

【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】平成30年3月1日(2018.3.1)

【公開番号】特開2016-184185(P2016-184185A)

【公開日】平成28年10月20日(2016.10.20)

【年通号数】公開・登録公報2016-060

【出願番号】特願2014-223834(P2014-223834)

【国際特許分類】

G 0 6 F 19/20 (2011.01)

C 1 2 Q 1/68 (2018.01)

C 1 2 N 15/09 (2006.01)

C 1 2 M 1/00 (2006.01)

【F I】

G 0 6 F 19/20

C 1 2 Q 1/68 A

C 1 2 N 15/00 F

C 1 2 M 1/00 A

【手続補正書】

【提出日】平成29年10月25日(2017.10.25)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0 0 2 8

【補正方法】変更

【補正の内容】

【0 0 2 8】

この好適な指標の一例を挙げれば、MAF(マイナーアレル頻度)5%以上のSNPについて、実験によりタイピングされた遺伝子型とインピュテーションにより推定された遺伝子型の相関係数の二乗の平均が0.94以上、好ましくは0.95以上、より好ましくは0.96以上となるような個数以上であることが挙げられる。この個数より少ないと、選択されたTag SNPのタイピング結果に基づいてインピュテーションした結果と、実際の遺伝子型との相関が従来品と比較して優れているとはいえず、本発明において期待される従来品に対する有用性を十分に発揮することが困難になる。さらに、MAF3~5%でのSNPのインピュテーションによる遺伝子型と実際の遺伝子型の相関係数の二乗の平均が0.82以上、好ましくは0.84以上、より好ましくは0.87以上とする指標や、MAF1~3%でのSNPのインピュテーションによる遺伝子型と実際の遺伝子型の相関係数の二乗の平均が0.73以上、好ましくは0.75以上、より好ましくは0.79以上とする指標を用いることも可能である。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 0

【補正方法】変更

【補正の内容】

【0 0 3 0】

上記の「30万個程度、40万個程度」等のSNPの個数を示す「程度」とは、「約」と同様であるが、特に、特定の個数、例えば「30万個」のTag SNPのインピュテーション性能は、ある程度の個数の幅の範囲内であれば実質的に変わらないことを示している。具体的には、特定のTag SNP数の1%以内、厳密には0.5%以内の違いであれば、実質的なインピュテーション性能に違いは無い。これは、一旦選択されたTag SNP群か

ら、いくつかのSNPを除く場合の目安となる値である。さらに、一旦選択されたTag SNP群から除かれるSNPが、実際にはインピュテーションに寄与しないものである場合には、SNPの抜去を行ってもインピュテーション性能における影響はさらに軽微になる。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0043

【補正方法】変更

【補正の内容】

【0043】

(a) Tag SNPとの間の連鎖不平衡の度合いが弱くインピュテーションで十分な精度で遺伝子型を推定しづらいSNP:

この分類の他のSNPには、Target SNPのうち、本発明のTag SNPとの間の r^2 連鎖不平衡値が低い(例えば $r^2 < 0.2$)SNPが該当する。それらの中から、タンパク質のアミノ酸配列に影響をあたえるようなSNPを選択することが実用上好ましい。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0060

【補正方法】変更

【補正の内容】

【0060】

本発明の選択方法では、

a) 当該ヒトゲノム情報中のSNP群を母集団として、その中でTag SNP候補となる各々のSNPの遺伝子座から一定範囲に定められた近傍に存在するSNPをTarget SNPとして、当該Tag SNP候補とこれらのTarget SNPとの間の相互情報量の和を算出する。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】0061

【補正方法】変更

【補正の内容】

【0061】

相互情報量とは、2つの確率変数 x と y が確率分布 $p(x)$ と $p(y)$ に従い、かつ x と y の同時確率が $p(x, y)$ に従うときに下記の式で定義される量である。

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0063

【補正方法】変更

【補正の内容】

【0063】

本発明においては、 x, y はそれぞれ2つの異なるSNPの遺伝子型であり、 $p(x)$ と $p(y)$ はその頻度に対応する。 $p(x, y)$ はそれらの遺伝子型が2つのSNPで同時に観察される頻度である。この定義に従って「Tag SNP候補とTarget SNPの相互情報量」を算出することができる。言い換えれば、相互情報量を算出する前提として、各々のTag SNP候補の遺伝子型の頻度の他に、各々のTag SNP候補の遺伝子座から一定範囲に定められた近傍内に存在する各々のTarget SNPの遺伝子型が同時に観察される頻度が算出されていることも必要となる。ただし、SNP群のハプロタイプの特定が行われている場合は、遺伝子型の頻度は遺伝子型を構成する対立遺伝子頻度と置き換え、2つのSNPで同時に遺伝子型が観察される頻度はハプロタイプの頻度と置き換えればよい。

