

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2009-504153

(P2009-504153A)

(43) 公表日 平成21年2月5日(2009.2.5)

(51) Int.Cl.

C 12 Q 1/68 (2006.01)
C 12 N 15/09 (2006.01)
G 06 F 17/30 (2006.01)

F 1

C 12 Q 1/68
C 12 N 15/00
G 06 F 17/30

Z N A A
A
1 7 O F

テーマコード(参考)

4 B 0 2 4
4 B 0 6 3
5 B 0 7 5

審査請求 未請求 予備審査請求 有 (全 79 頁)

(21) 出願番号 特願2008-525967 (P2008-525967)
(86) (22) 出願日 平成18年8月8日 (2006.8.8)
(85) 翻訳文提出日 平成20年4月14日 (2008.4.14)
(86) 国際出願番号 PCT/SG2006/000224
(87) 国際公開番号 WO2007/021250
(87) 国際公開日 平成19年2月22日 (2007.2.22)
(31) 優先権主張番号 11/202,023
(32) 優先日 平成17年8月12日 (2005.8.12)
(33) 優先権主張国 米国(US)

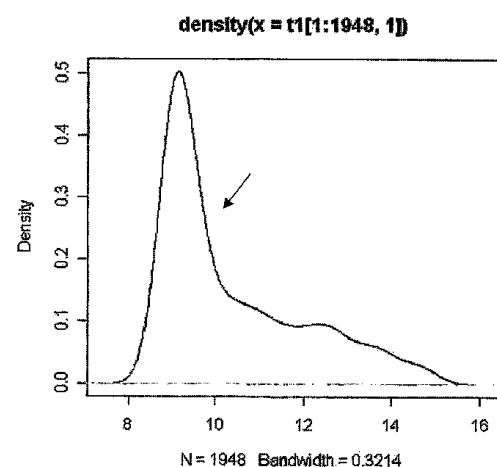
(71) 出願人 504354117
エイジエンシー・フォー・サイエンス、テ
クノロジー・アンド・リサーチ
シンガポール国, 138668, バイオポ
リス・ウェイ 20, #07-01 セン
トロス
(74) 代理人 100099623
弁理士 奥山 尚一
(74) 代理人 100096769
弁理士 有原 幸一
(74) 代理人 100107319
弁理士 松島 鉄男
(74) 代理人 100114591
弁理士 河村 英文

最終頁に続く

(54) 【発明の名称】オリゴヌクレオチド設計および/または核酸検出の方法および/または装置

(57) 【要約】

(I) 増幅される少なくとも1つの標的核酸の領域を同定および/または選択ステップと、ここで、前記領域が、平均増幅効率(AE)よりも高いAEを有し、(II)選択される領域にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドを設計するステップと、任意の順序で含む核酸検出のための少なくとも1つのオリゴヌクレオチドを設計する方法が提供される。(i)少なくとも1つの生体試料を提供するステップと、(ii)生体試料中に含まれる核酸を増幅するステップと、(iii)生体試料中に標的核酸が存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドを提供するステップと、(iv)前記オリゴヌクレオチドを増幅された核酸に接触させ、標的核酸にハイブリダイズしたオリゴヌクレオチドを検出するステップとを含む少なくとも1つの標的核酸を検出する方法もまた、提供される。具体的には、本方法は、少なくとも1つのヒト生体試料中の、少なくとも1つの病原体、例えばウイルスの存在を検出するためのものである。プローブは支持体、



【特許請求の範囲】

【請求項 1】

(I) 増幅される少なくとも 1 つの標的核酸の少なくとも 1 つの領域を同定および / または選択するステップと、ここで、領域が平均増幅効率 (A E) よりも高い A E を有し、

(II) 同定および / または選択された領域にハイブリダイズすることができる少なくとも 1 つのオリゴヌクレオチドを設計するステップと

を任意の順序で含む、核酸検出のための少なくとも 1 つのオリゴヌクレオチドを設計する方法。

【請求項 2】

選択される領域の A E が、フォワードプライマー r_i が位置 i に結合することができ、リバースプライマー r_j が標的核酸の位置 j に結合することができる確率である、増幅効率スコア (AES) として計算され、および | i - j | が、増幅されることが所望される標的核酸の領域である、請求項 1 に記載の方法。 10

【請求項 3】

| i - j | が 1 0 0 0 0 bp 以下である、請求項 2 に記載の方法。

【請求項 4】

| i - j | が 1 0 0 0 bp 以下である、請求項 2 に記載の方法。

【請求項 5】

| i - j | が 5 0 0 bp 以下である、請求項 2 に記載の方法。

【請求項 6】

ステップ (I) が、幾何学的増幅バイアスまたは標的核酸の各位置の効果を判定すること、および平均増幅効率 (A E) よりも高い A E を有する領域として増幅される少なくとも 1 つの領域を選択することを含む、請求項 1 ~ 5 のいずれか一項に記載の方法。 20

【請求項 7】

幾何学的増幅バイアスが PCR バイアスである、請求項 6 に記載の方法。

【請求項 8】

選択される領域にハイブリダイズすることができる少なくとも 1 つのオリゴヌクレオチドが、以下の基準：

(a) 選択されたオリゴヌクレオチドは、40% ~ 60% の CG 含有量を有する；

(b) オリゴヌクレオチドは、隣接モデルに基づいて計算される、最も高い自由エネルギーを有することによって選択される； 30

(c) 標的核酸 v_a および v_b のオリゴヌクレオチド s_a およびオリゴヌクレオチド s_b サブストリングを仮定すると、 s_a は s_b と標的核酸 v_b 由来の任意の長さ m のサブストリング s_b との間のハミング距離ならびに / または s_a およびオリゴヌクレオチド s_b の最も長い共通するサブストリングに基づいて選択される；

(d) 標的核酸 v_a に特異的な長さ m の任意のオリゴヌクレオチド s_a について、標的核酸と異なる核酸のいずれの領域ともいかなるヒットも有さない場合、オリゴヌクレオチド s_a が選択され、長さ m のオリゴヌクレオチド s_a が標的核酸と異なる核酸とヒットを有する場合、最も小さい最大アライメント長および / または最少の数のヒットを有する長さ m のオリゴヌクレオチド s_a が選択される；ならびに

(e) p_i が、増幅された標的核酸の位置 i にハイブリダイズすると予測される場合、標的核酸の位置 i のオリゴヌクレオチド p_i が選択される
の少なくとも 1 つに従って選択および設計される、請求項 1 ~ 7 のいずれか一項に記載の方法。 40

【請求項 9】

$P(p_i | v_a) >$ であり、式中 α が 0.5 であり、 $P(p_i | v_a)$ が、 p_i が標的核酸 v_a の位置 i にハイブリダイズする確率である場合、基準 (e) に基づいて、標的核酸 v_a の位置 I のオリゴヌクレオチド p_i が選択される、請求項 8 に記載の方法。

【請求項 10】

が 0.8 である、請求項 9 に記載の方法。 50

【請求項 1 1】

【数 1】

$$P(p_i | v_a) \approx P(X \leq x_i) = \frac{c_i}{k}$$

であり、式中 X は v_a の全てのオリゴヌクレオチドの増幅効率スコア (AES) 値を表す確率変数であり、k は v_a 中のオリゴヌクレオチドの数であり、 c_i は AES 値が x_i 以下であるオリゴヌクレオチドの数である、請求項 1 ~ 1 0 のいずれか一項に記載の方法。

【請求項 1 2】

選択および / または設計されるオリゴヌクレオチドを調製するステップをさらに含む、請求項 1 ~ 1 1 のいずれか一項に記載の方法。

【請求項 1 3】

(i) 少なくとも 1 つの生体試料を提供するステップと、
 (ii) 生体試料中に含まれる核酸を増幅するステップと、
 (iii) 生体試料中に標的核酸が存在する場合、少なくとも 1 つの標的核酸にハイブリダイズすることができる少なくとも 1 つのオリゴヌクレオチドを提供するステップと、ここで請求項 1 ~ 1 2 のいずれか一項に記載の方法に従って前記オリゴヌクレオチドが設計および / または調製され、

(iv) 前記オリゴヌクレオチドを増幅された核酸と接触させるステップおよび / または標的核酸にハイブリダイズしたオリゴヌクレオチドを検出するステップとを含む、少なくとも 1 つの標的核酸を検出する方法。

【請求項 1 4】

少なくとも 1 つのオリゴヌクレオチドが、少なくとも 1 つのオリゴヌクレオチドプライマーまたはプライマーである、請求項 1 ~ 1 3 のいずれか一項に記載の方法。

【請求項 1 5】

増幅ステップが、少なくとも 1 つのランダムフォワードプライマーおよび / または少なくとも 1 つのリバースランダムプライマーの存在下で行われる、請求項 1 ~ 1 4 のいずれか一項に記載の方法。

【請求項 1 6】

増幅ステップが RT - PCR である、請求項 1 ~ 1 5 のいずれか一項に記載の方法。

【請求項 1 7】

位置 i に結合しているフォワードランダムプライマーおよび標的核酸 v_a の位置 j に結合しているリバースランダムプライマーが、

【数 2】

$$AES_i = \sum_{j=i-Z}^i \left\{ P^f(j) \times \sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) \right\}$$

の標的核酸 v_a の各位置 i についての増幅効率スコア (AES_i) を有するプライマーの間で選択され、

式中、

【数 3】

$$\sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) = Pr(i+1) + Pr(i+2) + \dots + Pr(j+Z),$$

であり、 $P^f(i)$ および $P^r(i)$ は、それぞれ、フォワードプライマーおよびリバース

10

20

30

40

50

プライマーとしてランダムプライマー r_i が v_a の位置 i に結合する確率であり、1000
0 bp 以下である Z は、増幅されることが所望される v_a の領域である、請求項 2 ~ 16
のいずれか一項に記載の方法。

【請求項 18】

増幅ステップが、フォワードおよびリバースプライマーを含み、フォワードおよびリバースプライマーのそれぞれが、5' - 3' 配向で、不变のプライマー頭部および可変のプライマー尾部を含み、少なくとも可変の尾部が、標的核酸 v_a の一部にハイブリダイズする、請求項 1 ~ 17 のいずれか一項に記載の方法。

【請求項 19】

増幅ステップが、配列番号 1 のヌクレオチド配列またはそのバリエントもしくは誘導体を有するフォワードおよび / またはリバースランダムプライマーを含む、請求項 1 ~ 18 のいずれか一項に記載の方法。 10

【請求項 20】

検出される標的核酸が、生体試料の核酸にとって外因性の核酸である、請求項 1 ~ 19 のいずれか一項に記載の方法。

【請求項 21】

検出される標的核酸が、少なくとも、病原体ゲノムまたはその断片である、請求項 1 ~ 20 のいずれか一項に記載の方法。

【請求項 22】

病原体核酸が、少なくとも、ウイルス、寄生生物、もしくは細菌由来の核酸、またはその断片である、請求項 21 に記載の方法。 20

【請求項 23】

生体試料がヒトから得られ、標的核酸が生体試料中に存在する場合、前記標的核酸はヒト由来でない、請求項 1 ~ 22 のいずれか一項に記載の方法。

【請求項 24】

プローブが不溶性支持体上に配置される、請求項 1 ~ 23 のいずれか一項に記載の方法

。

【請求項 25】

不溶性支持体がマイクロアレイである、請求項 24 に記載の方法。

【請求項 26】

検出ステップ ($i v$) において、 v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高く、それによって生体試料中の v_a の存在が示される、請求項 13 ~ 25 のいずれか一項に記載の方法。 30

【請求項 27】

検出ステップ ($i v$) において、 v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有し v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む請求項 13 ~ 25 のいずれか一項に記載の方法であって、プローブ v_a のシグナル強度の密度分布が、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって生体試料中の v_a の存在が示される方法。 40

【請求項 28】

検出ステップ ($i v$) において、生体試料中の少なくとも 1 つの標的核酸の存在が、0.1 以下の t 検定の値および / または 0.05 以下のアンダーソン・ダーリン検定値および / または 1.0 以上の重み付きカルバック・ライブラー情報量の値によって示される、請求項 13 ~ 25 のいずれか一項に記載の方法。

【請求項 29】

t 検定値が 0.05 以下である、請求項 28 に記載の方法。

【請求項 30】

重み付きカルバック・ライブラー情報量の値が 5.0 以上である、請求項 28 または請

10

30

40

50

求項 2 9 に記載の方法。

【請求項 3 1】

検出ステップ(i v)が、重み付きカルバック・ライブラー(W K L)情報量スコア：
【数 4】

$$WKL(P_a | \bar{P}_a) = \sum_{j=0}^{k-1} \frac{Q_a(j) \log(\frac{Q_a(j)}{\bar{Q}_a(j)})}{\sqrt{\bar{Q}_a(j)[1-\bar{Q}_a(j)]}}$$

10

の分布を計算することによって、標的核酸 v_a についての各サインプローブセット(S P S)におけるプローブのシグナル強度を評価することを含み、式中、 $Q_a(j)$ は瓶 b_j に見られる P_a におけるプローブのシグナル強度の累積分布関数であり、

【数 5】

$$Q_a(j)$$

は瓶 b_j に見られる

【数 6】

$$\bar{P}_a$$

20

におけるプローブのシグナル強度の累積分布関数であり、 P_a はウイルス v_a のプローブのセットであり、

【数 7】

$$\bar{P}_a = P \cdot P_a$$

である、請求項 1 3 ~ 2 5 のいずれか一項に記載の方法。

【請求項 3 2】

標的核酸 v_a の非存在を表す各サインプローブセット(S P S)が、正規分布したシグナル強度および / または 5 より小さい重み付きカルバック・ライブラー(W K L)情報量スコアを有する、請求項 3 1 に記載の方法。

【請求項 3 3】

少なくとも 1 つの標的核酸 v_a の存在を表す各サインプローブセット(S P S)が、正に歪んでいるシグナル強度分布および / または 5 より大きい重み付きカルバック・ライブラー(W K L)情報量スコアを有する、請求項 3 2 に記載の方法。

【請求項 3 4】

W K L スコアの分布についてアンダーソン・ダーリン検定を行うことをさらに含み、 $P > 0.05$ の結果が、それによって標的核酸 v_a の非存在を示す、請求項 3 1 ~ 3 3 のいずれか一項に記載の方法。

【請求項 3 5】

W K L スコアの分布についてアンダーソン・ダーリン検定を行うことをさらに含み、 $P < 0.05$ の結果が、それによって、標的核酸 v_a の存在を示す、請求項 3 1 ~ 3 3 のいずれか一項に記載の方法。

【請求項 3 6】

さらなるアンダーソン・ダーリン検定が行われ、それによってさらなる重感染している標的核酸の存在が示される、請求項 3 5 に記載の方法。

【請求項 3 7】

少なくとも 1 つの標的核酸 v_a に対する少なくとも 1 つのオリゴヌクレオチドプローブ

40

30

10

50

のハイブリダイゼーションを検出することを含み、 v_a にハイブリダイズするプローブのシグナル強度の平均値が v_a に含まれないプローブの平均値よりも統計的に高く、それによって v_a の存在が示される、少なくとも1つの標的核酸 v_a の存在を判定する方法。

【請求項38】

v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高く、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含み、プローブ v_a のシグナル強度の密度分布が v_a に含まれないプローブのものより正に歪んでおり、それによって v_a の存在が示される、請求項37に記載の方法。

10

【請求項39】

少なくとも1つの生体試料中の少なくとも1つの標的核酸の存在が、0.1以下のt検定の値および/または0.05以下のアンダーソン・ダーリン検定値および/または1.0以上の重み付きカルバック・ライブラー情報量の値によって示される、請求項37または請求項38に記載の方法。

【請求項40】

t検定値が0.05以下である、請求項39に記載の方法。

【請求項41】

重み付きカルバック・ライブラー情報量の値が5.0以上である、請求項39または請求項40に記載の方法。

20

【請求項42】

(i) 少なくとも1つの生体試料を提供するステップと、
 (ii) 生体試料中に含まれる少なくとも1つの核酸を增幅するステップと、
 (iii) 生体試料中に標的核酸が存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドプローブを提供するステップと、
 (iv) プローブを増幅された核酸と接触させ、標的核酸にハイブリダイズしたプローブを検出するステップと、ここで、 v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高く、それによって生体試料中の v_a の存在が示されるステップと
 を含む、少なくとも1つの標的核酸を検出する方法。

30

【請求項43】

ステップ(iv)において、 v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む請求項42に記載の方法であって、プローブ v_a のシグナル強度の密度分布が、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって生体試料中の v_a の存在が示される、方法。

【請求項44】

ステップ(iv)において、少なくとも1つの生体試料中の少なくとも1つの標的核酸の存在が、0.1以下のt検定の値および/または0.05以下のアンダーソン・ダーリン検定値および/または1.0以上の重み付きカルバック・ライブラー情報量の値によって示される、請求項42または請求項43に記載の方法。

40

【請求項45】

t検定値が0.05以下である、請求項44に記載の方法。

【請求項46】

重み付きカルバック・ライブラー情報量の値が5.0以上である、請求項44または請求項45に記載の方法。

【請求項47】

50

検出される標的核酸が、生体試料の核酸にとって外因性の核酸である、請求項 37～46 のいずれか一項に記載の方法。

【請求項 48】

検出される標的核酸が、少なくとも 1 つの病原体ゲノムまたはその断片である、請求項 37～47 のいずれか一項に記載の方法。

【請求項 49】

病原体核酸が、ウイルス、寄生生物、もしくは細菌由来の少なくとも 1 つの核酸、またはその断片である、請求項 48 に記載の方法。

【請求項 50】

生体試料がヒトから得られ、標的核酸が、生体試料中に存在する場合、ヒトゲノム由来でない、請求項 37～47 のいずれか一項に記載の方法。 10

【請求項 51】

少なくとも 1 つのプローブが、不溶性支持体上に配置される、請求項 37～48 のいずれか一項に記載の方法。

【請求項 52】

支持体がマイクロアレイである、請求項 51 に記載の方法。

【請求項 53】

(I) 増幅される少なくとも 1 つの標的核酸の少なくとも 1 つの領域を同定および / または選択する、ここで、前記領域が平均増幅効率 (A E) よりも高い A E を有する、ならびに 20

(II) 同定および / または選択される領域にハイブリダイズすることができる少なくとも 1 つのオリゴヌクレオチドを設計する

ように構成された、核酸検出のための少なくとも 1 つのオリゴヌクレオチドを設計するための装置。

【請求項 54】

選択される領域の A E が、フォワードプライマー r_i が位置 i に結合することができ、リバースプライマー r_j が標的核酸の位置 j に結合することができる確率である、増幅効率スコア (AES) として計算され、および $|i - j|$ が、増幅されることが所望される標的核酸の領域である、請求項 53 に記載の装置。

【請求項 55】

$|i - j|$ が 1 0 0 0 0 b p 以下、1 0 0 0 b p 以下、または 5 0 0 b p 以下である、請求項 54 に記載の装置。 30

【請求項 56】

ステップ (I) が、幾何学的増幅バイアスまたは標的核酸の各位置の効果を判定すること、および平均増幅効率 (A E) よりも高い A E を有する領域として増幅される少なくとも 1 つの領域を選択することを含む、請求項 53～55 のいずれか一項に記載の装置。

【請求項 57】

幾何学的増幅バイアスが PCR バイアスである、請求項 56 に記載の装置。

【請求項 58】

選択される領域にハイブリダイズすることができるオリゴヌクレオチドが、以下の基準：

(a) 選択されたオリゴヌクレオチドが、40 %～60 % の C G 含有量を有する；
(b) オリゴヌクレオチドが、隣接モデルに基づいて計算される最も高い自由エネルギーを有することによって選択される；

(c) 標的核酸 v_a および v_b のオリゴヌクレオチド s_a およびオリゴヌクレオチド s_b サブストリングを仮定すると、 s_a が s_b と標的核酸 v_b 由来の任意の長さ m のサブストリング s_b との間のハミング距離および / または s_a およびオリゴヌクレオチド s_b の最も長い共通するサブストリングに基づいて選択される；

(d) 標的核酸 v_a に特異的な長さ m の任意のオリゴヌクレオチド s_a について、標的核酸と異なる核酸のいずれの領域ともいかなるヒットも有さない場合、オリゴヌクレオチド

10

20

30

40

50

s_a が選択され、長さ m のオリゴヌクレオチド s_a が標的核酸と異なる核酸とヒットを有する場合、最も小さい最大アライメント長および / または最少の数のヒットを有する長さ m のオリゴヌクレオチド s_a が選択される；ならびに

(e) 標的核酸の位置 i の少なくとも 1 つのオリゴヌクレオチド p_i が、 p_i が、増幅された標的核酸の位置 i にハイブリダイズすると予測される場合に選択されるの少なくとも 1 つに従って選択および / または設計される、請求項 53 ~ 57 のいずれか一項に記載の装置。

【請求項 59】

$P(p_i | v_a) >$ であり、式中 c_i が 0.5 であり、 $P(p_i | v_a)$ が、 p_i が標的核酸 v_a の位置 i にハイブリダイズする確率である場合に、基準 (e) に基づいて、標的核酸 v_a の位置 i のオリゴヌクレオチド p_i が選択される、請求項 58 に記載の装置。 10

【請求項 60】

が 0.8 である、請求項 59 に記載の装置。

【請求項 61】

【数 8】

$$P(p_i | v_a) \approx P(X \leq x_i) = \frac{c_i}{k}$$

であり、式中 X は v_a の全てのオリゴヌクレオチドの増幅効率スコア (AES) 値を表す確率変数であり、 k は v_a 中のオリゴヌクレオチドの数であり、 c_i は AES 値が x_i 以下であるオリゴヌクレオチドの数である、請求項 53 ~ 60 のいずれか一項に記載の装置。 20

【請求項 62】

構成が、選択および / または設計されたオリゴヌクレオチドを調製するステップをさらに含む、請求項 53 ~ 61 のいずれか一項に記載の装置。

【請求項 63】

少なくとも 1 つのオリゴヌクレオチドが、少なくとも 1 つのオリゴヌクレオチドプローブおよび / またはプライマーである、請求項 53 ~ 62 のいずれか一項に記載の装置。

【請求項 64】

(i) 少なくとも 1 つの生体試料を提供するステップと、 30

(ii) 生体試料中に含まれる核酸を増幅するステップと、

(iii) 生体試料中に標的核酸が存在する場合、少なくとも 1 つの標的核酸にハイブリダイズすることができる少なくとも 1 つのオリゴヌクレオチドを提供するステップと、ここで前記オリゴヌクレオチドが、請求項 53 ~ 63 のいずれか一項に記載の装置に従って構成されている装置に従って設計および / または調製され、

(iv) オリゴヌクレオチドを増幅された核酸と接触させる、および / または標的核酸にハイブリダイズしたオリゴヌクレオチドを検出するステップと
のいずれか 1 つのステップを含む、少なくとも 1 つの標的核酸を検出するよう構成された装置。

【請求項 65】

増幅ステップが、少なくとも 1 つのランダムフォワードプライマーおよび / または少なくとも 1 つのリバースランダムプライマーの存在下で行われる、請求項 53 ~ 64 のいずれか一項に記載の装置。 40

【請求項 66】

増幅ステップが RT - PCR である、請求項 53 ~ 65 のいずれか一項に記載の装置。

【請求項 67】

位置 i に結合しているフォワードランダムプライマーおよび標的核酸 v_a の位置 j に結合しているリバースランダムプライマーが、

【数9】

$$AES_i = \sum_{j=i-Z}^i \left\{ P^f(j) \times \sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) \right\}$$

の標的核酸 v_a の各位置 i についての増幅効率スコア (AES_i) を有するプライマーの間で選択され、式中、

【数10】

$$\sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) = Pr(i+1) + Pr(i+2) + \dots + Pr(j+Z)$$

10

であり、 $P^f(i)$ および $P^r(i)$ は、それぞれフォワードプライマーおよびリバースプライマーとして、ランダムプライマー r_i が v_a の位置 i に結合する確率であり、 10000 b p 以下である Z は増幅されることが所望される v_a の領域である、請求項 53～66 のいずれか一項に記載の装置。

【請求項68】

増幅ステップが、フォワードおよびリバースプライマーを含み、フォワードおよびリバースプライマーのそれぞれが、5' - 3' 配向で、不变のプライマー頭部および可変のプライマー尾部を含み、少なくとも可変の尾部は、標的核酸 v_a の一部にハイブリダイズする、請求項 53～67 のいずれか一項に記載の装置。

20

【請求項69】

増幅ステップが、配列番号1のヌクレオチド配列またはそのバリエントもしくは誘導体を有するフォワードおよび / またはリバースランダムプライマーを含む、請求項 53～68 のいずれか一項に記載の装置。

30

【請求項70】

検出される標的核酸が、生体試料の核酸にとって外因性である少なくとも 1 つの核酸である、請求項 53～69 のいずれか一項に記載の装置。

【請求項71】

検出される標的核酸が、少なくとも 1 つの病原体ゲノムまたはその断片である、請求項 53～69 のいずれか一項に記載の装置。

【請求項72】

病原体核酸が、ウイルス、寄生生物、もしくは細菌由来の少なくとも 1 つの核酸、またはその断片である、請求項 71 に記載の装置。

【請求項73】

生体試料がヒトから得られ、標的核酸が生体試料中に存在する場合、前記標的核酸がヒト由来でない、請求項 53～72 のいずれか一項に記載の装置。

40

【請求項74】

少なくとも 1 つのプローブを配置する少なくとも 1 つの不溶性支持体を含む、請求項 53～73 のいずれか一項に記載の装置。

【請求項75】

不溶性支持体がマイクロアレイである、請求項 74 に記載の装置。

【請求項76】

検出ステップ (i v) において、 v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高く、それによって生体試料中の v_a の存在が示される、請求項 64～75 のいずれか一項に記載の装置。

【請求項77】

検出ステップ (i v) において、 v_a にハイブリダイズするプローブのシグナル強度の

50

平均値が、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含むよう構成された請求項 6 4 ~ 7 5 のいずれか一項に記載の装置であって、プローブ v_a のシグナル強度の密度分布が v_a に含まれないプローブのものよりも正に歪んでおり、それによって生体試料中の v_a の存在が示される、装置。

【請求項 7 8】

検出ステップ(i v)において、生体試料中の標的核酸の存在が、0 . 1以下のt検定の値および/または0 . 0 5以下のアンダーソン・ダーリン検定値および/または1 . 0の重み付きカルバック・ライブラー情報量の値によって示される、請求項 6 4 ~ 7 5 のいずれか一項に記載の装置。 10

【請求項 7 9】

t検定値が0 . 0 5以下である、請求項 7 8に記載の装置。

【請求項 8 0】

重み付きカルバック・ライブラー情報量の値が5 . 0以上である、請求項 7 8または請求項 7 9に記載の装置。

【請求項 8 1】

検出ステップ(i v)が、重み付きカルバック・ライブラー(WKL)情報量スコアの分布:

【数 1 1】

$$WKL(P_a | \bar{P}_a) = \sum_{j=0}^{k-1} \frac{Q_a(j) \log(\frac{Q_a(j)}{Q_{\bar{a}}(j)})}{\sqrt{Q_{\bar{a}}(j)[1-Q_{\bar{a}}(j)]}}$$

