

【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】平成30年7月5日(2018.7.5)

【公表番号】特表2017-524374(P2017-524374A)

【公表日】平成29年8月31日(2017.8.31)

【年通号数】公開・登録公報2017-033

【出願番号】特願2017-515036(P2017-515036)

【国際特許分類】

C 1 2 Q 1/68 (2018.01)

G 0 6 F 19/22 (2011.01)

【F I】

C 1 2 Q 1/68 A

G 0 6 F 19/22

【手続補正書】

【提出日】平成30年5月23日(2018.5.23)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0230

【補正方法】変更

【補正の内容】

【0230】

図4Aは、そのような配列マスクを創出するための過程400のフローチャートを示しており、それを1つまたは複数の検査サンプルに適用して、コピー数の評価における検討から、関心対象の配列上のピンを除去することができる。過程は、複数の影響なしのトレーニングサンプルからの配列読み取りを含むトレーニングセットを提供することによって開始する。ブロック402。次いで、過程は、トレーニングセットの配列読み取りを、関心対象の配列を含む参照ゲノムにアラインメントし、それによってトレーニングサンプルに対するトレーニング配列タグが提供される。ブロック404。いくつかの態様において、非除外部位にマッピングされた、一意的にアラインメントされた非冗長のタグのみを、さらなる解析に用いる。過程は、参照ゲノムを複数のピンに分割する工程、および影響なしの各トレーニングに対して、各トレーニングサンプルに対する、各ピンにおけるトレーニング配列タグの被覆率を決定する工程を伴う。ブロック406。過程は、各ピンに対して、すべてのトレーニングサンプルにわたるトレーニング配列タグの予想被覆率も決定する。ブロック408。いくつかの態様において、各ピンの予想被覆率は、トレーニングサンプルにわたる中央値または平均である。予想被覆率は、包括的プロファイルをなす。次いで、過程は、包括的プロファイルにおける変動を除去することによって、各トレーニングサンプルに対して各ピンにおけるトレーニング配列タグの被覆率を調整し、それによって各トレーニングサンプルに対してピンにおけるトレーニング配列タグの包括的プロファイル除去被覆率を獲得する。ブロック410。いくつかの実践は、各トレーニングサンプルに存在する、GC含有量レベルと包括的プロファイル除去被覆率との間の関係に基づき、各トレーニングサンプルに対する包括的プロファイル除去被覆率を調整し、それによって各トレーニングサンプルに対してトレーニング配列タグのサンプルGC補正被覆率を獲得する工程を伴う。ブロック412。次いで、過程は、参照ゲノムにわたる、マスキングされていないおよびマスキングされたピンを含む配列マスクを創出する。マスキングされた各ピンは、マスキング閾値を超える分布特徴を有する。ブロック414。分布特徴は、トレーニングサンプルにわたるピンにおけるトレーニング配列タグの調整された被覆率に対して提供される。いくつかの実践において、マスキング閾値は、トレーニングサンプルにわたるピン内の、正規化された被覆率の観察される変動に関係し得る。サンプルにわたる正規化された被覆率に

についての高い変動係数または中央値絶対偏差を有するピンは、それぞれの測定基準の経験分布に基づいて同定され得る。いくつかの代替的な実践において、マスキング閾値は、トレーニングサンプルにわたるピン内の、正規化された被覆率の観察される変動に関係し得る。サンプルにわたる正規化された被覆率についての高い変動係数または中央値絶対偏差を有するピンは、それぞれの測定基準の経験分布に基づいてマスキングされ得る。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0237

