

【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6部門第3区分

【発行日】令和2年12月17日(2020.12.17)

【公表番号】特表2020-504357(P2020-504357A)

【公表日】令和2年2月6日(2020.2.6)

【年通号数】公開・登録公報2020-005

【出願番号】特願2019-523873(P2019-523873)

【国際特許分類】

G 1 6 B 10/00 (2019.01)

C 1 2 Q 1/6869 (2018.01)

【F I】

G 1 6 B 10/00

C 1 2 Q 1/6869 Z

【手続補正書】

【提出日】令和2年11月9日(2020.11.9)

【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】

細胞タイプを区別するための標的を同定する方法であつて、

(a) 複数の細胞の発現プロファイルを得ること、ここで前記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を幾つか含む、

(b) 前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの複数のクラスタを生成すること、ここで各クラスタは、(1)親クラスタ及び(2)2つ以上の子クラスタのうちの一方又は両方との1つ又は複数の関連性を有し、前記親クラスタは、前記クラスタによって表される前記複数の細胞の1つ又は複数の細胞の発現プロファイルを表し、前記クラスタは、前記2つ以上の子クラスタによって表される発現プロファイルを表す、

(c) 前記2つ以上の子クラスタを有する各クラスタについて、前記クラスタと前記2つ以上の子クラスタとの関連性が無効である場合、前記クラスタを統合クラスタセットに追加すること、

(d) 前記統合クラスタセット内の第1のクラスタのそれぞれについて、前記統合クラスタセット内の前記第1のクラスタと、前記第1のクラスタに最も近い前記統合クラスタセット内の第2のクラスタとの間の距離が、統合距離閾値内である場合、繰り返し、前記第1のクラスタ及び前記第2のクラスタを統合して、統合クラスタを生成すること、ここで前記統合クラスタは、前記第1のクラスタ及び前記第2のクラスタの発現プロファイルを含む、及び

(e) 前記統合クラスタセット内の各クラスタについて、細胞タイプを区別するための標的を、前記クラスタによって表される細胞の前記複数の標的の発現プロファイルに基づいて同定すること、

を含む、方法。

【請求項2】

前記複数の細胞の発現プロファイルを得ることは、標的カウントデータ構造を得ることを含み、

前記標的カウントデータ構造は、標的カウントマトリックスを含み、

前記標的カウントマトリックスの各行又は各列は、前記複数の細胞の異なる個々の細胞の発現プロファイルを含む、請求項1に記載の方法。

【請求項3】

前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルを発現プロファイルの前記複数のクラスタにクラスタリングすることは、

前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングして、前記複数の細胞の前記発現プロファイルを表す系統樹を生成することを含み、ここで前記系統樹は複数のクラスタを含み、前記複数のクラスタはルートクラスタ、複数のリーフクラスタ、及び複数の非ルート非リーフクラスタを含む、請求項1又は2に記載の方法。

【請求項4】

前記複数のリーフクラスタ及び前記複数の非ルート非リーフクラスタのそれぞれは、親クラスタとの関連性を有し、

前記ルートクラスタ及び前記複数の非ルート非リーフクラスタのそれぞれは、左子クラスタ及び右子クラスタとの関連性を有し、前記クラスタの前記左子クラスタ及び前記右子クラスタによって表される発現プロファイルを表し、

前記ルートクラスタは、前記複数の細胞の前記発現プロファイルを表す、請求項3に記載の方法。

【請求項5】

2つ以上の子クラスタを有する各クラスタについて、前記2つ以上の子クラスタとの前記クラスタ間の関連性が無効である場合、前記クラスタを統合クラスタセットに追加することは、

前記系統樹の前記ルートクラスタから前記系統樹の各クラスタを通して前記系統樹の前記複数のリーフクラスタにトラバースする間、

(1) 前記クラスタと前記クラスタの前記子クラスタとの関連性が有効であるか、それとも無効であるかを判断することと、

(2) 前記関連性が無効である場合、前記クラスタを統合クラスタセットに追加することと、

を含む、請求項3又は4に記載の方法。

【請求項6】

前記系統樹の前記複数のクラスタをトラバースする際、各クラスタにおいて、

前記関連性が有効である場合、前記クラスタから前記クラスタの前記左子クラスタ及び前記右子クラスタへのトラバースを継続することと、

前記関連性が無効である場合、前記クラスタから前記クラスタの前記左子クラスタ及び前記クラスタの前記右子クラスタへのトラバースを停止することと、

を含む、請求項5記載の方法。

【請求項7】

前記クラスタの前記子クラスタとの前記クラスタの前記関連性が有効であるか、それとも無効であるかを判断することは、

前記左子クラスタと前記右子クラスタとの間の距離が関連性閾値を超える場合、前記関連性を有効と判断することと、

その他の場合、無効と判断することと、

を含む、請求項5又は6に記載の方法。

【請求項8】

前記左子クラスタと前記右子クラスタとの間の距離は、

前記左子クラスタ及び前記右子クラスタによって表される発現プロファイル間の前記複数の標的的に対して実行される統計検定に基づいて決定されるか、或いは

前記左子クラスタによって表される発現プロファイルと前記右子クラスタによって表される各発現プロファイルとの間の前記複数の標的的に対して実行される前記統計検

定の最大 p 値に基づいて決定される、

請求項 7 に記載の方法。

【請求項 9】

前記系統樹の前記複数のクラスタをトラバースする際、各クラスタにおいて、(3)前記クラスタが1つの細胞の発現プロファイルを表す場合、前記クラスタを前記統合クラスタセットに追加することを含む、請求項 5 ~ 8 のいずれか一項に記載の方法。