を計算することによって、標的核酸についての各サインプローブセット(SPS)におけるプローブのシグナル強度を評価することを含み、式中 $Q_a(j)$ は瓶 b_j に見られる P_a におけるプローブのシグナル強度の累積分布関数であり、

【数 1 2】

$$Q_a(j)$$

は瓶 b_j に見られる

【数 1 3】

$$\bar{P}_a$$

におけるプローブのシグナル強度の累積分布関数であり、 P_a はウイルス v_a のプローブのセットであり、 40

【数 1 4】

$$\bar{P}_a = P \cdot P_a$$

である、請求項 6 4 ~ 7 5 のいずれか一項に記載の装置。

【請求項 8 2】

標的核酸の非存在を表す各サインプローブセット(SPS)が、正規分布したシグナル強度および/または5より小さい重み付きカルバック・ライブラー(WKL)情報量スコアを有する、請求項 8 1に記載の装置。

【請求項 8 3】

10

20

30

40

50

少なくとも 1 つの標的核酸の存在を表す各サインプローブセット (S P S) が、正に歪んでるシグナル強度分布および / または 5 より大きい重み付きカルバック・ライブラリー (W K L) 情報量スコアを有する、請求項 8 1 に記載の装置。

【請求項 8 4】

W K L スコアの分布についてアンダーソン・ダーリン検定を行うことをさらに含み、P > 0 . 0 5 の結果が、それによって標的核酸の非存在を示す、請求項 8 1 ~ 8 3 のいずれか一項に記載の装置。

【請求項 8 5】

W K L スコアの分布についてアンダーソン・ダーリン検定を行うことをさらに含み、P < 0 . 0 5 の結果が、それによって標的核酸の存在を示す、請求項 8 1 ~ 8 3 のいずれか一項に記載の装置。 10

【請求項 8 6】

さらなるアンダーソン・ダーリン検定が行われ、それによってさらなる重感染している標的核酸の存在が示される、請求項 8 5 に記載の装置。

【請求項 8 7】

少なくとも 1 つの電子記憶媒体上で構成が記憶される、請求項 5 3 ~ 8 6 のいずれか一項に記載の装置。

【請求項 8 8】

請求項 1 ~ 5 2 のいずれか一項に記載の方法を行うよう構成された、コンピュータプログラム製品。 20

【請求項 8 9】

少なくとも 1 つのオリゴヌクレオチドを設計するためおよび / または少なくとも 1 つの標的核酸を検出するために、W K L 情報量スコアおよび / またはアンダーソン・ダーリン検定を判定するよう構成されたソフトウェアを含む、コンピュータプログラム製品。

【請求項 9 0】

W K L 、アンダーソン・ダーリン検定、オリゴヌクレオチドプローブの設計、オリゴヌクレオチドプライマーの設計、および / または標的核酸の検出が、請求項 1 ~ 5 2 のいずれか一項に記載の方法に従って定義される、請求項 8 9 に記載のコンピュータプログラム製品。

【請求項 9 1】

請求項 1 ~ 5 2 のいずれか一項に記載の方法を行うよう構成されたソフトウェアを含む、取り外し可能な電子記憶媒体。 30

【請求項 9 2】

少なくとも 1 つのオリゴヌクレオチドプローブを設計するため、オリゴヌクレオチドプライマーを設計するため、および / または少なくとも 1 つの標的核酸を検出するために、W K L 情報量スコアおよび / またはアンダーソン・ダーリン検定を判定するよう構成されたソフトウェアを含む、取り外し可能な電子記憶媒体。

【請求項 9 3】

W K L 、アンダーソン・ダーリン検定、オリゴヌクレオチドプローブの設計、オリゴヌクレオチドプライマーの設計、および / または標的核酸の検出が、請求項 1 ~ 5 2 のいずれか一項に従って定義される、請求項 9 2 に記載の取り外し可能な電子記憶媒体。 40

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、オリゴヌクレオチド設計および / または核酸検出の分野に関する。本発明の方法、装置および / または製品は、病原体の検出、例えばウイルスの検出のために使用してもよい。

【背景技術】

【0 0 0 2】

ヒト患者および集団におけるウイルス性および細菌性病原体の正確かつ迅速な検出は、

50

医学的および疫学的に非常に重要である。歴史的に、診断技術は、細胞培養継代および種々の免疫学的アッセイまたは染色手順に依存してきた。感染性疾患因子の正確かつ感度の高い検出は、この分野における進歩の長い歴史にもかかわらず、現在でも依然として困難である。疾患症状から診断までの遅れの問題、およびこれらの手法による微生物の検出限界にもかかわらず、微生物学研究室において、培養および抗体ベースの検出の典型的な方法は、依然として中心的な役割を果たしている。感染のより速い診断によって、例えよう早期の適切な抗菌治療の実施によって、罹患率および死亡率が減少し得る。過去数十年間で、これを達成するための種々の方法が提案されており、PCRおよびマイクロアレイベースの技術を含む核酸検出に基づくものが、最も見込みがあると思われている。具体的には、PCRベースのアッセイが実施されて、より高い検出感度で疑いのある病原体をより迅速に診断することが可能になった。しかしながら、臨床上の実施において、病原因子は、しばしば未確認のままであり、無数の様式で検出を免れる。例えば、いくつかのウイルスは、培養を施行できない。またある時には、患者の試料は、従来の技術による病原体検出には、質が悪過ぎるか、または力値が不十分である場合がある。さらに、PCRベースおよび抗体ベースの手法はともに、PCRプライマー結合部位の変化および抗原連続変異を生じる天然の遺伝子多様化だけのために、疑いのある病原体を認識できない場合がある。

10

【0003】

複数の病原体を並行して検出する能力を有するDNAおよびオリゴヌクレオチドマイクロアレイが記載されてきた(Wang et al., 2002; Uriisman et al., 2005)。しかしながら、未解決の技術的問題によって、臨床設定におけるそれらの日常的な使用が妨げられている。例えば、増幅およびクロスハイブリダイゼーションによるアーチファクトに照らして病原体の「サイン」を比較するための最も情報をえるプローブをどのようにして選択するか?どのレベルの蛍光シグナルおよびサインプローブによって、病原体が検出されるか?最適化された検出アルゴリズムの精度および感度は何か?(Strikebel et al., 2003; Bodrossy and Sessitsch, 2004; Vora et al., 2004)。

20

【0004】

したがって、この技術の分野において、核酸の検出の代替的および改良された方法が必要である。具体的には、病原体の検出のための代替的および/または改良された診断方法が必要とされる。

30

【発明の開示】

【0005】

本発明は、上述の問題を扱い、具体的には、オリゴヌクレオチド設計の方法、装置および/または製品を提供する。具体的には、オリゴヌクレオチドプローブおよび/またはプライマー設計の方法、装置および/または製品が提供される。核酸検出の方法、装置および/または製品もまた提供される。

【0006】

第一の態様によると、本発明は、

(I) 増幅される少なくとも1つの標的核酸の少なくとも1つの領域を同定および/または選択するステップと、ここで、領域が平均増幅効率(AE)よりも高いAEを有し、

40

(II) 選択された領域にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドを設計するステップと

を任意の順序で含む、核酸検出のための少なくとも1つのオリゴヌクレオチドを設計する方法を提供する。

【0007】

少なくとも1つのオリゴヌクレオチドは、少なくとも1つのプローブおよび/またはプライマーであってもよい。具体的には、ステップ(I)において、AEのスコアは、標的核酸またはその少なくとも1つの領域の範囲上の各位置*i*について判定され、次いで、平均AEスコアが得られる。平均よりも高いAEスコアを示す領域を、増幅される標的核酸

50

の領域として選択してもよい。具体的には、選択される領域の A E は、增幅効率スコア(A E S)として計算してもよく、これは、フォワードプライマー r_i が位置 i に結合することができ、リバースプライマー r_j が標的核酸の位置 j に結合できる確率であり、 $|i - j|$ は、增幅されることが所望される標的核酸の領域である。具体的には、領域 $|i - j|$ は 1 0 0 0 0 b p 以下、より具体的には 5 0 0 0 b p 以下、または 1 0 0 0 b p 以下、例えば 5 0 0 b p 以下であってもよい。具体的には、フォワードおよびリバースプライマーは、ランダムプライマーであってもよい。

[0 0 0 8]

別の様態によると、ステップ(Ⅰ)は、標的核酸の各位置についての幾何学的増幅バイアスの効果を判定すること、および平均増幅効率(AE)よりも高いAEを有する領域として増幅される少なくとも1つの領域を選択することを含む。例えば、幾何学的増幅バイアスは、PCRバイアスである。

[0 0 0 9]

ステップ(Ⅰ)で選択される領域にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドを設計するステップ(Ⅱ)は、当分野で公知の任意のオリゴヌクレオチド設計技術に従って行ってもよい。具体的には、選択される領域にハイブリダイズすることができるオリゴヌクレオチドは、以下の基準：

(a) 選択されたオリゴヌクレオチドは、40%～60%のCG含有量を有する。

(b) 隣接モデルに基づいて計算される、最も高い自由エネルギーを有することによってオリゴヌクレオチドが選択される：

(c) 標的核酸 v_a および v_b のオリゴヌクレオチド s_a およびオリゴヌクレオチド s_b サブストリングを仮定すると、 s_a と任意の長さ m のサブストリング s_b との間のハミング距離に基づいて、ならびに / または s_a およびオリゴヌクレオチド s_b の最長の共通するサブストリングに基づいて s_c が選択される：

(d) 標的核酸 v_a に特異的な長さ m の任意のオリゴヌクレオチド s_a について、標的核酸と異なる核酸のいずれの領域ともいかなるヒットも有さない場合、オリゴヌクレオチド s_a が選択され、長さ m のオリゴヌクレオチド s_a が標的核酸と異なる核酸とのヒットを有する場合、最小の最大アライメント長および / または最少の数のヒットを有する長さ m のオリゴヌクレオチド s_a が選択される；ならびに

(e) p_i が、増幅された標的核酸の位置 i にハイブリダイズすると予測される場合、標的核酸の位置 i のオリゴヌクレオチド p_i が選択される。少なくとも 1 つに従って、選択および設計してもよい。

[0 0 1 0]

具体的には、オリゴヌクレオチドは、プローブおよび / またはプライマーであってよい。

[0 0 1 1]

したがって、オリゴヌクレオチドを設計するために、上述の基準の2つ以上を用いてよい。例えば、全ての基準（a）～（e）を適用することによって、オリゴヌクレオチドを設計してもよい。本明細書で言及されないが明らかに当業者の知識の範囲内である他の基準もまた、使用してもよい。

[0 0 1 2]

具体的には、基準 (e) に基づいて、 $P(p_i | v_a) >$ であり、式中 e が 0.5 であり、 $P(p_i | v_a)$ が、 p_i が標的核酸 v_a の位置 i にハイブリダイズする確率である場合、標的核酸 v_a の位置 i のオリゴヌクレオチド p_i が選択される。より具体的には、 e は 0.8 である。

[0 0 1 3]

具体的には、

【数1】

$$P(p_i | v_a) \approx P(X \leq x_i) = \frac{c_i}{k}$$

であり、式中、 X は、 v_a の全てのオリゴヌクレオチドの増幅効率スコア（AES）値を表す確率変数であり、 k は v_a 中のオリゴヌクレオチドの数であり、 c_i は、AES値が x_i 以下であるオリゴヌクレオチドの数である。

【0014】

本発明の別の態様によると、上述のようにオリゴヌクレオチドを設計する方法は、選択および設計されたオリゴヌクレオチドを調製するステップをさらに含む。オリゴヌクレオチドは、少なくとも1つのプローブおよび／またはプライマーであってもよく、当分野で公知の任意の標準的な方法、例えば、化学合成またはフォトリソグラフィーによって調製してもよい。10

【0015】

別の態様によると、本発明は、
 (i) 少なくとも1つの生体試料を提供するステップと、
 (ii) 生体試料中に含まれる核酸を増幅するステップと、
 (iii) 生体試料中に標的核酸が存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドを提供するステップと、ここで本明細書に記載される本発明の任意の態様による方法を用いることによって、前記オリゴヌクレオチドが設計および／または調製され。20

(iv) 前記オリゴヌクレオチドを増幅された核酸と接触させる、および／または標的核酸にハイブリダイズしたオリゴヌクレオチドを検出するステップとを含む、少なくとも1つの標的核酸を検出する方法を提供する。

【0016】

具体的には、オリゴヌクレオチドはプローブである。

【0017】

増幅ステップ(i)は、ランダムプライマーの存在下で行ってもよい。例えば、増幅ステップ(ii)は、少なくとも1つのランダムフォワードプライマー、少なくとも1つのランダムリバースプライマーおよび／または2つより多いランダムプライマーの存在下で行ってもよい。当分野で公知の任意の増幅方法を用いてもよい。例えば、増幅方法は、RT-PCRである。30

【0018】

具体的には、位置*i*に結合しているフォワードランダムプライマーおよび標的核酸 v_a の位置*j*に結合しているリバースランダムプライマーは、標的核酸 v_a の各位置*i*について、

【数2】

$$AES_i = \sum_{j=i-Z}^i \left\{ P^f(j) \times \sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) \right\}$$

の増幅効率スコア（AES_i）を有するプライマーの間から選択され、式中

【数3】

$$\sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) = Pr(i+1) + Pr(i+2) + \dots + Pr(j+Z),$$

40

50

であり、 $P^f(i)$ および $P^r(i)$ は、ランダムプライマー r_i が、それぞれフォワードプライマーおよびリバースプライマーとして、 v_a の位置 i に結合することができる確率であり、1 0 0 0 0 bp 以下である Z は、増幅されることが所望される v_a の領域である。より具体的には、 Z は 5 0 0 0 bp 以下、1 0 0 0 bp 以下、または 5 0 0 bp 以下であってもよい。

【0019】

増幅ステップは、フォワードおよびリバースプライマーを含んでもよく、フォワードおよびリバースプライマーのそれぞれは、5' - 3' 配向で、不变のプライマー頭部および可変のプライマー尾部を含んでもよく、少なくとも可変の尾部は、標的核酸 v_a の一部にハイブリダイズする。具体的には、増幅ステップは、配列番号 1 のスクレオチド配列またはそのバリエントもしくは誘導体を有するフォワードおよび / またはリバースランダムプライマーを含んでもよい。

10

【0020】

生体試料は、哺乳動物から、例えばヒトから採取された任意の試料であってもよい。生体試料は、組織、血清、鼻咽頭洗浄液、唾液、他の任意の体液、血液、尿、便等であってもよい。増幅ステップを行う前に、生体試料を処置して、生体試料中に含まれる核酸を取り除いてもよい。標的核酸は、検出されることが意図される任意の核酸であってもよい。検出される標的核酸は、少なくとも、生体試料にとって外因性の核酸であってもよい。したがって、生体試料がヒト由来である場合、検出される外因性標的核酸（生体試料中に存在する場合）は、ヒト由来でない核酸である。本発明の態様によると、検出される標的核酸は、少なくとも、病原体ゲノムまたはその断片である。病原体核酸は、少なくとも、ウイルス、寄生生物、もしくは細菌の核酸、またはその断片であってもよい。

20

【0021】

したがって、本発明は、少なくとも、存在する場合、生体試料中の標的核酸の検出の方法を提供する。方法は、生体試料中の病原体の存在の検出のための診断方法であってもよい。例えば、生体試料がヒトから得られる場合、標的核酸は、生体試料中に存在する場合、ヒト由来でない。

【0022】

本発明のいずれかの方法に従って設計および / または調製されたオリゴスクレオチドは、溶液中で使用するか、または不溶性支持体上に配置してもよい。例えば、当分野で公知の任意の技術に従って、オリゴスクレオチドプローブを、不溶性支持体上に塗布するか、スポットするか、または印刷してもよい。支持体は、マイクロアレイ、バイオチップ、膜 / 合成表面、固体支持体またはゲルであってもよい。

30

次いで、プローブを、生体試料の核酸と接触させ、存在する場合、標的核酸とプローブがハイブリダイズし、標的核酸の存在が検出される。具体的には、検出ステップ (iv)において、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、統計的に、 v_a に含まれないプローブの平均よりも高く、それによって、生体試料中の v_a の存在が示される。

【0023】

より具体的には、検出ステップに (iv) において、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、統計的に、 v_a に含まれないプローブの平均値よりも高く、方法は、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含み、プローブ v_a のシグナル強度の密度分布は、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって、生体試料中の v_a の存在が示される。

40

【0024】

例えば、検出ステップ (iv) において、そのプローブシグナル強度の密度分布が正規でない場合、すなわち、0.05 以下のアンダーソン・ダーリング検定値および / または 0.1 以下の t 検定の値および / または 1.0 以上、好ましくは 5.0 以上の重み付きカルバック・ライブラー情報量の値によって示すと、より正に歪んでいる場合に、生体試料中

50

の少なくとも 1 つの標的核酸が検出される。具体的には、t 検定値は 0 . 0 5 以下である。

【 0 0 2 5 】

より具体的には、検出ステップ (i v) の方法は、重み付きカルバック・ライラー (W K L) 情報量スコア :

【 数 4 】

$$WKL(P_a | \bar{P}_a) = \sum_{j=0}^{k-1} \frac{Q_a(j) \log(\frac{Q_a(j)}{Q_{\bar{a}}(j)})}{\sqrt{Q_a(j)[1-Q_a(j)]}}$$

10

の分布を計算することによって、標的核酸 v_a についての各病原体特異的サインプローブセット (S P S) におけるプローブのプローブシグナル強度を評価することをさらに含み、

式中、 $Q_a(j)$ は、瓶 b_j に見られる P_a におけるプローブのシグナル強度の累積分布関数であり、

【 数 5 】

$$Q_a(j)$$

20

は、瓶 b_j に見られる

【 数 6 】

$$\bar{P}_a$$

におけるプローブのシグナル強度の累積分布関数であり、

【 数 7 】

$$Q_{\bar{a}}(j)$$

30

は、瓶 b_j に見られる

【 数 8 】

$$\bar{P}_a$$

におけるプローブのシグナル強度の累積分布関数であり、 P_a は、ウイルス v_a のプローブのセットであり、

【 数 9 】

$$\bar{P}_a = P \cdot P_a$$

40

である。

【 0 0 2 6 】

例えば、標的核酸 v_a の非存在を表す各サインプローブセット (S P S) は、正規分布したシグナル強度 (0 . 0 5 以下のアンダーソン・ダーリン検定値によって評価される) および / または 5 より小さい重み付きカルバック・ライラー (W K L) 情報量スコアを有する。少なくとも 1 つの標的核酸 v_a の存在を表す各サインプローブセット (S P S) は、正に歪んでいるシグナル強度分布および / または 5 より大きい重み付きカルバック・ライラー (W K L) 情報量スコアを有する。

50

【0027】

本方法は、WKLスコアの分布に関してアンダーソン・ダーリン検定を行うことをさらに含み、 $P > 0.05$ の結果は、それによって、標的核酸 v_a の非存在を示すか、または、 $P < 0.05$ の結果は、それによって、標的核酸 v_a の存在を示す。また、さらなるアンダーソン・ダーリン検定を行って、それによって、さらなる重感染している標的核酸の存在を示してもよい。別の態様によると、本発明は、少なくとも1つの標的核酸 v_a への、少なくとも1つのオリゴヌクレオチドプローブ（プローブは、当分野で公知の任意の本発明の方法に必ずしも限定されない方法に従って、選択および設計される）のハイブリダイゼーションを検出することを含む、標的核酸 v_a の存在を判定する方法を提供し、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高く、それによって、 v_a の存在が示される。具体的には、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、本方法は、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む方法であって、プローブ v_a のシグナル強度の密度分布は、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって、 v_a の存在が示される方法である。より具体的には、生体試料中の標的核酸の存在は、0.1以下のt検定の値および/または0.05以下のアンダーソン・ダーリン検定値および/または1.0以上、好ましくは5.0以上の、重み付きカルバック・ライブラリー情報量の値によって示される。例えば、t検定値は0.05以下であってもよい。

10

20

20

【0028】

別の態様によると、本発明は、

- (i) 少なくとも1つの生体試料を提供するステップと、
- (ii) 生体試料中に含まれる少なくとも1つの核酸を増幅するステップと、
- (iii) 生体試料中に標的核酸が存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドプローブを提供するステップと、

30

(iv) 前記オリゴヌクレオチドを増幅された核酸と接触させ、標的核酸にハイブリダイズしたオリゴヌクレオチドを検出するステップと、ここで、 v_a にハイブリダイズするオリゴヌクレオチドのシグナル強度の平均値が、 v_a に含まれないオリゴヌクレオチドの平均値よりも統計的に高く、それによって生体試料中の v_a の存在が示されるステップとを含む、少なくとも1つの標的核酸を検出する方法を提供する。

40

【0029】

具体的には、オリゴヌクレオチドは、オリゴヌクレオチドプローブである。

【0030】

ステップ(iii)において、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、本方法は、高いシグナル強度を有する、検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む方法であって、プローブ v_a のシグナル強度の密度分布は、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって、生体試料中の v_a の存在が示される方法である。具体的には、ステップ(iv)において、生体試料中の少なくとも1つの標的核酸の存在は、0.1以下のt検定の値および/または0.05以下のアンダーソン・ダーリン検定値および/または1.0以上、好ましくは5.0以上の重み付きカルバック・ライブラリー情報量の値によって示される。t検定値は、0.05以下であってもよい。検出される核酸は、生体試料の核酸にとって外因性の核酸である。検出される標的核酸は、少なくとも1つの病原体ゲノムまたはその断片であってもよい。病原体核酸は、ウイルス、寄生生物、もしくは細菌由来の少なくとも1つの核酸、またはその断片であってもよい。具体的には、試料がヒトから得られる場合、標的核酸は、生体試料中に存在する

50

場合、ヒトゲノム由来でない。プローブは、不溶性支持体上に配置してもよい。支持体は、マイクロアレイ、バイオチップ、または膜／合成表面であってもよい。

【0031】

本発明は、本発明の方法を行うための装置を含む本発明の装置を提供する。具体的には、装置は、核酸検出および／または増幅のためにオリゴヌクレオチドを設計するためであってもよく、装置は、増幅される少なくとも1つの標的核酸の少なくとも1つの領域を同定および／または選択するよう構成され、ここで前記領域は、平均増幅効率(AE)よりも高いAEを有し、同定および／または選択される領域にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドが設計される。より具体的には、装置は、少なくとも1つの生体試料を提供するステップ；生体試料に含まれる核酸を増幅するステップ；生体試料中に存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドを提供し、オリゴヌクレオチドは、本発明に従って構成された装置に従って、設計および／または調製されるステップ；ならびにオリゴヌクレオチドを増幅された核酸と接触させる、および／または標的核酸にハイブリダイズしたオリゴヌクレオチドを検出するステップのいずれか1つを含む少なくとも1つの標的核酸を検出するよう構成してもよい。

10

【0032】

本発明はまた、本発明の方法を行うために構成された少なくとも1つのコンピュータプログラム製品を提供する。本発明の装置の構成を記憶する少なくとも1つの電子記憶媒体もまた提供される。一態様によると、本発明は、本発明の方法を行うよう構成されたソフトウェアを含む取り外し可能な電子記憶媒体を提供する。具体的には、取り外し可能な電子記憶媒体は、少なくとも1つのオリゴヌクレオチドプローブおよび／もしくはプライマーを設計するため、ならびに／または少なくとも1つの標的核酸を検出するためのWKL情報量スコアおよび／またはアンダーソン・ダーリン検定を判定するよう構成されたソフトウェアを含んでもよい。より具体的には、ソフトウェア構成を含む取り外し可能な電子記憶装置は、本発明に従って定義されるように、WKL、アンダーソン・ダーリン検定、プローブを設計することおよび／または標的核酸を検出することを含んでもよい。したがって、上述のように構成されたソフトウェアもまた提供される。

20

【発明を実施するための最良の形態】

【0033】

本明細書で言及される書誌参照は、便宜上、参考文献のリストの形態で列举し、実施例の最後に加えている。このような書誌参照の内容全体は、参照によって本明細書に組み込まれる。

30

【0034】

本発明は、従来技術の課題を扱い、具体的には、オリゴヌクレオチド設計の少なくとも1つの方法、装置および／または製品を提供する。具体的には、プローブおよび／またはプライマー設計の方法、装置および／または製品が提供される。核酸検出の方法、装置および／または製品もまた提供される。

【0035】

オリゴヌクレオチドハイブリダイゼーションマイクロアレイを病原体の存在を判定するための手段として使用する概念は提案されているが、相当な障害が残っており、したがって、これらのマイクロアレイの日常的な使用が妨げられている(Streibel, H.M., 2003)。これらの障害としては、プローブ設計およびデータ分析が挙げられる(Streibel, H.M., 2003; Bodrossy, L.およびSessitsch, A., 2004; Vora, G.J. et al., 2004)。本発明者らは、細部に至るプローブ選択にもかかわらず、コンピュータ内で設計された最良のプローブは、必ずしも患者試料に十分ハイブリダイズしないことを試験的なマイクロアレイにおいて観察した。本発明者らは、患者材料に一貫して十分ハイブリダイズするプローブを作製するために、最適な設計の予測の判断材料を判定するようなプローブ設計の新規なおび／または改良された方法を開発することが必要であることに気づいた。具体的には、実施

40

50

例の部分に記載されているように、本発明者らは、35のウイルスゲノム全体にタイリングされた重複する40塩基長のプローブを含む、マイクロアレイを作製した。しかしながら、本発明は、この特定の適用、プローブの長さ、および標的核酸の型に限定されない。

【0036】

本発明の特定の態様によると、本発明者らは、どのようにして、支持体、具体的にはマイクロアレイ基盤を標的核酸検出、具体的には病原体検出における実行可能な手段になるように最適化するかを説明している。本発明者らはまた、融解温度、プローブのG C含有量、二次構造、ハミング距離、ヒトゲノムに対する類似性、ランダムP C R増幅効率におけるP C Rプライマータグの効果、および／または配列多型の効果を含むプローブ設計の予測の判断材料を同定した。プローブおよび／またはプライマー設計の方法および基準の開発に、これらの結果は考慮され、および／または組み込まれた。より具体的な態様によると、本発明者らは、病原体であるかどうかわからない標的核酸の存在を正確に予測できる、データ分析アルゴリズムを開発した。例えば、病原体は、ウイルス、細菌および／または寄生生物であってもよいが、これらに限定されない。プローブが理想的に設計されなくて、アルゴリズムを使用してもよい。この検出アルゴリズムは、プローブ設計方法論と組み合わせて予測の信頼度レベルを有意に向上させる（表6および7参照）。