【補正方法】変更

【補正の内容】

【0237】

図5のブロック501に示されるように、解析法は、影響なしのサンプルのトレーニングセットにおける各ピンに対する被覆率を獲得する。トレーニングセットデータは、影響なしのサンプルをシーケンシングすることにより獲得された、各ピンに対する被覆率値（例えば、所定の100kbピン内に見出された非除外部位の数）を含む。シーケンシングは、当分野における、検査サンプルとともに用いられる同じ機器/プロトコールを用いる。ある特定の態様において、少なくとも100個、または少なくとも300個のトレーニングセットサンプルを用いる。ある特定の態様において、被覆率値を、マルチプレックスシーケンシング（例えば、12プレックスシーケンシング）から獲得する。いくつかの態様において、参照ゲノムを、それぞれ約100kbの等しいサイズのピンに分割する。本明細書における他の箇所で記載されるように、他のサイズが可能である。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0238

【補正方法】変更

【補正の内容】

【0238】

解析の初期部分は、症候群特異的領域におけるまたは症候群特異的領域の群におけるフィルタリングされていないすべてのピンが解析されるまで、一度に1つのピンに対して実施され、かつこうすることによって、解析は、症候群特異的領域におけるピンと大まかに同じやり方で変動する、症候群特異的領域の外側にあるピンの群を同定する。ブロック503に例証されるように、解析法は、症候群特異的領域における新たなピンを検討中のピンとして設定することによって、ピン解析ルーチンを制御する。ある特定の態様において、症候群特異的領域の境界を、本明細書において後で記載されるように、コンセンサス領域から選定する。いくつかの態様において、症候群特異的領域の境界を、本明細書において後で記載されるように、コンセンサス領域を越えた検索領域から選定する。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0239

【補正方法】変更

【補正の内容】

【0239】

次に、ブロック505に示されるように、解析過程は、検討中のピンと、あらゆるロバストな染色体における他のあらゆる利用可能なピンとの間の相関距離を決定する。一手法において、該過程は、すべてのサンプルからの正規化された被覆率値を用いることによって、相関距離を同定する。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】0240

【補正方法】変更

【補正の内容】

【0240】

ブロック505の作業（ブロック507の作業と連動して）は、検討中のピンと同じ様式で概して変動するピンを同定する。別の言い方をすれば、方法は、ピン被覆率を解析して、症候群領域において観察された体系的変性を共有するピンを同定する。それは、ロバストな染色体に属するすべての常染色体ピン（すなわち、すべてのヒト染色体、しかしながら第13、第18、および第21染色体を除いたもの由来のピン）の間のペアワイズ距離を表す距離行列を構築し得る。当然、上記で記載されるように、常染色体ピンは、NESなどをもたらすフィルタリングの後に残るものに限定され得る。

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0242

【補正方法】変更

【補正の内容】

【0242】

検討中のピンにさらに焦点を合わせて、解析過程は、相関距離に基づいて、（症候群特異的領域の外側にある）ロバストな染色体のピンを順位付けし、かつ検討中のピンに対して最も近い近隣ピン（SNB - 症候群近隣ピン）のセットに対するメンバーを同定する。ブロック507を参照されたい。この作業は、検討中のピンと最も高い相関を有する、症候群領域の外側にあるピンを収集する。いくつかの態様において、関心対象の症候群特異的領域における各100kbピンに対して、最も高度に相関したピンの約5%を選択する。他の態様において、種々の量のピンを選択し得る。これらのピンを、SNBセットへの包含のために確保する。

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】0243

【補正方法】変更

【補正の内容】

【0243】

これにより、症候群特異的領域における検討中のピンの処理が完結し、そうして解析は、他の任意の症候群特異的ピンが依然として検討されるべきかどうかを判定する。ブロック509を参照されたい。そうである場合、過程制御はブロック503に戻り、そこで、症候群特異的領域からの次のピンが検討中のピンとして設定される。このようにして、該解析は、症候群特異的領域に属するフィルタリングされていない各ピンに対して反復する。

