の合成マルチプレットプロファイルを作成する工程が、

前記複数の合成マルチプレットプロファイルの合成マルチプレットプロファイルについて、

(1) 前記複数のプロファイルのいくつか (a number of) のプロファイルを選択する工程と；

(2) (1) で選択された前記プロファイルを組合せ、前記合成マルチプレットプロファイルを作成する工程と、

を含む、前記〔 7 7 〕～〔 8 7 〕のいずれか一項に記載の方法。

[8 9] (1) で選択された前記プロファイルを組合せ、前記合成マルチプレットプロファイルを作成する工程が、

前記複数の標的の各々について、選択された前記プロファイルにおける対応する要素の値を組合せ、前記合成マルチプレットプロファイルにおける対応する要素の値を作成する工程を含む、前記〔 8 8 〕に記載の方法。

[9 0] 前記合成マルチプレットプロファイルにおける前記対応する要素の前記値が、選択された前記プロファイルにおける前記対応する要素の前記値の合計である、前記〔 8 8 〕に記載の方法。

[9 1] 前記合計が、選択された前記プロファイルにおける前記対応する要素の前記値の加重合計である、前記〔 9 0 〕に記載の方法。

[9 2] 前記合成マルチプレットプロファイルにおける前記標的の前記存在量が、選択された前記プロファイルにおける前記対応する要素の前記値の平均である、前記〔 8 8 〕に記載の方法。

[9 3] 前記平均が、選択された前記プロファイルにおける前記対応する要素の前記値の加重平均である、前記〔 9 2 〕に記載の方法。

[9 4] 前記複数の合成マルチプレットプロファイルの前記数が、およそ (a) で得られた前記複数のプロファイルの百分率である、前記〔 7 7 〕～〔 9 3 〕のいずれか一項に記載の方法。

[9 5] 前記百分率が約 10 パーセントである、前記〔 9 4 〕に記載の方法。

[9 6] 前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、(b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイル及び前記プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) プロファイルマルチプレットの同定を意図して機械学習モデルを (b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイル及び (a) で得られた前記複数のプロファイルの 1 つ以上のプロファイルから訓練する工程と；

(2) 前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、前記機械学習モデルを用いて前記プロファイルに基づいてシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔 7 7 〕～〔 9 5 〕のいずれか一項に記載の方法。

[9 7] 前記機械学習モデルを訓練する際に用いられる前記複数のプロファイルの前記 1 つ以上のプロファイルが、(b) で得られた前記複数のプロファイルの百分率を含む、前記〔 9 6 〕に記載の方法。

[9 8] 前記百分率が約 10 パーセントである、前記〔 9 7 〕に記載の方法。

[9 9] 前記機械学習モデルが分類モデルを含む、前記〔 9 6 〕～〔 9 8 〕のいずれか一項に記載の方法。

[1 0 0] 前記分類モデルが、教師あり分類モデル、半教師あり分類モデル、教師なし分類モデル、又はそれらの組合せを含む、前記〔 9 9 〕に記載の方法。

[1 0 1] 前記機械学習モデルが、ニューラルネットワーク、線形回帰モデル、ロジスティ

10

20

30

40

50

イック回帰モデル、決定木、サポートベクターマシン、ナイーブベイズネットワーク、k 近傍法 (KNN) モデル、k 平均法モデル、ランダムフォレストモデル、又はそれらの任意の組合せを含む、前記〔96〕～〔100〕のいずれか一項に記載の方法。

〔102〕前記機械学習モデルが、相関ルール学習モデル、帰納論理プログラミングモデル、強化学習モデル、特徴学習モデル、類似度学習モデル、スペース辞書学習モデル、遺伝的アルゴリズムモデル、ルールベース機械学習モデル、学習分類子システムモデル、又はそれらの任意の組合せを含む、前記〔96〕～〔101〕のいずれか一項に記載の方法。