【0037】

特定の態様によると、本発明の方法は、あり得る病原体の予測を必要としない場合があるが、ほとんどの公知のヒトウイルス、細菌および／または寄生生物、ならびにいくつかの新しい種を公平に検出することができる場合がある。ゲノムまたはその断片は、生物の染色体中の全遺伝物質として定義される。特定の生物の染色体中の遺伝物質由来のD N Aは、ゲノムD N Aである。ゲノムライブラリーは、生物の全ゲノムを表すランダムに生成された重複するD N A断片の組から作製されたクローンの収集物である。本発明のこの検出基盤の背後にいる論拠は、ウイルス、細菌および／または寄生生物の各種が、それらのゲノムの一次配列内に特有の分子特性を含むことである。これらの特徴的な領域の同定によって、個々の種、およびいくつかの場合、個々の系の、特異的検出のための合理的なオリゴヌクレオチドプローブ設計が可能になる。科および属のメンバーの間で最も高度に保存された領域を表すオリゴヌクレオチド（オリゴ）プローブの同時の設計および／または調製によって、いくつかの新しい病原体の検出および部分的な特徴づけが可能になる。さらに、単一の支持体中にこのようなプローブ全てを含むことによって、臨床的試料に同時に重感染している複数のウイルス、細菌および／または寄生生物の検出が可能になる場合がある。支持体は、不溶性支持体、具体的には固体支持体であってもよい（例えば、マイクロアレイまたはバイオチップアッセイ）。

【0038】

特定の態様によると、本発明は、オリゴヌクレオチドプローブが設計される様式、ならびに／またはマイクロアレイによって生成されるデータがどのように解釈および分析されるかに依存して、診断手段として使用してもよい。

【0039】

<増幅効率の判定>

第一の態様によると、本発明は、

(i) 増幅される少なくとも1つの標的核酸の少なくとも1つの領域を同定および／または選択するステップと、ここで、前記領域が平均増幅効率(A E)よりも高いA Eを有し、

(ii) 同定および／または選択された領域にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドプローブを設計するステップと
を任意の順序で含む、核酸検出のためのオリゴヌクレオチドプローブを設計する方法を提供する。

【0040】

具体的には、ステップ(i)において、A Eのスコアは、標的核酸の範囲またはその領域の各位置iについて判定され、平均A Eが得られる。増幅される標的核酸の領域として

10

20

30

40

50

、平均よりも高い A E を示す領域が選択される。具体的には、選択される領域の A E は、フォワードプライマー r_i が位置 i に結合することができ、リバースプライマー r_j が標的核酸の位置 j に結合することができる確率である、増幅効率スコア(A E)として計算してもよく、および $|i - j|$ は、増幅されることが所望される標的核酸の領域である。具体的には、領域 $|i - j|$ は、1 0 0 0 0 b p 以下、より具体的には 5 0 0 0 b p 以下、または 1 0 0 0 b p 以下、例えば 5 0 0 b p 以下であってもよい。具体的には、フォワードおよび / またはリバースプライマーはランダムプライマーであってもよい。

【 0 0 4 1 】

別の態様によると、増幅される標的核酸の領域の同定および / または選択のステップ(i)は、標的核酸の各位置についての幾何学的増幅バイアスの効果を判定すること、および平均増幅効率(A E)よりも高い A E を有する領域として増幅される領域を選択することを含む。幾何学的増幅バイアスは、核酸のいくつかの領域が、他の領域よりも効率的に増幅される可能性として定義してもよい。例えば、幾何学的増幅バイアスは、 P C R バイアスである。

10

【 0 0 4 2 】

< 増幅効率のモデリング >

患者試料内にどの標的核酸(例えば病原体) が存在するかは未知なので、増幅ステップおよび / または逆転写(R T) プロセスの間にランダムプライマーを用いて、存在する全ての R N A の、 D N A への公平な逆転写を確実にしてもよい。本発明の目的のために、当分野で公知の任意のランダム増幅方法を用いてもよい。本明細書において、ランダム増幅方法は、 R T - P C R であってもよい。

20

【 0 0 4 3 】

しかしながら、本発明の方法が R T - P C R に限定されないことが、当業者に明らかになる。具体的には、 R T - P C R 手法は、プライマー - 二量体結合によって引き起こされるシグナルの不正確さおよび R T - P C R プロセスにおける乏しい増幅効率の影響を受けやすい場合がある(Bustein, S. A et al., 2004)。この障害を克服するために、本発明者らは、ランダムプライマーを用いることによって、 R T - P C R プロセスをモデリングした。

【 0 0 4 4 】

本発明の特定の態様によると、増幅ステップは、フォワードおよびリバースプライマーを含み、フォワードおよびリバースプライマーのそれぞれは、 5' - 3' 配向で、不变のプライマー頭部および可変のプライマー尾部を含み、少なくとも可変の尾部は、標的核酸 v_a の一部にハイブリダイズする。固定されたプライマー頭部のサイズおよび可変のプライマー尾部のものは、本発明の方法の目的に適した、塩基長で任意のサイズであってもよい。固定された頭部は、 1 0 ~ 3 0 塩基長、具体的には 1 5 ~ 2 5 塩基長、例えば 1 7 塩基長であってもよい。可変の尾部は、 1 ~ 2 0 塩基長、具体的には 5 ~ 1 5 塩基長、例えば 9 塩基長であってもよい。これらのフォワードおよびリバースプライマーの例を、図 1 に示す。より具体的には、増幅ステップは、ヌクレオチド配列 5' - G T T T C C C A G T C A C G A T A N N N N N N N N N - 3' (配列番号 1) を有するフォワードおよび / またはリバースランダムプライマーを含んでもよく、 N は、 A 、 T 、 C 、および G またはその誘導体のいずれか 1 つである。

30

【 0 0 4 5 】

特定の実施形態によると、図 1 でも例示されているが、本発明者らは、以下のようにランダム R T - P C R プロセスをモデリングした。 v_a を、試料中の実際のウイルスとする。 R T - P C R プロセスにおいて使用したランダムプライマーは、好ましくは、形態(5' - G T T T C C C A G T C A C G A T A N N N N N N N N - 3') (配列番号 1 、および具体的には配列番号 2 ~ 7) の、固定された 1 7 塩基長の頭部および可変の 9 塩基長の尾部を有する 2 6 塩基長のプライマーであった。しかしながら、本発明のプライマーが配列番号 1 ~ 7 の配列および図 1 に限定されないことは、当業者に明らかである。実際、プライマーの、特にその頭部および可変の尾部のヌクレオチドサイズは、上述の範囲内で変

40

50

化および選択してもよい。 v_a の位置 i と j の間の領域における R T - P C R 産物を得るために、本発明者らは、(1) 位置 i に結合しているフォワードプライマー、(2) $|i-j| \leq 10000$ 、および(3) 位置 j に結合しているリバースプライマーを要した。具体的には、 $|i-j|$ は、増幅されることが所望される標的核酸の領域であり、500 b p 以下、より具体的には 1000 以下、例えば 500 b p 以下であってもよい。R T - P C R 産物の質は、フォワードプライマーおよび/またはリバースプライマーが v_a にどれだけよく結合するかに依存する。いくつかのランダムプライマーは、他のものよりもよく v_a に結合することができる。このようなプライマー、およびそれらがどこで v_a に結合するかの同定によって、 v_a の特定の領域がどの程度増幅されそうかの目安が提供される。この手法を用いて、 v_a の各位置についての増幅効率スコア (AES) を計算する増幅効率モデルが提供される。

10

【0046】

標的核酸 v_a の特定の位置 i について、 $P^f(i)$ および $P^r(i)$ は、ランダムプライマー r_i が、それぞれフォワードプライマーおよびリバースプライマーとして v_a の位置 i に結合することができる確率である。簡単にするために、ランダムプライマーの最後の 9 ヌクレオチドが v_a のリバース相補物のサブストリング (フォワードプライマー) または v_a のサブストリング (リバースプライマー) である場合に、ランダムプライマーが v_a にだけ結合することができると仮定する。これを図 1 に示す。確立したプライマー設計基準 (Wu, D. Y. et al., 1991) に基づいて、 r_i が有意なプライマー - 二量体を形成するか極度の融解温度を有する場合に、 $P^f(i)$ は低いと推定した。他方では、 r_i がいかなる有意なプライマー - 二量体を形成せず、最適な融解温度を有する場合、 $P^f(i)$ は高くなる。ランダムプライマーの頭部が v_a に類似している場合に、これによって結合が助けられ、より高い $P^f(i)$ を生じる場合があることに留意されたい。同様に、 $P^r(i)$ を計算した。

20

【0047】

フォワードプライマーとしての v_a の位置 i でのランダムプライマー r_i の結合は、位置 i の上流の少なくとも 10000 ヌクレオチドについて、R T - P C R 産物の質に影響を及ぼす。同様に、リバースプライマーとしての v_a の位置 i でのランダムプライマー r_i の結合は、位置 i の下流の少なくとも 10000 ヌクレオチドについて、R T - P C R 産物の質に影響を及ぼす。したがって、増幅効率スコア、 v_a の各位置 i についての AES _{i} は、これを増幅する全てのフォワードおよびリバースプライマー対の複合効果を考慮することによって計算することができ、

30

【数10】

$$AES_i = \sum_{j=i-Z}^i \left\{ P^f(j) \times \sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) \right\}$$

であり、式中、

【数11】

40

$$\sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) = Pr(i+1) + Pr(i+2) + \dots + Pr(j+Z)$$

であり、 $P^f(i)$ および $P^r(i)$ は、ランダムプライマー r_i が、それぞれフォワードプライマーおよびリバースプライマーとして v_a の位置 i に結合することができる確率であり、10000 b p 以下である Z は、増幅されることが所望される v_a の領域である。

【0048】

したがって、 Z は、10000 b p 以下、5000 b p 以下、1000 b p 以下または

50

500 bp 以下であってもよい。ウイルスの異なる領域によって提示されるシグナル強度の変化がそれらの対応する增幅効率スコアと直接相関を有するかどうかを検証するために、ヒトに影響を及ぼす一般的な病原体、ヒト呼吸器多核体ウイルス B (RSV - B) について、いくつかのマイクロアレイ実験（具体的な場合において、合計 5 つのマイクロアレイ実験）を行った。

【 0 0 4 9 】

＜增幅効率についての RT - PCR のモデリング＞

【 0 0 5 0 】

本発明者らのモデルにおいて、 v_a は、臨床試料中の病原体を示す。 v_a の位置 i および j によって定義されるゲノム中の任意の領域において、例えば $500 \sim 1000$ bp の少なくとも 1 つの PCR 産物を生成することは、 $500 = |i - j| = 1000$ 、および具体的には $500 = |i - j| = 100$ であるように、位置 I に結合しているフォワードプライマーおよびアンチセンス方向で位置 j に結合しているリバースプライマーを必要とする。プライマーの結合親和性は、少なくとも 2 つの要因、(1) プライマーニ量体形成、および(2) ウィルス v_a に対するプライマーのハイブリダイゼーション親和性によって判定される。 10000 ヌクレオチド、具体的には 1000 または 500 ヌクレオチド内で理想的なプライマー結合配置を有することによってうまく増幅される能够性があるゲノム領域は、 v_a の各位置の増幅効率スコア (AES) を計算することによって予測することができる (図 1)。

[0 0 5 1]

＜增幅効率スコア（AES）＞

v_a の各位置 i について、 $P^f(i)$ および $P^r(i)$ を、ランダムプライマー r_i が、それぞれフォワードプライマーおよびリバースプライマーとして v_a の位置 i に結合することができる確率とする。簡単にするために、本発明者らは、プライマーのランダムな尾部のスクレオチド（例えば、図 1 に示されるようなランダムプライマーの最後の 9 ヌクレオチド）が v_a のリバース相補物のサブストリング（フォワードプライマー）または v_a のサブストリング（リバースプライマー、図 1）である場合に、ランダムプライマーが v_a にのみ結合できると仮定する。確立したプライマー設計基準（Wu and Uguzzoli, 1991）に基づいて、本発明者らは、 r_i が有意なプライマー - 二量体を形成するか、極端な融解温度を有する場合に、 $P^f(i)$ が低いと推定した。他方では、 r_i が、いかなる有意なプライマー - 二量体も形成せず、最適な融解温度を有した場合、 $P^f(i)$ は高い。ランダムプライマーの固定されたオリゴヌクレオチドタグ（頭部）（例えば、図 1 に示されるような固定された 17 塩基長のタグ）は、 v_a と類似しており、これはまた、結合を助け、より高い $P^f(i)$ を生じる場合がある。同様に、 $P^f(i)$ を計算した。

〔 0 0 5 2 〕

フォワードプライマーとしての v_a の位置 i でのランダムプライマー r_i の結合は、位置Iの上流のヌクレオチド（例えば、位置 i の上流の500~1000ヌクレオチド）につ

いての R T - P C R 産物の質に影響を及ぼす。同様に、リバースプライマーとしての、 v_a の位置 i でのランダムプライマー r_i の結合は、位置 I の下流のヌクレオチド（例えば、位置 i の下流の 500 ~ 1000 ヌクレオチドについて）についての R T - P C R 産物の質および被覆に影響を及ぼす。 v_a の位置 x を仮定する。位置 i および j にある全ての効果的なプライマー対は、それぞれ x での R T - P C R 産物の質に寄与する。 $i = x = j$ および $i - j = 10000$ であることに留意されたい。例えば、500 ~ 1000 塩基対長である場合、本発明者らの R T - P C R 産物から $500 = i - j = 1000$ である。したがって、 v_a の各位置 x についての増幅効率スコア、AES_x は、これを増幅する全てのプライマー対の複合効果を考慮することによって計算することができる。

【数 1 2】

10

$$AES_x = \sum_{j=x-1000}^x \left\{ P^f(j) \times \sum_{k=\max(x+1, j+500)}^{j+1000} P^r(k) \right\}$$

【0053】

<成功する R T - P C R を予測する AES 閾値>

ウイルス v_a のためのプローブ選択のための増幅効率スコアの閾値を、AES 値 v_a の累積分布関数によって判定する。 X を、 v_a の全てのプローブの AES 値を表す確率変数とする。 k を、 v_a におけるプローブの数とする。このとき、本発明者らは、AES 値が x 以下である確率を

20

【数 1 3】

$$P(X \leq x) = c/k$$

と表し、式中、 c は、 x 以下の AES 値を有するプローブの数である。 v_a の位置 i のプローブ p_i について、 x_i を、その対応する AES 値とする。プローブのシグナル強度はその AES 値と高度に相関するので、本発明者らは、 v_a の存在下で p_i が高いシグナル強度を有する確率である $P(p_i | v_a)$ を、 $P(X = x_i)$ と推定した。したがって、

【数 1 4】

30

$$P(p_i | v_a) \approx P(X \leq x_i)$$

$$= \frac{c_i}{k}$$

であり、式中、 c_i は AES 値が x_i 以下である x_i と等しいプローブの数である。

プローブ選択のために、 $P(p_i | v_a) = 0.8$ である場合にプローブ p_i を選択する。本発明者らの実験において、本発明者らは、期待されるプローブの 50 % より多くが、異なる臨床試料に再現性良くハイブリダイズすることを観察した。より高い AES（例えば、上位 10 % AES）を有するプローブを用いることによって、再現性が向上するが、これによって、種のレベルで 10 未満までいくつかのゲノムに残っている特有のプローブの数が減少し、結果として、アレイが病原体を特異的に同定する能力が損なわれる。したがって、上位 20 % AES を用いた。

40

【0054】

<病原体検出マイクロアレイ上のクロスハイブリダイゼーション閾値の経験的判定>

<プローブ設計>

選択される領域にハイブリダイズすることができるオリゴヌクレオチドプローブの設計のステップ (i) は、当分野で公知のプローブ設計技術のいずれか 1 つに選択してもよい。以下の説明はプローブ設計に関するが、プライマーの設計、具体的には R T - P C R

50

のためのプライマーの設計にも、同じ原理が適用されることが当業者に明らかになる。

【0055】

例えば、各 $v_i \in V$ についての標的核酸（例えば、ウイルスゲノム） $V = \{v_1, v_2, \dots, v_n\}$ のセットを仮定すると、以下の条件を満たす長さ m のプローブのセット (v_i のサブストリングである) を、例えば以下の、

(a) 均質性、感度および特異性の確立されたプローブ設計基準 (Sung, W. K. et al., 2003, C S B)

(b) ヒトゲノムゲノムに対する有意な配列類似性はない

(c) 本明細書に記載されているように、例えば RT - PCR によって AE スコアを用いて効率的に増幅される

10

の少なくとも 1 つを考慮して設計してもよい。

【0056】

クロスハイブリダイゼーションによるアーチファクトによって引き起こされるノイズの多い信号は、具体的には核酸の複合混合物中に存在する稀な病原体配列の同定のためのマイクロアレイデータの解釈への主な障害を提示する。例えば、臨床材料において、宿主組織に由来するもの等の混入している核酸配列は、配列相補性のある閾値を超える病原体特異的マイクロアレイプローブとクロスハイブリダイズする。これは、誤った結論につながる偽陽性シグナルを生じる場合がある。同様に、病原体配列は、その特異的プローブに結合することに加えて、他の非標的プローブ（すなわち、他の病原体を検出するよう設計されている）とクロスハイブリダイズする場合がある。この後者の現象は、問題があるよう見えるがこのようなクロスハイブリダイゼーションを正確に予測することができる程度まで病原体同定のための有用な情報を提供することができた。アニーリング可能性および配列特異性を評価するための種々の評価指標で、マイクロアレイプローブは、昔から、（公知の標的への）最大の特異的ハイブリダイゼーションと（非特異的配列への）最小のクロスハイブリダイゼーションを確実にするよう設計されてきた。しかしながら、実際には、本発明者らは、多くのプローブは、最適なコンピュータ内でのパラメータを用いて設計されているが、不明な理由のために期待に従って機能しないことを見出した。

20

【0057】

アレイベースの病原体検出の動力学を体系的に調査するために、本発明者らは、Nimblegen アレイ合成技術 (Nuwaysir et al., 2002) を用いてオリゴヌクレオチドアレイを作製した。アレイを設計して、各ゲノムの全長にわたって平均 8 塩基の分解能でタイリングされた 40 塩基長のプローブを用いて 35 までの RNA ウィルスを検出した (53, 555 プローブ、図 6、表 1)。

30

【0058】

【表1A】

表1 病原体検出マイクロアレイ上に示されるゲノムのリスト。(列1) マイクロアレイ上で合成された各ゲノムについてのプローブの数。(列2) プローブ設計フィルターの適用の後に残っている各ゲノムについてのプローブの数。(列3) ゲノムに特有な各ゲノムについての、ヒトとクロスハイブリダイズしないプローブの数。

ゲノム	元々のプローブの数(1)	フィルタリングされたプローブの数(2)	特有のプローブ(3)	NCBI GI番号	参照タイプ	アクセッション番号	説明
1	1948	537	271	9629198	RefSeq	NC_001781.1	ヒト呼吸器多核体ウイルス、全ゲノム
2	1995	550	295	19718363	RefSeq	NC_003461.1	ヒトパラインフルエンザウイルス1株ワシントン/1964年、全ゲノム
3	2002	762	474	19525721	RefSeq	NC_003443.1	ヒトパラインフルエンザウイルス2、全ゲノム
4	1979	701	345	10937870	RefSeq	NC_001796.2	ヒトパラインフルエンザウイルス3、全ゲノム
5	3805	588	444	30468042	Genbank	AY283794.1	SARSコロナウイルスSIN2500、全ゲノム
6	3937	604	356	38018022	RefSeq	NC_005147.1	ヒトコロナウイルスOC43、全ゲノム
7	3495	182	112	12175745	RefSeq	NC_002645.1	ヒトコロナウイルス229E、全ゲノム
8	1705	292	177	46852132	RefSeq	NC_004148.2	ヒトメタニューモウイルス、全ゲノム
9	296	118	101	8486138	RefSeq	NC_002023.1	インフルエンザAウイルスRNAセグメント1、完全な配列
10	282	69	42	8486136	RefSeq	NC_002022.1	インフルエンザAウイルスRNAセグメント3、完全な配列
10	296	81	54	8486134	RefSeq	NC_002021.1	インフルエンザAウイルスRNAセグメント2、完全な配列
10	110	69	57	8486131	RefSeq	NC_002020.1	インフルエンザAウイルスRNAセグメント8、完全な配列
10	196	71	62	8486129	RefSeq	NC_002019.1	インフルエンザAウイルスRNAセグメント5、完全な配列
10	177	75	59	8486127	RefSeq	NC_002018.1	インフルエンザAウイルスRNAセグメント6、完全な配列
10	225	70	51	8486125	RefSeq	NC_002017.1	インフルエンザAウイルスRNAセグメント4、完全な配列
10	300	105	48	8486164	RefSeq	NC_002204.1	インフルエンザBウイルスRNA-1、完全な配列
10	293	113	74	8486148	RefSeq	NC_002205.1	インフルエンザBウイルスRNA-2、完全な配列
10	279	94	59	8486150	RefSeq	NC_002206.1	インフルエンザBウイルスRNA-3、完全な配列

10

20

30

40

【表1B】

10	237	70	53	8486152	RefSeq	NC_002207.1	インフルエンザBウイルスRNA-4、完全な配列
10	232	90	82	8486154	RefSeq	NC_002208.1	インフルエンザBウイルスRNA-5、完全な配列
10	195	64	32	8486156	RefSeq	NC_002209.1	インフルエンザBウイルスRNA-6、完全な配列
10	150	47	37	8486159	RefSeq	NC_002210.1	インフルエンザBウイルスRNA-7、完全な配列
10	136	59	50	8486161	RefSeq	NC_002211.1	インフルエンザBウイルスRNA-8、完全な配列
11	1401	85	54	11528013	RefSeq	NC_001563.2	ウエストナイルウイルス、全ゲノム
12	1389	145	123	9627244	RefSeq	NC_002031.1	黄熱病ウイルス、全ゲノム
13	2335	235	171	13559808	RefSeq	NC_002728.1	ニパーウイルス、全ゲノム
14	1943	244	211	11545722	RefSeq	NC_002617.1	ニューカッスル病ウイルス、全ゲノム
15	1174	208	128	9629357	RefSeq	NC_001802.1	ヒト免疫不全ウイルス1、全ゲノム
16	409	134	106	21326584	RefSeq	NC_003977.1	B型肝炎ウイルス、全ゲノム
17	1011	169	135	9627257	RefSeq	NC_001576.1	ヒトパピローマウイルス10型、全ゲノム
18	1036	325	299	10445391	RefSeq	NC_002554.1	口蹄疫ウイルスC、全ゲノム
19	1246	211	209	9790308	RefSeq	NC_001545.1	風疹ウイルス、全ゲノム
20	955	309	172	9626732	RefSeq	NC_001489.1	A型肝炎ウイルス、全ゲノム
21	834	103	29	38371716	RefSeq	NC_005222.1	ハンタンウイルス、全ゲノム
22	837	188	98	38371727	RefSeq	NC_005217.1	シンノンブレウイルス、全ゲノム
23	430	100	86	23334588	RefSeq	NC_004294.1	リンパ球性脈絡髄膜炎ウイルスセグメントS、完全な配列
23	853	455	286	23334585	RefSeq	NC_004291.1	リンパ球性脈絡髄膜炎ウイルスセグメントL、完全な配列
24	1404	204	122	9626460	RefSeq	NC_001437.1	日本脳炎ウイルス、ゲノム
25	1370	284	91	51850386	DNA Database of Japan	AB189128.1	デング熱ウイルス3型 ゲノムRNA、全ゲノム、株: 98902890 DF DV-3
26	1361	130	57	12659201	Genbank	AF326573.1	デング熱ウイルス4型 株814669、全ゲノム

10

20

30

40

【表1C】

27	1370	142	21	19744844	Genbank	AF489932.1	デング熱ウイルス2型 株BR64022、 全ゲノム
28	1370	152	52	323660	Genbank	M87512.1	DENT1SEQ デング熱 ウイルス1型 全ゲノム
29	944	175	87	9626436	RefSeq	NC_001430.1	ヒトエンテロウイルス D、全ゲノム
30	945	183	122	9626433	RefSeq	NC_001428.1	ヒトエンテロウイルス C、全ゲノム
31	946	196	148	9627719	RefSeq	NC_001612.1	ヒトエンテロウイルス A、全ゲノム
32	945	364	154	21363125	RefSeq	NC_003986.1	ヒトエンテロウイルス 1、全ゲノム
33	944	94	12	9626677	RefSeq	NC_001472.1	ヒトエンテロウイルス B、全ゲノム
34	913	283	190	9627730	RefSeq	NC_001617.1	ヒトライノウイルス 89、全ゲノム
35	920	426	291	9626735	RefSeq	NC_001490.1	ヒトライノウイルスB、 全ゲノム

10

20

【0059】

各ウイルスプローブについての7つの複製、ならびにアレイ合成およびハイブリダイゼーションのための対照配列とともに(下記のように)、アレイは、合計390,482のプローブを含んだ。

【0060】

<均質性、感度および特異性>

均質性は、類似した融解温度を有するプローブの選択を要する。低いC G含有量を有するプローブが、信頼できるハイブリダイゼーションシグナル強度を生じないこと、および高いC G含有量を有するプローブが非特異的結合を通して高いシグナル強度を生じる性向を有するがわかった。したがって、選択されたプローブのC G含有量が40%~60%であるべきであることが確立できた。

30

【0061】

したがって、本発明は、40%~60%のC G含有量を有するプローブを選択することを含む核酸検出のためのオリゴヌクレオチドプローブを設計する方法を提供する。

【0062】

用語「ハイブリダイゼーション」は、オリゴプローブが標的核酸またはその一部に非共有結合して、安定した二本鎖を形成するプロセスをいう。三重ハイブリダイゼーションもまた、理論的に可能である。

【0063】

ハイブリダイゼーションプローブは、標的核酸の相補鎖に塩基特異的に結合することができるオリゴヌクレオチドである。特異的にハイブリダイズすることは、分子が、DNAまたはRNAの複合混合物(例えば、全細胞)中に配列が存在する場合に、ストリンジエントな条件下で、実質的に、または唯一、特定のヌクレオチド配列に、結合、二重化、またはハイブリダイズすることをいう。ハイブリダイゼーション、例えば対立遺伝子特異的プローブハイブリダイゼーションは、一般に、ストリンジエントな条件下で行われる。例えば、塩濃度が約1モル以下(M)、温度が少なくとも25°C、例えば、750mM NaCl、50mM リン酸Na、5mM EDTA、pH 7.4(5×SSPE)および約25°C~約30°Cの温度である条件。ハイブリダイゼーションは、通常、ストリンジエントな条件下で、例えば1M以下の塩濃度および少なくとも25°Cの温度で行われる。ストリンジエントな条件下については、例えば、その全体が上述の全ての目的のために参照

40

50

によって本明細書に組み込まれる、Sambrook and Russell, Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York (2001) も参照されたい。

【0064】

感度は、少量のmRNAを検出するために、有意な二次構造を形成することができないプローブが選択されることを要する。したがって、隣接モデルに基づいて計算される最も高い自由エネルギーを有するプローブが選択される (Santa Lucia, J., Jr., et al., 1996)。

【0065】

したがって、本発明は、核酸検出のための少なくとも1つのオリゴヌクレオチドプローブを設計する方法を提供し、プローブは、隣接モデルに基づいて計算される最も高い自由エネルギーを有することによって選択される。