【請求項 10】

前記系統樹の前記複数のクラスタをトラバースする際、各クラスタにおいて、クラスタ標識を前記クラスタに割り当てるなどを含む、請求項 5 ~ 9 のいずれか一項に記載の方法。

【請求項 11】

前記クラスタが1つの細胞の発現プロファイルを表す場合、前記クラスタの前記クラスタ標識は、1つの細胞名称を含み、

その他の場合、前記クラスタが前記親クラスタの前記左子クラスタであるとき、前記クラスタの前記クラスタ標識は、前記親クラスタの前記クラスタ標識及び左名称を含み、

その他のとき、前記クラスタの前記クラスタ標識は、前記親クラスタの前記クラスタ標識及び右名称を含む、請求項 10 に記載の方法。

【請求項 12】

前記統合クラスタセット内の各クラスタについて、前記細胞タイプを区別するための前記標的を、前記クラスタによって表される前記細胞の前記複数の標的の前記発現プロファイルに基づいて同定することは、

前記細胞タイプを区別するための前記標的に関連付けられた別個の配列を有する分子標識数において、前記クラスタによって表される発現プロファイルと、前記統合クラスタセット内の別のクラスタによって表される発現プロファイルとの差が有意な閾値よりも大きいと判断すること

を含む、請求項 1 ~ 11 のいずれか一項に記載の方法。

【請求項 13】

(b)において、前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの複数のクラスタを生成する前に、

(i) 前記複数の細胞の前記発現プロファイルの距離データ構造を決定すること、ここで、前記距離データ構造は、前記複数の細胞の前記発現プロファイルの距離マトリックスを含み、

(b)において、前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの前記複数のクラスタを生成することは、前記距離マトリックスに基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの前記複数のクラスタを生成することを含む

を含む、請求項 1 ~ 12 のいずれか一項に記載の方法。

【請求項 14】

(i)において前記距離データ構造を決定する前、前記標的カウントデータ構造を対数変換済み標的カウントデータ構造に対数変換することを含み、

前記標的カウントデータ構造の要素の前記距離データ構造を決定することは、前記対数変換済み標的カウントデータ構造の前記距離データ構造を決定することを含み、

(b)において、前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの前記複数のクラスタを生成することは、前記対数変換済み標的カウントデータ構造及び前記距離データ構造に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、前記複数のクラスタを生成することを含む、請求項 13 に記載の方法。

【請求項 15】

(b)において、前記複数の細胞の前記発現プロファイル間の距離に基づいて前記複数の細胞の前記発現プロファイルをクラスタリングすることは、

前記複数の細胞の各発現プロファイルを前記複数のクラスタにおける異なるリーフクラスタに割り当てることと、

第2のクラスタが前記複数のクラスタ中で第1のクラスタに最も近いクラスタである場合、繰り返し、前記複数のクラスタの前記第1のクラスタ及び前記第2のクラスタを組み合わせて、前記第1のクラスタ及び前記第2のクラスタの親クラスタを生成することと、を含む、請求項1～14のいずれか一項に記載の方法。

【請求項16】

前記第1のクラスタと前記第2のクラスタとの間の前記距離は、前記第1のクラスタによって表される任意の発現プロファイルと、前記第2のクラスタによって表される任意の発現プロファイルとの間の最大距離である、請求項15に記載の方法。

【請求項17】

前記第1のクラスタのクラスタ内相関及び前記第2のクラスタのクラスタ内相関は、前記第1のクラスタと前記第2のクラスタとのクラスタ間相関よりも高い、請求項15又は16に記載の方法。

【請求項18】

前記第1のクラスタのクラスタ内相関及び前記第2のクラスタのクラスタ内相関の指示は、前記第1のクラスタと前記第2のクラスタとのクラスタ間相関よりも高い、請求項15～17のいずれか一項に記載の方法。

【請求項19】

前記第1のクラスタのクラスタ内相関及び前記第2のクラスタのクラスタ内相関の指示は、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内最大相関、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内平均相関、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内メジアン相関、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内最小相関、及び

