

(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2012-525147

(P2012-525147A)

(43) 公表日 平成24年10月22日(2012.10.22)

(51) Int. Cl.	F I	テーマコード (参考)
C 1 2 Q 1/68 (2006.01)	C 1 2 Q 1/68 Z N A A	4 B O 2 4
G O 1 N 33/53 (2006.01)	G O 1 N 33/53 M	4 B O 6 3
C 1 2 N 15/09 (2006.01)	C 1 2 N 15/00 A	

審査請求 未請求 予備審査請求 未請求 (全 80 頁)

(21) 出願番号 特願2012-508490 (P2012-508490)
 (86) (22) 出願日 平成22年4月30日 (2010. 4. 30)
 (85) 翻訳文提出日 平成23年12月12日 (2011. 12. 12)
 (86) 国際出願番号 PCT/US2010/001293
 (87) 国際公開番号 W02010/126614
 (87) 国際公開日 平成22年11月4日 (2010. 11. 4)
 (31) 優先権主張番号 61/179, 358
 (32) 優先日 平成21年5月18日 (2009. 5. 18)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 61/178, 923
 (32) 優先日 平成21年5月15日 (2009. 5. 15)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 61/174, 470
 (32) 優先日 平成21年4月30日 (2009. 4. 30)
 (33) 優先権主張国 米国 (US)

(71) 出願人 511262946
 グッド スタート ジェネティクス, インコーポレイテッド
 アメリカ合衆国 マサチューセッツ 02215, ポストン, セント メアリーズ ストリート 8, スイート 621
 (74) 代理人 100078282
 弁理士 山本 秀策
 (74) 代理人 100062409
 弁理士 安村 高明
 (74) 代理人 100113413
 弁理士 森下 夏樹

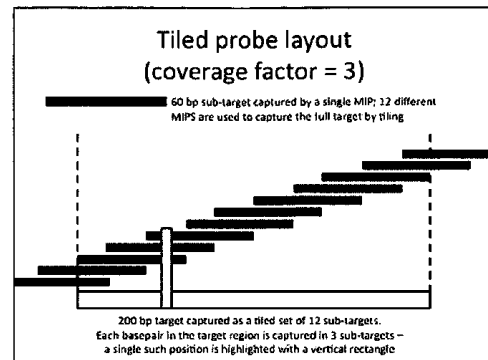
最終頁に続く

(54) 【発明の名称】 遺伝マーカーを評価するための方法および組成物

(57) 【要約】

本発明の局面は、遺伝子座の多重分析のバイアスを減少させるためおよび遺伝子座の多重分析の再現性を高めるために有用な方法および組成物に関する。いくつかの形態では、所定の調製工程および/または核酸配列分析法が、複数のサンプル中の複数の遺伝子座に対する多重分析において用いられる。一局面において、複数の遺伝子座を分析する方法が提供され、該方法は：複数の標的核酸の各々をプローブセットと接触させる工程、複数の核酸を単離する工程、ならびに該単離された核酸を分析する工程を包含する。

Figure 1



【特許請求の範囲】

【請求項 1】

複数の遺伝子座を分析する方法であって、該方法は：

複数の標的核酸の各々をプローブセットと接触させる工程であって、

ここで、各プローブセットは、複数の異なるプローブを含み、各プローブは、該標的核酸の複数の部分領域のうちの一つの同じ鎖に隣接する核酸と相補的である、5'領域および3'領域に隣接する中央領域を有し、ここで、該標的核酸の部分領域は、異なるものであり、各部分領域は、少なくとも一つの他の部分領域と重複している、工程、

複数の核酸を単離する工程であって、該複数の核酸の各々は、該複数の標的核酸の各々に関して異なる部分領域の核酸配列を有する、工程、ならびに

該単離された核酸を分析する工程

を包含する、方法。

【請求項 2】

少なくとも一つの部分領域が、少なくとも一つの他の部分領域と完全に重複する、請求項 1 に記載の方法。

【請求項 3】

各プローブの前記 5'領域および前記 3'領域の配列が、それぞれ、他の各プローブの 5'領域および 3'領域の配列と重複していない、請求項 1 に記載の方法。

【請求項 4】

各プローブの前記 5'領域および前記 3'領域の配列が、それぞれ、他の各プローブの 5'領域および 3'領域の配列と異なる、請求項 1 に記載の方法。

20

【請求項 5】

複数の遺伝子座を分析する方法であって、該方法は：

複数の標的核酸の各々をプローブセットと接触させる工程であって、

ここで、各プローブセットは、複数の異なるプローブを含み、各プローブは、該標的核酸の複数の部分領域のうちの一つの同じ鎖に隣接する核酸と相補的である、5'領域および3'領域に隣接する中央領域を有し、ここで、該標的核酸の部分領域は、異なるものであり、プローブの該 5'領域の一部および該 3'領域の一部は、それぞれ、異なるプローブの 5'領域および 3'領域の配列を有する、工程、

複数の核酸を単離する工程であって、該複数の核酸の各々は、該複数の標的核酸の各々に関して異なる部分領域の核酸配列を有する、工程、ならびに

30

該単離された核酸を分析する工程

を包含する、方法。

【請求項 6】

前記標的核酸が、表 1 から選択される、請求項 1 または 5 に記載の方法。

【請求項 7】

前記分析工程が、前記標的核酸の多型の遺伝子タイピングを行う工程を包含する、請求項 1 または 5 に記載の方法。

【請求項 8】

前記多型が、表 2 から選択される、請求項 7 に記載の方法。

40

【請求項 9】

被験体の遺伝子タイピングを行う方法であって、該方法は：

独立して単離された少なくとも閾値数の核酸の配列を決定する工程を包含し、ここで、単離された核酸の各々の配列は、標的核酸配列および識別タグ配列を含み、

ここで、該閾値数は、標的核酸配列と識別タグ配列との独特の組み合わせの数であり、ここで、該単離された核酸は、標的核酸配列と識別タグ配列との独特の組み合わせを含む場合、独立して単離されたと同定され、ここで、該標的核酸配列は、被験体のゲノム遺伝子座の配列である、方法。

【請求項 10】

前記単離された核酸が、分子反転プローブによる捕捉産物である、請求項 9 に記載の方法

50

。

【請求項 1 1】

被験体の遺伝子タイピングを行う方法であって、該方法は：

分子反転プローブによる捕捉産物の配列を決定する工程であって、該分子反転プローブによる捕捉産物の各々は、分子反転プローブおよび標的核酸を含み、該分子反転プローブの配列は、識別タグ配列および必要に応じてプライマー配列を含み、該標的核酸は、被験体の捕捉されたゲノム遺伝子座である、工程、ならびに

分子反転プローブによる捕捉産物の標的核酸配列と識別タグ配列との少なくとも閾値数の独特の組み合わせの配列に基づいて、該捕捉されたゲノム遺伝子座について該被験体の遺伝子タイピングを行う工程