【0066】

特異性は、ウイルスゲノムに最も特有のプローブの選択を要する。これは、他の非標的核酸（例えば、ウイルスゲノム）とのプローブのクロスハイブリダイゼーションを最小限にすることである。標的核酸 v_a および v_b のプローブ s_a およびプローブ s_b サブストリングを仮定すると、 s_a が、 s_a と標的核酸 v_b 由来の任意の長さ m のサブストリング s_b との間のハミング距離ならびに／または s_a およびプローブ s_b の最も長い共通するサブストリングに基づいて選択される。具体的には、 s_a および s_b を、それぞれウイルスゲノム v_a および v_b 由来の長さ m のサブストリングとし、ここで($v_a - v_b$)である。

【0067】

設計されるプローブの長さは、本発明の目的のために有用な任意の長さであってよい。プローブは、100塩基長未満、例えば20～80塩基長、25～60塩基長、例えば40塩基長であってもよい。ハミング距離および／または最も長い共通するサブストリングもまた変化してもよい。

【0068】

Kaneの基準によると(Kane, M. D. et al., 2000)、 s_a は、(a) s_a とウイルスゲノム v_b 由来の任意の長さ m のサブストリング s_b との間のハミング距離が0.25mより多く、

(b) s_a および s_b の最も長い共通するサブストリングが15未満である場合、 v_a に特異的である。

【0069】

ハミング距離についてのカットオフ値は、所望されるストリンジェンシーに従って選択してもよい。所望される特定のストリンジェンシーに従ってハミング距離カットオフをどのようにして選択するかは当業者に明らかになる。本明細書に記載されるプローブ設計の特定の実施例によると、本発明者らは、特定のプローブについての他の標的核酸に関して10より大きく、および保存されたプローブについては10より小さい、好ましくは5より小さいハミング距離カットオフを使用した。特定のプローブでは、特定の標的核酸にのみハイブリダイズするプローブが示されるが、保存されたプローブでは、標的核酸のファミリーの任意のメンバーにハイブリダイズすることができるプローブが示される。

【0070】

したがって、本発明はまた、核酸検出のためのオリゴヌクレオチドプローブを設計する方法を提供し、生体試料中に含まれる標的核酸 v_a および v_b のプローブ s_a および s_b サブストリングを仮定すると、 s_a は、 s_a と標的核酸 v_b 由来の任意の長さ m のサブストリング s_b との間のハミング距離が0.25mより多く、 s_a およびプローブ s_b の最も長い共通するサブストリングが15未満である場合に選択される。

【0071】

ヒトRNA由来のクロスハイブリダイゼーションの複雑性なしにアレイハイブリダイゼーション動力学を研究するために、SARSコロナウイルスおよびデング熱血清型1ウイ

ルス R N A を感染細胞系の培地から精製し、逆転写し、ウイルス特異的プライマーを用いて P C R 増幅した (W o n g e t a l . , 2 0 0 4)。各ゲノム c D N A を、(配列決定によって確認されるように)その全体を増幅し、C y 3 で標識し、マイクロアレイ上で別々にハイブリダイズさせた。S A R S 試料は、検出閾値を十分超えた蛍光 (C y 3) シグナルを提示している (平均アレイシグナル強度よりも標準偏差の 2 倍を超える強度のプローブシグナル強度によって判定される; 図 7 A) 全 3 , 8 0 5 の S A R S 特異的プローブで、S A R S タイリングプローブに十分ハイブリダイズした。コロナウイルス科の他のメンバーならびにピコナウイルス科およびパラミクソウイルス科のいくつかの種についてのみ観察すると、他の病原体プローブセットとのクロスハイブリダイゼーションは最小であり、S A R S が他の公知のウイルスと配列相同性をほとんど共有していないという観察と一致した (K s i a z e k e t a l . , 2 0 0 3)。他方では、1型デング熱のハイブリダイゼーションパターンは、より複雑であった (図 7 B)。まず、本発明者らは、配列多型のために、1型デング熱プローブセットへのハイブリダイゼーションが部分的に不完全である (すなわち、シグナルのない領域) ことを観察した。アレイ上でハイブリダイズされた1型デング熱試料を、1944年のハワイの分離株 (A T C C カタログ番号 V R - 1 2 5 4) から培養したが、アレイプローブセットは、1990年にシンガポールで単離された S 2 7 5 / 9 0 株の配列 (F u e t a l . , 1 9 9 2) に基づく。c D N A 標識とハイブリダイズできなかった1型デング熱プローブは、それぞれ、標的配列との少なくとも 3 つのミスマッチ (15 塩基の範囲内に) を含んだ。第二に、本発明者らは、アレイ上に存在するほとんど全てのウイルスプローブセット、具体的には他のフラビウイルス科メンバーのプローブと、ある程度のクロスハイブリダイゼーションが起こったことを観察し、4型デング熱血清型が 6 0 ~ 7 0 % の相同性を共有していることと一致した。ハイブリダイゼーションシグナル出力とアニーリング特異性との間の関係を理解するために、本発明者らは、まず、類似性の 2 つの尺度: プローブハミング距離 (H D) および最大連続マッチ (M C M) を用いて、全てのプローブ配列を各ウイルスゲノムと比較した。H D は、類似した配列については低いスコアで、2つの配列の全体の類似性距離を評価する (H a m m i n g , 1 9 5 0)。M C M は、類似した配列については高いスコアで、正確な適合である連続した塩基の数を評価する (K a n e e t a l . , 2 0 0 0)。

【0072】

本発明者らは、ハワイの1型デング熱分離株に関連して、各プローブについて H D および M C M スコアを計算し、これらのスコアが、それぞれプローブシグナル強度に対して反比例および直接の相関があることを観察した。ハワイの1型デング熱ゲノムに対して高い類似性を有する、すなわち H D = 2 (n = 9 4 2) または M C M = 2 7 (n = 6 2 7) のアレイ上の全てのプローブは、バックグラウンドより 3 対数分高く、メジアンシグナル強度でハイブリダイズした。プローブの 9 8 % は 0 ~ 4 の低 H D 範囲または 1 8 ~ 4 0 の高い M C M 範囲で検出可能であったが、メジアンプローブシグナル強度は、配列距離の各増分で減少した。メジアンシグナル強度は、それぞれ 4 3 % および 4 6 % の検出可能なプローブで、H D = 7 および M C M = 1 5 でバックグラウンドレベルまで激減した。プローブの大部分 (> 9 6 %, n > 5 1 , 0 0 0) は、8 ~ 2 1 の間の H D スコアおよび / または 0 ~ 1 5 の間の M C M スコアを有し、それぞれその 1 . 2 3 % および 1 . 5 7 % が検出可能であった。

【0073】

理想的な、クロスハイブリダイゼーション類似性閾値は、特定の病原体を同定する全てのプローブが、病原体配列の多型の存在下であっても、常にバックグラウンドノイズよりも高い検出可能なシグナル強度を有するものである。最適な類似性閾値 H D = 4 および M C M = 1 8 で、9 8 % を超えるプローブが、バックグラウンドよりも 2 対数分高いメジアンシグナル強度で検出することができたが、閾値を H D = 5 および M C M = 1 7 まで 1 段階下に調節することによって、わずか約 8 5 % のプローブ検出、およびバックグラウンドよりも約 1 . 2 対数分高いメジアンシグナル強度が生じた (図 8)。

【0074】

10

20

30

40

50

これらの最適なH D およびM C M 閾値を用いてクロスハイブリダイゼーションを予測して、本発明者らは、全てのプローブを、所与の病原体を最も検出しそうなグループに分けた。本発明者らは、これらのグループを、特異的サインプローブセット（S P S）と呼び、アレイ上に示される35の病原体ゲノムのそれぞれについてのS P Sを定義した（表2）。

【0075】

【表2A】

各病原体サインプローブセット（S P S）は、AESが上位20パーセンタイルにある、そのプローブを含む[列(1)]。40～60%の間のGC[列(2)]またはヒトゲノムに対する高い類似性[列(3)]を有さないプローブを除去した。HDおよびMCMに基づいて病原体にクロスハイブリダイズする他の病原体由来のプローブ[列(4)]を、S P Sに加えた[列(5)]。

	病原体	科	全タイリングプローブ	AES (1)	GC含有量 フィルター (2)	ヒトゲノム フィルター (3)	残っている、 フィルタリングされた プローブの 数	予測される、クロ スハイブリダイ ゼーションして いるプローブの 数 (HD≤4およ びMCM≥18) (4)	SPS中の プローブ の数 (5)
1	LCMV	アレナウイルス科	1283	574	1	18	555	0	555
2	ハンタン	ブニヤウイルス科	834	131	6	22	103	2	105
3	シンノン ブレ	ブニヤウイルス科	837	225	8	29	188	3	191
4	229E	コロナウイルス科	3495	196	2	12	182	2	184
5	OC43	コロナウイルス科	3937	663	16	43	604	3	607
6	SARS	コロナウイルス科	3805	672	6	78	588	3	591
7	デング熱 血清型1	フラビウイルス科	1370	201	2	47	152	50	202
8	デング熱 血清型2	フラビウイルス科	1370	178	0	36	142	71	213
9	デング熱 血清型3	フラビウイルス科	1370	336	1	51	284	69	353
10	デング熱 血清型4	フラビウイルス科	1361	172	1	41	130	44	174
11	日本脳炎	フラビウイルス科	1404	274	6	64	204	40	244
12	ウエスト ナイル	フラビウイルス科	1401	111	4	22	85	22	107
13	黄熱病	フラビウイルス科	1389	151	0	6	145	10	155
14	B型肝炎	ヘパドナウイルス科	409	146	2	10	134	0	134
15	インフル エンザA	オルトミクスウイルス科	1582	601	2	46	553	0	553
16	インフル エンザB	オルトミクスウイルス科	1822	718	7	69	642	2	644
17	ヒトバビ ローマウ イルス10 型	パピローマウイルス科	1011	177	1	7	169	0	169
18	hMPV	パラミクソウイルス科	1705	375	23	60	292	8	300
19	ニューカ ッスル病	パラミクソウイルス科	1943	252	0	8	244	0	244
20	ニパー	パラミクソウイルス科	2335	274	22	17	235	0	235
21	パライン フルエン ザ1	パラミクソウイルス科	1995	625	13	62	550	3	553
22	パライン フルエン ザ2	パラミクソウイルス科	2002	838	31	45	762	0	762
23	パライン フルエン ザ3	パラミクソウイルス科	1979	834	29	104	701	9	710

10

20

30

40

【表2B】

24	RSV B	パラミクソウイルス科	1948	655	52	66	537	4	541
25	エコーウイルス1	ピコナウイルス科	945	439	3	72	364	59	423
26	エンテロウイルスA	ピコナウイルス科	946	205	0	9	196	21	217
27	エンテロウイルスB	ピコナウイルス科	944	109	0	15	94	47	141
28	エンテロウイルスC	ピコナウイルス科	945	202	0	19	183	31	214
29	エンテロウイルスD	ピコナウイルス科	944	191	0	16	175	15	190
30	口蹄疫	ピコナウイルス科	1036	356	26	5	325	0	325
31	A型肝炎	ピコナウイルス科	955	355	9	37	309	0	309
32	ライノウイルスA (89型)	ピコナウイルス科	913	333	2	48	283	13	296
33	ライノウイルスB	ピコナウイルス科	920	464	3	35	426	11	437
34	HIV 1	レトロウイルス科	1174	229	4	17	208	0	208
35	風疹	トガウイルス科	1246	748	534	3	211	0	211
		合計	53555				10955		11497

10

20

30

【0076】

各病原体のS P Sは、そのゲノム配列に由来するタイリングプローブ(H D = 0、M C M = 4 0)、ならびに他の病原体に由来するクロスハイブリダイズしているプローブを含んだ(H D = 4、M C M = 1 8)。

【0077】

本発明者らは、次に、本発明者らのS P Sプローブの性能に影響を及ぼす確率のある他の非特異的ハイブリダイゼーション現象を考慮した。例えば、本発明者らは、プローブシグナルと% G C含有量との間の一般的な関係を観察した。以前の観察と一致して、本発明者らは、40%未満のG Cのプローブが、減少したシグナル強度を生じるが、60%を超えるG C含有量のプローブは、より高いシグナル強度を示すことを見出した(W o n g et al . , 2 0 0 4 ; M a s k o s a n d S o u t h e r n , 1 9 9 3)。したがって、本発明者らは、さらなる選択フィルターとして% G C含有量を利用し、それによって、最適なH DおよびM C M値にもかかわらず、40%未満のG Cおよび60%を超えるG Cを有するプローブを本発明者らのS P Sから排除した。

【0078】

<ヒトゲノムに対する配列類似性>

検出される標的核酸がヒトから(例えば、ウイルスゲノムを含むヒト試料)抽出される場合、ヒトゲノムに対して高い相同意を有するプローブもまた避けられるべきである。したがって、標的核酸 v_a に特異的な長さmの任意のプローブ s_a について、標的核酸と異なる核酸のいずれの領域ともいかなるヒットも有さない場合、プローブ s_a が選択され、長さmのプローブ s_a が標的核酸と異なる核酸とヒットを有する場合、最も小さい最大アライメント長および/または最少の数のヒットを有する長さmのプローブ s_a が選択される。具体的には、任意の長さmのプローブ s_a について、B L A S Tアルゴリズムで、ヒトゲノムとの s_a のヒットが見られる(A l t s c h u l , S . F . e t a l . , 1 9 9 7)。(W = 1 5)のB L A S T文字列サイズおよび1 0 0の期待値を用いて、全てのヒットを見出した。ヒトゲノムといかなるヒットも有さない、すなわち、 v_a に特異的である場合、 s_a を選択する。しかしながら、 v_a の長さmの全てのサブストリングがヒトゲノ

40

50

ムとヒットを有する場合、最も小さい最大アライメント長および最少の数のヒットを有するものを選択した。

【0079】

さらに、ヒト配列とのクロスハイブリダイゼーションはまた、結果を複雑にする可能性があったので、本発明者らは、15の文字列サイズを用いたBLASTによって(Alt schul et al., 1997)、全てのプローブをヒトゲノム集合(17を構築)(International Human Genome Sequencing Consortium. Initial sequencing and analysis is of the human genome. Nature 409 (6822)、860-921 (2001))と比較した。SPSから100の期待値を有するプローブをさらにフィルタリングした(上記の表2参照)。
10

【0080】

したがって、本発明は、核酸検出のためのオリゴヌクレオチドプローブを設計する方法を提供し、標的核酸 v_a に特異的な長さ m の任意のプローブ s_a について、標的核酸と異なる核酸のいずれの領域ともいかなるヒットも有さない場合、プローブ s_a が選択され、長さ m のプローブ s_a が標的核酸と異なる核酸とヒットを有する場合、最も小さい最大アライメント長および/または最少の数のヒットを有する長さ m のプローブ s_a が選択される。
。

【0081】

さらに、オリゴヌクレオチドプローブの設計はまた、本発明のAESによって行ってよい。具体的には、本発明は、 p_i が増幅された標的核酸の位置 i にハイブリダイズすると予測される場合に標的核酸の位置 i のプローブ p_i が選択される、プローブを選択および/または設計する方法を提供する。
20

【0082】

具体的には、選択される領域にハイブリダイズすることができるオリゴヌクレオチドプローブは、以下の基準:

(a) 選択されたプローブが、40%~60%のCG含有量を有する;

(b) 隣接モデルに基づいて計算される最も高い自由エネルギーを有することによってプローブが選択される;

(c) 標的核酸 v_a および v_b のプローブ s_a およびプローブ s_b サブストリングを仮定すると、 s_a と標的核酸 v_b 由来の任意の長さ m のサブストリング s_b の間のハミング距離および/もしくは s_a およびプローブ s_b の最も長い共通するサブストリングに基づいて、 s_a が選択される;
30

(d) 標的核酸 v_a に特異的な長さ m の任意のプローブ s_a について、標的核酸と異なる核酸のいずれの領域ともいかなるヒットも有さない場合、プローブ s_a が選択され、長さ m のプローブ s_a が標的核酸と異なる核酸とヒットを有する場合、最も小さい最大アライメント長および/もしくは最少の数のヒットを有する長さ m のプローブ s_a が選択される;ならびに/または

(e) p_i が増幅された標的核酸の位置 i にハイブリダイズすると予測される場合、標的核酸の位置 i のプローブ p_i が選択される
40

の少なくとも1つに従って、選択および/または設計してもよい。

【0083】

本発明の特定の態様によると、オリゴヌクレオチドプローブを設計するために、上述の基準の2つ以上を用いてよい。例えば、全ての基準(a)~(e)を適用することによって、プローブを設計してもよい。本明細書で明らかに言及されていないが当業者に明らかな他の基準もまた用いてよい。

【0084】

具体的には、基準(e)に基づいて、 $P(p_i | v_a) > \theta$ であり、式中 θ が 0.5 であり、 $P(p_i | v_a) < \theta$ が、 p_i が標的核酸 v_a の位置 i にハイブリダイズしなければならない確率である場合、標的核酸 v_a の位置 i のプローブ p_i が選択される。より具体的には、
50

は 0 . 8 である。

【 0 0 8 5 】

別の態様によると、本発明は、

【 数 1 5 】

$$P(p_i | v_a) \approx P(X \leq x_i) = \frac{c_i}{k}$$

であり、式中 X は v_a の全てのプローブの増幅効率スコア (AES) 値を表す確率変数であり、 k は v_a におけるプローブの数であり、 c_i は AES 値が x_i 以下であるプローブの数である、上述のような方法を提供する。

10

【 0 0 8 6 】

別の態様によると、AES を用いて、ランダムプライマータグを設計して、ランダム PCR によって試料のランダム増幅を容易にすることもできる（病原体の検出、遺伝子発現の検出、クローン DNA ライブラリーの構築、および当業者がランダム PCR を採用する他の用途等の用途における使用のために）。

20

【 0 0 8 7 】

< 支持体上のオリゴヌクレオチドプローブの合成 >

本発明の別の態様によると、上述のように少なくとも 1 つのオリゴヌクレオチドプローブを選択および / または設計する方法は、選択および / または設計されたプローブを調製するステップをさらに含む。プローブを設計することは、その配列を理解すること、および / または任意の適した手段によって、例えばソフトウェアを使用することによってそれを設計することを含む。プローブを調製するステップは、その物理的調製を含む。プローブは、当分野で公知の任意の標準的な方法に従って調製してもよい。例えば、プローブは、化学的に合成するか、またはクローニングによって調製してもよい。例えば、Sambrook and Russell, 2001 に記載されているように。

20

【 0 0 8 8 】

本発明の任意の実施形態に従って調製された、支持体、例えばマイクロアレイまたはバイオチップもまた提供される。

30

【 0 0 8 9 】

本発明の任意の方法に従って設計および調製されたプローブは、溶液中で使用するか、または不溶性支持体上に配置してもよい。例えば、当分野で公知の任意の技術に従って、不溶性支持体に塗布するか、スポットするか、または印刷してもよい。支持体は、固体支持体またはゲルであってもよい。プローブを塗布された支持体は、マイクロアレイまたはバイオチップであってもよい。

40

【 0 0 9 0 】

より具体的には、本発明は、最初の組織試料から調製された PCR 増幅 cDNA (具体的には、ランダム PCR 増幅 cDNA) からの、病原体、例えばウイルスおよび / または細菌病原体の迅速な検出および同定のためのオリゴマイクロアレイハイブリダイゼーションベースの手法を提供する。

40

【 0 0 9 1 】

以下の説明において、プローブの調製は、マイクロアレイに特に関連して行われる。しかしながら、支持体、ならびにプローブは、本出願の内容全体にわたる任意の記載に従って調製してもよい。具体的には、「アレイ」は、合成または生合成のいずれで調製されてもよい分子の意図的に作られた収集物である。アレイ中の分子は、互いに同一でも異なってもよい。アレイは、種々の形式、例えば、可溶性分子のライブラリー、樹脂ビーズ、シリカチップ、または他の固体支持体に繋ぎ止められた化合物のライブラリーの形をとってもよい。アレイプレートまたはプレートは、ウェルと呼ばれる、液体の通過に耐性があつて領域または空間を形成する物理的障壁によって、各アレイが他のアレイから分離されている複数のアレイを有する物体である。

50

【0092】

<試料調製およびマイクロアレイへのハイブリダイゼーション>

生体試料は、哺乳動物、例えばヒトから採取される、任意の試料であってもよい。生体試料は、血液、体液、唾液、尿、便等であってもよい。生体試料は、増幅ステップを行う前に生体試料中に含まれる核酸を取り除くよう処理してもよい。標的核酸は、検出されることが意図される任意の核酸であってもよい。検出される標的核酸は、少なくとも、生体試料の核酸にとって外因性である核酸であってもよい。したがって、生体試料がヒトに由来する場合、検出される外因性標的核酸（生体試料中に存在する場合）は、ヒト由来でない核酸である。本発明の態様によると、検出される標的核酸は、少なくとも、病原体ゲノムまたはその断片である。病原体核酸は、少なくとも、ウイルス、寄生生物、もしくは細菌由来の核酸、またはその断片であってもよい。

10

【0093】

本発明の態様によると、標的核酸検出分析の方法が提供される。検出されることが所望される生体試料由来の標的核酸は、任意の標的核酸、RNAおよび/またはDNAであってもよい（例えば、mRNAおよび/またはcDNA）。より具体的には、検出される標的核酸は、病原体または非病原体であってもよい。例えば、少なくとも1つのウイルス、少なくとも1つの細菌および/または少なくとも1つの寄生生物のゲノムまたはその断片であってもよい。選択および/または調製されるプローブは、当業者に公知の任意の標準的な技術に従って支持体上に配置、塗布および/または固定されてもよい。支持体は、不溶性支持体、例えば固体支持体であってもよい（具体的には、マイクロアレイおよび/またはバイオチップ）。

20

【0094】

特定の実施例によると、確立されたプロトコルおよび商業的キットを用いて、患者試料、例えば組織、血清、鼻咽頭洗浄液、便からRNAおよびDNAを抽出した。例えば、核酸抽出のためのQiagenキットを用いてもよい。あるいは、フェノール/クロロホルムもまた、DNAおよび/またはRNAの抽出のために用いてもよい。例えばSambrookおよびRussel, 2001に記載されているような、当分野で公知の任意の技術を用いてもよい。Bohlander et al., 1992およびWang et al., 2003によって記載されたプロトコルに基づいて、タグを付けたランダムプライマーを用いてRNAをcDNAに逆転写した。次いで、ランダムPCRによってcDNAを増幅した。マイクロアレイへの試料の断片化、標識およびハイブリダイゼーションは、Wong et al., 2004によって記載されているように行った。

30

【0095】

<マイクロアレイ合成>

実施例部分に記載されている特定の実験によると、本発明者らは、シンガポールにおけるウイルス疾患の最も一般的な原因を代表するいくつかのウイルスゲノムを選択した。Genbankからダウンロードした全ゲノム配列を使用して、ゲノム全体にわたってタイミングされたオーバーラップ5塩基の解像度の40塩基長のプローブを生成した。Nimblegen技術(Nuwaysir, E. F. et al., 2002)を用いて、各ウイルスプローブの7つの複製を、マイクロアレイ上に直接合成した。マイクロアレイ上でプローブをランダムに分配して、ハイブリダイゼーションのアーチファクトの効果を最小限にした。試料の、プローブへの非特異的ハイブリダイゼーションを制御するために、10,000のオリゴヌクレオチドプローブをマイクロアレイ上に設計および合成した。これらの10,000のオリゴヌクレオチドは、ヒトゲノムに対して、または病原性ゲノムに対していかなる配列類似性も有さなかった。これらは40~60%のCG含有量を有するランダムプローブであった。これらのプローブは、バックグラウンドシグナル強度を評価した。陽性対照として、免疫応答において公知または推定される機能を有するヒト遺伝子に対する400のオリゴヌクレオチドプローブをアレイ上で合成した。合計およそ380,000のプローブについて、植物ウイルス、PMMVを陰性対照として含めた。以下の説明において、病原体検出チップ分析(PDCとも呼ばれる)に関連して、本発明が

40

50

、より具体的に説明される。しかしながら、分析（方法）は、この特定の実施形態に限定されず、本出願の内容全体にわたって記載されるような本発明のいくつかの態様を包含する。

【0096】

<標的核酸を検出する方法>

別の態様によると、本発明は、

(i) 生体試料を提供するステップと、

(ii) 生体試料中に含まれる核酸を増幅するステップと、

(iii) 生体試料中に存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドプローブを提供するステップと、ここで、前記プローブが、本明細書に記載される本発明の任意の態様による方法を用いることによって調製され、

(iv) プローブを増幅された核酸と接触させ、少なくとも1つの標的核酸にハイブリダイズしたプローブを検出するステップと

を含む、少なくとも1つの標的核酸を検出する方法を提供する。

【0097】

増幅ステップ(iii)は、ランダムな、部分的にランダムな（すなわち、固定された部分およびランダムな部分を含む）または特異的なプライマーの存在下で行ってもよい。具体的には、増幅ステップ(iv)は、少なくとも1つのランダムプライマーの存在下で行ってもよい。より具体的には、少なくとも1つのランダムフォワードプライマーおよび/または少なくとも1つのランダムリバースプライマーの存在下で。例えば、増幅ステップ(iii)は、2つより多いランダムプライマーの存在下で行ってもよい。当分野で公知の任意の増幅方法を用いてもよい。例えば、増幅方法は、RT-PCRである。

【0098】

具体的には、本発明者らは、増幅効率スコア(AES)に基づいて、標的核酸にハイブリダイズしたプローブを検出する方法を開発した。これは、本明細書で、本発明のアルゴリズムとも呼ばれる。具体的には、

【数16】

$$AES_i = \sum_{j=i-Z}^i \left\{ P^f(j) \times \sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) \right\}$$

の標的核酸 v_a の各位置 i についての増幅効率スコア(AES_i)を有するプライマーの間で、位置 i に結合しているフォワードランダムプライマーおよび標的核酸 v_a の位置 j に結合しているリバースプライマーが選択され、式中、

【数17】

$$\sum_{k=\max(i+1, j+500)}^{j+Z} P^r(k) = Pr(i+1) + Pr(i+2) + \dots + Pr(j+Z)$$

であり、 $P^f(i)$ および $P^r(i)$ は、ランダムプライマー r_i が、それぞれフォワードプライマーおよびリバースプライマーとして v_a の位置 i に結合することができる確率であり、10000bp以下である Z は、増幅されることが所望される v_a の領域である。より具体的には、 Z は5000bp以下、1000bp以下、または500bp以下であってもよい。

【0099】

増幅ステップは、フォワードおよびリバースプライマーを含んでもよく、フォワードおよびリバースプライマーのそれぞれは、5' - 3'配向で、不变のプライマー頭部および

10

20

30

40

50

可変のプライマー尾部を含んでもよく、少なくとも可変の尾部は、標的核酸 v_a の一部にハイブリダイズする。具体的には、増幅ステップは、配列番号 1 ~ 7 のいずれかのヌクレオチド配列、またはそのバリエントもしくは誘導体を有するフォワードおよび / またはリバースランダムプライマーを含んでもよい。