【手続補正8】

【補正対象書類名】明細書

【補正対象項目名】0244

【補正方法】変更

【補正の内容】

【0244】

検討中の各症候群特異的ピンに対して同定されかつ確保された近隣ピンは、症候群領域の外側にある一意的ピンの収集物としてプールされかつ確保される。近隣ピンは、検討中の各症候群ピンに対して行われる過程繰り返しにわたって収集される。すべての症候群ピンが解析された後にもたらされる収集物がSNBセットである。この作業は、ブロック511によって例証される。この点から先へ、解析は、SNBからの被覆率値のみが検討される第2の相に着手する。いくつかの態様において、SNBセットは約6000個のピンを含む。いくつかの態様において、ゲノム全体の代表的部分集団を含むSNBセットを有することが有利である。例えば、ゲノム全体の約10%を占めるSNBセットを有することが有利である。

【手続補正9】

【補正対象書類名】明細書

【補正対象項目名】 0 2 4 5

【補正方法】 変更

【補正の内容】

【 0 2 4 5 】

第2の相を始めると、作業513は、SNB被覆率値を次元として含有するベクトルを用いて、トレーニングセットのあらゆる2つのメンバー間の相関距離を決定する。この過程は、ブロック515の過程と連動して、SNBセットにおけるすべてのピンにわたって同じ様式で概して変動するトレーニングセットメンバー（影響なしのサンプル）を同定する。次いで、そのようなトレーニングセットメンバーを用いて、検討中のピンに対して、多ウェーブ補正過程において単一ウェーブである補正を同定する。

【手続補正 1 0】

【補正対象書類名】 明細書

【補正対象項目名】 0 2 4 9

【補正方法】 変更

【補正の内容】

【 0 2 4 9 】

次に、ブロック517によって示されるように、解析過程は、クラスターメンバーのピン被覆率値を用いて、補正ウェーブを創出しかつ確保する。この作業は、参照配列における各非除外ピンに対して別個の補正值を作成する。該作業は、いくつかの技法のうちのいずれか1つを採用して、クラスターメンバーのピン値から補正值をもたらし得る。一例において、該作業は、ピンの補正值に対する個々の被覆率値の中心的傾向（例えば、中央値）をクラスターメンバーから選ぶ。すべてのピンに対して、同じ過程が採用される。ゆえに、この例において、補正ウェーブはピン被覆率値の収集物であり、それぞれは検討中のピンに対する被覆率値の中央値を表す。

【手続補正 1 1】

【補正対象書類名】 明細書

【補正対象項目名】 0 2 5 0

【補正方法】 変更

【補正の内容】

【 0 2 5 0 】

この時点で、症候群特異的領域の各非除外ピン、および症候群特異的領域の外側にある高度に相関したピンに対して、補正值が創出されている。これらの値は、補正の「第1のウェーブ」を表す。描かれたフローチャートにおいて、解析過程は、この補正のウェーブを、トレーニングセット被覆率値に適用する。ブロック519を参照されたい。該過程は、多くの利用可能な技法のうちのいずれかを用いて、補正ウェーブを適用し得る。一手法では、ピンごとに、サンプルの被覆率値を補正值で割る。別の手法では、ピンごとに、サンプル値から補正值を差し引く。さらに別の手法では、ピンごとに、サンプル値からピン予測値を差し引く。

【手続補正 1 2】

【補正対象書類名】 明細書

【補正対象項目名】 0 2 5 3

【補正方法】 変更

【補正の内容】

【 0 2 5 3 】

過程は、次に、任意のより多くの症候群特異的補正ウェーブを適用するべきかどうかを判定する。ブロック521を参照されたい。いくつかの実践において、該過程は、単一補正ウェーブのみを生成する。より典型的には、それは多数の補正ウェーブを生成し、各補正はサンプル解析をさらに向上させる；すなわち、体系的変動を低下させかつ変動の係数を向上させる。いくつかの態様において、十分に大きな数のウェーブ、例えば10～20個のウェーブを導き出す。ウェーブの性能は、症候群CVをウェーブの数の関数として算出するこ

とによって、独立した検査セットにおいて評価される。ウェーブの所望の数は、データを検査することにおいてCVを統計的に有意に低下させるすべてのウェーブを見出すことによって決定される。