を包含する、方法。

【請求項 1 2】

被験体の遺伝子タイピングを行う方法であって、該方法は：

分子反転プローブによる捕捉産物を得る工程であって、各捕捉産物は、分子反転プローブおよび標的核酸を含み、ここで、該分子反転プローブの配列は、識別タグ配列および必要に応じてプライマー配列を含み、ここで、該標的核酸は、該被験体の捕捉されるゲノム遺伝子座である、工程、

該分子反転プローブによる捕捉産物を増幅する工程、ならびに

各標的核酸について、分子反転プローブによる捕捉産物の標的核酸配列と識別タグ配列との少なくとも閾値数の独特の組み合わせの配列を決定することによって該被験体の遺伝子タイピングを行う工程を包含する、方法。

【請求項 1 3】

前記得る工程は、分子反転プローブを用いて前記被験体のゲノムサンプルから標的核酸を捕捉する工程を包含し、該分子反転プローブの各々は、独特の識別タグ配列を含む、請求項 1 2 に記載の方法。

【請求項 1 4】

前記捕捉工程が、標的と識別タグ配列との同一の組み合わせを用いて 2 つ以上の分子反転プローブによる捕捉産物を得る尤度が所定の値と等しいかまたはそれ未満である条件下で行われ、必要に応じて該所定の値が約 0 . 0 5 である、請求項 1 2 または 1 3 に記載の方法。

【請求項 1 5】

特定の標的核酸配列に対する前記閾値数が、遺伝子型に対する所望の統計学的信頼度に基づいて選択される、請求項 1 2 ~ 1 4 のいずれか 1 項に記載の方法。

【請求項 1 6】

標的核酸配列と識別タグ配列との独特の組み合わせの数に基づいて、遺伝子型に対する統計学的信頼度を決定する工程をさらに包含する、請求項 1 2 ~ 1 4 のいずれか 1 項に記載の方法。

【請求項 1 7】

複数の遺伝子座を分析する方法であって、該方法は：

分子反転プローブによる複数の捕捉産物を得る工程であって、該複数の捕捉産物の各々は、分子反転プローブおよび標的核酸を含み、ここで、該分子反転プローブの配列は、識別タグ配列および必要に応じてプライマー配列を含む、工程、

該分子反転プローブによる複数の捕捉産物を増幅する工程、

増幅された複数のものにおける、分子反転プローブによる捕捉産物の標的核酸配列と識別タグ配列との組み合わせの出現数を測定する工程、ならびに

標的核酸配列と識別タグ配列との特定の組み合わせの出現数が、所定の値を超える場合、該特定の組み合わせを含む該分子反転プローブの増幅におけるバイアスを検出する工程を包含する、方法。

【請求項 1 8】

前記複数のものにおける標的配列の遺伝子タイピングを行う工程をさらに包含し、ここで

10

20

30

40

50

、該遺伝子タイピング工程は、バイアスが検出された場合に該バイアスを補正する工程を包含する、請求項 17 に記載の方法。

【請求項 19】

前記標的核酸配列が、表 1 から選択される遺伝子の配列である、請求項 9 ~ 18 のいずれかに記載の方法。

【請求項 20】

標的核酸が異常な長さを有するか否かを決定するための方法であって、該方法は、被験体由来の生物学的サンプル中の標的核酸の捕捉効率を評価する工程を包含し、ここで、基準となる捕捉効率と異なる捕捉効率は、該生物学的サンプル中の異常な長さを有する標的核酸の存在を示す、方法。

10

【請求項 21】

欠失または挿入を有すると疑われる標的領域の捕捉効率を測定する工程、ならびに該捕捉効率を、正常な捕捉効率を示す参照と比較する工程を包含する、請求項 20 に記載の方法。

【請求項 22】

前記捕捉効率が、前記基準となる捕捉効率よりも低い、請求項 20 ~ 21 に記載の方法。

【請求項 23】

前記被験体が、前記標的領域に挿入を有すると同定される、請求項 22 に記載の方法。

【請求項 24】

前記捕捉効率が、前記基準となる捕捉効率よりも高い、請求項 20 ~ 21 に記載の方法。

20

【請求項 25】

前記被験体が、前記標的領域に欠失を有すると同定される、請求項 24 に記載の方法。

【請求項 26】

前記被験体が、前記挿入についてヘテロ接合であると同定される、請求項 23 に記載の方法。

【請求項 27】

前記被験体が、前記欠失についてヘテロ接合であると同定される、請求項 25 に記載の方法。

【請求項 28】

サブターゲット核酸が、分子反転プローブを用いて前記標的核酸から捕捉される、前述のいずれかの請求項に記載の方法。

30

【請求項 29】

前記分子反転プローブが、その 5' 末端に第 1 の標的化アームを含み、その 3' 末端に第 2 の標的化アームを含み、ここで、該第 1 の標的化アームは、前記サブターゲット核酸の一方の末端に隣接する第 1 の領域と特異的にハイブリダイズすることができ、該第 2 の標的化アームは、該標的核酸の同じ鎖における該サブターゲット核酸の他方の末端に隣接する第 2 の領域と特異的にハイブリダイズすることができる、請求項 28 に記載の方法。

【請求項 30】

前記第 1 の標的化アームおよび前記第 2 の標的化アームが、約 10 から約 100 ヌクレオチド長である、請求項 29 に記載の方法。

40

【請求項 31】

前記第 1 の標的化アームおよび前記第 2 の標的化アームが、約 10 ~ 20、20 ~ 30、30 ~ 40 または 40 ~ 50 ヌクレオチド長である、請求項 30 に記載の方法。

【請求項 32】

前記第 1 の標的化アームおよび前記第 2 の標的化アームが、約 20 ヌクレオチド長である、請求項 31 に記載の方法。

【請求項 33】

前記第 1 の標的化アームおよび前記第 2 の標的化アームが、同じ長さを有する、請求項 29 ~ 31 のいずれか 1 項に記載の方法。

【請求項 34】

50

前記第 1 の標的化アームおよび前記第 2 の標的化アームのハイブリダイゼーション T m が、同様のものである、請求項 2 9 ~ 3 3 のいずれか 1 項に記載の方法。

【請求項 3 5】

前記第 1 の標的化アームおよび前記第 2 の標的化アームのハイブリダイゼーション T m が、互いの 2 ~ 5 以内である、請求項 3 4 に記載の方法。

【請求項 3 6】

前記第 1 の標的化アームおよび前記第 2 の標的化アームのハイブリダイゼーション T m が、同一である、請求項 3 4 に記載の方法。

【請求項 3 7】

前記サブターゲット核酸が、核酸リピートを含む、請求項 2 8 ~ 3 6 のいずれか 1 項に記載の方法。

10

【請求項 3 8】

前記核酸リピートが、ジヌクレオチドリピートまたはトリヌクレオチドリピートである、請求項 3 7 に記載の方法。

【請求項 3 9】

前記サブターゲット核酸が、核酸リピートの異常な増加または減少の非存在下において、1 0 ~ 1 0 0 コピーの前記核酸リピートを含む、請求項 3 7 に記載の方法。