【0100】

生体試料は、哺乳動物、例えばヒトから採取された任意の試料であってもよい。生体試料は、組織、血清、鼻咽頭洗浄液、唾液、任意の他の体液、血液、尿、便等であってもよい。増幅ステップを行う前に、生体試料を処置して、生体試料中に含まれる核酸を取り除いてもよい。標的核酸は、検出されることが意図される任意の核酸であってもよい。検出される標的核酸は、少なくとも、生体試料の核酸にとって外因性である核酸であってもよい。したがって、生体試料がヒト由来である場合、検出される外因性標的核酸（生体試料中に存在する場合）は、ヒト由来でない核酸である。

10

【0101】

本発明の態様によると、検出される標的核酸は、少なくとも、病原体ゲノムまたはその断片である。病原体核酸は、少なくとも、ウイルス、寄生生物、もしくは細菌由来の核酸、またはその断片であってもよい。

【0102】

したがって、本発明は、少なくとも、生体試料中に存在する場合、標的核酸の検出の方法を提供する。方法は、生体試料中の病原体の存在の検出のための診断方法であってもよい。例えば、生体試料がヒトから得られる場合、標的核酸は、生体試料中に存在する場合、ヒト由来でない。

20

【0103】

本発明の任意の方法に従って設計および / または調製されたプローブは、溶液中で用いてもよく、または、不溶性支持体上に配置してもよい。例えば、当分野で公知の任意の技術に従って不溶性支持体上に、塗布するか、スポットするか、または印刷してもよい。プローブを塗布された支持体は、固体支持体またはゲルであってもよい。具体的には、マイクロアレイまたはバイオチップであってもよい。

【0104】

次いで、プローブを生体試料の核酸に接触させ、存在する場合、標的核酸とプローブがハイブリダイズし、標的核酸の存在が検出される。具体的には、検出ステップ（i v）において、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高く、それによって、生体試料中の v_a の存在が示される。

30

【0105】

より具体的には、検出ステップ（i v）において、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、本方法は、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む方法であって、プローブ v_a のシグナル強度の密度分布は、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって、生体試料中の v_a の存在が示される方法である。

40

【0106】

例えば、検出ステップ（i v）において、生体試料中の標的核酸の存在は、0 . 1 以下の t 検定の値および / または 0 . 05 以下のアンダーソン・ダーリング検定値および / または 1 . 0 以上、好ましくは 5 . 0 以上の重み付きカルバック・ライブラー情報量の値によって示される。具体的には、t 検定値は 0 . 05 以下である。

【0107】

別の態様によると、本発明は、標的核酸 v_a へのプローブのハイブリダイゼーションを検出することを含む標的核酸 v_a の存在を判定する方法を提供し、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的

50

に高く、それによって、 v_a の存在が示される。具体的には、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、本方法は、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む方法であって、プローブ v_a のシグナル強度の密度分布は、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって、 v_a の存在が示される方法である。より具体的には、生体試料中の標的核酸の存在は、0.1以下のt検定の値および/または0.05以下のアンダーソン・ダーリン検定値および/または1.0以上、好ましくは5.0以上の重み付きカルバック・ライブラー情報量の値によって示される。例えば、t検定値は、0.05以下であってもよい。

10

【0108】

別の態様によると、本発明は、

- (i) 少なくとも1つの生体試料を提供するステップと、
- (ii) 生体試料中に含まれる核酸を増幅するステップと、
- (iii) 生体試料中に存在する場合、少なくとも1つの標的核酸にハイブリダイズすることができる少なくとも1つのオリゴヌクレオチドプローブを提供するステップと、
- (iv) プローブを増幅された核酸と接触させ、標的核酸にハイブリダイズしたプローブを検出するステップと、ここで、 v_a にハイブリダイズするプローブのシグナル強度の平均値が、 v_a に含まれないプローブの平均値よりも統計的に高く、それによって生体試料中の v_a の存在が示されるステップと

20

を含む、少なくとも1つの標的核酸を検出する方法を提供する。

【0109】

ステップ(iiv)において、 v_a にハイブリダイズするプローブのシグナル強度の平均値は、 v_a に含まれないプローブの平均値よりも統計的に高いことを特徴とし、本方法は、高いシグナル強度を有する検出方法において使用されるプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算するステップをさらに含む方法であって、プローブ v_a のシグナル強度の密度分布は、 v_a に含まれないプローブのものよりも正に歪んでおり、それによって、生体試料中の v_a の存在が示される方法である。具体的には、ステップ(iv)において、生体試料中の標的核酸の存在は、0.1以下のt検定の値および/または0.05以下のアンダーソン・ダーリン検定値および/または1.0以上、好ましくは5.0以上の重み付きカルバック・ライブラー情報量の値によって示される。t検定値は0.05以下であってもよい。検出される核酸は、生体試料の核酸にとって外因性の核酸である。検出される標的核酸は、少なくとも、病原体ゲノムまたはその断片であってもよい。病原体核酸は、少なくとも、ウイルス、寄生生物、もしくは細菌由来の核酸、またはその断片であってもよい。具体的には、試料がヒトから得られる場合、標的核酸は、生体試料中に存在する場合、ヒトゲノム由来でない。プローブは、不溶性支持体上に配置してもよい。支持体は、マイクロアレイまたはバイオチップであってもよい。

30

【0110】

<RSV-Bの鑄型配列を用いた試験>

40

ウイルスの異なる領域によって提示されるシグナル強度の変動がそれらの対応する増幅効率スコアと直接的な相関を有するかどうかを検証するために、ヒトに影響を及ぼす一般的な病原体、ヒト呼吸器多核体ウイルスB(RSV-B)に関して、合計5つのマイクロアレイ実験を行った。

【0111】

次に、NCBIから得られたRSV-Bの鑄型配列(NC_001781)に、上述のようなプローブ設計基準を適用した。これは、各マイクロアレイ上に斑点をつけられた1948のプローブが得られた。RSV-Bについての増幅効率マップもまた、実際の実験の前に計算し、図2に示す。この図は、平均AESよりも高いAESを有するピークを示し、より高い増幅の確率を有するRSV-Bの領域が示される。

50

【 0 1 1 2 】

ヒト呼吸器多核体ウイルスB（RSV-B）を含む5つの試料を用いて、独立したマイクロアレイ実験を行った。1つのこのような実験についての得られたシグナル強度を、図3に示す。

【 0 1 1 3 】

各実験について、1948のプローブのシグナル強度を、減少順で順位付けし、それらの対応するAES値と関連付けた。p値は、平均で 2.2×10^{-16} 未満であるとわかった。これは、RSV-Bの位置iのプローブのシグナル強度とAES_iとの間の相関が全くランダムでないことを示す。さらなる調査によって、5つ全ての実験において一貫して高いシグナル強度を生じる約300のプローブが、90パーセンタイルレベルで増幅効率スコアを有することが明らかになった。10

【 0 1 1 4 】

説明された増幅効率モデルはRSV-Bゲノムに関して功を奏してきたが、本発明の方法が他のウイルスゲノムにも拡張することができることを示すことが望まれた。ヒトメタニュームウイルス（HMPV）に関して、別のマイクロアレイ実験を行った。今回は、マイクロアレイ上に1705のプローブがあった。ここでも、HMPVについての増幅効率マップを計算した。この実験において、シグナル強度と増幅効率スコアとの間の相関試験によって、 1.335×10^{-9} のp値が示された。

【 0 1 1 5 】

したがって、本発明の増幅効率モデルは、説明された実験設定において、ウイルスゲノムの異なる領域によって生成されたシグナルの相対的強度を予測することができる。低い増幅効率スコアを有する領域由来のプローブは、シグナル強度を生じないか、または低いシグナル強度を生じる傾向が高い。これは、マイクロアレイ上で偽陰性を生じる。このようなプローブは、マイクロアレイデータの分析を複雑にし、これは、低いシグナル強度を有するプローブはその標的ゲノムが存在しないか、または単に増幅されなかつたためである場合があるので、これがさらに複雑になる。そのようなものとして、適度に高い増幅効率スコアを有する領域におけるプローブを選択して、ランダムプライマーを用いたRT-PCRプロセスによって引き起こされる不正確性を最小限にするべきである。20

【 0 1 1 6 】

ウイルスv_aのプローブ選択のための増幅効率スコアの閾値は、AES値v_aの累積分布関数によって判定される。Xを、全てのプローブv_aのAES値を表す確率変数とする。kを、v_aにおけるプローブの数とする。次いで、本発明者らは、AES値がx以下である確率を、 $P(X \leq x) = c / k$ とし、式中cは、x以下であるAES値を有するプローブの数である。v_aの位置iのプローブp_iについて、x_iを、その対応するAES値とする。プローブのシグナル強度はそのAES値と高度に相関があるので、本発明者らは、v_aの存在下でp_iが高いシグナル強度を有する確率であるP(p_i | v_a)を、P(x _i | x_i)であると推定した。したがって、30

【 数 1 8 】

$$P(p_i | v_a) \approx P(X \leq x_i)$$

$$= \frac{c_i}{k}$$

であり、式中c_iは、AES値がx_i以下であるプローブの数である。

【 0 1 1 7 】

プローブ選択のために、 $P(p_i | v_a) > c/k$ の場合に、プローブp_iが選択される。本実験において、c/k = 0.8として設定される。

【 0 1 1 8 】

したがって、本発明はまた、プローブ設計および/または標的核酸検出の方法を提供し50

、 $P(p_i | v_a) >$ の場合、標的核酸 v_a の位置 i のプローブ p_i が選択され、式中、
は 0.75 であり、 $P(p_i | v_a)$ は、 p_i が v_a の存在下で高いシグナル強度を有する確率である。より具体的には、

【数19】

$$P(p_i | v_a) \approx P(X \leq x_i) = \frac{c_i}{k}$$

であり、式中 X は、 v_a の全てのプローブの增幅効率スコア(AES)値を表す確率変数であり、 k は v_a におけるプローブの数であり、 c_i は、 x_i 以下であるAES値を有するプローブの数である。

10

【0119】

<標的核酸検出分析>

以下の説明において、病原体検出チップ分析(PDCとも呼ばれる)に関連して、本発明をより具体的に説明する。しかしながら、分析(方法)は、この特定の実施形態に限定されず、本出願の内容全体にわたって記載されるような、本発明のいくつかの態様を包含する。したがって、具体的には、ウイルスゲノムのセット $V = \{v_1, v_2, \dots, v_n\}$ のために設計される長さ m のプローブのセット $P = \{p_1, p_2, \dots, p_m\}$ を有するPDC仮定すると、病原体検出チップ分析の問題は、チップデータに基づいて試料中に存在するウイルスを検出することである。本明細書で、チップデータは、PDC上のプローブシグナルによって提供される集合的な情報をいう。したがって、チップデータ $D = \{d_1, d_2, \dots, d_x\}$ は、PDC上のプローブセット P の対応するシグナルのセットである。

20

【0120】

試料を考慮すると、存在する場合、試料中に何の病原体が存在するか、いくつの異なる病原体が存在するのかは、全く不明である。しかしながら、ウイルス v_a が実際に試料中にある場合、 v_a のプローブのシグナル強度は、他のウイルス由来のプローブのシグナル強度と有意に異なるべきである。特に、他のウイルスと比較して v_a のより高い割合のプローブが、高いシグナル強度を有するべきである。したがって、 v_a におけるプローブのシグナル強度の平均値が、 v_a に含まれないプローブのものよりも統計的に高いべきであると期待される。

30

【0121】

したがって、本発明は、 v_a にハイブリダイズするプローブのシグナル強度の平均値が v_a に含まれないプローブの平均値よりも統計的に高く、これが生体試料中の v_a の存在を示す方法を提供する。

【0122】

しかしながら、統計的により高い平均値を有することは、試料中に v_a が存在すると結論付けるには、依然として不十分である場合がある。好ましくは、さらなるステップが必要とされる場合がある。本発明者らは、高いシグナル強度を有するPDC上のプローブの割合に対する、高いシグナル強度を有する v_a に含まれないプローブの割合の相対的差異を計算する必要がある。これは、 v_a に含まれるプローブのシグナル強度の分布が、 v_a に含まれないプローブのものよりも正に歪んでいるという観察に基づく(図4Aの矢印参照。比較のために、図4B参照)。

40

【0123】

上述の観察に基づいて、ウイルスの存在に関するチップデータ D を、以下のように分析した。各ウイルス $v_a \in V$ について、本発明者らは、片側 t 検定(Gouleden, C.H., 1956)を用いて、 v_a に含まれるプローブのシグナル強度の平均値が、 v_a に含まれないプローブのシグナル強度のものよりも統計的に高いかどうかを判定した。したがって、 t 統計を計算した：

【数20】

$$t_i = \frac{\mu_a - \mu_{a'}}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_{a'}^2}{n_{a'}}}}$$

式中、 μ_a 、 σ_a^2 および n_a は、それぞれ v_a に含まれるプローブのシグナル強度の、平均値、分散、およびサイズであり、 $\mu_{a'}$ 、 $\sigma_{a'}^2$ および $n_{a'}$ は、それ respective v_a に含まれないプローブのシグナル強度の平均値、分散、およびサイズである。

【0124】

差異の有意性を試験するために、有意性のレベルを0.05に設定した。これは、 v_a に含まれるプローブのシグナル強度の平均値が v_a に含まれないプローブのシグナル強度のものよりも高いという仮説が、 t_a のp値が0.05未満である場合にのみ許容されることを意味する。この場合、 v_a は、試料中に存在する見込みがある。

【0125】

ウイルスのシグナル強度の分布が他のウイルスのものと異なるかどうかを本発明者らが知ることができる t 検定のみでは、特定のウイルスが試料中にあるかどうかを判定するのに十分でない。2つの分布がどれだけ類似または相違しているかを知ることも必須である。真の分布とモデル分布との間の類似性を測定するのに用いることができる尺度は、カルバック・ライブラー情報量(Kullback and Leibler, 1951)(相対エントロピーとしても公知)である。この適用において、 v_a におけるプローブのシグナル強度の確率分布は真の分布であるが、 P における全てのプローブのシグナル強度の確率分布はモデル分布である。 P_a を v_a におけるプローブのセットとする。 P_a および P のシグナル強度の確率分布のカルバック・ライブラー(KL)情報量は、

【数21】

$$KL(P_a \| P) = \sum_{\mu \leq x \leq \max(D)} f_a(x) \log \left(\frac{f_a(x)}{f(x)} \right)$$

であり、式中、 μ は P におけるプローブの平均シグナル強度であり、 $f_a(x)$ はシグナル強度 x の P_a におけるプローブの比率であり、 $f(x)$ はシグナル強度 x の P におけるプローブの比率である。したがって、 $KL(P_a \| P) = 0$ の場合、 P_a の確率分布は、 P のものとちょうど同じということになる。そうでなければ、それらは異なる。

【0126】

試料中に存在するウイルスは、母集団のものよりも高いシグナル強度を有するので、これは、 $KL(P_a \| P) > 0$ の場合に v_a が試料中に存在する確率を有することを含意する。したがって、 $KL(P_a \| P)$ の値が大きいほど、2つの確率分布はより異なり、 v_a が実際に試料中に存在する見込みが高くなる。

【0127】

2つの確率分布の全ての x にわたるカルバック・ライブラー情報量が集合的な差異であることに留意することが重要である。したがって、カルバック・ライブラー情報量は確率分布の変化を見出すのには優れているが、拡散を見出すのには常にそれほど優れているわけではなく、これは確率分布の尾部に、より影響を及ぼす。図4(A、B)に説明されているように、確率分布の尾部は、ウイルスが試料中に存在するかどうかについてのほとんどの情報を提供する。したがって、カルバック・ライブラー情報量統計は、このような観察をより正確に反映するように改良されなければならない。

【0128】

その感度を尾部の外に増大させるために、本発明者らは、カルバック・ライブラー情報量に安定化および重み付きされた統計、アンダーソン・ダーリング統計を導入した(Stephens, M.A. (1974). EDF Statistics for Good

10

20

30

40

50

ness of Fit and Some Comparisons, Journal of the American Statistical Association, Vol 69, pp. 730 - 737)。したがって、重み付きカルバック・ライブラ－情報量(WKL)は、

【数22】

$$WKL(P_a \parallel P) = \sum_{\mu < x \leq \max(D)} \frac{f_a(x) \log \frac{f_a(x)}{f(x)}}{\sqrt{Q(x)[1-Q(x)]}}$$

10

であり、式中、Q(x)は、Pにおけるプローブのシグナル強度の累積分布関数である。

【0129】

経験的試験から、ウイルスがない試料中で、0.05の有意性レベルでt検定を通過するウイルスが5.0未満のWKLを有することが示される。実際にウイルスが存在する試料中で、実際のウイルスは、0.05の有意性レベルでt検定を通過するだけでなく、5.0以上のWKLを有する唯一のウイルスである。したがって、本発明者らは、試料中に存在するウイルスについての重み付きカルバック・ライブラ－情報量閾値を、5.0に設定した。この解析フレームワークを図5に示す。

【0130】

<本発明の方法を行う装置および／または製品>

本発明において提供されるアルゴリズムおよび／または方法を行うことできるソフトウェアをどのようにして構成するかは、当業者に周知である。したがって、本発明はまた、本発明の任意の実施形態によるアルゴリズムおよび／または方法を行うよう構成されたソフトウェアおよび／またはコンピュータプログラム製品も提供する。少なくとも1つの電子記憶媒体もまた提供される。電子記憶媒体は、コンピュータハードドライブ、CD-R ROM、フラッシュメモリー素子(例えば、USBメモリ)、フロッピー(登録商標)ディスク、または当分野の任意の他の電子記憶媒体であってもよい。ソフトウェアは、パソコン、メインフレーム、および任意の計算処理装置上で実行されてもよく、具体的な構成は、当業者に公知である。

【0131】

本発明が例によってのみ説明されてきたこと、ならびに本発明の精神および範囲から逸脱することなく設計における種々の変更を行ってもよいことは、当業者に理解される。

【0132】

ここで、本発明を一般的に説明してきたが、例示によって提供され、本発明を限定することを意図されない、以下の実施例を参照して、より容易にこれが理解される。

【実施例】

【0133】

当分野で公知で、具体的に説明されていない標準的な分子生物学技術は、一般に、Sambrook and Russell, Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York (2001)に記載されているように従った。

【0134】

<マイクロアレイ合成>

本発明者らは、シンガポールにおけるウイルス疾患の最も一般的な原因を代表する35のウイルスゲノムを選択した(上記の表1参照)。

【0135】

NCBI生物分類データベース(<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>)から全ゲノム配列をダウンロードして、ゲノム全体にわたってタイリングされたオーバーラップ平均8塩基の解像度の40塩基長のプローブ配列を生成した。Nimblegen専売特許の技術(

20

30

40

50

Nuwaysir et al. , 2002) を用いて、マイクロアレイ上に直接、各ウイルスプローブの 7 つの複製を合成した。マイクロアレイ上にプローブをランダムに分布させて、ハイブリダイゼーションアーチファクトを最小限にした。プローブへの試料の非特異的ハイブリダイゼーションを制御するため、およびバックグラウンドシグナルを測定するために、マイクロアレイ上に 10,000 のオリゴヌクレオチドプローブを設計および合成した。それらはヒトゲノムまたは病原体ゲノムに対する配列類似性がなく、40 ~ 60 % の G C 含有量を有するランダムプローブである。陽性対照として、免疫応答において公知または推測される機能を有するヒト遺伝子に対する 400 のオリゴヌクレオチドプローブを、アレイ上で合成した。植物ウイルス、PMMV を、合計 390,482 のプローブで陰性対照として含めた。

10

【0136】

< 試料調製、マイクロアレイハイブリダイゼーションおよび染色 >

デング熱細胞系 (ATCC #VR-1254) を ATCC 推奨のように培養し、Sinn850 SARS 細胞系を記載されているように (Vega et al. , 2004) 培養した。臨床材料（鼻咽頭洗浄液）を、インドネシアの小児科集団から得て、RNA ゾル (Leedo Medical Laboratories, Inc., Friendswood, TX) 中、-80 °C で保存した。全ては特定の呼吸器疾患の臨床兆候を示す、肺炎の疑いのある 7 ~ 38 月齢の間の患者であった。製造業者の使用説明書に従って (Smallling et al. , 2002; Tang et al. , 1999) 、RNA ゾルで RNA を抽出した。抽出された RNA を、RNA 保存溶液 (Ambion、米国) に再懸濁し、必要になるまで -80 °C で保存した。Bohlander et al. and Wang et al. (Wang et al. , 2002; Bohlander et al. , 1992) によって記載されたプロトコルに基づいて、タグを付けられたランダムプライマーを用いて RNA を cDNA に逆転写した。次いでこれまでに記載されているように (Wong et al. , 2002) 、cDNA をランダム PCR によって増幅し、断片化し、ビオチン標識で末端標識し、マイクロアレイ上にハイブリダイズさせ、染色した。本発明者らの最初の実験において本発明者らは、プローブ G C 含有量によって、シグナル強度がプローブ G C 含有量に正比例して増加する、シグナル強度測定におけるアーチファクトが生じる可能性があることを見出した。0.82 M の TMAc を Nimblegen の専売特許の TMAc ハイブリダイゼーションバッファーに加えることによって、このアーチファクトが排除された。

20

【0137】

< RSV および hMPV についてのリアルタイム診断用 RT-PCR >

2 μl の精製された患者 RNA、5 U の MuLV 逆転写酵素、8 U の組換え RNase 阻害剤、UNG を含まない 10 μl の 2 X ユニバーサル PCR マスター・ミックス (全て Applied Biosystems から)、0.9 μM のプライマーおよび 0.2 μM のプローブを含有する 20 μl の反応混合物。ABI Prism 7900 HT 配列検出システム (Applied Biosystems) において、リアルタイム RT-PCR 反応を行った。48 °C で 30 分間 RT を行い、続いて DNA ポリメラーゼの活性化のために 95 °C で 10 分間行った。95 °C で 15 秒間および 60 °C で 1 分間の 40 サイクルによって RT 産物の増幅を行った。陰性対照およびプラスミドクローニングの段階希釈 (陽性対照) を各 PCR アッセイに含んだ。増幅の間、各熱サイクルで蛍光発光をモニタリングした。閾値 (CT) は、有意な蛍光が最初に検出されるサイクルを表す。既知の濃度の対照プラスミドを用いることによって、CT 値をコピー数に変換した。RSV については、 2.61×10^9 コピーが 11.897 の CT 値を有したが、hMPV については、7.51 × 10⁹ コピーが 10.51 の CT 値を有した。

30

【0138】

< コロナウイルスおよびライノウイルスについての一段階診断用 RT-PCR >

陽性対照としての使用のために、ヒトコロナウイルス OC43、229E およびライノウイルス 16 の凍結した生きた培養物を、ATCC (Cat# VR-1558, VR-

40

50

740, VR-283)から購入した。製造業者の使用説明書に従ってRNAミニキット(Qiagen, Germany)を用いて、これらの培養物からRNAを抽出した。以下の診断用プライマー対:パンコロナウイルス(Cor-FW, Cor-RV)、OC43(OC43-FW, OC43-RV)、229E(229E-FW, 229E-RV)、ライノウイルス(アンプリマ-1, アンプリマ-2)を用いて、これまでに記載されているように試料を増幅した(Moes et al., 2005; Delfernez et al., 2004)。

【0139】

<病原体マイクロアレイデータの分析>

本発明者らの病原体マイクロアレイは、35のウイルスゲノム $V = \{v_1, v_2, \dots, v_{35}\}$ について異なるプローブハイブリダイゼーションサインに分けられた、40塩基長のプローブのセット $P = \{p_1, p_2, \dots, p_s\}$ を含有する。病原体核酸のハイブリダイゼーションの際に、プローブセット P に対応するプローブシグナル強度データのセット $D = \{d_1, d_2, \dots, d_s\}$ が生じる。

【0140】

<片側T検定>

ウイルス v_a が存在する場合、次いでそのハイブリダイゼーションサインを含むプローブ(v_a に含まれるプローブ)は、t統計(片側T検定)によって判定すると、 v_a に含まれないプローブよりも統計的に高いシグナル強度を有するべきである。

【数23】

10

20

$$t_i = \frac{\mu_a - \mu_{a'}}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_{a'}^2}{n_{a'}}}}$$

式中、 μ_a 、 $\mu_{a'}$ および n_a は、それぞれ v_a に含まれるプローブのシグナル強度の平均値、分散、およびサイズであり、 $\mu_{a'}$ 、 $\mu_{a'}$ および $n_{a'}$ はそれぞれ v_a に含まれないプローブのシグナル強度の平均値、分散、およびサイズである。

【0141】

有意性のレベルを0.05に設定した。これは、 v_a に含まれるプローブのシグナル強度の平均値が v_a に含まれないプローブのシグナル強度のものよりも高いという仮説を、 t_a のp値が0.05未満である場合にのみ許容することを意味する。この場合 v_a は、試料中に存在する見込みがある。しかしながらT検定の検出の方法は、多くの偽陽性信号を生じる。

30

【0142】

<PDA v. 1>

PDA v. 1は、重み付きカルバック・ライブラー検定およびZスコア変換(WKLスコア)で始まり、正規性についてのアンダーソン・ダーリン検定が続く、一連の統計試験を含む。

【0143】

ウイルス v_a を考慮する。 P_a をウイルス v_a のプローブのセットとし、

【数24】

40

$$\overline{P_a} = P \cdot P_a$$

である。 $[r_{low}, r_{high}]$ をシグナル強度範囲とする。本発明者らはこれを、 $j = 0, 1, \dots, c - 1$ についてのc個の瓶

【数25】

$$[r_{low} + j\left(\frac{r_{high} - r_{low}}{c}\right), r_{low} + (j+1)\left(\frac{r_{high} - r_{low}}{c}\right)]$$

に分割した。修正されていないカルバック・ライナー情報量は、

【数26】

$$KL(P_a | \bar{P}_a) = \sum_{j=0}^{c-1} f_a(j) \log\left(\frac{f_a(j)}{\bar{f}_a(j)}\right)$$

10

によって計算してもよく、式中 n_a^j および

【数27】

$$n_a^j$$

は、それぞれ瓶 b_j に含有される P_a におけるプローブおよび

【数28】

20

$$\bar{P}_a$$

におけるプローブの数である。

【数29】

$$f_a(j) = \frac{n_a^j}{\sum_{h=0}^{c-1} n_a^h}$$

30

は、瓶 b_j に見られる P_a におけるプローブの比率であり、

【数30】

$$\bar{f}_a(j) = \frac{n_a^j}{\sum_{h=0}^{c-1} n_a^h}$$

は、瓶 b_j に見られる

【数31】

$$\bar{P}_a$$

40

におけるプローブの比率である。

【0144】

確率分布の尾部のシグナル差異を比較するために、本発明者らは、

【数32】

$$r_{low} = \bar{\mu}_a$$

【数33】

$$\overline{P}_a$$

におけるプローブの平均シグナル強度、および r_{high} = 最大シグナル強度に設定した。本発明者らは、瓶のデフォルト数を $c = 20$ に設定した。

【0145】

確率分布の尾部でのカルバック・ライブラー情報量の感度をさらに安定化および／または増大させるために、2つの修正を行った。まず本発明者らは、アンダーソン・ダーリン型重み付き関数をカルバック・ライブラー情報量に導入した。これによって分布の中央部よりも尾部に、より多くの重み付けが生じた。次いで本発明者らは、それらの確率密度関数の代わりに、2つの対応する累積分布関数に統計を適用した。本発明者らは、本発明者らの改善されたカルバック・ライブラー情報量を重み付きカルバック・ライブラー情報量（WKLスコア）と呼ぶ。