【手続補正 13】

【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】

1種または複数種のゲノムの核酸を含む検査サンプルにおける、関心対象の配列のコピー数の評価のための、1つまたは複数のプロセッサおよびシステムメモリーを含むコンピューターシステムで実践される方法であって、該方法は、

(a) 検査サンプルにおけるDNAをシーケンシングすることによって獲得された配列読み取りを受け取る工程；

(b) 検査サンプルの配列読み取りを、関心対象の配列を含む参照ゲノムにアラインメントし、それによって検査配列タグを提供する工程であって、該参照ゲノムは複数のピンに分割されており、該関心対象の配列は、コピー数変動が遺伝的症候群と関連している染色体部分ゲノム領域 (sub-chromosomal genomic region) にある、工程；

(c) 参照ゲノムにおけるピンについて、検査配列タグの被覆率を決定する工程；

(d) 検査サンプルと実質的に同じ様式でシーケンシングされかつアラインメントされた、影響なしのトレーニングサンプルのトレーニングセットの部分集団から獲得された関心対象の配列における複数のピンにおける複数の予想被覆率を採用することによって、関心対象の配列における複数のピンにおける複数の被覆率を調整する工程であって、該複数のピンにおける該複数の予想被覆率は、関心対象の配列内にあるピンの被覆率と相関することが見出された、関心対象の配列の外側にあるピンの被覆率を用いて獲得された、工程；および

(e) (d) からの調整された被覆率に基づき、検査サンプルにおける関心対象の配列のコピー数を評価する工程

を含む、方法。

【請求項 2】

ゲノムのうちの1種または複数種が染色体異数性を有するかどうかを(e) に基づいて決定する工程をさらに含む、請求項1記載の方法。

【請求項 3】

(d) の前に、トレーニングセットから獲得された包括的ウェーブプロファイルを適用することによって検査配列タグの被覆率を調整する工程をさらに含み、該包括的ウェーブプロファイルは、トレーニングセットにわたって平均化された、参照ゲノムにおけるピンの被覆率を含む、前記請求項のいずれか一項記載の方法。

【請求項 4】

(d) の前に、検査サンプルのピン間のGC含有量レベルと被覆率との間の関係に基づき、検査配列タグの被覆率を調整する工程をさらに含む、請求項3記載の方法。

【請求項 5】

1種または複数種の染色体のコピー数を評価する工程は、検査サンプルの1種または複数種の染色体のそれぞれに対して配列量 (sequence dose) を算出する工程を含み、該配列量は、1種または複数種の染色体のすべてのピンにおける検査配列タグの検査サンプル合計被覆率を、正規化配列のすべてのピンにおける検査配列タグの合計被覆率で割ることによって算出される、請求項2記載の方法。

【請求項 6】

配列量をトレーニングセットの配列量の標準偏差で割ることによって、正規化された配列値を獲得する工程、および正規化された配列値を1つまたは複数の閾値と比較する工程

をさらに含む、請求項5記載の方法。

【請求項7】

(c)においてピンに対する被覆率を決定する工程は、すべてのピンにわたる配列タグの総数に対して、ピンごとのタグの計数を正規化する工程を含み、かつ(d)において調整される被覆率は、正規化された被覆率である、請求項1記載の方法。

【請求項8】

(d)において用いられる、関心対象の配列の外側にあるピンは、第13、第18、および第21染色体以外の1つまたは複数のヒト常染色体におけるピンである、請求項1記載の方法。

【請求項9】

関心対象の配列の外側にあるピンは、関心対象の配列内にある検討中のピンにおける被覆率と、関心対象の配列の外側にあるピンにおける被覆率との間の相関距離を決定することによって同定される、請求項1記載の方法。