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】0065

【補正方法】変更

【補正の内容】

【0065】

そして、b) 全Tag SNP候補の中から、前記相互情報量の総和の値が大きいTag SNP候補を、上記のインピュテーションするための手段として用いる核酸プローブ中に存在させるTag SNPとして、当該総和の大きい順に選択することで、本発明の選択方法を行うことができる。

【手続補正8】

【補正対象書類名】明細書

【補正対象項目名】0076

【補正方法】変更

【補正の内容】

【0076】

ステップS2では、前記第一の機能により読み出された(a)～(d)の情報に基づき、個々のTag SNP候補毎に対応するTarget SNPとの間の相互情報量の和を計算し、これらの中で当該和をスコアとする機能が記述されている。相互情報量とは、前述した内容の数値計算により算出される情報概念であり、算出の前提として、各々のTag SNP候補の遺伝子型の頻度の他に、各々のTag SNP候補の遺伝子座から一定範囲に定められた近傍内に存在するTarget SNP各々における、当該Tag SNP候補とTarget SNP候補の遺伝子型の組み合わせの頻度が算出されていることも必要であり、これらの頻度計算はこのステップS2において行われることが好適である。

【手続補正9】

【補正対象書類名】明細書

【補正対象項目名】0079

【補正方法】変更

【補正の内容】

【0079】

ステップS2～4は、スコアの計算を行うか否かの判断を行うステップであることを示している。Tag SNP候補「i」と、それと組で精査されるTarget SNP「j」の組み合わせにおいて、「 $L[i, j] \leq L_0$ 」とは、Tag SNP候補「i」とTarget SNP「j」とのゲノム上の距離(bp)である「 $L[i, j]$ 」が特定値以下であることを示している。すなわち「 L_0 」とは、Tag SNP候補の遺伝子座から一定範囲に定められた近傍内の距離を示している。当該距離については上述した通りである。また、「 $R[i, j] \geq R_0$ 」とは、Tag SNP候補「i」とTarget SNP「j」との間の r^2 連鎖不平衡値が閾値「 R_0 以上」であることを示している。当該閾値についても上述した通りである。T[j]は、精査済みのTarget SNP「j」が1つ以上のTag SNP候補で既にカバーされている場合は1、カバーされていない場合には0を示すものである。すなわちT[j]=0であれば、選択されたTarget SNP「j」は、組となるTag SNP候補「i」によってカバーされていないことを示している。このステップ2～4は、その判断ボックス内の条件について「Yes」であれば、次のステップS2～5に進み、「No」であれば、再びステップS2～3(1)に戻る判断がなされるステップとして記述されている。

【手続補正10】

【補正対象書類名】明細書

【補正対象項目名】0082

【補正方法】変更

【補正の内容】

【0082】

図1に示すステップS3は、「ステップS2において算出されたスコアが最大のTag S

N P 候補を 1 つ選び出す」ステップである。このステップ S 3 には、本発明のプログラムの第二の機能の後半が記述されており、図 2 に示すステップ S 3 - 1 、 S 3 - 2 (1) 、 S 3 - 3 、及び、 S 3 - 2 (2) に該当する。ステップ S 3 - 2 (1) / (2) は、一組のループ端である。

【手続補正 1 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項 4

【補正方法】変更

【補正の内容】

【請求項 4】

相互情報量以外の指標が、前記 Tag S N P 候補の遺伝子座から一定範囲に定められた近傍に存在する Target S N P 群との連鎖不平衡値であることを特徴とする、請求項 3 に記載の Tag S N P の選択方法。

【手続補正 1 2】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項 1 3

【補正方法】変更

【補正の内容】

【請求項 1 3】

請求項 1 ~ 1 2 のいずれかに記載の Tag S N P の選択方法に従い選択された Tag S N P に対応する核酸プローブが搭載されていることを特徴とする、DNA マイクロアレイ。

【手続補正 1 3】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項 1 4

【補正方法】変更

【補正の内容】

【請求項 1 4】

下記の工程 (1) 及び (2) を含むことを特徴とする、DNA マイクロアレイの生産方法。

(1) 請求項 1 ~ 1 2 のいずれかに記載の選択方法に従い、Tag S N P を選択する第 1 工程；

(2) 第 1 工程により選択された Tag S N P に基づいて、検体中のヒトゲノム中の当該 Tag S N P の遺伝子型を検出するための核酸プローブを、DNA マイクロアレイに搭載する第 2 工程。

【手続補正 1 4】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項 1 5

【補正方法】変更

【補正の内容】

【請求項 1 5】

複数個人の遺伝子型が特定された S N P 群の情報が含まれるヒトゲノム情報を用いて、ヒトゲノムの S N P 情報をインピュテーションするための手段として用いる Tag S N P に対応する核酸プローブ群を構成するために当該 Tag S N P を選択するコンピュータシステムであって、記録部と演算処理部とを備え；