それらの任意の組合せ

の少なくとも1つに基づいており、

前記第1のクラスタの前記クラスタ内相関は、

前記第1のクラスタのクラスタ内最大相関、

前記第1のクラスタのクラスタ内平均相関、

前記第1のクラスタのクラスタ内メジアン相関、

前記第1のクラスタのクラスタ内最小相関、及び

それらの任意の組合せ

の少なくとも1つに基づいており

前記第2のクラスタの前記クラスタ内相関は、

前記第2のクラスタのクラスタ内最大相関、

前記第2のクラスタのクラスタ内平均相関、

前記第2のクラスタのクラスタ内メジアン相関、

前記第2のクラスタのクラスタ内最小相関、及び

それらの任意の組合せ

の少なくとも1つに基づいており、且つ

前記第1のクラスタと前記第2のクラスタとの前記クラスタ間相関は、

前記第1のクラスタと前記第2のクラスタとのクラスタ間最大相関、

前記第1のクラスタと前記第2のクラスタとのクラスタ間平均相関、

前記第1のクラスタと前記第2のクラスタとのクラスタ間メジアン相関、

前記第1のクラスタと前記第2のクラスタとのクラスタ間最小相関、及び

それらの任意の組合せ

の少なくとも1つに基づいている、

請求項 1 8 に記載の方法。【請求項 2 0】

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と、
命令が記憶された非一時的メモリであって、前記命令は、前記ハードウェアプロセッサ
によって実行されると、前記プロセッサに、請求項 1 ~ 1 9 のいずれか一項に記載の方法
を実行させる、非一時的メモリと、
を備える、コンピュータシステム。

【請求項 2 1】

請求項 1 ~ 1 9 のいずれか一項に記載の方法を実行する方法を実行するためのコードを
含むコンピュータ可読媒体。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 3 3 3

【補正方法】変更

【補正の内容】

【0 3 3 3】

様々な態様及び実施形態が本明細書に開示されたが、他の態様及び実施形態が当業者に
明らかになろう。本明細書に開示される様々な態様及び実施形態は、例示を目的とし、限
定を意図せず、真の範囲及び趣旨は以下の特許請求の範囲によって示される。

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) 細胞タイプを区別する標的を識別する方法であって、

(a) 標的カウントデータ構造を受信することであって、前記標的カウントデータ構造
は、複数の細胞の発現プロファイルを含み、前記複数の細胞の前記発現プロファイルは、
前記複数の細胞の各細胞の複数の標的の各標的を幾つか含む、受信することと、

(b) 前記標的カウントデータ構造及び前記複数の細胞の前記発現プロファイル間の距
離に基づいて、前記複数の細胞の発現プロファイルを階層的にクラスタリングして、前記
複数の細胞の前記発現プロファイルを表す系統樹を生成することであって、前記系統樹は
複数のノードを含み、前記複数のノードは、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、前記複数のリーフノードの各リーフノードは、前記
複数の細胞の異なる細胞の発現プロファイルを表し、前記ルートノードは、前記複数の細
胞の発現プロファイルを表す、階層的にクラスタリングして、系統樹を生成することと、

(c) 前記系統樹の前記複数のノードの各ノードを通して前記系統樹の前記ルートノ
ードから前記系統樹の前記複数のリーフノードまでトラバースする間、

(1) 前記ノードを前記ノードの子ノードに分割することが有効であるか、それとも
無効であるかを判断することと、

(2) 前記ノードを前記ノードの子ノードに分割することが無効である場合、前記ノ
ードを統合クラスタセットに追加することと、

(d) 前記統合クラスタセット内の第 1 のノードのそれぞれについて、繰り返し、
前記統合クラスタセット内の前記第 1 のノードと、前記第 1 のノードに最も近い前記
統合クラスタセット内の第 2 のノードとの間の距離が、統合距離閾値内である場合、前記
第 1 のノードを前記第 2 のノードと統合して、前記第 1 のノード及び前記第 2 のノードに
よって表される発現プロファイルを含む統合ノードを生成することと、

(e) 前記統合クラスタセット内のノードのそれぞれについて、前記ノードによって表
される細胞の前記複数の標的の発現プロファイルに基づいて、細胞タイプを区別する標的
を識別することと、

を含む、方法。

(2) 前記標的カウントデータ構造は、標的カウントマトリックスを含む、前記〔 1 〕に
記載の方法。

(3) 前記標的カウントマトリックスの各行又は各列は、前記複数の細胞の異なる個々の

細胞の複数の標的を幾つか含む、前記〔2〕に記載の方法。

〔4〕前記複数のリーフノード及び前記複数の非ルート非リーフノードのそれぞれには、親ノードが関連付けられ、

前記ルートノード及び前記複数の非ルート非リーフノードのそれぞれには、左子ノード及び右子ノードが関連付けられ、前記ルートノード及び前記複数の非ルート非リーフノードのそれぞれは、前記ノードの前記左子ノード及び前記右子ノードによって表される発現プロファイルを表す、前記〔1〕に記載の方法。

〔5〕(a)において、前記標的カウントデータ構造を受信する前、

(f)複数のバーコードを使用して前記複数の細胞内の前記複数の標的をバーコーディングして、複数のバーコード付き標的を作成することであって、前記複数のバーコードのそれぞれは、細胞標識及び分子標識を含み、異なる細胞の標的から作成されるバーコード付き標的は、異なる細胞標識を有し、前記複数の細胞の1つの細胞の標的から作成されるバーコード付き標的は、異なる分子標識を有する、バーコーディングして複数のバーコード付き標的を作成することと、

