

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第1部門第1区分

【発行日】令和3年4月30日(2021.4.30)

【公表番号】特表2020-513837(P2020-513837A)

【公表日】令和2年5月21日(2020.5.21)

【年通号数】公開・登録公報2020-020

【出願番号】特願2019-552517(P2019-552517)

【国際特許分類】

C 1 2 Q 1/6813 (2018.01)

C 1 2 Q 1/6851 (2018.01)

C 1 2 Q 1/6869 (2018.01)

【F I】

C 1 2 Q 1/6813 Z

C 1 2 Q 1/6851 Z

C 1 2 Q 1/6869 Z

【手続補正書】

【提出日】令和3年3月22日(2021.3.22)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【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；

(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；

(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、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；

(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；及び

(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、を含む、方法。

【請求項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、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って、

(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おける複数の標的の各標的の発生率を含む；

(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。

【請求項4】

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、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発現プロファイルの数を選択する工程と；

(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。

【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発生率を組合せて、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発生率を作成する工程を含む、

請求項5に記載の方法。

【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発生率が、

(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発生率の合計であるか、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発生率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発生率の平均であるか、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発生率の加重平均である、

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前記数が、(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。

【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。

【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、

(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、
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、
請求項12に記載の方法。

【請求項14】

(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

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、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、

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、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、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、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、

前記複数の標的に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、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。

【請求項22】

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 2 3】

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
使用説明書が格納されている非一時的なメモリー
を含み、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
請求項 1 ~ 1 7、2 1 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、コン
ピュータシステム。

【請求項 2 4】

請求項 1 ~ 1 7、2 1 又は 2 2 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。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 2 8 0

【補正方法】変更

【補正の内容】

【0 2 8 0】

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
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關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關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。

[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關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。

[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關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；

(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關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發

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。

〔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、スパース辞書学習モデル、遺傳
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。

〔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、前記複数の発現ブ
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ブ
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ブ
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(tSNE)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、

(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。

〔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、

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、

(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発生率を含む工程と；

(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、
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〔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発現プロファイルの数を選択する工程と；

(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発生率を組合せて、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発生率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発生率が、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発生率の合計である、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発生率の
加重合計である、前記〔43〕に記載の方法。

[4 5] 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発生率が、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発生率の平均である、前記〔42〕に記載の方法。

[4 6] 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発生率の加重平均である、前記〔45〕に記載の方法。

[4 7] 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。

[4 8] 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。

[4 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。

[5 0] 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。

[5 1] 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。

[5 2] 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〔52〕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(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〔49〕～〔53〕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〔49〕～〔54〕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

ロファイルから訓練する工程と、

を含む、前記〔49〕～〔55〕のいずれか一項に記載の方法。

〔57〕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〔56〕に記載の方法。

〔58〕前記低次元空間が二次元空間である、前記〔56〕～〔57〕のいずれか一項に記載の方法。

〔59〕(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(tSNE)法を用いる工程を含む、前記〔56〕～〔58〕のいずれか一項に記載の方法。

〔60〕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、

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(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〔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作成する工程が、ポリメラーゼ連鎖反応(PCR)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、前記[77]に記載の方法。

[79] 前記mRNA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mRNA分子の発生率を含む、前記[78]に記載の方法。

[80] 各遺伝子の前記発生率が、前記遺伝子の絶対発生率、前記遺伝子の正規化された発生率、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発生率が、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(RPKM)の単位又は閾値数(Ct)の単位で決定される、前記[80]に記載の方法。

[82] 前記mRNA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(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発生率を含む、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。

[87] 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[77]～[86]のいずれか一項に記載の方法。

[88] 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、プロファイルの数を選択する工程と；

(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[77]～[87]のいずれか一項に記載の方法。

[89] (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[88]に記載の方法。

[90] 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[88]に記載の方法。

[91] 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[90]に記載の方法。

[92] 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発生率が、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[88]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〔 9 6 〕～〔 1 0 0 〕のいずれか一項に記載の方法。

[1 0 2] 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〔 9 6 〕～〔 1 0 1 〕のいずれか一項に記載の方法。

[1 0 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、(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〔 9 6 〕～〔 1 0 2 〕のいずれか一項に記載の方法。

[1 0 4] 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

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

使用説明書が格納されている非一時的なメモリー

を含み、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導く、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。