〔103〕プロファイルマルチプレットの同定を意図して前記機械学習モデルを、(b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイル及び (a) で得られた前記複数のプロファイルの 1 つ以上のプロファイルから訓練する工程が、

(1) (b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイルをプロファイル空間から低次元投影空間に投影して、前記複数の合成マルチプレットプロファイルの投影されたプロファイルを作成する工程と；

(2) (a) で得られた前記複数のプロファイルの前記 1 つ以上のプロファイルを前記プロファイル空間から前記低次元投影空間に投影して、前記複数のプロファイルの 1 つ以上の投影されたプロファイルを作成する工程と；

(3) プロファイルマルチプレットの同定を意図して前記機械学習モデルを、(1) からの前記複数の合成マルチプレットプロファイルの前記投影されたプロファイル及び (1) における前記複数のプロファイルの前記 1 つ以上の投影されたプロファイルから訓練する工程と、

を含む、前記〔96〕～〔102〕のいずれか一項に記載の方法。

〔104〕前記複数の細胞の前記細胞に関連付けられた前記複数の前記複数の前記プロファイルを投影して、前記複数のプロファイルの投影されたプロファイルを作成する工程を含み、

ここで前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、前記機械学習モデルを用いて前記プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、前記機械学習モデルを用いて前記複数のプロファイルの前記投影されたプロファイルに基づいてシングレット又はマルチプレットとして同定する工程を含む、

前記〔103〕に記載の方法。

〔105〕前記低次元空間が二次元空間である、前記〔103〕～〔104〕のいずれか一項に記載の方法。

〔106〕(b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイルを前記プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレットプロファイルの前記投影されたプロファイルを作成する工程が、(b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイルを前記プロファイル空間から前記低次元投影空間に投影して、前記複数の合成マルチプレットプロファイルの前記投影されたプロファイルを作成する工程を含み、t 分布型確率的近傍埋め込み (t SNE) 法を用いる工程を含む、前記〔103〕～〔105〕のいずれか一項に記載の方法。

〔107〕前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、(b) で作成された前記複数の合成マルチプレットプロファイルの前記プロファイル及び前記プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

前記細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルと前記複数のプロファイルの少なくとも 1 つのプロファイルとの間の第 1 の距離、及び

前記細胞に関連付けられた前記複数の発現プロファイルの前記プロファイルと前記複数の合成マルチプレット発現プロファイルの少なくとも 1 つの合成マルチプレット発現プロ

10

20

30

40

50

口ファイルとの間の第2の距離

に基づいてシングレット又はマルチプレットとして同定する工程を含む、前記〔77〕～〔106〕のいずれか一項に記載の方法。

〔108〕前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、(b)で作成された前記複数の合成マルチプレットプロファイルの前記プロファイル及び前記プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数のプロファイルをプロファイルの第1のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの第2のクラスターにクラスター化する工程と；

(3) 前記細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、

前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルとプロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルと合成マルチプレットプロファイルの前記第2のクラスターとの間の第2の距離

に基づいてシングレット又はマルチプレットとして同定する工程と、

前記〔77〕～〔106〕のいずれか一項に記載の方法。

〔109〕前記複数の細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、(b)で作成された前記複数の合成マルチプレットプロファイルの前記プロファイル及び前記プロファイルに基づいてシングレット又はマルチプレットとして同定する工程が、

(1) 前記複数のプロファイルをプロファイルの第1のクラスターにクラスター化する工程と；

(2) 前記複数の合成マルチプレットプロファイルを合成マルチプレットプロファイルの複数の第2のクラスターにクラスター化する工程と；

(3) 前記細胞の前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルを、

前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルとプロファイルの前記第1のクラスターとの間の第1の距離、及び

前記細胞に関連付けられた前記複数のプロファイルの前記プロファイルと合成マルチプレットプロファイルの前記複数の第2のクラスターの1つ以上のクラスターとの間の第2の距離