(g)前記複数のバーコード付き標的の配列データを取得することと、

(h)前記複数の細胞のそれれについて、

(1)前記細胞の前記配列データ内の前記複数の標的の各標的に関連付けられた別個の配列を有する分子標識の数をカウントすることと、

(2)(h)(1)においてカウントされた前記配列データ内の前記標的に関連付けられた別個の配列を有する分子標識の数に基づいて、前記細胞の前記複数の標的の各標的の数を推定することと

を含む、前記〔1〕～〔4〕のいずれか一項に記載の方法。

〔6〕前記標的カウントデータ構造を受信することは、

(h)(2)において推定された前記細胞の前記複数の標的の各標的の数から、標的カウントデータ構造を生成することであって、前記複数の細胞のうちの前記細胞の前記発現プロファイルは、(h)(2)において推定された前記細胞の前記複数の標的の各標的の数を含む、生成することを含む、前記〔5〕に記載の方法。

(7)(b)において、前記標的カウントデータ構造及び前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングして、前記複数の細胞の前記発現プロファイルを表す前記系統樹を生成する前、

(i)前記標的カウントデータ構造の要素の距離データ構造を特定することであって、前記距離データ構造は、前記複数の細胞の前記発現プロファイル間の距離を含む、特定すること

を含む、前記〔1〕～〔6〕のいずれか一項に記載の方法。

〔8〕前記距離データ構造は距離マトリックスを含む、前記〔7〕に記載の方法。

〔9〕前記距離マトリックスの各対角線要素は0の値を有する、前記〔8〕に記載の方法。

。

(10)(b)において、前記標的カウントデータ構造及び前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングして、前記複数の細胞の前記発現プロファイルを表す前記系統樹を生成することとは、前記標的カウントデータ構造及び前記距離データ構造に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングすることを含む、前記〔7〕～〔9〕のいずれか一項に記載の方法。

(11)前記複数の細胞の前記発現プロファイル間の前記距離は、前記複数の細胞の前記発現プロファイル間のペア毎の相関距離を含む、前記〔7〕～〔10〕のいずれか一項に記載の方法。

(12)(i)において前記標的カウントデータ構造の要素の距離データ構造を特定する前、前記標的カウントデータ構造を対数変換して、対数変換済み標的カウントデータ構造にすることを含み、

前記標的カウントデータ構造の要素の前記距離データ構造を特定することは、前記対数

変換済み標的カウントデータ構造の前記距離データ構造を特定することを含み、

(b)において、前記標的カウントデータ構造及び前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングすることは、前記対数変換済み標的カウントデータ構造及び前記距離データ構造に基づいて前記複数の細胞の前記発現プロファイルを階層的にクラスタリングして、前記系統樹を生成することを含む、前記〔7〕～〔11〕のいずれか一項に記載の方法。

〔13〕前記標的カウントデータ構造を前記対数変換済み標的カウントデータ構造に対数変換することは、前記標的カウントデータ構造の各要素の値を増分によって増大させることを含む、前記〔12〕に記載の方法。

〔14〕前記増分は1である、前記〔13〕に記載の方法。

〔15〕(b)において、前記標的カウントデータ構造及び前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングすることは、

前記複数の細胞の各発現プロファイルを異なるリーフノードに割り当てることと、
前記第2のノードが前記第1のノードへの前記複数のノードの最近傍ノードである場合、
前記複数のノードの前記第1のノード及び前記第2のノードを繰り返し結合して、前記第1のノード及び前記第2のノードの親ノードを生成することと、
を含む、前記〔1〕～〔14〕のいずれか一項に記載の方法。

〔16〕前記第1のノードと前記第2のノードとの間の前記距離は、前記第1のノードによって表される発現プロファイルを有する任意の細胞と、前記第2のノードによって表される発現プロファイルを有する任意の細胞との間の最大距離である、前記〔15〕に記載の方法。

〔17〕前記第1のノードのノード内相関及び前記第2のノードのノード内相関の少なくとも一方は、前記第1のノードと前記第2のノードとのノード間相関よりも高い、前記〔15〕又は〔16〕に記載の方法。

〔18〕前記第1のノードのノード内相関及び前記第2のノードのノード内相関の指示は、前記第1のノードと前記第2のノードとのノード間相関よりも高い、前記〔15〕又は〔16〕に記載の方法。

〔19〕前記第1のノードのノード内相関及び前記第2のノードのノード内相関の指示は、

前記第1のノード及び前記第2のノードのノード内最大相関、

前記第1のノード及び前記第2のノードのノード内平均相関、

前記第1のノード及び前記第2のノードのノード内メジアン相関、

前記第1のノード及び前記第2のノードのノード内最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔18〕に記載の方法。