【請求項 4 0】

前記サブターゲット核酸が、核酸リピートを含む脆弱 X 遺伝子座の領域である、請求項 3 7 に記載の方法。

20

【請求項 4 1】

一方または両方の標的化アームが、核酸リピートの領域のすぐ隣である、前記標的核酸における領域とハイブリダイズする、請求項 2 9 ~ 3 6 のいずれか 1 項に記載の方法。

【請求項 4 2】

一方または両方の標的化アームが、任意の核酸リピートを含まない領域によって核酸リピートの領域から分離された前記標的核酸における領域とハイブリダイズする、請求項 2 9 ~ 3 6 のいずれか 1 項に記載の方法。

【請求項 4 3】

前記分子反転プローブが、捕捉された前記サブターゲット核酸ならびに必要なに応じて前記第 1 の標的化アームおよび / または前記第 2 の標的化アームを配列決定するために使用され得るプライマー結合領域をさらに含む、請求項 2 9 ~ 3 6 のいずれか 1 項に記載の方法。

30

【請求項 4 4】

生物学的サンプル中の複数の異なる標的核酸が分析される、前述の請求項のいずれかに記載の方法。

【請求項 4 5】

前記複数の標的核酸が、複数の異なる分子反転プローブを用いて分析される、請求項 4 4 に記載の方法。

【請求項 4 6】

異なる分子反転プローブの各々が、3 ' および 5 ' 末端の各々に第 1 の標的化アームと第 2 の標的化アームとの異なる対を含む、請求項 4 5 に記載の方法。

40

【請求項 4 7】

異なる分子反転プローブの各々が、同じプライマー結合配列を含む、請求項 4 6 に記載の方法。

【請求項 4 8】

前記生物学的サンプルが、血液サンプルである、前述のいずれかの請求項に記載の方法。

【請求項 4 9】

前記生物学的サンプルが、組織サンプルである、前述のいずれかの請求項に記載の方法。

【請求項 5 0】

前記捕捉効率が、捕捉された標的核酸の量を測定することによって評価される、前述のい

50

ずれかの請求項に記載の方法。

【請求項 5 1】

捕捉された標的核酸の前記量が、独立して捕捉された標的核酸分子の数を測定することによって測定される、請求項 5 0 に記載の方法。

【請求項 5 2】

捕捉された標的核酸の前記量が、捕捉された核酸の基準となる量と比較される、請求項 5 0 ~ 5 1 に記載の方法。

【請求項 5 3】

前記基準となる量が、独立して捕捉された基準となる核酸の分子の数を測定することによって測定される、請求項 5 2 に記載の方法。

10

【請求項 5 4】

前記基準となる核酸が、欠失または挿入を含むとは疑われない生物学的サンプル中の異なる遺伝子座の核酸である、請求項 5 3 に記載の方法。

【請求項 5 5】

前記基準となる核酸が、捕捉反応に加えられる、既知サイズおよび既知量の核酸である、請求項 5 3 に記載の方法。

【請求項 5 6】

前記捕捉効率が、前記基準となる捕捉効率と統計学的に有意に異なる場合、被験体は、前記遺伝子座の 1 つ以上の対立遺伝子に挿入または欠失を有すると同定される、前述のいずれかの請求項に記載の方法。

20

【請求項 5 7】

標的核酸の長さを推定するための方法であって、該方法は：

(i) 該標的核酸に対する検出プローブのハイブリダイゼーションを可能にする条件下で複数の該検出プローブと該標的核酸を接触させる工程であって、

ここで、各検出プローブは、該標的核酸の第 1 の領域にハイブリダイズする第 1 のアームおよび該標的核酸の第 2 の領域にハイブリダイズする第 2 のアームを含むポリヌクレオチドであり、

ここで、該第 1 の領域および該第 2 の領域は、該標的核酸の共通の鎖に存在し、

ここで、該第 1 の領域の 5 ' 末端と該第 2 の領域の 3 ' 末端との間のヌクレオチド配列は、サブターゲット核酸のヌクレオチド配列である、工程；ならびに

30

(i i) 複数の該検出プローブとハイブリダイズした複数のサブターゲット核酸を捕捉する工程；ならびに

(i i i) 該複数のサブターゲット核酸におけるあるサブターゲット核酸の出現頻度を測定する工程であって、ここで、該複数のサブターゲット核酸における該サブターゲット核酸の出現頻度は、該サブターゲット核酸の長さを示す、工程

を包含する、方法。

【請求項 5 8】

核酸検出アッセイの感度を高める方法であって、該方法は：

生物学的サンプルに対して第 1 の調製方法を用いて標的核酸の第 1 の調製物を得る工程、該生物学的サンプルに対して第 2 の調製方法を用いて標的核酸の第 2 の調製物を得る工程

40

、第 1 の核酸調製物と第 2 の核酸調製物との両方において得られた配列をアッセイする工程

、第 1 の核酸調製物と第 2 の核酸調製物との両方からの配列情報を用いることにより、該生物学的サンプル中の該標的核酸の遺伝子型を決定する工程

を包含し、ここで、該第 1 の調製方法および該第 2 の調製方法は、異なる体系的な配列バイアスを有する、方法。

【請求項 5 9】

前記第 1 の核酸調製物および前記第 2 の核酸調製物が、配列アッセイを行う前に混合される、請求項 5 8 に記載の方法。

50

【請求項 6 0】

別個の配列アッセイが、前記第 1 の核酸調製物および前記第 2 の核酸調製物において行われ、両方のアッセイからの配列情報が、組み合わせられることにより、前記生物学的サンプル中の前記標的核酸の遺伝子型が決定される、請求項 5 8 に記載の方法。

【請求項 6 1】

前記第 1 の調製方法が、PCR ベース、ハイブリダイゼーションベースまたは環状プローブベースの調製方法である、請求項 5 8 に記載の方法。

【請求項 6 2】

前記第 2 の調製方法は、PCR ベース、ハイブリダイゼーションベースまたは環状プローブベースの調製方法であるが、但し、該第 2 の調製方法は、前記第 1 の方法と異なる、請求項 6 1 に記載の方法。

10

【請求項 6 3】

生物学的サンプル中の核酸を代表する核酸調製物を得る方法であって、該方法は、生物学的サンプルに対して第 1 の調製方法を用いて第 1 の標的核酸調製物を得る工程、該生物学的サンプルに対して第 2 の調製方法を用いて第 2 の核酸調製物を得る工程、ならびに