に基づいたシングレット又はマルチプレットとして同定する工程と、

を含む、前記〔77〕～〔106〕のいずれか一項に記載の方法。

〔110〕標的の前記数を決定するためのコンピュータシステムであって、

ハードウェア・プロセッサー；及び

命令(instructions)が保存されている非一時的なメモリーであって、前記ハードウェア・プロセッサーによって実行されるとき、前記プロセッサーが前記〔1〕～〔109〕のいずれか一項に記載の方法を実施するように導く、前記非一時的なメモリーを含む、コンピュータシステム。

〔111〕前記〔1〕～〔109〕のいずれか一項に記載の方法を実施するためのコードを含む、コンピュータで読み込み可能なメディア。

10

20

30

40

50

【図面】
【図 1】

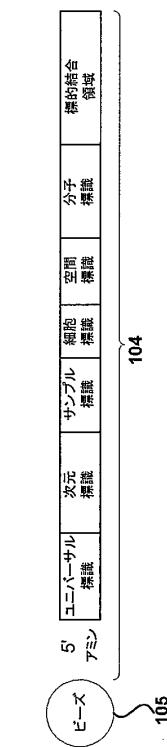


図 1

【図 2】

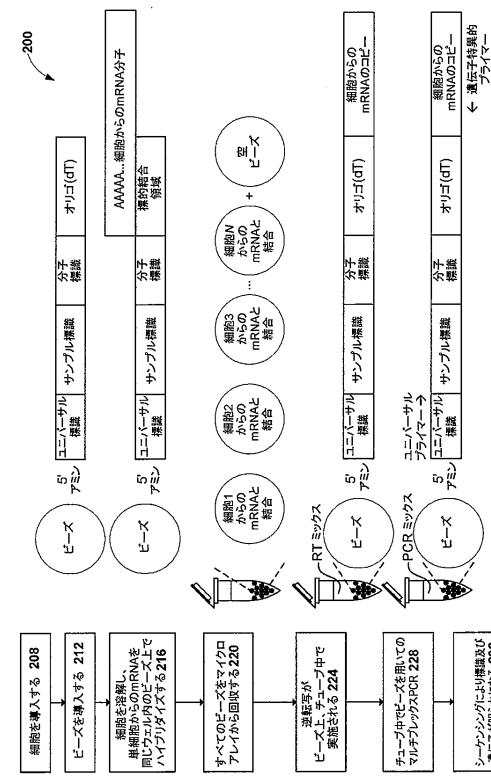


図 2

【図 3】

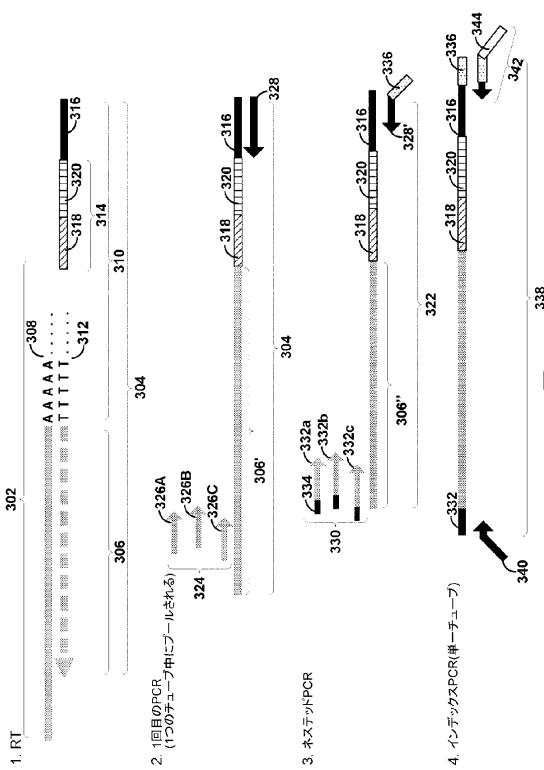


図 3

【図 4】

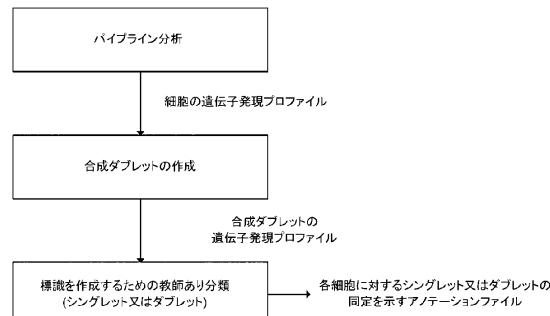
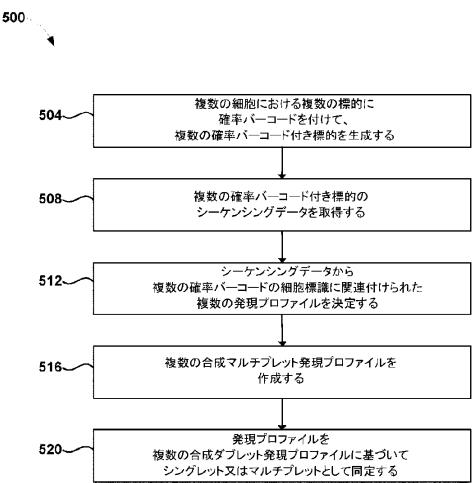
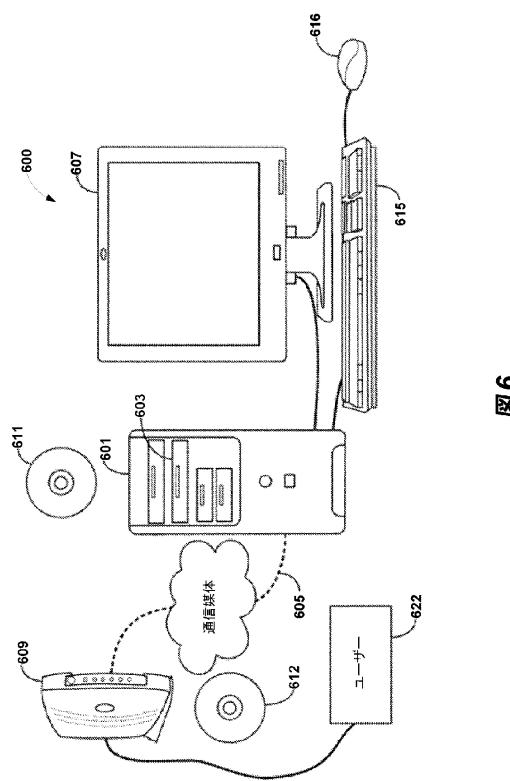


図 4

【図5】



【図6】



10

20

図6

図5

【図7】

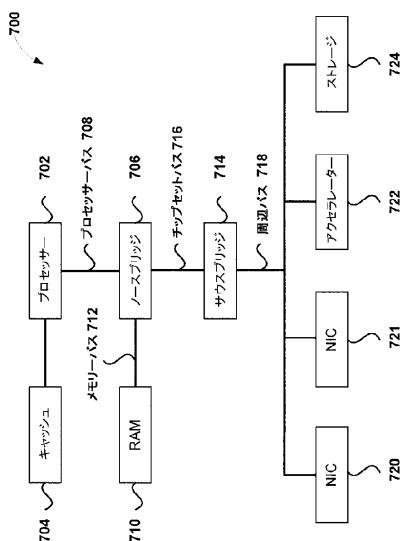
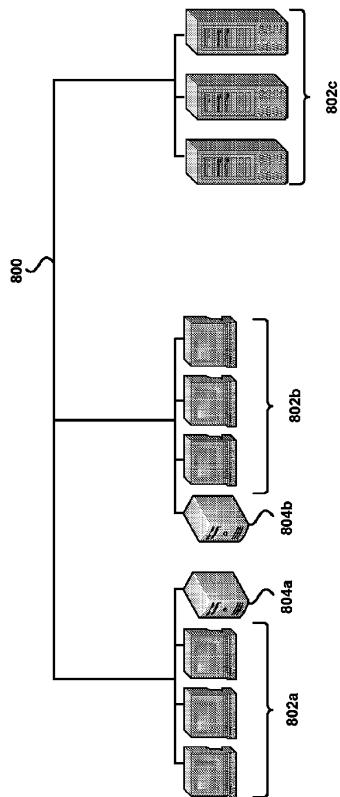