(A) 当該記録部には、当該ヒトゲノム情報から読み出された Tag S N P 候補の情報、及び、それらの Tag S N P 候補の遺伝子座から一定範囲に定められた近傍に存在する S N P の情報を Target S N P 情報として、

(1) 各々の Tag S N P 候補のヒトゲノム上の遺伝子座、

(2) 個々のヒトゲノム情報における Tag S N P 候補の遺伝子型、

(3) Target S N P のヒトゲノム上の遺伝子座、

(4) 個々のヒトゲノム情報におけるTarget S N P の遺伝子型、
が少なくとも記録されており；

(B) 当該演算処理部は、前記記録部から(A)の(1)～(4)の情報に基づいて個々のTag S N P 候補毎に対応するTarget S N Pとの間の相互情報量の和を計算し、これらの中で当該和が最大のTag S N P 候補を選択して、第一のTag S N Pとして選択を行い；

(C) これまでに選択されたTag S N P と対応するTarget S N P 群の情報が抜去された、前記Tag S N P 情報及びTarget S N P 情報を基にして、再度前記(B)工程により最大の相互情報量の和を伴うTag S N P 候補を選択して、第二のTag S N Pとして選択を行い；

(D) 前記工程(B)、(C)を繰り返して、この繰り返し工程を第M(Mは自然数)のTag S N P の選択のために行い、この自然数Mの値が、定められたインピュテーションするための手段として用いるTag S N P の予定数に達するまで、残りM-2回の当該繰り返し工程を行う；

ことを特徴とする、Tag S N P を選択するコンピュータシステム。

【手続補正15】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項19

【補正方法】変更

【補正の内容】

【請求項19】

相互情報量以外の指標が、前記Tag S N P 候補の遺伝子座から一定範囲に定められた近傍に存在するTarget S N P 群との連鎖不平衡値であることを特徴とする、請求項18に記載のTag S N P を選択するコンピュータシステム。

【手続補正16】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項26

【補正方法】変更

【補正の内容】

【請求項26】

複数個人の遺伝子型が特定されたS N P 群の情報が含まれるヒトゲノム情報を用いて、ヒトゲノムのS N P 情報をインピュテーションするための手段として用いるTag S N P に対応する核酸プローブ群を構成するために当該Tag S N P を選択するコンピュータプログラムであって、コンピュータに、

(A) ヒトゲノム情報から読み出された当該Tag S N P 候補の情報、及び、それらのTag S N P 候補の遺伝子座から一定範囲に定められた近傍に存在するS N P の情報をTarget S N P 情報として、

(1) 各々のTag S N P 候補のヒトゲノム上の遺伝子座、

(2) 個々のヒトゲノム情報におけるTag S N P 候補の遺伝子型、

(3) Target S N P のヒトゲノム上の遺伝子座、

(4) 個々のヒトゲノム情報におけるTarget S N P の遺伝子型、

が記録されている記録部から、これら(1)～(4)の情報を演算処理部における処理のために読み出す、第一の機能；

(B) 前記第一の機能により読み出された(1)～(4)の情報に基づき、個々のTag S N P 候補毎に対応するTarget S N Pとの間の相互情報量の和を計算し、これらの中で当該和が最大のTag S N P 候補を選択して第一のTag S N Pとして選択を行う、第二の機能；

(C) これまでに選択されたTag S N P と対応するTarget S N P 群の情報が抜去された、前記Tag S N P 情報及びTarget S N P 情報を基にして、再度前記第二の機能により、最大の相互情報量の和を伴うTag S N P 候補を選択して、第二のTag S N Pとして選択を行い、以降工程(B)、(C)を繰り返して、この繰り返し工程を第M(Mは自然数)のTag S N P の選択のため残りM-2回行い、この自然数Mの値が、定められたインピュテーションするための手段として用いるTag S N P の予定数に達するまで行う、第三の機能；

を実現させるアルゴリズムが含まれることを特徴とする、コンピュータプログラム。

【手続補正17】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項29

【補正方法】変更

【補正の内容】

【請求項29】

前記第二の機能を実現させるアルゴリズムの前段階において、相互情報量以外の指標によりTarget SNP候補を選択して、前記第二の機能を行う対象となるTarget SNP候補群の予備的な絞り込みを実現させるアルゴリズムが設けられていることを特徴とする、請求項26～28のいずれかに記載のコンピュータプログラム。

【手続補正18】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項30

【補正方法】変更

【補正の内容】

【請求項30】

相互情報量以外の指標が、前記Tag SNP候補の遺伝子座から一定範囲に定められた近傍に存在するTarget SNP群との連鎖不平衡値であることを特徴とする、請求項29に記載のコンピュータプログラム。