10

【0146】

【数34】

$$WKL(P_a | \overline{P}_a) = \sum_{j=0}^{k-1} \frac{Q_a(j) \log(\frac{Q_a(j)}{Q_a^-(j)})}{\sqrt{Q_a^-(j)[1-Q_a^-(j)]}}$$

20

式中、 $Q_a(j)$ は、瓶 b_j に見られる P_a におけるプローブのシグナル強度の累積分布関数であり；

【数35】

$$Q_a^-(j)$$

は、瓶 b_j に見られる

【数36】

$$\overline{P}_a$$

30

におけるプローブのシグナル強度の累積分布関数である。

【0147】

したがってハイブリダイズした各試料については、本発明者らは V に含まれる各ウイルス v_a の WKL スコアを計算した。次に本発明者らは、 V に含まれる全てのウイルス v_a の WKL スコアの分布が試料中にウイルスが存在しない場合、およそ正規であることを主張した。本発明者らはブートストラッププロセスによって、本発明者らの主張が正しいかどうかを経験的に検証した： n を V におけるウイルスの数とする。 $k = 1, \dots, n$ である、 V に含まれる各ウイルス v_k について本発明者らは、置換によって v_k の「乱された」シグナル強度分布を形成して、実際のデータセット D からランダムに $|v_k|$ プローブシグナル強度を選択した。このような分布は v_k が試料 D 中に存在しない状況を模倣することができる。その後 n 個のウイルスのセットについて、 n 個の WKL スコアが生じる。次に本発明者らは、95%信頼区間での正規性についてのアンダーソン・ダーリン検定によって、 n 個の WKL スコアが正規分布に従うかどうかを調べた。ブートストラップを 100,000 回繰り返した。分布は 99% を超える回で正規であることがわかった。（NB：本発明者らのマイクロアレイ上に示される 35 のウイルスゲノムがあったので、 $n = 35$ ）

40

【0148】

上述の議論に基づいて本発明者らは、以下の帰無仮説および対立仮説を作成することに

50

よって、試料がウイルスを含有するかどうかを試験することができる。

H_0 : WKLスコアの分布は正規であり、すなわちウイルスは試料中に存在しない。

H_1 : WKLスコアの分布は正規でなく、すなわち少なくとも1つのウイルスが試料中に存在する。

<定義>

アンダーソン・ダーリン検定は、

H_0 : データは特定分布に従う。

H_a : データは特定分布に従わない。

検定統計：アンダーソン・ダーリン検定統計は、

$$A^2 = -N - S$$

10

として定義され、

式中、

【数37】

$$S = \sum_{i=1}^N \frac{(2i-1)}{N} [\ln F(Y_i) + \ln (1 - F(Y_{N+1-i}))]$$

であり、Fは特定分布の累積分布関数である。 Y_1 が順序付けられたデータであることに留意されたい。

20

有意性レベル：

臨界領域：アンダーソン・ダーリン検定についての臨界値は、試験されている特定分布に依存する。いくつかの特定分布（正規、対数正規、指數関数、ワイブル、ロジスティック、極値分布1型）についての表集計値および式は、公開されている（Stephens, 1974, 1976, 1977, 1979）。検定は片側検定であり、分布が特定の形態であるという仮説は、検定統計Aが臨界値よりも大きい場合に拒絶される。

として定義される。

【0149】

本発明者らは、WKLスコアの分布に関する正規性についてのアンダーソン・ダーリン検定の適用を進め、95%の信頼区間を有する H_0 を拒絶する。WKLスコアの分布が正規でない場合本発明者らは、次いで範囲外のWKLスコアを有するウイルスを排除し、再びアンダーソン・ダーリン検定を適用する。 H_0 が許容されるまでこのプロセスを繰り返す（重感染している病原体の存在を同定するため）。

30

【0150】

本発明者らは H_0 がバックグラウンドWKL分布として許容される場合に、WKLスコアの分布を表示する。したがって排除されたウイルスは、それらのWKLスコアがバックグラウンドWKL分布に従わないので、試料中に存在する可能性が非常に高い。

【0151】

本発明者らの実験において本発明者らは、ウイルスを含有する試料中で、所与のWKLスコアで非正規分布が偶然によって起こる可能性であるPが非常に低い、すなわち $P < 1.0 \times 10^{-6}$ であることを観察した（WKLスコアのZスコア変換によって得られる）。本発明者らのウイルス検出アルゴリズムについての擬似コードを以下に示す。

40

【0152】

<ウイルス検出アルゴリズム>

ウイルスセットVおよびプローブセットPを有する病原体マイクロアレイデータDを仮定し、

$V_{present} = F$ とし、

Vに含まれる全てのvについて、 D_{WKL} を

【数38】

$$WKL(P_v \| P_{\bar{v}})$$

のセットとし、

1. 正規性についてのアンダーソン・ダーリン検定で D_{WKL} の正規性を判定する。 D_{WKL} が有意性レベル 0.05 の正規分布である場合、 $V_{present}$ を返す。そうでなければステップ 2 に進む。

2. 最も高い $WKL(P_a \parallel P_{\bar{a}})$ を有するウイルス v_a を D_{WKL} から見つける。

$V_{present} = V_{present} \setminus \{v_a\}$; $D_{WKL} = D_{WKL} - \{WKL(P_a \parallel P_{\bar{a}})\}$ とする; ステップ 1 に進む。 10

3. 検出された SPS を除去し、 WKL 分布が正規であることを検証する。

4. 分布が正規でない場合、ステップ 2 に戻って重感染している病原体を見つける。

【0153】

<ゲノム規模での増幅バイアスの予測>

臨床材料中の未知の病原体を同定するためには、プライマー特異的増幅よりもランダムプライマー増幅が好ましい。しかしながら、既知の病原体を同定するためのランダムプライミング増幅を用いた最初の実験において本発明者らは、しばしば配列多型によって説明されない、ゲノム領域に渡る不完全なハイブリダイゼーションを観察した(図 7C)。ゲノム二次構造、プローブ二次構造およびプローブGC含有量もまた、これらの低シグナル強度プローブを説明することができなかった。したがって本発明者らは、不完全なハイブリダイゼーションが、ランダムプライマーが逆転写(RT)ステップでウイルスゲノムに結合する能力の差異から生じるPCRバイアスによるかもしれないという仮説を立てた。本発明者らの実験において使用されるランダムプライマーは、固定された 17 塩基長の配列(5' - G T T T C C C A G T C A C G A T A)(配列番号 1)でタグを付けられたランダムノナマー(3')で構成される 26 塩基長であり(図 1も参照)、ここで固定された 5' タグの目的は、RT 産物の PCR を容易にして、10000 bp 未満の PCR 断片、具体的には 500 ~ 1000 bp の PCR 断片を生じることであった(Pang et al., 2005; Wang et al., 2002; Wang et al., 2003)。この現象を研究するために本発明者らは、アルゴリズム(AES)を設計して実験データを用いて RT-PCR プロセスをモデリングした。成功する RT-PCR はプライマーが鋳型に結合する能力に依存する。鋳型との間の二量体およびヘアピン形成等のプライマー内二次構造形成。タグとノナマーとの間の二量体およびヘアピン形成等のプライマー内二次構造形成、ならびにプローブ融解温度は、結合効率に影響を及ぼすことが公知である(Nguyen and Southern, 2000; Ratushna et al., 2005)。 20

【0154】

ランダムプライマーミックス中のノナマーがウイルスゲノムの配列を完璧に補完すると仮定して、アルゴリズムは 500 ~ 1000 bp の産物が、ゲノム中の可能性のある各開始位置から生じることができる確率を判定する。したがって 1000 塩基のスライディングウインドウの各ヌクレオチドについて、これがうまく増幅される確率は、その増幅効率スコア(AES; 上述の増幅効率スコア参照)に反映される。アルゴリズムを検証するために本発明者らは、ハイブリダイゼーションシグナル強度を RSV ゲノムについての全 1,948 の SPS プローブについて順位付けし、それらをそれらの AES 値と比較した。RSV ゲノム全体で本発明者らは、AES がハイブリダイゼーションシグナル強度と極めて関連していることを観察し(Fisher の直接確率法 $P = 2.2 \times 10^{-16}$)、AES とプローブ検出との間の強い相関が示された(図 12)。メタニュームウイルスについての 1,705 の SPS プローブを用いた別の比較から類似した結果、 $P = 1.3 \times 10^{-9}$ が示された。臨床試料中の SPS プローブ検出の予測における AES の重要性を図 10 に示す。とりわけ本発明者らは、AES のより高い値が特に上位 20% の AES 値において 40

20

30

40

50

て、より大きい割合の検出可能なプローブと相關していたことを観察した。したがってH D、M C M、% G C および配列独自性は、プローブ性能の有益なパラメータであるが、それらはP C Rバイアスを考慮せず、したがってA E Sなしで考えた場合、プローブ性能の不十分な予測の判断材料である。病原体S P Sの選択における第一のフィルターとしての上位20パーセンタイルのA E Sを用いることによって、より高いW K Lスコアおよび偽陽性信号の排除によって明らかなように病原体予測が有意に向上した(表3)。

【0155】

【表3】

平均プローブシグナル強度のみを用いて病原体を検出すること(T検定)によって、多数の偽陽性信号が生じる。最適化されたハイブリダイゼーションサインおよびヒトゲノムにクロスハイブリダイズするプローブの除去(フィルタリングされる)によって、偽陽性信号は減少するが、検出精度のために十分ではない。P D A v. 1は、フィルタリングされていないプローブセット全体を用いて、正確な診断を行うことができる。W K Lスコアが5を超える場合、ウイルスは「検出される」。最適化されたハイブリダイゼーションサインを用いること(フィルタリングされる)によって、W K Lスコアが増大し、診断の信頼度の増大に対応する。ウイルスC T値:ウイルスが検出されたリアルタイムP C Rサイクル(上記参照)。

		PDA v.1を用いた検出			ウイルス CT値	ウイルス コピー数
チップ 番号	病原体	最大WKLスコア (フィルタリング なし)	最大WKLスコア (フィルタリング された)	検出された ウイルスの 数		
32272	純粋なSARS	5.007	5.803	1	--	--
34959	純粋なデング 熱	14.351	20.373	1	--	--
35259	RSV患者324	18.288	20.611	1	21.4366	9.8×10^7
35179	hMPV患者122	1.747	8.439	1	25.5388	50384
35253	RSV患者841	12.056	12.069	1	20.8619	14×10^7
36042	RSV患者412	16.466	17.531	1	23.5804	2.5×10^7
36053	RSV患者483	12.089	12.168	1	24.8340	1.2×10^7
35915	非肺炎患者(陰 性対照)	3.916	4.284	0	0	0

アレイ上でハイブリダイズした全ての患者試料についてのデータを、下記の表4に示す。

【0156】

10

20

30

【表4A】

病原体マイクロアレイ上にハイブリダイズした臨床患者の完全なリスト

アレイ	患者ID	WKL	P値	PDA v.1診断	臨床診断*	最初のPCR診断	PCR CT値	ウイルスコピー数	RT-PCRプライマー
35179	122	8.439 216	1.34×10^{-71}	hMPV	LRTI	hMPV	24.8	5.0×10^4	A1
35887	122	18.31 2077	2.98×10^{-22}	hMPV	LRTI	hMPV	24.8	5.0×10^4	A2
71180	133	17.35 9597	2.42×10^{-37}	hMPV	LRTI	hMPV	25.1 159	4.0×10^4	A2
66691	165	8.567 86	1.84×10^{-4}	hMPV	肺炎	hMPV	27.9	3.9×10^3	A2
70935	254	21.34 8515	8.70×10^{-30}	hMPV	LRTI	hMPV	21.9 518	5.4×10^5	A2
63781	283	16.68 0752	3.97×10^{-12}	hMPV	肺炎	未知			A2
73067	769	24.00 6323	1.34×10^{-51}	hMPV	LRTI	hMPV	25.6 715	2.5×10^4	A2
66690	853			検出されたものはない	肺炎	hMPV	36	0.5	A2
68359	892	12.53 4284	5.66×10^{-5}	ライノウイルス属	肺炎	hMPV	33.8	27	A2
35915	111			検出されたものはない	陰性対照	なし			A1
70927	818			検出されたものはない	陰性対照	なし			A2
66701	312			検出されたものはない	肺炎	RSVA	33.7	44	A2
71006	321			検出されたものはない	肺炎	RSVA	31.1	340	A2
66702	368			検出されたものはない	肺炎	未知			A2
71025	414	25.40 6289	3.80×10^{-24}	RSV B	肺炎	RSVA	22.3	3.9×10^5	A2
71027	478			検出されたものはない	肺炎	RSVA	34.8	18	A2
73068	832	59.27 5233	1.91×10^{-10}	RSV属	LRTI	RSVA	23.7 681	1.2×10^5	A2
71028	913	25.89 7084	3.23×10^{-30}	RSV B	肺炎	RSVA	19.1	4.7×10^8	A2
66703	924	12.67 3149	9.71×10^{-6}	RSV属	肺炎	RSVA	31.5	250	A2
35259	324	20.611 47	3.55×10^{-94}	RSV B	LRTI	RSV B	21.4 366	3.0×10^6	A1
35662	355	17.99 9418	2.97×10^{-40}	RSV B	LRTI	RSV B	20.2 642	6.7×10^6	A1

10

20

30

40

【表4B】

66695	374			検出されたものはない	肺炎	RSV B	34.1	500	A2
70933	378	13.81 578	7.77×10^{-17}	RSV B	LRTI	RSV B	23.9 204	5.4×10^5	A2
36042	412	17.53 1234	4.58×10^{-55}	RSV B	LRTI	RSV B	23.5 804	6.9×10^5	A1
35890	412	17.21 4556	1.05×10^{-43}	RSV B	LRTI	RSV B	23.5 804	6.9×10^5	A2+A 3
36053	483	12.16 8025	1.47×10^{-12}	RSV B	LRTI	RSV B	24.8 34	2.9×10^5	A1
70997	554	76.54 7183 54.01 3223	1.83×10^{-11} 9 2.45×10^{-61}	ライノ ウイル ス属;エ ンテロ ウイル ス科	肺炎	RSV B	35.1	240	A2
35253	841	12.06 9138	4.86×10^{-26}	RSV B	肺炎	RSV B	20.8 619	4.4×10^6	A1
73070	841	22.10 857 5.708 560	6.80×10^{-50} 5.66×10^{-6}	RSV B, hMPV 重感染	肺炎	RSV B/ hMPV	20.8 619 35.4	4.4×10^6 8	A2
68360	841	21.36 9516 9.647 188	2.09×10^{-25} 1.23×10^{-8}	RSV B, hMPV 重感染	肺炎	RSV B/ hMPV	20.8 619 35.4	4.4×10^6 8	A2
66696	185			検出さ れたもの はない	肺炎	未知			A2
66697	261			検出さ れたもの はない	肺炎	未知			A2
66698	331			検出さ れたもの はない	肺炎	未知			A2
71189	393			検出さ れたもの はない	肺炎	未知			A2
66699	461			検出さ れたもの はない	肺炎	未知			A2
66700	573	41.39 7051 27.44 4893	3.97×10^{-23} 1.34×10^{-11}	ライノ ウイル ス属;エ ンテロ ウイル ス科	肺炎	未知			A2
71182	639			検出さ れたもの はない	肺炎	未知			A2
71007	699			検出さ れたもの はない	肺炎	未知			A2
71188	859			検出さ れたもの はない	肺炎	未知			A2

* L R T I : 下気道感染

10

20

30

40

50

【 0 1 5 7 】

A E S の重要性から、最適化された R T - P C R プライマーラグを用いることによって、増幅効率およびそれに続くプローブ検出が向上することができる事が示唆された。したがって本発明者らは、ランダムに生じた 17 塩基長のタグ配列を用いて A E S スコアを計算し、最も大きい A E S スコアの全体的増大を生じる上位 3 つの最も相違するプライマーを選択した（図 13）。A E S 最適化プライマーを用いて本発明者らは、向上した P C R 効率および検出感度で、臨床試料由来のメタニューモウイルスおよび R S V を増幅した（図 14、表 5）。

【 0 1 5 8 】

【表5】

患者試料#412および#122に関するE-PredictおよびPDA v. 1アルゴリズムの比較。結果に記載されている元々のPCRプライマーを用いて、アレイ35179を増幅した。プライマーA2を用いてアレイ36731および35887を増幅し、プライマーA2およびA3の両方を用いてアレイ35890を増幅した。PDA v. 1は、全ての場合において、正しい病原体のみを返した。E-Predictの著者らは、それらの基盤において、 $P < 0.01$ を有意性カットオフとして使用する(Urisman et al. 2005)。より低いカットオフは、このアルゴリズムを用いて本発明者らのアレイデータを分析する場合に、必要であるとわかる。PCRモデリングによって設計された新しいプライマーは、いずれかのアルゴリズム(アレイ35179対35887)を用いて、よりよい予測スコアを生じる。PCRプロセスの間に第二のプライマーを有することによって、WKLスコアおよびP値の漸進的な改善が提供された(アレイ36731対35890)。

アレイ	患者	PCR 増幅プラ イマー	E-Predict アルゴリズム			GISPathogen アルゴリズム	
			ゲノム	類似性スコア	P 値	ゲノム	WKL
36042	412 (RSV)	元々のプ ライマー A1	RSV	0.35128	0	RSV	21.526316
			OC43 コロナウイルス	0.350264	6.84E-20		
			229E コロナウイルス	0.323503	1.77E-10		
			B型肝炎	0.134825	3.03E-04		
			SARS コロナウイルス	0.338911	0.00299		
			A型肝炎	0.229589	0.00847		
36731	412 (RSV)	A2	RSV	0.335389	0	RSV	21.836754
			OC43 コロナウイルス	0.348043	2.29E-13		
			229E コロナウイルス	0.322055	2.00E-09		
			B型肝炎	0.135222	1.02E-06		
			風疹	0.164332	0.00919		
35890	412 (RSV)	A2 + A3	RSV	0.334602	0	RSV	22.093258
			OC43 コロナウイルス	0.348969	3.63E-23		
			229E コロナウイルス	0.322805	3.20E-14		
			B型肝炎	0.13436	6.74E-04		
			SARS コロナウイルス	0.338609	0.03060		
35179	122 (hMPV)	元々のプ ライマー A1	hMPV	0.260110695	5.01E-28	hMPV	9.763149
			風疹	0.164784981	1.20E-17		
			口蹄 C	0.206747816	4.66E-11		
			日本脳炎	0.201347222	1.65E-04		
			B型肝炎	0.133407622	1.98E-04		
			黄熱病	0.200500564	0.00567		
			エコーウィルス 1	0.222002025	0.01740		
35887	122 (hMPV)	A2	hMPV	0.299655	0	hMPV	39.677149
			風疹	0.169626	3.40E-19		
			B型肝炎	0.137703	5.84E-12		
			OC43 コロナウイルス	0.347685	5.06E-10		
			229E コロナウイルス	0.321702	1.72E-06		
			SARS コロナウイルス	0.340504	1.76E-06		
40			口蹄 C	0.2075	1.31E-04		
			ニューカッスル	0.234481686	0.01820		
40				0.23453	0.04310		

【0159】

< PDA v. 1 - 病原体を検出するためのアルゴリズム >

臨床材料はしばしば、ゲノム増幅のために準最適である。それらは低いウイルス価を有

10

20

30

40

50

するか、アレイ上の参考菌株由来の配列多型を有するか、または重感染している病原体を有する場合がある。マイクロアレイはまた、非特異的ハイブリダイゼーションおよび他のアーチファクト由来の内部雑音を有する。したがってマイクロアレイデータを解釈することは、プローブシグナル強度プロフィールを S P S に適合させるか、または単純な統計的手法（例えば、T 検定、A N O V A 等）を用いる簡単な問題ではない。この問題を扱うために本発明者らは、コンピュータ内で予測される S P S と関連してプローブシグナル強度の分布を分析して、ハイブリダイズした試料中に存在する病原体を同定する、ロバスト統計ソフトウェア、P D A v . 1 を確立した（上記参照）。

【 0 1 6 0 】

アレイ上の全てのプローブについてのシグナル強度は正規分布に分類されるが、試料中に存在する病原体 S P S を含むプローブの大部分は右に歪んだ分布を生じる非常に強力なシグナル強度を有するという本発明者らの観察に基づいて、本発明者らは、プローブシグナル強度の分布を分析することによって、病原体の存在を検出することができると推測した（図 9 A）。各 S P S についてシグナル強度分布の尾部を調べることによっても、試料中で重感染している病原体の存在を同定することが可能になる。

【 0 1 6 1 】

したがって、P D A v . 1 は、2つの部分を含む。（1）各病原体 S P S におけるプローブのプローブシグナル強度を評価するための重み付きカルバック・ライブラー情報量（W K L；本発明者らの、改良されたカルバック・ライブラー検定）、および（2）各 S P S についての W K L スコアの分布が正規であるかどうかを判定するためのアンダーソン・ダーリン検定。

【 0 1 6 2 】

元々のカルバック・ライブラーは確率分布の尾部の差異を確実に判定することができず、プローブ / ゲノムの数および各シグナル強度瓶のサイズに高度に依存している（K u l l b a c k a n d L e i b l e r , 1 9 5 1）。本発明者らはアンダーソン・ダーリン統計を組み込んで、各分布の尾部により多くの重みを与えることによって、および元々の確率分布の代わりに累積分布関数を用いることによって、これらの欠点を克服した（A n d e r s o n a n d D a r l i n g , 1 9 5 2）。本発明者らは本発明者らの改良された K L 情報量を、重み付きカルバック・ライブラー情報量（W K L）と呼ぶ。

【 0 1 6 3 】

【 数 3 9 】

$$WKL(P_a | \bar{P}_a) = \sum_{j=0}^{k-1} \frac{Q_a(j) \log(\frac{Q_a(j)}{\bar{Q}_a(j)})}{\sqrt{\bar{Q}_a(j)[1-\bar{Q}_a(j)]}}$$

式中、 $Q_a(j)$ は、瓶 b_j に見られる P_a におけるプローブのシグナル強度の累積分布関数であり；

【 数 4 0 】

$$\bar{Q}_a(j)$$

は、瓶 b_j に見られる

【 数 4 1 】

$$\bar{P}_a$$

におけるプローブのシグナル強度の累積分布関数である。病原体の非存在を表す S P S は、正規のシグナル強度分布、および従って比較的低い W K L スコアを有するべきであるが、病原体の存在を表すものは高い、統計的に有意な、範囲外の W K L スコアを有するべき

10

20

30

40

50

である(図9B)。PDA v.1の第二の部分において、WKLスコアの分布を、正規性についてのアンダーソン・ダーリング検定に供する。 $P < 0.05$ の場合WKL分布は、正規でないと見なされ、範囲外のWKLスコアを有する病原体が存在することが含意される。病原体の同定の際に、そのWKLスコアなしで別々のアンダーソン・ダーリング検定を行って、重感染している病原体の存在について試験する。このようにして正規分布だけが残るまで(すなわち、 $P > 0.05$; 上記の表3および表4参照)、手順を反復して適用する。PDA v.1は非常に速く、ハイブリダイズしたマイクロアレイから約10秒で診断を行うことができる。

【0164】

<33の臨床患者試料に関する病原体診断>

10

本発明者らは図11に示されるワークフローに従って、33の臨床材料を本発明者らの病原体マイクロアレイ基盤にハイブリダイズすることによって、本発明者らの基盤を評価した。これらのうち27の標本は、予めRSV A、RSV-Bまたはメタニューモウイルスと診断されていた。本発明者らの基盤は21/27の試料から病原体を正確に検出した。ウイルスが検出されなかった6つの試料(偽陰性)は、リアルタイムPCRによる検出限界(<10ウイルスコピー/反応)であり、このような低いウイルス量は患者の深刻な疾患に関与する病原因子である可能性は低かった。これらのうち2つは、マイクロアレイによってライノウイルスに感染していると正しく診断された。未知の病原体によって引き起こされた深刻な呼吸器疾患を有する別の6人の患者のスクリーニングにおいて、マイクロアレイは試料のうちの1つにおいて、病原因子(ライノウイルス)を同定した(上記の表4)。リアルタイムPCRによってこれらの結果を検証した。期待されたように本発明者らは、非ウイルス性病因を有する肺炎患者から抽出された試料をハイブリダイズさせた場合、いかなる病原体も検出しなかった。

20

【0165】

<データ解析>

Axon 4000bスキャナーおよびGenepix 4ソフトウェア(Axon Instruments)を用いて、5 μm分解能でマイクロアレイをスキャンした。NimbleScan 2.1ソフトウェア(NimbleGen Systems)を用いてシグナル強度を抽出した。自動化スクリプトを使用して本発明者らは、各プローブの7つの複製からメジアンシグナル強度(ハイブリダイゼーションアーチファクトを排除するため)および標準偏差を計算した。プローブシグナル強度をゲノムによって分類し、配列順に配置し、次いでJava(登録商標)Treeview(htt://jtreeview.sourceforge.net)でシグナル強度を図式的に見るために、CDT形式に再フォーマットした。並行してPDA v.1を用いてプローブメジアンシグナル強度を分析して、どの病原体が存在するか、および関連する予測の信頼水準を判定した。本発明者らは実験を行って、実験結果に対するプローブ設計の効果を示し、次いで本発明による分析アルゴリズムのロバスト性を示した。

30

【0166】

<実験結果に対するプローブ設計の効果>

40

ヒトに影響を及ぼす35のウイルス由来の53555の40塩基長のプローブを含有するPDCを、4つの独立したマイクロアレイ実験に使用した。これらの53555のプローブを、各ウイルスの5bpのタイリングに基づいて選択し、プローブ設計基準のいずれにも供さなかった。したがって本発明者らは、CG含有量、クロスハイブリダイゼーションおよび不十分な増幅によって生じる誤差が、うまく設計されたプローブを有するPDCのものよりも有意に高いと期待した。本発明者らは4つの実験についてのこのような不利な設定において、本発明者らの分析アルゴリズムを試験した。

【0167】

本実施例において、未知の病原体を含むヒト試料をランダムプローブを用いたRT-PCRプロセスによって増幅し、次いでPDC上にハイブリダイズさせた。本発明者らは、本発明者らのPDC上の35のウイルスのそれぞれについてのプローブを、0.05の有

50

意性レベルの片側 t 検定に供し、チップ上の全てのプローブのシグナル強度に対してそれらのシグナル強度の重み付きカルバック・ライブラリー (WKL) 情報量を計算して、各実験について試料中にどのウイルスがあるかを判定した。ウェットラボ PCR によって本発明者らのプログラムによる分析の精度の確認を行って、試料中の実際のウイルスを同定した。表 6 の 4 つの実験についての本発明者らの分析の結果およびそれらの対応する PCR 検証を、表 6 に示す。