図7

【図8】



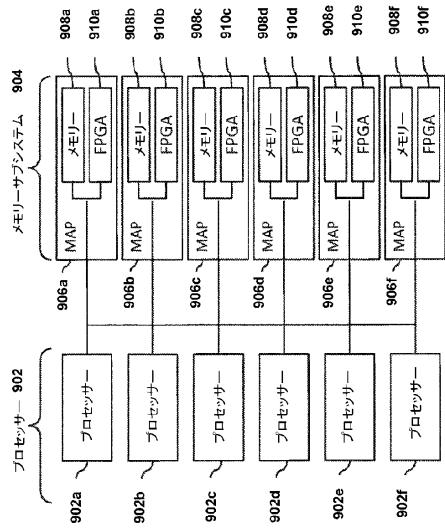
30

40

FIG. 8

50

【図 9】



【図 10】

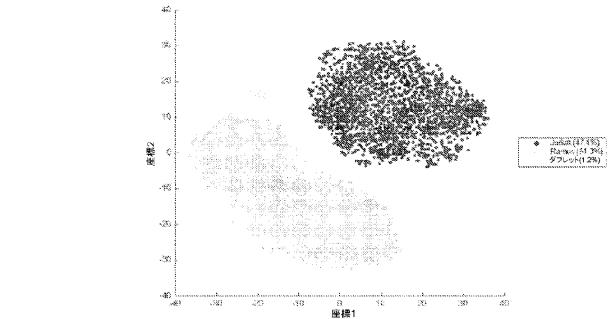
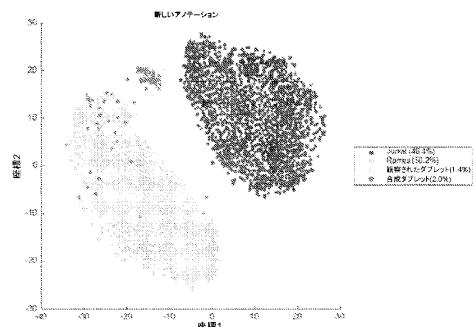