〔20〕前記第1のノードの前記ノード内相関は、

前記第1のノードのノード内最大相関、

前記第1のノードのノード内平均相関、

前記第1のノードのノード内メジアン相関、

前記第1のノードのノード内最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔17〕～〔19〕のいずれか一項に記載の方法。

〔21〕前記第2のノードの前記ノード内相関は、

前記第2のノードのノード内最大相関、

前記第2のノードのノード内平均相関、

前記第2のノードのノード内メジアン相関、

前記第2のノードのノード内最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔17〕～〔20〕のいずれか一項に記載の方法。

[2 2] 前記第 1 のノードと前記第 2 のノードとの前記ノード間相関は、

前記第 1 のノードと前記第 2 のノードとのノード間最大相関、

前記第 1 のノードと前記第 2 のノードとのノード間平均相関、

前記第 1 のノードと前記第 2 のノードとのノード間メジアン相関、

前記第 1 のノードと前記第 2 のノードとのノード間最小相関、

それらの任意の組合せ

の少なくとも 1 つに基づく、前記 [1 7] ~ [2 1] のいずれか一項に記載の方法。

[2 3] 前記系統樹の前記複数のノードをトラバースする際、各ノードにおいて、

前記分割が有効である場合、前記ノードから前記左子ノード及び前記ノードの前記右子ノードに引き続きトラバースすることと、

前記分割が無効である場合、前記ノードから前記左子ノード及び前記ノードの前記右子ノードへのトラバースを停止することと、

を含む、前記 [4] ~ [1 6] のいずれか一項に記載の方法。

[2 4] 前記ノードの前記子ノードを有する前記ノードの前記分割が有効であるか、それとも無効であるかを判断することは、

前記左子ノードと前記右子ノードとの間の距離が分割閾値を超える場合、前記分割を有効と判断することと、

その他の場合、無効と判断することと、

を含む、前記 [4] ~ [2 3] のいずれか一項に記載の方法。

[2 5] 前記左子ノードと前記右子ノードとの間の距離は、前記左子ノード及び前記右子ノードによって表される発現プロファイル間の前記複数の標的的各標的で実行される統計検定に基づいて特定される、前記 [2 4] に記載の方法。

[2 6] 前記統計検定はウェルチ t 検定を含む、前記 [2 5] に記載の方法。

[2 7] 前記左子ノードと前記右子ノードとの間の距離は、前記左子ノードによって表される各発現プロファイルと前記右子ノードによって表される各発現プロファイルとの間の前記複数の標的的に対して実行される前記統計検定の最大 p 値に基づいて特定される、前記 [2 5] 又は [2 6] に記載の方法。

[2 8] 前記系統樹の前記複数のノードをトラバースする際、各ノードにおいて、(3) 前記ノードが 1 つの細胞の発現プロファイルを表す場合、前記ノードを前記統合クラスタセットに追加することを含む、前記 [1] ~ [2 7] のいずれか一項に記載の方法。

[2 9] 前記系統樹の前記複数のノードをトラバースする際、各ノードにおいて、前記ノード標識を前記ノードに割り当てるなどを含む、前記 [4] ~ [2 8] のいずれか一項に記載の方法。

[3 0] 前記ノードが 1 つの細胞の発現プロファイルを表す場合、前記ノードの前記ノード標識は、1 つの細胞名称を含み、

その他の場合、前記ノードが前記親ノードの前記左子ノードであるとき、前記ノードの前記ノード標識は、前記親ノードの前記ノード標識及び左名称を含み、

その他のとき、前記ノードの前記ノード標識は、前記親ノードの前記ノード標識及び右名称を含む、前記 [2 9] に記載の方法。

[3 1] 前記統合クラスタセット内の各ノードについて、前記ノードによって表される前記細胞の前記複数の標的的前記発現プロファイルに基づいて前記細胞タイプを区別する前記標的を識別することは、

前記細胞タイプを区別する前記標的に関連付けられた別個の配列を有する分子標識数を単位として、前記ノードによって表される発現プロファイルと、前記統合クラスタセット内の別のノードによって表される発現プロファイルとの差が有意な閾値よりも大きいと判断すること

を含む、前記 [1] ~ [3 0] のいずれか一項に記載の方法。

[3 2] (d) において前記第 1 のノードを前記第 2 のノードと統合して、前記統合ノードを生成する前、

第 3 のノードと第 4 のノードとの間の距離がノード距離閾値内にある場合、1 つの細胞

の発現プロファイルを表す前記統合クラスタセット内の前記第3のノードのそれを前記統合クラスタセット内の前記第4のノードと統合することを含む、前記〔1〕～〔31〕のいずれか一項に記載の方法。

〔33〕前記細胞の発現プロファイルを表す前記統合クラスタセット内の前記ノードに基づいて前記複数の細胞を分類することを含む、前記〔1〕～〔32〕のいずれか一項に記載の方法。

〔34〕識別された細胞タイプを区別する前記標的に基づいて全体トランスクリプトームアッセイを指定することを含む、前記〔1〕～〔33〕のいずれか一項に記載の方法。

〔35〕識別された細胞タイプを区別する前記標的に基づいて、標的とされるトランスクリプトームアッセイを指定することを含む、前記〔1〕～〔33〕のいずれか一項に記載の方法。