【請求項10】

相関距離は、トレーニングセットのサンプルから創出された、ピン被覆率のベクトル間の距離として算出される、請求項9記載の方法。

【請求項11】

(i)トレーニングセットの部分集団として、関心対象の配列の外側にあるピンにおけるそれらの被覆率において互いに相関する、トレーニングセット中のトレーニングサンプルを同定し、かつ(ii)関心対象の配列における複数のピンにおける部分集団の被覆率から予想被覆率を獲得することによって、予想被覆率を獲得した、請求項1記載の方法。

【請求項12】

サンプルの群を同定する工程は、該サンプルのクラスターを同定する工程を含む、請求項11記載の方法。

【請求項13】

予想被覆率を獲得する工程は、トレーニングセットの部分集団の被覆率の中心的傾向を判定する工程を含む、請求項11記載の方法。

【請求項14】

何回か繰り返して(d)を反復する工程をさらに含み、各繰り返しは、先行の繰り返しからの調整された被覆率を、最新の繰り返しにおいて調整される対象となる被覆率として用い、かつ各繰り返しは、影響なしのサンプルの異なる部分集団から獲得された予想被覆率を採用する、請求項1記載の方法。

【請求項15】

作業(d)においてピンの検査配列タグの被覆率を調整する工程は、関数をデータ点に適合させる工程であって、各データ点は、予想被覆率を、ピンにおける検査サンプルに対する対応する被覆率に関連付けする、工程；ピンにおける被覆率を該関数に適用することによって、関心対象の配列のピンにおける被覆率を調整する工程を含む、請求項1記載の方法。

【請求項16】

関数は線形関数である、請求項15記載の方法。

【請求項17】

作業(d)において関心対象の配列のピンにおける検査配列タグの被覆率を調整する工程は、関心対象の配列のピンに対する測定された被覆率値から予想値を差し引く工程を含む、請求項1～13のいずれか一項記載の方法。

【請求項18】

関心対象の配列として、症候群特異的領域の開始点および終了点を決定するセグメント化を実施する工程をさらに含む、前記請求項のいずれか一項記載の方法。

【請求項19】

検査サンプルは、2種の異なるゲノム由来の核酸の混合物を含む、前記請求項のいずれ

か一項記載の方法。

【請求項 20】

DNAはcfDNA分子を含む、前記請求項のいずれか一項記載の方法。

【請求項 21】

検査サンプルは、胎児および母体の細胞フリー核酸を含む、前記請求項のいずれか一項記載の方法。

【請求項 22】

検査サンプルは、同じ対象由来の癌性細胞および影響なしの細胞由来の核酸を含む、前記請求項のいずれか一項記載の方法。

【請求項 23】

シーケンサーを用いて検査サンプル由来の核酸をシーケンシングし、それによって検査サンプルの配列読み取りを生成する工程をさらに含む、前記請求項のいずれか一項記載の方法。

【請求項 24】

検査配列タグおよびトレーニング配列タグの被覆率は非除外部位計数（NES計数）として提供され、NES計数は、非除外部位にマッピングされた非冗長の配列タグの数である、前記請求項のいずれか一項記載の方法。

【請求項 25】

シーケンシング読み取りは初回のマルチプレックスシーケンシングによって獲得され、検査サンプルを、第1の閾値よりも高い、症候群分類またはコピー数変動をコールするための第1の値を有すると判定する工程；
初回のマルチプレックスシーケンシングよりも深いシーケンシング深度において、検査サンプルを再シーケンシングして、再シーケンシングされたデータを獲得する工程；および再シーケンシングされたデータを用いて症候群分類またはコピー数変動を判定する工程をさらに含む、前記請求項のいずれか一項記載の方法。

【請求項 26】

再シーケンシングされたデータを用いて症候群分類またはコピー数変動を判定する工程は、
再シーケンシングされたデータから、症候群分類またはコピー数変動をコールするための第2の値を獲得する工程；および
該第2の値を第2の閾値と比較する工程であって、該第2の閾値は第1の閾値よりも高い、工程を含む、請求項25記載の方法。