【 0 1 6 8 】

【表 6 A】

プローブ設計基準を適用せずにPDC上で行われた分析結果。各実験について試験された試料中で本発明者らの分析アルゴリズムによって実際のウイルスであると判定されたウイルスを、薄灰色で強調する。

	実験 1		実験 2		実験 3		実験 4	
試料名	35259_324		35179_122		35253_841		35915_111	
D	53555		53555		53555		53555	
ウイルス (アクセスショ ン番号)	t 検定 p 値	WKL	t 検定 p 値	WKL	t 検定 p 値	WKL	t 検定 p 値	WKL
NC_001781.1	0	16.391	1	NA	0	10.85635	1	NA
NC_003461.1	1	NA	1	NA	1	NA	1	NA
NC_003443.1	0.999324	NA	0.873017	NA	0.99802	NA	0.999961	NA
NC_001796.2	1	NA	1	NA	1	NA	1	NA
AY283794.1	0	0.5435	0.108141	NA	0	0.775959	0	0.435427
NC_005147.1	0	1.2896	1	NA	0	1.399591	0	1.762912
NC_002645.1	0	1.2943	0.999847	NA	0	1.655888	0	2.079334
NC_004148.2	1	NA	0.002733	5.762907	1	NA	1	NA
NC_002023.1	1	NA	0.579561	NA	1	NA	1	NA
NC_002022.1								
NC_002021.1								
NC_002020.1								
NC_002019.1								
NC_002018.1								
NC_002017.1	1	NA	1	NA	1	NA	1	NA
NC_002204.1								
NC_002205.1								
NC_002206.1								
NC_002207.1								
NC_002208.1								
NC_002209.1	1	NA	0.999062	NA	1	NA	1	NA
NC_002210.1								
NC_002211.1								
NC_001563.2								
NC_002031.1								
NC_002728.1								
NC_002617.1	0.999994	NA	0	0.571844	1	NA	0.769098	NA
NC_001802.1	1	NA	0.999966	NA	1	NA	1	NA
NC_003977.1	0	2.7424	0	2.189827	0	3.978747	0	1.490665
NC_001576.1	0.371224	NA	0.004643	0.94841	0.009599	1.257041	0	3.961532
NC_002554.1	0.000062	0.7146	0	1.527292	0.299334	NA	0.000002	0.166239
NC_001545.1	0	1.4545	0	2.438558	0	0.869782	0	0.989592
NC_001489.1	0	1.7088	0.319125	NA	0	2.593065	0	1.510399
NC_005222.1	0.999757	NA	0.646314	NA	0.773912	NA	0.807875	NA
NC_005217.1	0.60477	NA	0.999903	NA	0.354358	NA	0.000871	0.626818
NC_004294.1	0	1.8411	0.000523	0.902399	0	2.43215	0.000007	0.537531
NC_004291.1	0.662386	NA	0.954137	NA	0.255422	NA	0.099148	NA
NC_001437.1	1	NA	0	0.593093	1	NA	1	NA
AB189128.1	1	NA	0.906213	NA	1	NA	1	NA
AF326573.1	1	NA	0.038503	0.539783	1	NA	1	NA
AF489932.1	1	NA	0.899797	NA	1	NA	1	NA
M87512.1	1	NA	0.759668	NA	1	NA	1	NA
NC_001430.1	1	NA	0.912496	NA	1	NA	0.999912	NA

10

20

30

40

【表6B】

NC_001428.1	0.999988	NA	0.284792	NA	0.999346	NA	0.957164	NA
NC_001612.1	0.970379	NA	0.000001	0.557865	0.998878	NA	0.061226	NA
NC_003986.1	1	NA	0.000012	0.604474	1	NA	0.997945	NA
NC_001472.1	0.999999	NA	0.0046	0.455194	0.999579	NA	0.143404	NA
NC_001617.1	0.721465	NA	0.98373	NA	0.178733	NA	0.414209	NA
NC_001490.1	0.999808	NA	0.995029	NA	0.997369	NA	0.859025	NA
推測 ウイルス	NC_001781.1 (RSV)		NC_004148.2 (HMPV)		NC_001781.1 (RSV)		なし	
確認 ウイルス (PCR)	NC_001781.1 (RSV)		NC_004148.2 (HMPV)		NC_001781.1 (RSV)		なし	

10

【0169】

本結果から分析アルゴリズムが、最初の3つの実験において試験された試料中の実際のウイルスを正確に予測することが示される（上記の表6に示される結果）。さらに本発明者らは最後の実験において、試料がウイルスを有さないことを推測することができた。本発明者らが有意性のレベルが0.05のt検定を使用しただけの場合、各試料について存在すると検出されたウイルスの数が下記の表7に示されることに留意されたい。

【0170】

20

【表7】

t検定のみを用いた、ウイルスの偽陽性検出

試料名	35259_324	35179_122	35253_841	35915_111
T検定を用いて検出されたウイルス	9	14	9	10
偽陽性	8	13	8	10
最大KL 発散(>5.0)	16.391	5.76	10.85	-
KL発散の後T検定を用いて検出されたウイルス	1	1	1	0

30

【0171】

t検定を通過するウイルスの重み付きカルバック・ライブラー情報量を用いることによって本発明者らは、全ての偽陽性ウイルスを除去して、実際のウイルスを同定することができた。したがって本発明者らの分析アルゴリズムは、高レベルの雑音下でウイルスを確実に判定することができる。

【0172】

次に本発明者らは、本発明者らの分析結果に対する、プローブ設計基準を適用したPDCを用いることの効果を調べた。まず35のウイルスのそれぞれについての増幅効率マップを計算した。次いで元々のPDC上のちょうど53555のプローブを、プローブ設計基準に供した。極端なレベルのCG含有量、ヒトおよび非標的ウイルスに対する高い類似性、ならびに低い増幅効率スコアを有するプローブをチップから除去した。合計10955のプローブを、第二の組の実験のために保持した。第一の組の実験で使用された試料を用いて、本発明者らは新しいチップで、下記の表8の4つの実験を繰り返した。実験結果を表8に示す。

40

【0173】

【表 8 A】

プローブ設計基準を適用してPDC上で行われた分析結果。各実験について試験された試料中で本発明者らの分析アルゴリズムによって実際のウイルスと判定されたウイルスを、薄灰色で強調する。

	実験 1		実験 2		実験 3		実験 4	
試料名	35259_324		35179_122		35253_841		35915_111	
D	10955		10955		10955		10955	
ウイルス (アクセスショ ン番号)	t 検定 p 値	WKL						
NC_001781.1	0	18.54859	1	NA	0	11.17914	1	NA
NC_003461.1	1	NA	1	NA	1	NA	1	NA
NC_003443.1	0.548718	NA	0.53727	NA	0.002783	0.837121	0.020436	0.603552
NC_001796.2	1	NA	0.999907	NA	1	NA	1	NA
AY283794.1	0	1.347801	0.024116	0.858364	0	1.523272	0	1.128637
NC_005147.1	0	1.604381	0.999697	NA	0	2.150019	0	2.893555
NC_002645.1	0	2.802742	0.999895	NA	0	4.612482	0	3.635771
NC_004148.2	1	NA	0.000003	9.324785	1	NA	1	NA
NC_002023.1								
NC_002022.1								
NC_002021.1								
NC_002020.1	1	NA	0.124517	NA	1	NA	0.999163	NA
NC_002019.1								
NC_002018.1								
NC_002017.1								
NC_002204.1								
NC_002205.1								
NC_002206.1								
NC_002207.1								
NC_002208.1								
NC_002209.1								
NC_002210.1								
NC_002211.1								
NC_001563.2	0.986443	NA	0.428418	NA	0.76002	NA	0.112011	NA
NC_002031.1	0.998103	NA	0.003435	2.52162	0.278672	NA	0.409527	NA
NC_002728.1	0.999375	NA	0.30951	NA	0.969492	NA	0.297244	NA
NC_002617.1	0.63418	NA	0.003578	0.965856	0.247148	NA	0.025188	0.861163
NC_001802.1	1	NA	0.998118	NA	1	NA	1	NA
NC_003977.1	0	3.062956	0.000028	3.027442	0	4.574591	0	3.277708
NC_001576.1	0.579342	NA	0.101093	NA	0.155219	NA	0.026417	3.280335
NC_002554.1	0.6722	NA	0	2.289379	0.80654	NA	0.106683	NA
NC_001545.1	0	2.225817	0	2.794877	0.000019	1.674329	0	1.97064
NC_001489.1	0.099427	NA	0.999985	NA	0.000366	1.829543	0.000006	3.023235
NC_005222.1	0.999735	NA	0.294141	NA	0.974031	NA	0.356952	NA
NC_005217.1	0.916186	NA	0.994358	NA	0.600759	NA	0.032616	2.105628
NC_004294.1	0.867625	NA	0.235197	NA	0.100961	NA	0.052759	NA
NC_004291.1	0.992032	NA	0.964128	NA	0.714211	NA	0.206422	NA
NC_001437.1	1	NA	0.001058	1.563913	1	NA	0.857228	NA
AB189128.1	1	NA	0.732737	NA	0.999997	NA	0.98859	NA
AF326573.1	1	NA	0.435629	NA	0.999986	NA	0.905393	NA
AF489932.1	1	NA	0.322655	NA	0.999996	NA	0.996837	NA
M87512.1	0.999617	NA	0.057346	NA	0.999758	NA	0.937937	NA
NC_001430.1	1	NA	0.865038	NA	1	NA	0.882339	NA

10

20

30

40

【表 8 B】

NC_001428.1	1	NA	0.522986	NA	0.999351	NA	0.749412	NA
NC_001612.1	0.991708	NA	0.751091	NA	0.990929	NA	0.257635	NA
NC_003986.1	0.999997	NA	0.02014	0.93616	0.937996	NA	0.708985	NA
NC_001472.1	0.99959	NA	0.977242	NA	0.957869	NA	0.692936	NA
NC_001617.1	0.435562	NA	0.474076	NA	0.028549	1.699567	0.079676	NA
NC_001490.1	1	NA	0.90881	NA	0.996231	NA	0.518662	NA
推測 ウイルス	NC_001781.1 (RSV)		NC_004148.2 (HMPV)		NC_001781.1 (RSV)		なし	
確認 ウイルス(PCR)	NC_001781.1 (RSV)		NC_004148.2 (HMPV)		NC_001781.1 (RSV)		なし	

10

【0174】

以下の組の実験において分析アルゴリズムは、3つの試料、および陰性試料においても、実際のウイルスを正確に検出した。本発明者らのチップのための優れたプローブを設計した後、実験1、2および3における実際のウイルスの重み付きカルバック・ライブラ－情報量は、プローブ設計のない対応する実験のものよりも大きかった。これは実際のウイルス由来のシグナル強度が、PDCにおけるバックグラウンドノイズよりも比較的高かったことを意味する。このことから、本発明者らのプローブ設計基準によってPDCからいくつかの不良なプローブが除去され、それによってより正確な分析が生じたことが示された。

20

【0175】

ここでも有意性のレベルが0.05のt検定を用いただけの場合、下記の表9に示される4つの実験の結果を示す。このとき各試料について存在すると検出されたウイルスの数を表9に示す。

【0176】

【表9】

プローブ設計のあるPDCにおいてt検定のみを用いた、ウイルスの偽陽性検出。

試料名	35259_324	35179_122	35253_841	35915_111
T検定を用いて検出されたウイルス	6	9	9	10
偽陽性	5	8	8	10
最大KL発散(>5.0)	18.54859	9.324785	11.17914	-
KL発散の後にT検定を用いて検出されたウイルス	1	1	1	0

30

【0177】

表9から、プローブ設計によって試料35259_324および35179_122についてのt検定によって検出された偽陽性ウイルスの数が減少したことがわかる。より重要な点は、実際のウイルスについての重み付きカルバック・ライブラ－情報量が4つ全ての試料において増大したことである。これはPDCにプローブ設計基準が適用された場合に、実際のウイルスのシグナルがバックグラウンドシグナルよりも差異があったことを意味する。

40

【0178】

結論として本発明者らは、0.05の有意性レベルの片側t検定を用い、それに続いて各ウイルスのシグナル強度についての重み付きカルバック・ライブラ－情報量を計算して、本発明者らがPDC上のデータを正確に分析し、高い確率で試料中の実際の病原体を判定することができたことを示した。分析アルゴリズムは高レベルのノイズ下でもうまく機能するが、本発明者らは、分析の精度が上述のプローブ設計基準を用いることによって改良されて、PDCについてのプローブの優れたセットが選択されることを示した。

【0179】

50

<プローブ設計および病原体検出のための代替的方法>

マイクロアレイ上のクロスハイブリダイゼーションを予測するのに利用可能なアルゴリズムはほとんどなく、1つのアルゴリズム、E - Predictだけが、マイクロアレイ上で病原体を検出することについて報告および検証されている(Urisman et al., 2005; Li et al., 2005)。E - Predictは、ハイブリダイゼーションサインを各マイクロアレイプローブについてのハイブリダイゼーションの理論的自由エネルギーに由来する、予測されるサインと適合させる。しかしながらE - predictを用いて本発明者らのマイクロアレイを分析すると、多くの偽陽性信号を生じた(上記の表5参照)。例えばE - Predictは、RSV患者412においてコロナウイルスを検出した(図15)。パンコロナウイルスプライマーならびにOC43および229Eコロナウイルスのための特定の診断用プライマーを用いた診断用PCRによって、患者412のコロナウイルスの非存在が確認された(上記の表4参照)。本発明者らは、E - Predictを用いた偽陽性信号が、ヒトまたはRSVゲノムとクロスハイブリダイズするコロナウイルスプローブから生じたという仮説を立てた。実際に最も高いシグナル強度を有する50のコロナウイルスプローブの85%がヒトゲノムとクロスハイブリダイズすると予測され、65%がRSVに関して17未満のHDを有し、これは家族性クロスハイブリダイゼーションについての本発明者らの12のHD閾値のすぐ上である。さらにE - Predictを最適化して、ヒトゲノムへのクロスハイブリダイゼーションが重要な考慮すべき事柄であるタイリングアレイの代わりに、ウイルスゲノム領域の間で高度に保存されたプローブを含有するマイクロアレイ上で機能させた。したがってこれらの2つの要因 - 異なるマイクロアレイ設計戦略およびヒトゲノムへのクロスハイブリダイゼーションは、本発明者らの基盤に関してE - predictの貧弱な性能に寄与するようである。本発明者らのE - predictでの経験から、PDA v1はそれらは異なるプローブ長について設計され、他の適用および基盤について最適化されているので、他のアルゴリズムと比較するのは公正ではない。

【0180】

<結論>

クロスハイブリダイゼーション閾値を経験的に判定することによって、本発明者らは、臨床試料中に存在する特定のウイルスとよくハイブリダイズするプローブのみを含むコンピュータ内での病原体サインプローブセットを作製した。AESアルゴリズムによってユニバーサルプライマータグを設計して、ウイルスゲノム全体を効率的に増幅することが可能になった。PDA v. 1検出アルゴリズムとともに、臨床試料由来のマイクロアレイ上に示される病原体のいずれかを確実に同定することができる。この手法によって、各病原体ハイブリダイゼーションサインの経験的な検証の必要性が排除され、病原体同定の強力な診断の基盤となる、10000を越える病原体についてのプローブを含有する将来のマイクロアレイが可能になる。

本発明者らは、病原体検出マイクロアレイの設計および分析を最適化し、病院におけるそれらの使用を容易にした。本発明者らは、ランダムPCRにおいて日常的に使用されるプライマータグが偏っており、病原体ゲノムの不均一な増幅を生じることを発見した。この偏りは、本発明者らのAESアルゴリズムを用いてプライマーを設計することによって避けることができる。本発明者らのコンピュータ内でのサインプローブセットによって、どのプローブがアレイ上に示されるいずれかの病原体にハイブリダイズするかを正確に予測することができる。PDA v. 1検出アルゴリズムとともに、この手法によって、各病原体ハイブリダイゼーションサインの経験的な検証の必要性が排除され、病原体同定のための強力な診断の基盤となる10000を越える病原体についてのプローブを含有する将来のマイクロアレイが可能になる。

【0181】

ここで本発明者らは、ウイルス増幅効率、ハイブリダイゼーションシグナル出力、標的 - プローブアニーリング特異性、および特注設計のマイクロアレイ基盤を用いた病原体検出の再現性の間の、複雑な関係の体系的調査の結果を報告する。本発明者らの発見によっ

10

20

30

40

50

て、最適な病原体サインプローブセット（S P S）のコンピュータ内での予測の新規の方法論の基礎が形成され、（マイクロアレイハイブリダイゼーションの前の）ウイルス増幅効率を左右する要因に光明が投じられ、ウイルス増幅効率スコア（A E S）と最適なプローブ選択との間の重要な関連が示される。最後に本発明者らは、ウイルス価の範囲全体で臨床材料中の病原体を迅速かつ再現性良く同定することができる、新しい統計ベースの病原体検出アルゴリズム（P D A）を説明する。

【 0 1 8 2 】

本発明者らは、公的に利用可能なデータベースから得られたウイルスゲノム配列を用いて、少なくとも4000のウイルスコピーが存在するかどうかの高い程度の確実性で、臨床試料中のウイルスを検出することの実行可能性を示した（上記の表3参照）。その感度は抗原検出方法のものに近づき、それによって臨床的に意義のある検出手段になる（Liuet al., 2005; Marrara et al., 2003）。コンピュータ内で病原体ハイブリダイゼーションサインを正確に予測できることによって、最初にアレイを純粋な病原体試料にハイブリダイズさせることによる経験的な検証を必要とする現在のマイクロアレイ方法に対し、相当な進展が提供される。アレイ上に示される病原体の特異的同定に加えて、P D A v. 1によってアレイ上に特異的に示されないそれらのゲノムについて、病原体の綱、科または属の同定が可能になる（H DおよびM C Mの閾値を緩めることによって）。この情報はしばしば、病院における治療法の決定に十分である。A E S最適化タグで、本発明者らは非A E S最適化タグを用いて増幅した場合に、以前は検出することができなかつた臨床試料由来のウイルスを同定することができた。したがってA E Sによるタグの選択によって、P C R効率および検出の感度が増大した。本発明のアルゴリズムは、D N Aライブラリーの生成および再配列決定のためのR N Aの強化等の、他のタグベースのP C R用途に適用してもよい。

10

20

【 0 1 8 3 】

【表 10 A】

参考文献

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
- Anderson TW, Darling DA (1952) Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Annals of Mathematical Statistics* 23: 192-212.
- Bodrossy L, Sessitsch A (2004) Oligonucleotide microarrays in microbial diagnostics. *Curr Opin Microbiol* 7: 245-254.
- Bohlander SK, Espinosa I, Rafael, Le Beau MM, Rowley JD, Diaz MO (1992) A method for the rapid sequence-independent amplification of microdissected chromosomal material. *Genomics* 13: 1322-1324.
- Bustin, S.A. & Nolan, T. (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech* 15, 155-166
- Deffernez C, Wunderli W, Thomas Y, Yerly S, Perrin L, et al. (2004) Amplicon Sequencing and Improved Detection of Human Rhinovirus in Respiratory Samples 10.1128/JCM.42.7.3212-3218.2004. *J Clin Microbiol* 42: 3212-3218.
- Fu J, Tan BH, Yap EH, Chan YC, Tan YH (1992) Full-length cDNA sequence of dengue type 1 virus (Singapore strain S275/90). *Virology* 188: 953-958.
- Goulden, C.H. *Methods of Statistical Analysis*, Edn. 2nd. (John Wiley & Sons, Inc., New York; 1956).
- Hamming RW (1950) Error Detecting and Error Correcting Codes. *Bell System Technical Journal* 29: 147-160.
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860-921 (2001).
- Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, et al. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28: 4552-4557.
- Kane, M.D. et al. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28, 4552-4557 (2000).
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, et al. (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348: 1953-1966.
- Kullback S, Leibler RA (1951) On information and sufficiency. *Annals of Mathematical Statistic* 22: 79-86.
- Li X, He Z, Zhou J (2005) Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucl Acids Res* 33: 6114-6123.
- Liu J, Lim SL, Ruan Y, Ling AE, Ng LF, et al. (2005) SARS transmission pattern in Singapore reassessed by viral sequence variation analysis. *PLoS Med* 2(2), 162-168.
- Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, et al. (2003) The Genome sequence of the SARS-associated coronavirus. *Science* 300: 1399-1404.
- Maskos U, Southern EM (1993) A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support. *Nucleic Acids Res* 21: 4663-4669.
- Moës E, Vijgen L, Keyaerts E, Zlateva K, Li S, et al. (2005) A novel pan-coronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium. *BMC Infect Dis* 5: 6.
- Nguyen HK, Southern EM (2000) Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridisation. *Nucleic Acids Res* 28: 3904-3909.

10

20

30

40

【表10B】

- Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, et al. (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* 12: 1749-1755.
- Pang XL, Preiksaitis JK, Lee B (2005) Multiplex real time RT-PCR for the detection and quantitation of norovirus genogroups I and II in patients with acute gastroenteritis. *J Clin Virol* 33: 168-171.
- Ratushna VG, Weller JW, Gibas CJ (2005) Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. *BMC Genomics* 6: 31.
- Ruan YJ, Wei CL, Ee AL, Vega VB, Thoreau H, et al. (2003) Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* 361: 1779-1785.
- Sambrook and Russel, (2001) Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York
- SantaLucia, J., Jr., Allawi, H.T. & Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 35, 3555-3562.
- Smalling TW, Sefers SE, Li H, Tang YW (2002) Molecular approaches to detecting herpes simplex virus and enteroviruses in the central nervous system. *J Clin Microbiol* 40: 2317-2322.
- Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association*, Vol. 69, pp. 730-737.
- Striebel HM, Birch-Hirschfeld E, Egerer R, Foldes-Papp Z (2003) Virus diagnostics on microarrays. *Curr Pharm Biotechnol* 4: 401-415.
- Sung, W.K. & Lee, W.H. Fast and Accurate Probe Selection Algorithm for Large Genomes. CSB (2003).
- Sung, W.K. & Lee, W.H. (2003) in IEEE Computational Systems Bioinformatics ConferenceStanford University, Stanford, CA.
- Urisman A, Fischer KF, Chiu CY, Kistler AL, Beck S, et al. (2005) E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns. *Genome Biol* 6: R78.
- Vega VB, Ruan Y, Liu J, Lee WH, Wei CL, et al. (2004) Mutational dynamics of the SARS coronavirus in cell culture and human populations isolated in 2003. *BMC Infect Dis* 4: 32.
- Vora GJ, Meador CE, Stenger DA, Andreadis JD (2004) Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl Environ Microbiol* 70: 3047-3054.
- Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, et al. (2002) Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* 99: 15687-15692.
- Wang D, Urisman A, Liu YT, Springer M, Ksiazek TG, et al. (2003) Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 1: E2.
- Wong CW, Albert TJ, Vega VB, Norton JE, Cutler DJ, et al. (2004) Tracking the Evolution of the SARS Coronavirus Using High-Throughput, High-Density Resequencing Arrays. *Genome Res* 14: 398-405.
- Wu, D.Y., Ugozzoli, L., Pal, B.K., Qian, J. & Wallace, R.B. (1991) The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. *DNA Cell Biol* 10, 233-238
- 10
- 20
- 30
- 40

【図面の簡単な説明】

【0184】

【図1】ウイルス配列（配列番号1～9）上のランダムプライマーの対のRT-PCR結合プロセスを示す。図1についての標識は以下のとおりである。A：逆転写（RT）。プライマーが鋳型に結合する。B：タグを付けたRT産物が生じる（詳細には、仮説上のウイルス配列鋳型および仮説上の特異的ランダムプライマーで）。C：タグを組み込んで、第二の鎖合成を完了する。D：PCRプライマーGTTTCCCGAGTCACGAATA（配列番号8）を使用した、タグを付けたRT産物の増幅。

【図2】RSV-Bゲノムについての増幅効率スコアリング（AES）マップを示す。

50

【図3】RSV-Bについての一実験のオリゴヌクレオチドプローブシグナル強度を示す。

【図4】図4Aは、試験した試料中のウイルスのシグナル強度の密度分布を示す。矢印は分布の正の歪みを示す。これは、ノイズはあるが有意な量の実信号もあることを示す。図4Bは、試料中にはウイルスのシグナル強度の密度分布を示す。これはノイズ領域である。

【図5】病原体検出チップデータの解析フレームワークを示す。

【図6】オリゴヌクレオチドプローブ設計スキーマ。これは、NC_001781ヒト呼吸器多核体ウイルス(RSV)のゲノム全体で作製されたタイリングプローブを示す。数字は各プローブの開始および終止部分を表す。1948のプローブを合成して、1522 10
5bpのRSVゲノム全体をカバーした。このプロセスを、残りの34のウイルスゲノムについて繰り返した。

【図7】マイクロアレイバーの標識の手掛けり：

【表 11】

ウイルス科	ウイルス属/種
オルトミクスウイルス科	Sars Sin2500 OC41 229E
コロナウイルス科	Flu A Flu B
ピコナウイルス科	エンテロD エンテロC エコー1 エンテロB エンテロA ライノ89 ライノB Hep A 口蹄C
ブニヤウイルス科	ハンタン シンノンブレ
フラビウイルス科	ウエストナイル 日本脳炎 デング熱3 デング熱1 デング熱2 デング熱4 黄熱病
パラミクソウイルス科	Paraflu 1 Paraflu 3 ニパー Paraflu 2 ニューカッスル RSV (B1) メタニューモウ イルス
その他	HPV 10型 HIV 1 Hep B 風疹 LCMV-S LCMV-L PMMV ヒト対照

10

20

30

40

S A R S S i n 8 5 0 感染細胞系 (A) または 1 型デング熱感染細胞系 (B) から単離された R N A を、それぞれ S A R S 特異的または 1 型デング熱特異的 R T - P C R の後、病原体マイクロアレイ上にハイブリダイズした。S A R S は他のコロナウイルス科ゲノム、具体的には高度に保存されたゲノムの中間部分にクロスハイブリダイズした（黒色で示す）(Ruan et al., 2003)。1 型デング熱はそれらの配列類似性に基づいて、フラビウイルス科由来のプローブおよび他のゲノムにクロスハイブリダイズした。ハミング距離 (H D) および最大連続マッチ (M C M) スコアを調べることによって、クロスハイブリダイゼーションが起こるかどうかを予測する閾値を確立し、この情報を利用してコンピュータ内でのハイブリダイゼーションサインを生成した。(C) R S V を有すると診断された臨床患者から単離された R N A をランダム R T - P C R を用いて増幅し、

50

病原体マイクロアレイ上にハイブリダイズさせた。

【図8】プローブハミング距離(HD)、プローブ最大連続マッチ(MCM)およびプローブシグナル強度の間の関係。平均プローブシグナル強度は、HDが増大しMCMが減少すると減少した。これは、検出可能なプローブのパーセンテージの減少と相関がある(シグナル強度 > 平均値 + 2SD)。最適なクロスハイブリダイゼーション閾値HD = 4またはMCM = 18(影付き)で、98%を超えるプローブを検出することができる。HD = 5またはMCM = 17で、検出率は85%に落ちる。