〔36〕細胞タイプを区別する標的を識別する方法であって、

(a) 複数の細胞の発現プロファイルを受信することであって、前記発現プロファイルは、前記複数の細胞の各細胞の複数の標的の各標的を幾つか含む、受信することと、

(b) 前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの複数のクラスタを生成することであって、各クラスタは、(1)親クラスタ及び(2)2つ以上の子クラスタのうちの一方又は両方との1つ又は複数の関連性を有し、前記親クラスタは、前記クラスタによって表される前記複数の細胞の1つ又は複数の細胞の発現プロファイルを表し、前記クラスタは、前記2つ以上の子クラスタによって表される発現プロファイルを表す、クラスタリングして生成することと、

(c) 前記2つ以上の子クラスタを有する各クラスタについて、前記クラスタと前記2つ以上の子クラスタとの関連性が無効である場合、前記クラスタを統合クラスタセットに追加することと、

(d) 前記統合クラスタセット内の第1のクラスタのそれぞれについて、前記統合クラスタセット内の前記第1のクラスタと、前記第1のクラスタに最も近い前記統合クラスタセット内の第2のクラスタとの間の距離が、統合距離閾値内である場合、繰り返し、前記第1のクラスタ及び前記第2のクラスタを統合して、統合クラスタを生成することであって、前記統合クラスタは、前記第1のクラスタ及び前記第2のクラスタの発現プロファイルを含む、繰り返し統合することと、

(e) 前記統合クラスタセット内の各クラスタについて、前記クラスタによって表される細胞の前記複数の標的の発現プロファイルに基づいて細胞タイプを区別する標的を識別することと、
を含む、方法。

〔37〕前記複数の細胞の発現プロファイルを受信することは、標的カウントデータ構造を受信することを含む、前記〔36〕に記載の方法。

〔38〕前記標的カウントデータ構造は、標的カウントマトリックスを含む、前記〔37〕に記載の方法。

〔39〕前記標的カウントマトリックスの各行又は各列は、前記複数の細胞の異なる個々の細胞の発現プロファイルを含む、前記〔38〕に記載の方法。

〔40〕前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルを発現プロファイルの前記複数のクラスタにクラスタリングすることは、

前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルを階層的にクラスタリングして、前記複数の細胞の前記発現プロファイルを表す系統樹を生成することであって、前記系統樹は複数のクラスタを含み、前記複数のクラスタはルートクラスタ、複数のリーフクラスタ、及び複数の非ルート非リーフクラスタを含む、階層的にクラスタリングして、系統樹を生成することを含む、前記〔36〕～〔39〕のいずれか一項に記載の方法。

〔41〕前記複数のリーフクラスタ及び前記複数の非ルート非リーフクラスタのそれぞ

は、親クラスタとの関連性を有し、

前記ルートクラスタ及び前記複数の非ルート非リーフクラスタのそれぞれは、左子クラスタ及び右子クラスタとの関連性を有し、前記クラスタの前記左子クラスタ及び前記右子クラスタによって表される発現プロファイルを表し、

前記ルートクラスタは、前記複数の細胞の前記発現プロファイルを表す、前記〔40〕に記載の方法。

〔42〕2つ以上の子クラスタを有する各クラスタについて、前記2つ以上の子クラスタとの前記クラスタ間の関連性が無効である場合、前記クラスタを統合クラスタセットに追加することは、

前記系統樹の前記ルートクラスタから前記系統樹の各クラスタを通して前記系統樹の前記複数のリーフクラスタにトラバースする間、

(1) 前記クラスタと前記クラスタの前記子クラスタとの関連性が有効であるか、それとも無効であるかを判断することと、

(2) 前記関連性が無効である場合、前記クラスタを統合クラスタセットに追加することと、

を含む、前記〔40〕又は〔41〕に記載の方法。

〔43〕(a)において、前記複数の細胞の前記発現プロファイルを受信する前、

(f) 複数のバーコードを使用して前記複数の細胞内の前記複数の標的をバーコーディングして、複数のバーコード付き標的を作成することであって、前記複数のバーコードのそれぞれは、細胞標識及び分子標識を含み、異なる細胞の標的から作成されるバーコード付き標的は、異なる細胞標識を有し、前記複数の細胞の1つの細胞の標的から作成されるバーコード付き標的は、異なる分子標識を有する、バーコーディングして複数のバーコード付き標的を作成することと、

(g) 前記複数のバーコード付き標的の配列データを取得することと、

(h) 前記複数の細胞のそれぞれについて、

(1) 前記細胞の前記配列データ内の前記複数の標的の各標的に関連付けられた別個の配列を有する分子標識の数をカウントすることと、

(2) (h)(1)においてカウントされた前記配列データ内の前記標的に関連付けられた別個の配列を有する分子標識の数に基づいて、前記細胞の前記複数の標的の各標的の数を推定することと、