該第 1 の核酸調製物および該第 2 の核酸調製物を混合することにより、該生物学的サンプル中の核酸を代表する混合された調製物を得る工程

を包含する、方法。

【請求項 6 4】

第 1 の調製方法および第 2 の調製方法と異なる第 3 の調製方法を用いて第 3 の核酸調製を行う工程をさらに包含し、ここで、該第 1 の調製方法、該第 2 の調製方法および該第 3 の調製方法のすべてが、異なる体系的な配列バイアスを有する、請求項 1 または 6 に記載の方法。

20

【請求項 6 5】

前記異なる調製方法が、前記生物学的サンプル中の複数の異なる遺伝子座に対して用いられることにより、多重核酸分析の感度が高められる、前述の請求項のいずれかの請求項に記載の方法。

【請求項 6 6】

生物学的サンプル中の核酸の遺伝子タイピングを行う方法であって、該方法は：
生物学的サンプルに対してある調製方法を用いて核酸調製物を得る工程、
該核酸調製物の標的核酸をシーケンシングする工程、ならびに
該生物学的サンプルに対して分子反転プローブによる捕捉反応を行う工程であって、ここで、該生物学的サンプル中の該標的核酸の分子反転プローブによる捕捉は、該標的核酸における多型の存在を示す、工程、
該シーケンシングおよび該捕捉反応の結果に基づいて該標的核酸を遺伝子タイピングする工程
を包含する、方法。

30

【請求項 6 7】

前記標的核酸が、表 1 から選択される遺伝子の配列を有する、前述の請求項のいずれかに記載の方法。

40

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

この出願は、米国特許法 § 119 (e) の下、2009 年 4 月 30 日に提出された米国仮出願第 61 / 174, 470 号、2009 年 5 月 15 日に提出された米国仮出願第 61 / 178, 923 号、2009 年 5 月 18 日に提出された米国仮出願第 61 / 179, 358 号および 2009 年 5 月 28 日に提出された米国仮出願第 61 / 182, 089 号 (各々の全体の内容は、参考として本明細書に援用される) の利益を主張する。

50

【 0 0 0 2 】

発明の背景

本発明は、患者サンプル中の遺伝子型を決定するための方法および組成物に関する。

【 背景技術 】

【 0 0 0 3 】

発明の背景

被験体の遺伝子型に関する情報は、多くの疾患、障害および生理学的特徴に対する遺伝的根拠がさらに解明されるにつれて、一連の健康管理の判断にとってますます重要になり、かつそのような判断に直接関係してくる。医学的な助言は、次第に個人向けになってきており、個別の判断および勧告は、特定の遺伝情報に基づく。1つ以上の遺伝子座における対立遺伝子の型および数に関する情報は、健康管理について考慮すべき他の事柄の中でも疾患リスク、予後、治療法の選択肢および遺伝カウンセリングに影響を及ぼす。

10

【 0 0 0 4 】

費用効果が高く、かつ信頼できる医学的カウンセリングおよび生殖カウンセリングを大規模に行うためには、多くの被験体の多くの異なる遺伝子座についての対立遺伝子の状態を正確かつ明白に同定できることが重要である。

【 0 0 0 5 】

生物学的サンプルから核酸配列を検出するためおよび分析するために、数多くの技術が開発されている。これらの技術を用いることにより、被験体の遺伝子タイピングを行うことおよび目的の任意の遺伝子座の対立遺伝子の状態を決定することができる。しかしながら、多数の患者の多くの遺伝子座の信頼できるハイスループット分析にむけてスケールアップするには、それらの技術は、十分に頑強ではなく、費用効果の高いものではない。現在の技術は、不正確またはあいまいなコール (c a l l) の頻度が高すぎて、不確かさを解消するためおよびアウトプット情報の信頼性を提供するために高価かつ時間のかかる工程を要せずに多数の患者サンプルを管理することはできない。

20

【 発明の概要 】

【 課題を解決するための手段 】

【 0 0 0 6 】

発明の要旨

本発明の局面は、遺伝子型を評価するため、特に、被験体の1つ以上の遺伝子座の対立遺伝子の同一性 (または二倍体生物における同一性) を決定するための、調製および分析の方法および組成物に関する。

30

【 0 0 0 7 】

本発明の局面は、遺伝分析におけるあいまいさおよびエラーの種々の起源の同定に部分的に基づき、また、遺伝分析の種々の段階におけるこれらのエラーおよびあいまいさを回避するため、減少させるため、認識するため、および / または解消するための1つ以上のアプローチの同定に部分的に基づく。

【 0 0 0 8 】

本発明の局面によると、ある特定のタイプの遺伝情報は、調製段階 (例えば、捕捉、増幅など)、決定段階 (例えば、対立遺伝子特異的検出、配列決定など)、データ解釈段階 (例えば、アッセイ情報が被験体をホモ接合またはヘテロ接合同定するのに十分であるか否かを決定する段階) および / または他の段階のいずれかにおける、確率的変動と系統誤差 (s y s t e m a t i c b i a s) との組み合わせに起因して、遺伝分析において提示不足であり得るかまたは提示過剰であり得る。

40

【 0 0 0 9 】

本発明の局面によると、エラーまたはあいまいさは、遺伝分析において明らかであり得るが、追加のサンプリングまたはより高価なアッセイを行わずに容易に解決されることはない (例えば、アレイベースのアッセイは、ノイズの多いシグナル / 低レベルのシグナルに起因してコールなし (n o - c a l l) を報告し得る)。本発明のさらなる局面によると、エラーまたはあいまいさは、遺伝分析において説明されないかもしれず、不正確なべ

50

ースコール (base call) は、それらについての証拠が限られているときおよび / または統計学的に有意でないときでさえ、生じ得る (例えば、次世代シーケンシング技術は、それらについての証拠が統計学的に有意でない場合でさえ、ベースコールを報告し得る)。本発明のさらなる局面によると、エラーまたはあいまいさは、明らかであるが、分析の1つ以上の工程では容易に解決されず、また、分析の他の工程では明らかでないかまたは説明されないので、多段階の遺伝分析にとって問題となり得る。

【0010】

いくつかの実施形態において、1つ以上の工程におけるエラーおよびあいまいさの起源は、生成される遺伝情報の複雑度および / または忠実度に影響を及ぼし得る任意の系統誤差または確率的影響を克服するように設計された重複プローブの1つ以上のセットを用いて目的の各標的遺伝子座を捕捉することおよび / または調べることによって対処され得る。

10

【0011】

いくつかの実施形態において、1つ以上の工程におけるエラーおよびあいまいさの起源は、少なくとも1セットのプローブを用いて目的の各標的遺伝子座を捕捉することおよび / または調べることによって対処され得、ここで、種々のプローブは、アッセイ反応を追跡するため、および生成される情報においてある特定のタイプの遺伝情報が提示不足であるかまたは提示過剰であるかを決定するために使用され得る種々の識別子で標識されている。