【請求項 27】

前記検査サンプルは、あらかじめ設定された値よりも低い第1の値を有し、該あらかじめ設定された値は第1の閾値よりも高く、かつ第1の閾値よりも低いサンプルは影響なしであると判定され、あらかじめ設定された値よりも高いサンプルは影響ありであると判定され、かつ第1の閾値からあらかじめ設定された値に及ぶサンプルは再シーケンシングについて同定される、請求項25記載の方法。

【請求項 28】

遺伝的症候群は、1p36欠失症候群、ウォルフ・ヒルシュホーン症候群、猫鳴き症候群、アンジェルマン症候群、ウィリアムズ症候群、およびディジョージ症候群からなる群より選択される、前記請求項のいずれか一項記載の方法。

【請求項 29】

コンピューターシステムの1つまたは複数のプロセッサによって実行される場合に遺伝的症候群に関係した関心対象の配列のコピー数の評価のための方法を該コンピューターシステムに実践させるプログラムコードを保存している、非一時的な機械可読媒体を含むコンピュータープログラム製品であって、該プログラムコードは、
(a) 検査サンプルにおけるDNAをシーケンシングすることによって獲得された配列読み取りを受け取る工程；

(b) 検査サンプルの配列読み取りを、関心対象の配列を含む参照ゲノムにアラインメントし、それによって検査配列タグを提供する工程であって、該参照ゲノムは複数のピンに分割されており、該関心対象の配列は、コピー数変動が遺伝的症候群と関連している染色体部分ゲノム領域にある、工程；

(c) 参照ゲノムにおけるピンについて、検査配列タグの被覆率を決定する工程；

(d) 検査サンプルと実質的に同じ様式でシーケンシングされかつアラインメントされた、影響なしのトレーニングサンプルのトレーニングセットの部分集団から獲得された関心対象の配列における複数のピンにおける複数の予想被覆率を採用することによって、関心対象の配列における複数のピンにおける複数の被覆率を調整する工程であって、該複数のピンにおける該複数の予想被覆率は、関心対象の配列内にあるピンの被覆率と相関することが見出された、関心対象の配列の外側にあるピンの被覆率を用いて獲得された、工程；および

(e) (d) からの調整された被覆率に基づき、検査サンプルにおける関心対象の配列のコピー数を評価する工程

のためのコードを含む、コンピュータープログラム製品。

【請求項 30】

1種または複数種のゲノムの核酸を含む検査サンプルを用いた、遺伝的症候群に関係した関心対象の配列のコピー数の評価のためのシステムであって、該システムは、

サンプルからの核酸配列情報を提供する検査サンプル由来の核酸を受け取るためのシーケンサー；

(a) 検査サンプルにおける細胞フリーDNAをシーケンシングすることによって獲得された配列読み取りを受け取る作業；

(b) 検査サンプルの配列読み取りを、関心対象の配列を含む参照ゲノムにアラインメントし、それによって検査配列タグを提供する作業であって、該参照ゲノムは複数のピンに分割されており、該関心対象の配列は、コピー数変動が遺伝的症候群と関連している染色体部分ゲノム領域にある、作業；

(c) 参照ゲノムにおけるピンに対して、検査配列タグの被覆率を決定する作業；

(d) 検査サンプルと実質的に同じ様式でシーケンシングされかつアラインメントされた、影響なしのトレーニングサンプルのトレーニングセットから獲得された予想被覆率を採用することによって、関心対象の配列における複数のピンにおける複数の被覆率を調整する作業であって、該複数のピンにおける該複数の予想被覆率は、関心対象の配列内にあるピンの被覆率と相関することが見出された、関心対象の配列の外側にあるピンの被覆率を用いて獲得された、作業；および

(e) (d) からの調整された被覆率に基づき、検査サンプルにおける関心対象の配列のコピー数をコンピューターシステムによって評価する作業

を実行するまたは引き起こすように設計されたまたは構成された論理回路を含む、システム。

【手続補正 14】

【補正対象書類名】図面

【補正対象項目名】図 5

【補正方法】変更

【補正の内容】

【図 5】