を含む、を含む、前記〔36〕～〔42〕に記載の方法。

〔44〕前記複数の細胞のうちの前記細胞の前記発現プロファイルは、(h)(2)において推定された前記細胞の前記複数の標的の各標的の数を含む、前記〔43〕に記載の方法。

〔45〕(b)において、前記複数の細胞の前記発現プロファイル間の距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの複数のクラスタを生成する前、

(i) 前記複数の細胞の前記発現プロファイルの距離データ構造を特定することを含む、前記〔36〕～〔44〕のいずれか一項に記載の方法。

〔46〕前記距離データ構造は、前記複数の細胞の前記発現プロファイルの距離マトリックスを含む、前記〔45〕に記載の方法。

〔47〕前記距離マトリックスの各対角線要素は0の値を有する、前記〔46〕に記載の方法。

〔48〕(b)において、前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの前記複数のクラスタを生成することは、前記距離マトリックスに基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの前記複数のクラスタを生成することを含む、前記〔46〕又は〔47〕に記載の方法。

〔49〕前記複数の細胞の前記発現プロファイル間の前記距離は、前記複数の細胞の前記発現プロファイル間のペア毎の相関距離である、前記〔45〕～〔48〕のいずれか一項

に記載の方法。

[5 0] (i) において前記距離データ構造を特定する前、前記標的カウントデータ構造を対数変換済み標的カウントデータ構造に対数変換することを含み、

前記標的カウントデータ構造の要素の前記距離データ構造を特定することは、前記対数変換済み標的カウントデータ構造の前記距離データ構造を特定することを含み、

(b) において、前記複数の細胞の前記発現プロファイル間の前記距離に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、発現プロファイルの前記複数のクラスタを生成することは、前記対数変換済み標的カウントデータ構造及び前記距離データ構造に基づいて、前記複数の細胞の前記発現プロファイルをクラスタリングして、前記複数のクラスタを生成することを含む、前記〔45〕～〔49〕のいずれか一項に記載の方法。

[5 1] 前記標的カウントデータ構造を前記対数変換済み標的カウントデータ構造に対数変換することは、前記標的カウントデータ構造の各要素の値を増分によって増大させることを含む、前記〔50〕に記載の方法。

[5 2] 前記増分は1である、前記〔51〕に記載の方法。

[5 3] (b) において、前記複数の細胞の前記発現プロファイル間の距離に基づいて前記複数の細胞の前記発現プロファイルをクラスタリングすることは、

前記複数の細胞の各発現プロファイルを異なるリーフクラスタに割り当てることと、第2のクラスタが第1のクラスタへの前記複数のクラスタの最近傍クラスタである場合、前記複数のクラスタの前記第1のクラスタ及び前記第2のクラスタを繰り返し結合して、前記第1のクラスタ及び前記第2のクラスタの親クラスタを生成することと、

を含む、前記〔36〕～〔52〕のいずれか一項に記載の方法。

[5 4] 前記第1のクラスタと前記第2のクラスタとの間の前記距離は、前記第1のクラスタによって表される任意の発現プロファイルと、前記第2のクラスタによって表される任意の発現プロファイルとの間の最大距離である、前記〔53〕に記載の方法。

[5 5] 前記第1のクラスタのクラスタ内相関及び前記第2のクラスタのクラスタ内相関の少なくとも一方は、前記第1のクラスタと前記第2のクラスタとのクラスタ間相関よりも高い、前記〔53〕又は〔54〕に記載の方法。

[5 6] 前記第1のクラスタのクラスタ内相関及び前記第2のクラスタのクラスタ内相関の指示は、前記第1のクラスタと前記第2のクラスタとのクラスタ間相関よりも高い、前記〔15〕又は〔16〕に記載の方法。

[5 7] 前記第1のクラスタのクラスタ内相関及び前記第2のクラスタのクラスタ内相関の指示は、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内最大相関、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内平均相関、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内メジアン相関、

前記第1のクラスタ及び前記第2のクラスタのクラスタ内最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔17〕又は〔18〕に記載の方法。

[5 8] 前記第1のクラスタの前記クラスタ内相関は、

前記第1のクラスタのクラスタ内最大相関、

前記第1のクラスタのクラスタ内平均相関、

前記第1のクラスタのクラスタ内メジアン相関、

前記第1のクラスタのクラスタ内最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔17〕～〔19〕のいずれか一項に記載の方法。

[5 9] 前記第2のクラスタの前記クラスタ内相関は、

前記第2のクラスタのクラスタ内最大相関、

前記第2のクラスタのクラスタ内平均相関、

前記第2のクラスタのクラスタ内メジアン相関、

前記第2のクラスタのクラスタ内最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔17〕～〔20〕のいずれか一項に記載の方法。