【0012】

いくつかの実施形態において、多数の配列リポートを含む領域の分析に関連するエラーおよびあいまいさは、アッセイの特定の段階 (例えば、捕捉、シーケンシングまたは検出の段階) においてある特定の核酸の出現頻度を体系的に分析することによって対処される。そのような手法は、多くの異なる遺伝子座を、各遺伝子座に対して異なるアッセイ手順を必要とすることなく並行して評価できるように設計された標準化プロトコルにおいて特に有用であり得ることが認識されるだろう。いくつかの実施形態において、複数のタイプの遺伝子病変 (例えば、点変異、挿入 / 欠失、長さの多型) をアッセイするために単一の検出様式 (例えば、シーケンシング) を使用することは、臨床の状況において都合がよい。本発明のいくつかの実施形態において、並行した複数のサンプル調製工程の使用を促進する方法が提供され、その方法には、配列検出の後に複数の分析プロセスが連結される。よって、本発明のいくつかの実施形態において、複数の患者の複数の遺伝子座にわたって種々のタイプの遺伝子病変を同時にアッセイするときのエラーおよび不確かさを減少させる改善されたワークフローが提供される。

20

30

【0013】

いくつかの実施形態において、本発明の局面は、2つ以上の手法 (各々が、異なるバイアス (bias) (例えば、1つ以上のタイプの配列の提示不足または提示過剰に向かう公知のバイアス) を有する) を組み合わせ、そして得られたデータを用いることにより、より高い信頼性で被験体に対する遺伝的コール (genetic call) を決定することによって、調製および / または分析のバイアスを克服するための方法を提供する。

【0014】

いくつかの実施形態において、本発明の局面は、多重診断方法 (multiplex diagnostic method) に関することが認識されるだろう。いくつかの実施形態において、多重診断方法は、複数の遺伝子座 (例えば、表1からの1つ以上の遺伝子座) を並行して捕捉する工程を包含する。いくつかの実施形態において、それらの遺伝子座は、1つ以上の多型 (例えば、表2からの1つ以上の多型) を有し、その遺伝子型は、疾患を引き起こす対立遺伝子に対応する。したがって、いくつかの実施形態において、本開示は、複数の遺伝性障害を並行して評価するための方法を提供する。いくつかの実施形態において、着床前、出生前、出生時または出生後の段階において複数の遺伝性障害を並行して診断するための方法が提供される。いくつかの実施形態において、本開示は、患者サンプル (例えば、血液、着床前胚、絨毛または羊水サンプル) から、または本発明の

40

50

局面はこの点に限定されないので、他のサンプル（例えば、他の生体液または組織サンプル、例えば、生検サンプル）から複数の遺伝子座（例えば、表 1 から選択される複数の標的核酸）を分析するための方法を提供する。

【0015】

他のサンプルとしては、腫瘍組織または循環腫瘍細胞が挙げられ得る。いくつかの実施形態において、患者サンプル（例えば、腫瘍組織サンプルまたは腫瘍細胞サンプル）は、目的の 1 つ以上の変異に対する寄せ集めであるので、生殖細胞系における変異の分析に必要な感度よりも高い感度を必要とし得る。いくつかの実施形態において、サンプルは、宿主生物に由来しない細胞（例えば、ヒト被験体内の細菌感染物またはウイルス感染物）または環境モニタリング用のサンプル（例えば、土壌、水または大気サンプルの細菌、ウイルス、真菌の組成物）を含む。

10

【0016】

したがって、いくつかの実施形態において、本明細書中に開示される方法の局面は、標的核酸の多型の遺伝子タイピングに関する。いくつかの実施形態において、遺伝子タイピングは、標的核酸の 1 つ以上の対立遺伝子がヘテロ接合であるかまたはホモ接合であるかを決定することを包含し得る。さらなる実施形態において、遺伝子タイピングは、多型の配列を決定すること、およびその配列を、疾患リスクを示唆するコントロール配列と比較することを包含し得る。いくつかの実施形態において、多型は、表 1 または表 2 における遺伝子座から選択される。しかしながら、目的の疾患または状態に関連する任意の遺伝子座が使用され得ることが認識されるだろう。

20

【0017】

いくつかの実施形態において、被験体の 1 つ以上の遺伝子座において決定された遺伝子型に基づいて（例えば、被験体から得られた生物学的サンプルの分析に基づいて）、診断、予後または疾患リスクの評価が、その被験体に対して提供される。いくつかの実施形態において、1 つ以上の遺伝子座における夫婦のそれぞれの遺伝子型に基づいて、ある疾患または状態に関連する遺伝子型（例えば、ある疾患または状態に関連するホモ接合またはヘテロ接合の遺伝子型）を有する 1 人以上の小児を有するリスクに関する評価がその夫婦に提供される。いくつかの実施形態において、被験体または夫婦は、本発明の実施形態に従って決定された遺伝子型に関連して遺伝カウンセリングまたは生殖カウンセリングを求め得る。いくつかの実施形態において、腫瘍または循環腫瘍細胞からの遺伝情報は、予後

30

【0018】

本明細書中に記載される任意の方法または組成物は、本明細書中に記載されるような 1 つ以上の遺伝子座に関連する任意の医学的評価と組み合わせて使用され得ることが認識されるだろう。

【0019】

いくつかの実施形態において、本発明の局面は、多重的なゲノムの捕捉および/または分析（配列決定分析を含む）における系統誤差（バイアス）および/または確率的影響に関連する課題を克服するために有効な方法を提供する。いくつかの実施形態において、本発明の局面は、1 つ以上のサンプリング工程および/または分析工程におけるばらつきを回避するため、減少させるためおよび/または説明するために有用である。例えば、いくつかの実施形態において、標的核酸の提示のばらつきおよび捕捉された標的核酸プール内のヘテロ接合性対立遺伝子の不均等なサンプリングが、克服され得る。

40

【0020】

したがって、いくつかの実施形態において、本開示は、多重捕捉方法における標的核酸の検出においてばらつきを減少させる方法を提供する。他の実施形態において、その方法は、捕捉プールにおける対立遺伝子の提示を改善し、ゆえに、相違する検出結果を改善する。ある特定の実施形態において、本開示は、従来技術の方法と比べて、捕捉プール内の標的核酸（例えば、遺伝子座）のより均一な提示を達成するための、標的核酸の重複領域を捕捉する複数のプローブ（例えば、分子反転プローブ（molecular inverse