図 9

10

図 10A



20

図 10B

【図 11】

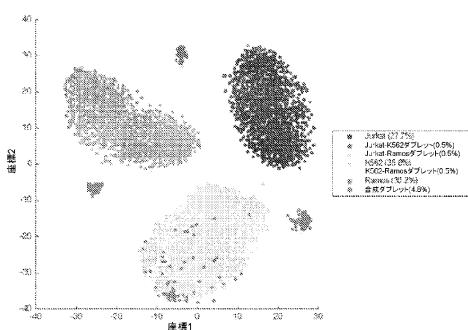


図 11

【図 12】

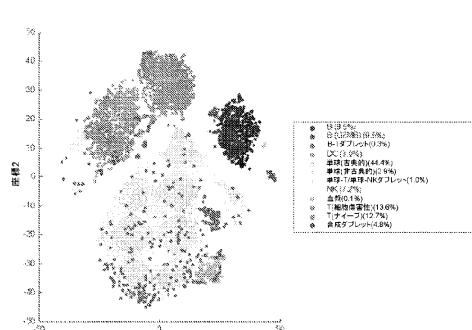


図 12

30

40

50

【図 1 3】

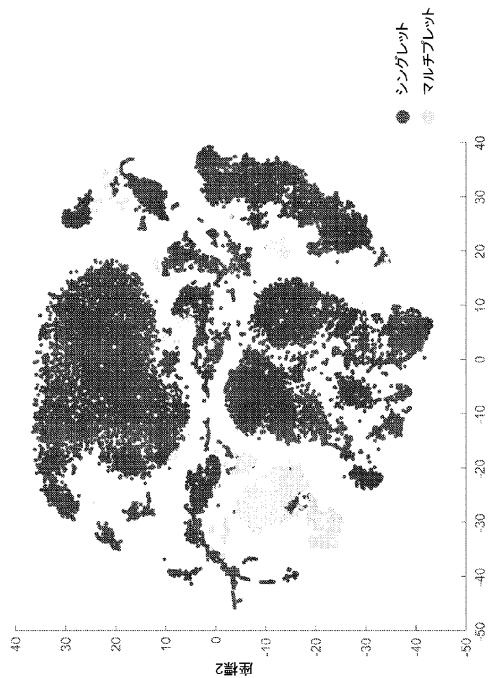


図13

【図 1 4】

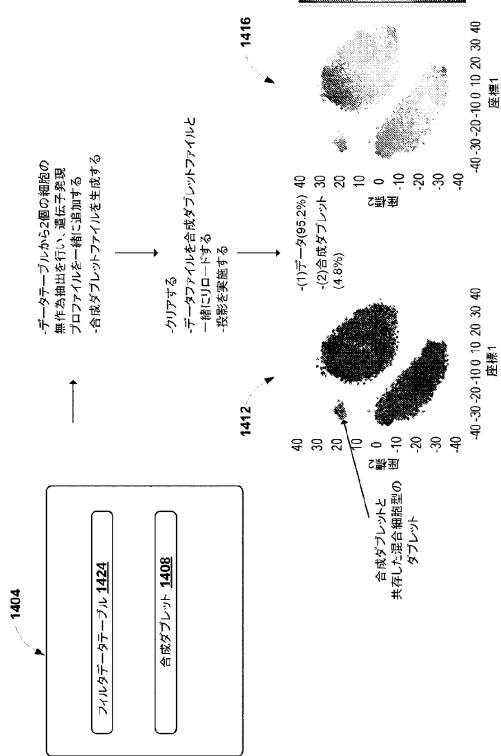


図14

フロントページの続き

(74)代理人 100103610
弁理士 吉田 和彦

(74)代理人 100109070
弁理士 須田 洋之

(74)代理人 100119013
弁理士 山崎 一夫

(74)代理人 100123777
弁理士 市川 さつき

(74)代理人 100111796
弁理士 服部 博信

(74)代理人 100168631
弁理士 佐々木 康匡

(72)発明者 ファン ジュエ
アメリカ合衆国 カリフォルニア州 94025 メンロー パーク キャンベル アベニュー 404
0 スイート 110

(72)発明者 ファン クリストイーナ
アメリカ合衆国 カリフォルニア州 94025 メンロー パーク キャンベル アベニュー 404
0 スイート 110

(72)発明者 ローゼンフェルド デイヴィッド
アメリカ合衆国 カリフォルニア州 94025 メンロー パーク キャンベル アベニュー 404
0 スイート 110

審査官 中野 あい

(56)参考文献 国際公開第2016/040476 (WO, A1)
特表2016-533187 (JP, A)
国際公開第2016/118915 (WO, A1)
国際公開第2018/226293 (WO, A1)
EVAN Z MACOSKO, HIGHLY PARALLEL GENOME-WIDE EXPRESSION PROFILING OF INDIVIDUAL CELLS USING NANOLITER DROPLETS, CELL, 2015年05月, VOL:161, NR:5, PAGE(S):1202-1214, <http://dx.doi.org/10.1016/j.cell.2015.05.002>

(58)調査した分野 (Int.Cl., DB名)
C12Q 1/00 - 3/00
C12N 15/00 - 15/90
Caplus / MEDLINE / EMBASE / BIOSIS (STN)