【図9】RSV感染患者から単離されたRNAを、病原体検出アレイにハイブリダイズさせた。(A)全53, 555のプローブのプローブシグナル強度の分布から、正規分布(灰色の実線)が示される。非RSVプローブはゲノム特異的レベルで調べた場合、例えばパラインフルエンザ-1(灰色の点線)からも、正規分布が示される。RSV特異的プローブのシグナル強度は正の歪みを有し、分布の尾部のシグナル強度がより高い(黒色の実線)。(B)大多数が-5~3の間の範囲である35のSPSについてのWKLスコアの分布頻度。しかしながら、RSVゲノムのWKLスコアは17であり、そのため分布は正規でない(アンダーソン・ダーリン検定によって、P < 0.05)。異常値ゲノムを排除することによって正規分布が生じた。この計算から、ハイブリダイズした試料中にRSVが存在すると結論付ける。

【図10】AESはプローブ増幅効率の指標である。高いAESを有するより高い割合のプローブが、5つの実験にわたってシグナル強度閾値より高く検出可能である。

【図11】マイクロアレイを用いた病原体検出に必要なプロセスを示すスキーマ。

【図12】ハイブリダイゼーションシグナル強度は増幅効率スコア(AES)と相関があり、 $P = 2.2 \times 10^{-16}$ である。RSV患者試料をマイクロアレイ上にハイブリダイズさせ、各プローブのシグナル強度を、計算したAESとともにプロットした。典型的なアレイ上の高信頼度検出のためのシグナル閾値を緑色の線によって示す。

【図13】ランダムRT-PCRのためのAES最適化プライマータグを用いることによって、AESが10~30倍増大する。最適化プライマーは、マイクロアレイ上に示される全35のゲノム全体で同じ性能を有すると予測された。AES最適化プライマーA2を用いてほとんどの患者試料を増幅した。

【表12】

10

20

30

配列番号	プライマー	ヌクレオチド配列
10	A1	GTTTCCCCAGTCACGATA
11	A2	GATGAGGGAAAGATGGGG
12	A3	CTCATGCACGACCCAAA
13	A4	AGATCCATTCCACCCCA

【図14】マイクロアレイバーの標識の手掛けり：

【表13】

ウイルス科	ウイルス属/種
オルトミクスウイルス科	Sars Sin2500 OC41 229E
コロナウイルス科	Flu A Flu B
ピコナウイルス科	エンテロD エンテロC エコー1 エンテロB エンテロA ライノ89 ライノB Hep A 口蹄C
ブニヤウイルス科	ハンタン シンノンブレ
フラビウイルス科	ウエストナイル 日本脳炎 デング熱3 デング熱1 デング熱2 デング熱4 黄熱病
パラミクソウイルス科	Paraflu 1 Paraflu 3 ニパー Paraflu 2 ニューカッスル RSV (B1) メタニューモウ イルス
その他	HPV 10型 HIV 1 Hep B 風疹 LCMV-S LCMV-L PMMV ヒト対照

10

20

30

40

ランダム R T - P C R におけるプライマータグの選択は、P C R 効率に対して有意な効果を有する。(A) Bo h l a n d e r e t a l . , 1 9 9 2 によって記載された元々のプライマー、または(B)マイクロアレイ上に示された全てのゲノム(高AES)が確実に効率的に増幅されるようにP C R モデリングに従って設計されたプライマーを用いたR T - P C R の後、臨床h M P V 試料にハイブリダイズしているプローブを示すヒートマップ。

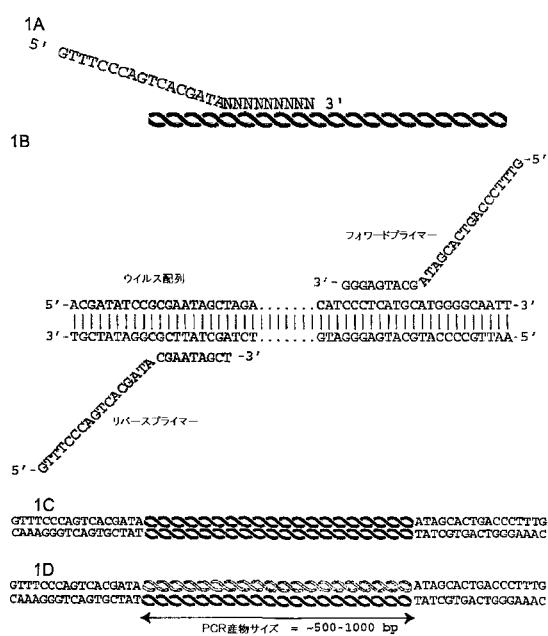
【図15】R S V 患者#412についての診断P C R 結果から、患者がコロナウイルス感染を有さないことが確認される。(A)パンコロナウイルスプライマーを用いたP C R。レーン1:O C 4 3 コロナウイルス陽性対照、レーン2:2 2 9 E コロナウイルス陽性対照、レーン3:R S V 患者#412、レーン4:P C R プライマーおよび試薬のみの陰性

50

対照。1 kb ラダー。(B) O C 4 3 特異的プライマーを用いたPCR。レーン1: O C 4 3 コロナウイルス陽性対照、レーン2: R S V 患者#412、レーン3: A T C C 由来の精製されたR S V、レーン4: PCR陰性対照。50 bp ラダー。(C) 2 2 9 E 特異的プライマーを用いたPCR。レーン1: 2 2 9 E コロナウイルス陽性対照、レーン2: R S V 患者#412、レーン3: PCR陰性対照。1 kb ラダー。

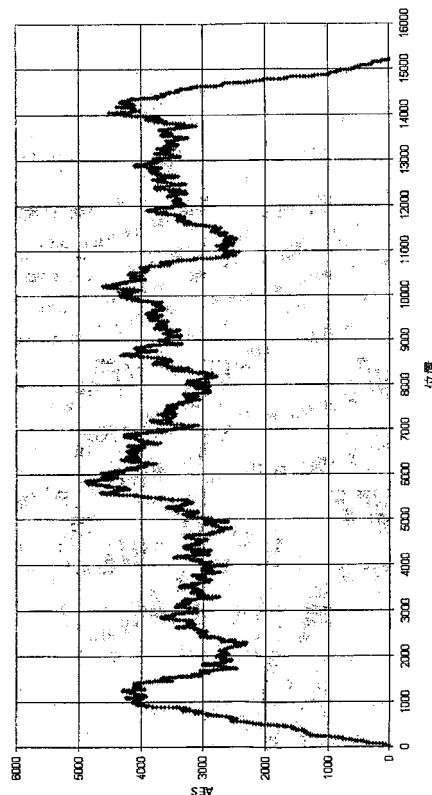
【図1】

FIGURE 1



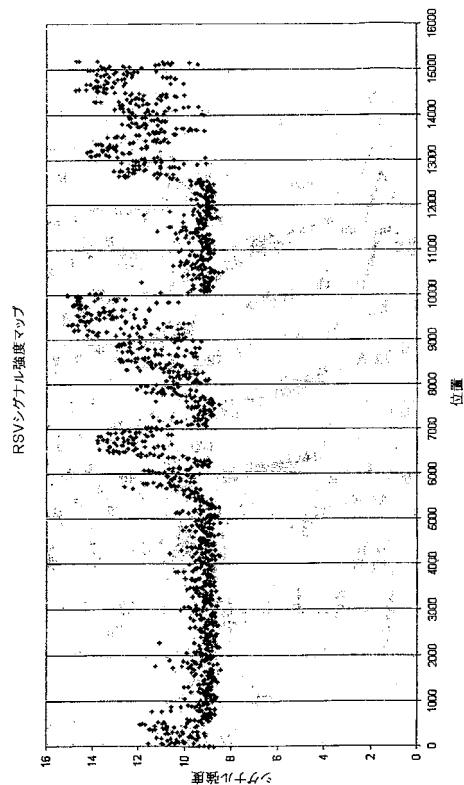
【図2】

FIGURE 2



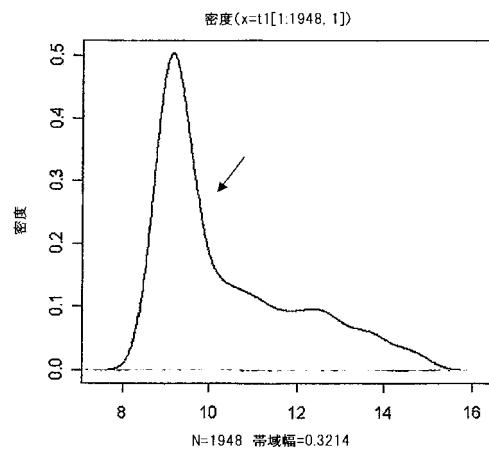
【図3】

FIGURE 3



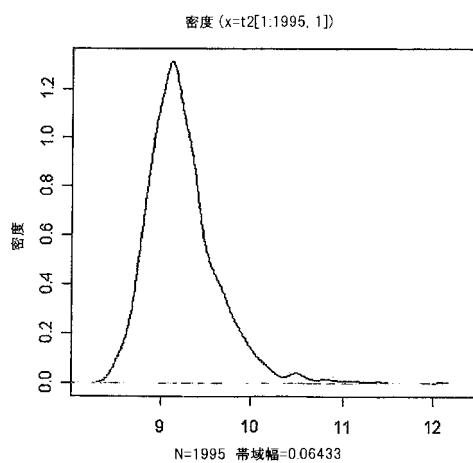
【図4A】

FIGURE 4A



【図4B】

FIGURE 4B



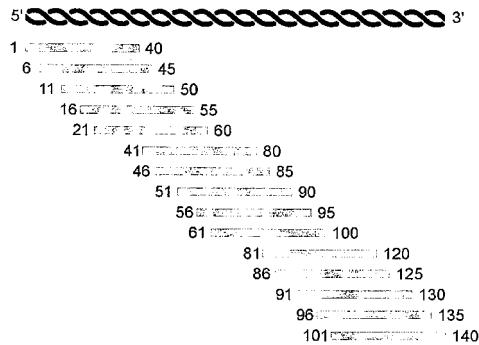
【図5】

FIGURE 5

ウイルスセットVおよびプローブセットPでのPDCデータDを仮定し、
 $V = \emptyset$ とし、
各 $v_a \in V$ について、
・ v_a におけるプローブの有意水準0.05で片側検定を計算する
(t_a のp値 < 0.05)の場合
・ $KL(P_a \| P) \geq 0.1$; $V' = V \cup \{v_a\}$ を許容し、
・ そうでなければ拒絶し;
そうでなければ拒絶し;
Vを返す

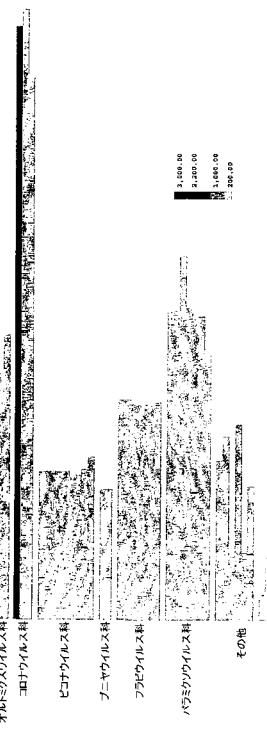
【図6】

FIGURE 6



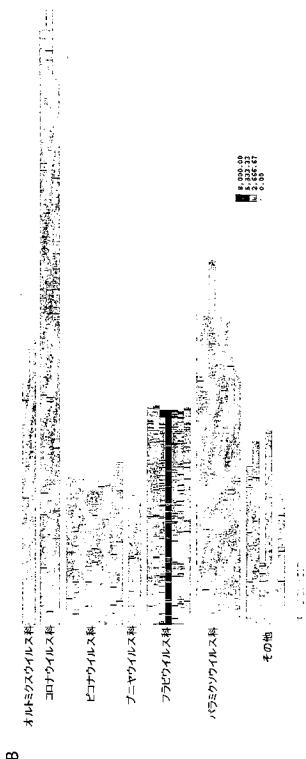
【図7A】

FIGURE 7A



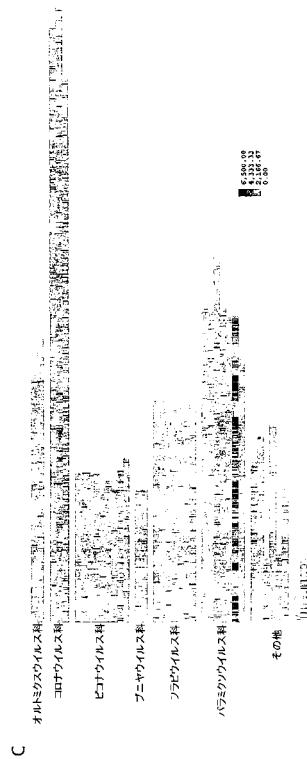
【図7B】

FIGURE 7B

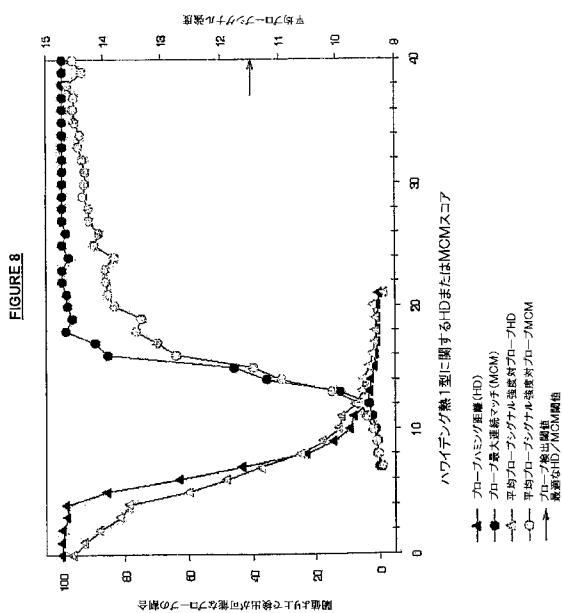


【図7C】

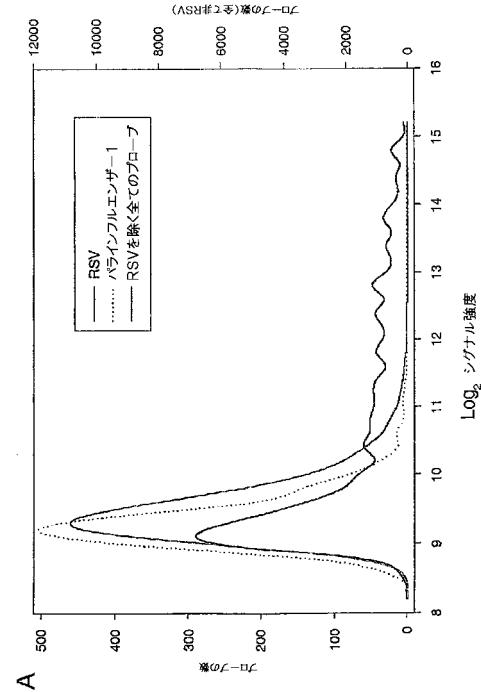
FIGURE 7C



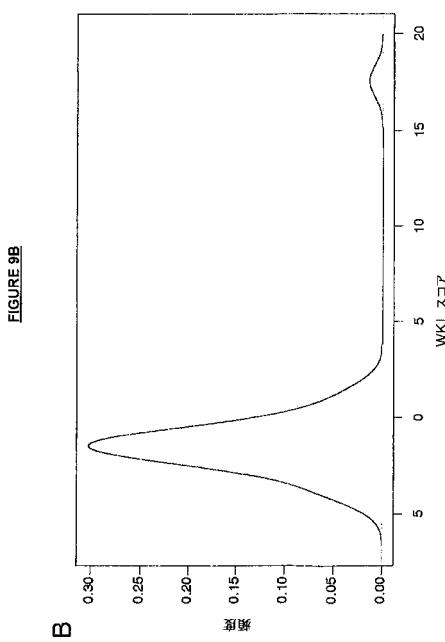
【図 8】



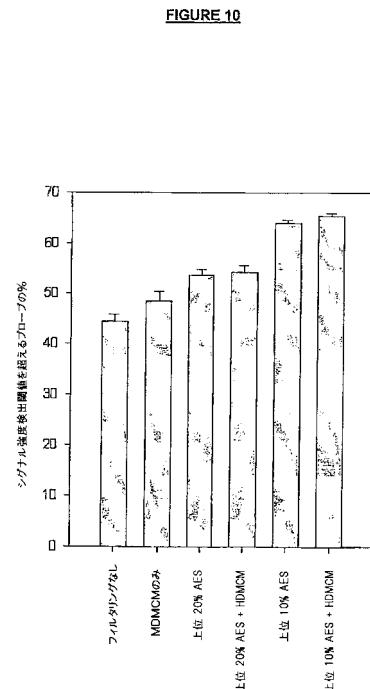
【図 9 A】



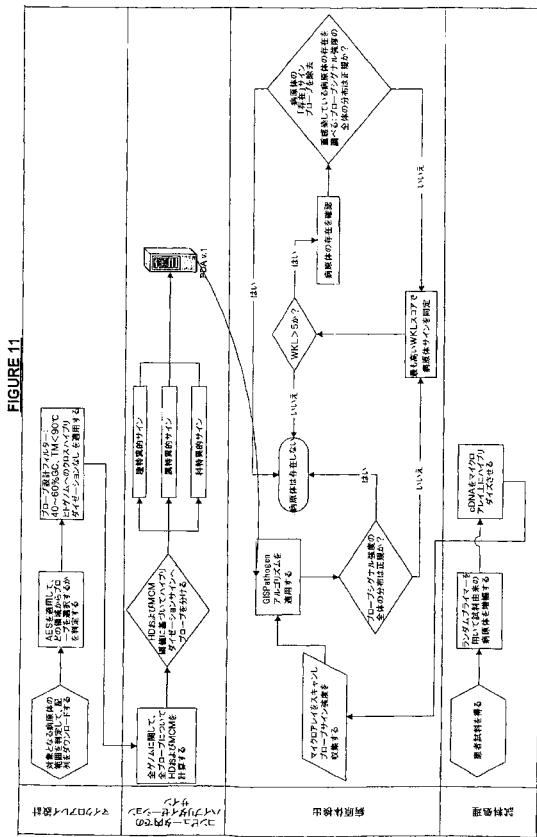
【図 9 B】



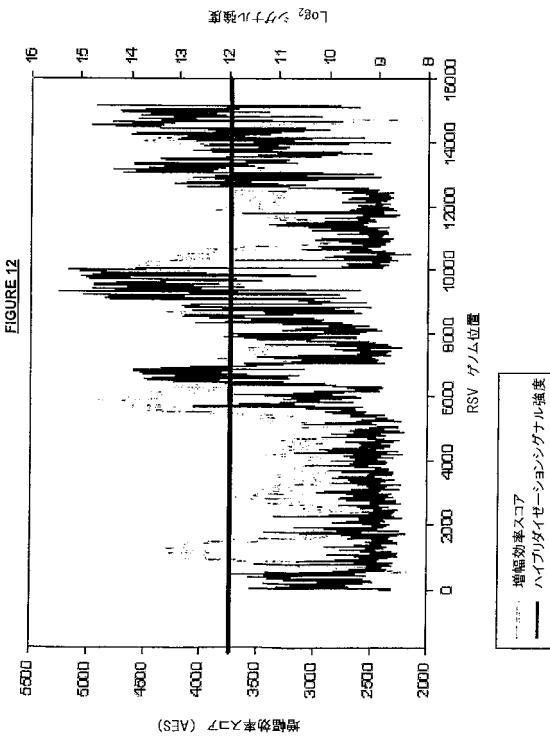
【図 10】



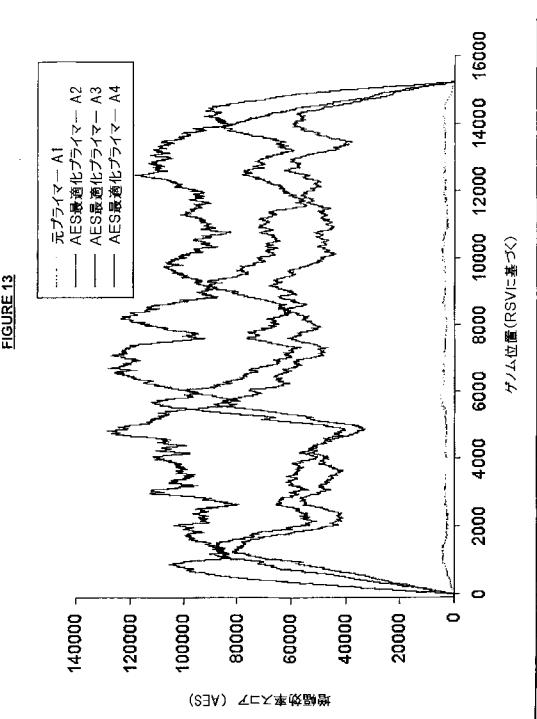
【図 1 1】



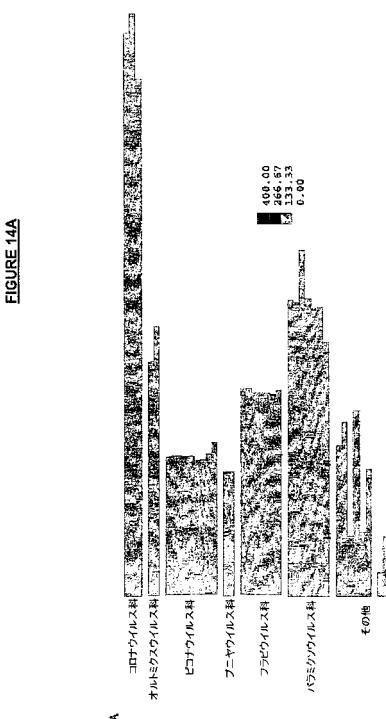
【図 1 2】



【図 1 3】

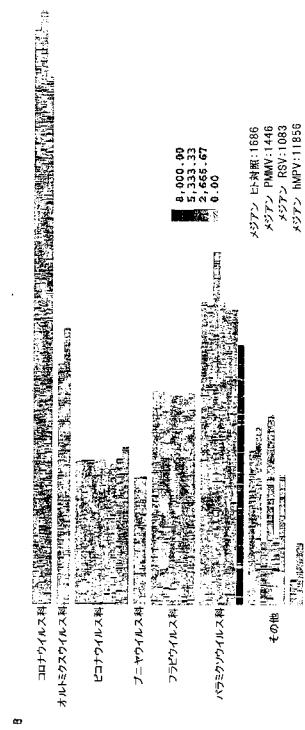


【図 1 4 A】



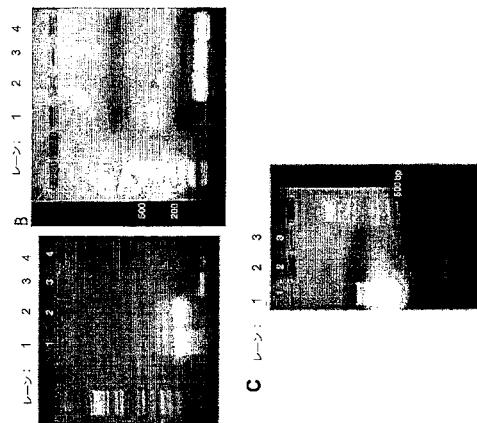
【図 14B】

FIGURE 14B



【図 15】

FIGURE 15



【配列表】

2009504153000001.xml

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2006/000224

A. CLASSIFICATION OF SUBJECT MATTER		
C12Q 1/68 (2006.01) G06F 19/00 (2006.01)		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) WPIDS, Caplus, Medline, Biosis: (keywords) probe, oligonucleotide, microarray, array, design, select, primer, tag, tail, header, PCR, geometric, amplification, bias		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SUNG, W-K. et al., 'Fast and accurate probe selection algorithm for large genomes', Proceedings IEEE Computer Society Bioinformatics Conference, 2003, 2: 65-74 See the whole document	1-17, 19-93
A	US 2004/0259124 A1 (BEKIRANOV et al) 23 December 2004 See the whole document	1-93
A	US 2003/0236633 A1 (MEI et al) 25 December 2003 See the whole document	1-93
<input type="checkbox"/> Further documents are listed in the continuation of Box C		<input checked="" type="checkbox"/> See patent family annex
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"B" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the International filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search 23 April 2007	Date of mailing of the international search report - 9 MAY 2007	
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer LEXIE PRESS AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No : (02) 6283 2677	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2006/000224

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: 37-52, 89, 92 (in part)
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
See supplemental sheet

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No. PCT/SG2006/000224

Supplemental Box

(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: II

Claims 37-52 relate to downstream diagnostic applications of the oligonucleotide probes designed by the methods disclosed in the specification and defined in claims 1-36. Claims 89 and 92 define computer storage material which presents information for use in the design method. The specification only supports the use of oligonucleotides designed by the disclosed method, and the search of claims 37-52 has been limited to applications and methods that necessarily use the oligonucleotide probes designed by the said method. Similarly claim 89 and 92 are searched only in so far as they relate to the design of oligonucleotides as defined in claim1.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2006/000224

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member			
US	2004259124				
US	2003236633	AU	41501/02	US	2003096986
		US	2005158790	WO	0242485

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LS,MW,MZ,NA,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HU,IE,IS,IT,LT,LU,LV,MC,NL,PL,PT,RO,SE,SI,SK,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BW,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,HN,HR,HU,ID,IL,IN,IS,JP,KE,KG,KM,KN,KP,KR,KZ,LA,L,C,LK,LR,LS,LT,LU,LV,LY,MA,MD,MG,MK,MN,MW,MX,MZ,NA,NG,NI,NO,NZ,OM,PG,PH,PL,PT,RO,RS,RU,SC,SD,SE,SG,SK,SL,SM,SY,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,ZA,ZM,ZW

(74)代理人 100118407
弁理士 吉田 尚美
(74)代理人 100125380
弁理士 中村 紗子
(74)代理人 100130960
弁理士 岡本 正之
(74)代理人 100125036
弁理士 深川 英里
(74)代理人 100142996
弁理士 森本 聰二
(72)発明者 ワン,クリストファー・ワイン・チョン
シンガポール国,シンガポール 138672,バイオポリス・ストリート 60,ゲノム・イン
スティテュート・オヴ・シンガポール
(72)発明者 スン,ワイン キン
シンガポール国,シンガポール 138672,バイオポリス・ストリート 60,ゲノム・イン
スティテュート・オヴ・シンガポール
(72)発明者 リー,チャーリー
シンガポール国,シンガポール 138672,バイオポリス・ストリート 60,ゲノム・イン
スティテュート・オヴ・シンガポール
(72)発明者 ミラー,ランス・デイヴィッド
シンガポール国,シンガポール 138672,バイオポリス・ストリート 60,ゲノム・イン
スティテュート・オヴ・シンガポール
F ターム(参考) 4B024 AA13 AA14 CA04 CA09 HA14
4B063 QA01 QA18 QQ06 QQ07 QQ10 QQ42 QR32 QR39 QR40 QR56
QR62 QR75 QR76 QR79 QS25 QS34 QX02
5B075 UU18

【要約の続き】

例えばマイクロアレイ上に配置してもよい。