50

rsion probes) MIPs)の種々のセットの使用を含む、標的核酸を捕捉するための調製方法を提供する。他の実施形態において、その方法は、例えば、診断目的での、遺伝子座の大規模な並行捕捉に関連するバイアスまたはバイアスのリスクを低下させる。他の実施形態において、複数の遺伝子座の並行検出の再現性を高めるための(例えば、標的核酸の捕捉に対する多型の影響を低下させることによって)方法が提供される。さらなる実施形態において、複数の遺伝子座の並行した分析に対する、プローブ合成および/またはプローブ増幅のばらつきの影響を減少させるための方法が提供される。

【0021】

いくつかの局面によると、複数の遺伝子座を分析する方法が提供される。いくつかの実施形態において、その方法は、複数の標的核酸の各々をプローブセットと接触させる工程(ここで、各プローブセットは、複数の異なるプローブを含み、各プローブは、標的核酸の複数の部分領域のうちの1つの同じ鎖に隣接する核酸と相補的な5'領域および3'領域に隣接する中央領域を有し、ここで、その標的核酸の部分領域は、異なり、各部分領域は、少なくとも1つの他の部分領域と重複する)、複数の核酸(各々、複数の標的核酸の各々に関して異なる部分領域の核酸配列を有する)を単離する工程、ならびに単離された核酸を分析する工程を包含する。

10

【0022】

他の実施形態において、方法は、複数の標的核酸の各々をプローブセットと接触させる工程(ここで、各プローブセットは、複数の異なるプローブを含み、各プローブは、標的核酸の複数の部分領域のうちの1つの同じ鎖に隣接する核酸と相補的な5'領域および3'領域に隣接する中央領域を有し、ここで、その標的核酸の部分領域は、異なり、プローブの5'領域の一部および3'領域の一部は、それぞれ、異なるプローブの5'領域の配列および3'領域の配列を有する)、複数の核酸(各々、複数の標的核酸の各々に関して異なる部分領域の核酸配列を有する)を単離する工程、ならびに単離された核酸を分析する工程を包含する。

20

【0023】

本開示の局面は、ゲノムの捕捉、増幅およびシーケンシングの方法における系統誤差および偶然誤差(バイアス)、すなわち、核酸の捕捉および増幅、ならびにシーケンシングライブラリー内のヘテロ接合対立遺伝子の不均衡な提示における大きなばらつきに関連する問題を克服するための方法の発見に部分的に基づく。したがって、いくつかの実施形態において、本開示は、核酸の捕捉および増幅におけるばらつきに関連するエラーを減少させる方法を提供する。他の実施形態において、その方法は、シーケンシングライブラリー内の対立遺伝子の提示を改善し、ゆえに、相違する検出結果を改善する。ある特定の実施形態において、本開示は、個別の核酸分子を独特にタグ化する識別タグ配列(differentiator tag sequence)の使用を含む、標的核酸(例えば、遺伝子座)を捕捉するための調製方法を提供する。いくつかの実施形態において、その識別タグ配列のおかげで、シーケンシング反応において観察される識別タグ配列と標的配列との組み合わせの出現に基づいたバイアスの検出が可能になる。他の実施形態において、その方法は、例えば、診断目的での、遺伝子座の捕捉、増幅およびシーケンシングに関連するバイアスによって引き起こされるエラー、またはバイアスのリスクを減少させる。

30

40

【0024】

本発明の局面は、アッセイにおいて同定された標的核酸配列が独立して単離された標的核酸に由来するかまたは複数のコピー数の同じ標的核酸分子に由来するかを決定する(例えば、調製工程、例えば、増幅におけるバイアスに起因して)ために有用である配列タグ(識別タグと呼ばれる)を提供することに関する。この情報は、信頼できる配列情報を得るために、生物学的サンプルから独立して単離された閾値数(threshold number)の標的核酸の分析を助けるために使用され得、また、所望の程度の信頼度で遺伝子型の結論(例えば、コール)をもたらすために使用され得る。この情報は、1つ以上の核酸調製工程におけるバイアスを検出するためにも使用され得る。

50

【0025】

いくつかの実施形態において、本明細書中に開示される方法は、例えば、ゲノムの単離、増幅、シーケンシングに関連する、バイアスを減少させることが重要な任意の用途にとって有用である。例えば、不均一な組織サンプルにおける癌変異の検出、母体に循環する胎児DNAにおける変異の検出、および着床前遺伝子診断法において単離された細胞における変異の検出。

【0026】

したがって、いくつかの局面において、被験体の遺伝子タイピングを行う方法が提供される。いくつかの実施形態において、その方法は、独立して単離された少なくとも閾値数の核酸の配列を決定する工程を包含し、ここで、単離された各核酸の配列は、標的核酸配列および識別タグ配列を含み、閾値数は、標的核酸配列と識別タグ配列との独特の組み合わせの数であり、単離された核酸は、標的核酸配列と識別タグ配列との独特の組み合わせを含む場合、独立して単離されたと同定され、ここで、標的核酸配列は、被験体のゲノム遺伝子座の配列である。

10

【0027】

いくつかの実施形態において、単離された核酸は、環状化選択(circularization selection)ベースの調製方法の産物、例えば、分子反転プローブによる捕捉産物である。他の実施形態において、単離された核酸は、増幅ベースの調製方法の産物である。他の実施形態において、単離された核酸は、ハイブリダイゼーションベースの調製方法の産物である。

20

【0028】

環状化選択ベースの調製方法は、目的の領域(標的核酸)を共有結合的に閉じた環状分子に選択的に変換し、次いでその環状分子を、代表的には、任意の環状化されていない直鎖状核酸の除去(通常、例えばエキソヌクレアーゼを用いた、酵素的な除去)によって単離する。目的の領域(標的核酸)に隣接する末端および必要に応じてプライマー部位、例えば、シーケンシングプライマー部位を有するオリゴヌクレオチドプローブ(例えば、分子反転プローブ)を設計する。それらのプローブは、ゲノム標的にハイブリダイズすることができ、酵素を用いることにより、第1に(必要に応じて)プローブの末端間の任意のギャップが埋められ、第2に、プローブが閉じてライゲートされる。環状化の後、代表的には、任意の残りの(非標的)直鎖状核酸が除去され、標的核酸が単離(捕捉)される。環状化選択ベースの調製方法には、分子反転プローブによる捕捉反応および「セクター(selector)」による捕捉反応が含まれる。いくつかの実施形態において、分子反転プローブによる標的核酸の捕捉は、標的核酸における多型の存在を示す。