〔60〕前記第1のクラスタと前記第2のクラスタとの前記クラスタ間相関は、

前記第1のクラスタと前記第2のクラスタとのクラスタ間最大相関、

前記第1のクラスタと前記第2のクラスタとのクラスタ間平均相関、

前記第1のクラスタと前記第2のクラスタとのクラスタ間メジアン相関、

前記第1のクラスタと前記第2のクラスタとのクラスタ間最小相関、

それらの任意の組合せ

の少なくとも1つに基づく、前記〔17〕～〔21〕のいずれか一項に記載の方法。

〔61〕前記系統樹の前記複数のクラスタをトラバースする際、各クラスタにおいて、

前記関連性が有効である場合、前記クラスタから前記クラスタの前記左子クラスタ及び前記クラスタの前記右子クラスタに引き続きトラバースすることと、

前記関連性が無効である場合、前記クラスタから前記クラスタの前記左子クラスタ及び前記クラスタの前記右子クラスタへのトラバースを停止することと、

を含む、前記〔42〕～〔54〕のいずれか一項に記載の方法。

〔62〕前記クラスタの前記子クラスタとの前記クラスタの前記関連性が有効であるか、それとも無効であるかを判断することは、

前記左子クラスタと前記右子クラスタとの間の距離が関連性閾値を超える場合、前記関連性を有効と判断することと、

その他の場合、無効と判断することと、

を含む、前記〔42〕～〔61〕のいずれか一項に記載の方法。

〔63〕前記左子クラスタと前記右子クラスタとの間の距離は、前記左子クラスタ及び前記右子クラスタによって表される発現プロファイル間の前記複数の標的の各標的で実行される統計検定に基づいて特定される、前記〔62〕に記載の方法。

〔64〕前記統計検定はウェルチt検定を含む、前記〔63〕に記載の方法。

〔65〕前記左子クラスタと前記右子クラスタとの間の距離は、前記左子クラスタによって表される発現プロファイルと前記右子クラスタによって表される各発現プロファイルとの間の前記複数の標的の各標的に対して実行される前記統計検定の最大p値に基づいて特定される、前記〔63〕又は〔64〕に記載の方法。

〔66〕前記系統樹の前記複数のクラスタをトラバースする際、各クラスタにおいて、(3)前記クラスタが1つの細胞の発現プロファイルを表す場合、前記クラスタを前記統合クラスタセットに追加することを含む、前記〔42〕～〔65〕のいずれか一項に記載の方法。

〔67〕前記系統樹の前記複数のクラスタをトラバースする際、各クラスタにおいて、前記クラスタ標識を前記クラスタに割り当てるなどを含む、前記〔42〕～〔66〕のいずれか一項に記載の方法。

〔68〕前記クラスタが1つの細胞の発現プロファイルを表す場合、前記クラスタの前記クラスタ標識は、1つの細胞名称を含み、

その他の場合、前記クラスタが前記親クラスタの前記左子クラスタであるとき、前記クラスタの前記クラスタ標識は、前記親クラスタの前記クラスタ標識及び左名称を含み、

その他のとき、前記クラスタの前記クラスタ標識は、前記親クラスタの前記クラスタ標識及び右名称を含む、前記〔67〕に記載の方法。

〔69〕前記統合クラスタセット内の各クラスタについて、前記クラスタによって表される前記細胞の前記複数の標的の前記発現プロファイルに基づいて前記細胞タイプを区別する前記標的を識別することは、

前記細胞タイプを区別する前記標的に関連付けられた別個の配列を有する分子標識数を単位として、前記クラスタによって表される発現プロファイルと、前記統合クラスタセット内の別のクラスタによって表される発現プロファイルとの差が有意な閾値よりも大きいと判断すること

を含む、前記〔36〕～〔68〕のいずれか一項に記載の方法。

〔70〕(d)において前記第1のクラスタを前記第2のクラスタと統合して、前記統合クラスタを生成する前、

第3のクラスタと第4のクラスタとの間の距離がクラスタ距離閾値内にある場合、1つの細胞の発現プロファイルを表す前記統合クラスタセット内の前記第3のクラスタのそれを前記統合クラスタセット内の前記第4のクラスタと統合すること

を含む、前記〔36〕～〔69〕のいずれか一項に記載の方法。

〔71〕前記細胞の発現プロファイルを表す前記統合クラスタセット内の前記クラスタに基づいて前記複数の細胞を分類することを含む、前記〔36〕～〔70〕のいずれか一項に記載の方法。

〔72〕識別された細胞タイプを区別する前記標的に基づいて全体トランスクリプトームアッセイを指定することを含む、前記〔36〕～〔71〕のいずれか一項に記載の方法。

〔73〕識別された細胞タイプを区別する前記標的に基づいて、標的とされるトランスクリプトームアッセイを指定することを含む、前記〔36〕～〔71〕のいずれか一項に記載の方法。

〔74〕標的の数を特定する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と、

命令が記憶された非一時的メモリであって、前記命令は、前記ハードウェアプロセッサによって実行されると、前記プロセッサに、前記〔1〕～〔73〕のいずれか一項に記載の方法を実行させる、非一時的メモリと、

を備える、コンピュータシステム。

〔75〕前記〔1〕～〔73〕のいずれか一項に記載の方法を実行する方法を実行するコードを含むコンピュータ可読媒体。