30

【0029】

増幅ベース(例えば、PCRベースまたはLCRベースなど)の調製方法において、ゲノム遺伝子座(標的核酸)は、1つ以上のオリゴヌクレオチドプライマーを用いて各遺伝子座を選択的に増幅するポリメラーゼ連鎖反応またはリガーゼ連鎖反応(または他の増幅方法)を用いて直接単離される。プライマーは、標的核酸とハイブリダイズするのに十分かつ標的核酸の増幅をプライミングするのに十分、標的配列と相補的であることが理解されるべきである。様々な分野において公知の方法のいずれか1つが、プライマーの設計および合成のために利用され得る。1つ以上のプライマーが、標的配列と完全に相補的であり得る。縮重プライマーもまた使用され得る。プライマーは、標的配列と相補的でないが下流の用途を容易にする追加の核酸(例えば、制限酵素認識部位および識別タグ配列を含む)も含み得る。増幅ベースの方法は、単一標的核酸の増幅および多重増幅(複数の標的核酸の並行した増幅)を含む。

40

【0030】

ハイブリダイゼーションベースの調製方法は、さらなる操作のために標的核酸を選択的に固定化する工程を包含し得る。識別タグ配列を含み、かつ15~170ヌクレオチド長であり得る1つ以上のオリゴヌクレオチド(固定化オリゴヌクレオチド)が使用され、そのオリゴヌクレオチドが、遺伝子座の標的領域にわたってハイブリダイズすることにより

50

、その標的領域が固定化されることが理解されるべきである。いくつかの実施形態において、固定化オリゴヌクレオチドは、ハイブリダイゼーションが行われる前に固定化されるか（例えば、Roche/Nimblegen「配列捕捉」）、または例えば、ストレプトアビジンでコーティングされたマイクロビーズに結合することによってハイブリダイゼーションの後に標的核酸を選択的に固定化するために使用され得る部分（例えば、ビオチン）を含むように調製される（例えば、Agilent「SureSelect」）。

【0031】

本明細書中に記載される環状化、増幅および/またはハイブリダイゼーションベースの方法のいずれかが、本明細書中に記載されるタイリング（tiling）/スタガリング（staggering）、タギング、サイズ検出および/または感度増強アルゴリズムのうちの1つ以上に関連して使用され得ることが認識されるだろう。

10

【0032】

いくつかの実施形態において、本明細書中に開示される方法は、分子反転プローブによる捕捉産物（各捕捉産物は、分子反転プローブおよび標的核酸を含む）の配列を決定する工程（ここで、その分子反転プローブの配列は、識別タグ配列および必要に応じてプライマー配列を含み、標的核酸は、被験体の捕捉されるゲノム遺伝子座である）および分子反転プローブによる捕捉産物の標的核酸配列と識別タグ配列との少なくとも閾値数の独特の組み合わせの配列に基づいて、捕捉されたゲノム遺伝子座において被験体の遺伝子タイピングを行う工程を包含する。

20

【0033】

いくつかの実施形態において、本明細書中に開示される方法は、分子反転プローブによる捕捉産物（各捕捉産物は、分子反転プローブおよび標的核酸を含む）を得る工程（ここで、その分子反転プローブの配列は、識別タグ配列および必要に応じてプライマー配列を含み、ここで、標的核酸は、被験体の捕捉されるゲノム遺伝子座である）、分子反転プローブによる捕捉産物を増幅する工程、ならびに各標的核酸について、分子反転プローブによる捕捉産物の標的核酸配列と識別タグ配列との少なくとも閾値数の独特の組み合わせの配列を決定することによって被験体の遺伝子タイピングを行う工程を包含する。ある特定の実施形態において、上記の得る工程は、分子反転プローブ（各々が独特の識別タグ配列を含む）を用いて被験体のゲノムサンプルから標的核酸を捕捉する工程を包含する。特定の実施形態において、捕捉工程は、標的配列と識別タグ配列との同一の組み合わせを用いて2つ以上の分子反転プローブによる捕捉産物が得られる尤度が所定の値と等しいかまたはそれ未満であり、必要に応じてその所定の値が約0.05である条件下で行われる。

30

【0034】

1つの実施形態において、特定の標的核酸配列に対する閾値数は、その遺伝子型に対する所望の統計学的信頼度に基づいて選択される。いくつかの実施形態において、上記方法は、標的核酸配列と識別タグ配列との独特の組み合わせの数に基づいて、遺伝子型に対する統計学的信頼度を決定する工程をさらに包含する。

【0035】

いくつかの局面によると、複数の遺伝子座を分析する方法が提供される。いくつかの実施形態において、その方法は、分子反転プローブによる複数の捕捉産物（各々、分子反転プローブおよび標的核酸を含む）を得る工程（ここで、その分子反転プローブの配列は、識別タグ配列および必要に応じてプライマー配列（例えば、シーケンシング用または他の伸長反応用のプライマーとして使用される核酸の配列と相補的である配列）を含む）、その分子反転プローブによる複数の捕捉産物を増幅する工程、増幅された複数のものにおける、分子反転プローブによる捕捉産物の標的核酸配列と識別タグ配列との組み合わせの出現数を測定する工程、ならびに標的核酸配列と識別タグ配列との特定の組み合わせの出現数が、所定の値を超える場合、その特定の組み合わせを含む分子反転プローブの増幅におけるバイアスを検出する工程を包含する。いくつかの実施形態において、その方法は、その複数のものにおける標的配列の遺伝子タイピングを行う工程をさらに包含し、ここで、その遺伝子タイピングは、バイアスが検出された場合にそのバイアスを補正する工程を

40

50

包含する。

【0036】

いくつかの実施形態において、標的核酸は、表1から選択される遺伝子（またはその一部）である。いくつかの実施形態において、遺伝子タイピングは、被験体のゲノム（二倍体ゲノム）の1つ以上（両方）の対立遺伝子における標的核酸（例えば、多型配列）の配列を決定することを含む。ある特定の実施形態において、遺伝子タイピングは、被験体の二倍体ゲノムの両方の対立遺伝子における標的核酸の配列を決定することを含み、ここで、その標的核酸は、表1、表2の配列または他の目的の遺伝子座を含むか、またはそれらからなる。

【0037】

いくつかの実施形態において、本発明の局面は、目的のゲノム領域のヌクレオチド配列を決定することなく、これらの領域における核酸の挿入または欠失を同定するための方法および組成物を提供する。本発明の局面は、核酸配列リピート（例えば、ジヌクレオチドリピートまたはトリヌクレオチドリピート）を含むゲノム領域において核酸の挿入または欠失を検出するために特に有用である。しかしながら、本発明は、核酸リピートを分析することに限定されず、任意の目的の標的核酸における挿入または欠失を検出するためにも使用され得る。本発明の局面は、多重アッセイにおいて複数の遺伝子座を分析するために特に有用である。

【0038】

いくつかの実施形態において、本発明の局面は、ゲノム捕捉アッセイにおいて捕捉される標的核酸の量が、予想よりも多いまたは少ないことを決定することに関する。いくつかの実施形態において、予想される量からの統計学的に有意な偏差（例えば、多いまたは少ない）は、目的のゲノム領域における核酸の挿入または欠失の存在を示す。いくつかの実施形態において、その量は、捕捉される核酸分子の数である。いくつかの実施形態において、その量は、サンプル中の独立して捕捉された核酸分子の数である。捕捉される核酸は、サンプルから文字通り捕捉されてもよいし、それらの配列は、サンプル中の元の核酸を実際に捕捉することなく捕捉されてもよいことが認識されるだろう。例えば、核酸配列は、サンプル中の、目的の領域を有する鋳型ベースの核酸伸長を含むアッセイにおいて捕捉され得る。

【0039】

本発明の局面は、ある特定の捕捉手法の効率が、捕捉される核酸の長さに影響されるという認識に基づく。したがって、標的核酸の長さが長くなることまたは短くなること（例えば、反復配列の挿入または欠失に起因して）によって、その核酸の捕捉効率が増加し得る。いくつかの実施形態において、標的核酸の捕捉効率の差（例えば、捕捉効率の統計学的に有意な差）は、標的核酸における挿入または欠失を示唆する。標的核酸に対する捕捉効率は、コントロールの量に対する（例えば、捕捉されたコントロール核酸の量に基づく）、捕捉された核酸の量（例えば、捕捉された核酸分子の数）に基づいて評価され得ることが認識されるだろう。しかしながら、本発明は、この点に限定されず、捕捉効率を評価するための他の手法を使用してもよい。

【0040】

本発明の局面によると、リピート領域全体の配列を決定することに対立するものとして捕捉効率を評価することは、リピート領域のシーケンシングに関連するエラーを減少させる。リピート配列は、種々の条件下において複数のシーケンシング反応を行うことなくおよび結果を慎重に分析することなく標的領域内のリピートの数を正確に決定することを困難にするスタッター（*stutter*）またはスキップをシーケンシング反応において引き起こすことが多い。そのような手順は、煩わしく、標的核酸のハイスループット分析と一致する様式として容易に拡張できない。いくつかの実施形態において、リピート領域は、個別の配列リード（*read*）の長さよりも長いことがあり、単一のリードに基づいた長さの決定を不可能にする。例えば、次世代シーケンシングを用いるとき、リピート領域は、個別の配列リードの長さよりも長いことがあり、単一のリードに基づいた長

10

20

30

40

50

さの決定を不可能にする。したがって、本発明の局面は、標的領域、特に、反復配列を含む標的領域における挿入または欠失を検出する感度を高めるために有用である。

【0041】

いくつかの実施形態において、本発明の局面は、分子反転プローブ（例えば、MIPまたはパドロックプローブ）法を用いてゲノム核酸配列を捕捉すること、および捕捉された配列の量（例えば、数）が予想よりも多いまたは少ないことを決定することに関する。いくつかの実施形態において、捕捉された配列の量（例えば、数）は、コントロールアッセイにおいて捕捉された配列の量（例えば、数）と比較される。そのコントロールアッセイは、既知の配列長（例えば、既知の核酸リピート数）を有する同じ遺伝子座由来の核酸を含むコントロールサンプルを分析することを含み得る。しかしながら、コントロールは、いかなる挿入または欠失も含むと予想されない第2の（例えば、異なる）遺伝子座を分析することを含み得る。その第2の遺伝子座は、調べられている遺伝子座と同じサンプルまたはその長さが予め測定されている異なるサンプルにおいて分析され得る。その第2の遺伝子座は、核酸リピートの存在を特徴としない（ゆえに、リピート配列の挿入または欠失を含むと予想されない）遺伝子座であり得る。

10

【0042】

いくつかの実施形態において、評価されている標的核酸領域は、標的領域（またはその配列）を捕捉するために設計されたプローブの標的化アーム（targeting arm）の同一性によって決定され得る。例えば、MIPプローブの標的化アームは、挿入または欠失を含むと疑われる標的領域に隣接するゲノム領域に対して相補的（例えば、選択的なハイブリダイゼーションおよび/またはポリメラーゼ伸長および/またはライゲーションにとって十分に相補的）であるように設計され得る。2つの標的化アームが、ゲノム核酸の一方の鎖において配列リピートの領域のすぐ隣（例えば、それぞれすぐ5'および3'）である2つの隣接領域に対して相補的（例えば、選択的なハイブリダイゼーションおよび/またはポリメラーゼ伸長および/またはライゲーションにとって十分に相補的）であるように設計され得ることが認識されるだろう。しかしながら、一方または両方の標的化アームは、捕捉される配列が、リピート領域の片側または両側に独特のゲノム配列の領域を含むように、リピート領域から数塩基（例えば、1~5、5~10、10~25、25~50またはそれ以上）上流または下流とハイブリダイズするように設計され得る。次いで、この独特の領域を用いることにより、捕捉された標的を同定することができる（例えば、配列情報またはハイブリダイゼーション情報に基づいて）。

20

30

【0043】

いくつかの実施形態において、2つ以上（例えば、2、3、4、5、6、7、8、9、10個またはそれ以上）の異なる遺伝子座が、単一アッセイ（例えば、多重アッセイ）において並行して調べられ得る。いくつかの実施形態において、各遺伝子座に対する捕捉された核酸の比を用いることにより、核酸の挿入または欠失が、他の遺伝子座と比べて1つの遺伝子座に存在するか否かが決定され得る。例えば、その比は、コントロール配列（例えば、正常である配列またはそれらの遺伝子座について健常表現型に関連すると知られている配列）と比べて挿入または欠失を有しないときの2つの遺伝子座を代表するコントロール比と比較され得る。しかしながら、捕捉された核酸の量は、本明細書中で述べられるような任意の適当なコントロールと比較されてもよい。

40

【0044】

捕捉された配列の遺伝子座は、欠失または挿入を含むと疑われる標的核酸内のリピート領域に対して5'および/または3'の独特の配列の一部を決定することによって同定され得る。これは、捕捉されたリピート領域自体のシーケンシングを必要としない。しかしながら、本発明の局面は、この点に限定されないため、リピート領域の一部または全部が配列決定され得る。

【0045】

本発明の局面は、1つ以上の配列ベースのアッセイ（例えば、SNP検出アッセイ）と組み合わせる（例えば多重形式で）ことにより、被験体の1つ以上の領域の遺伝子型を決

50

定し得る。

【0046】

いくつかの実施形態において、生物学的サンプル中の核酸において多型を検出する方法が提供される。いくつかの実施形態において、その方法は、1つ以上の遺伝子座における捕捉の効率を評価する工程、ならびにその遺伝子座における一方または両方の対立遺伝子がコントロール遺伝子座（例えば、健常表現型に関連するリピート配列の長さを示す遺伝子座）と比べて挿入または欠失を含むか否かを決定する工程を包含する。

【0047】

したがって、本発明の局面は、被験体由来の生物学的サンプル中の標的核酸の捕捉効率を評価することによって、標的核酸が異常な長さを有するか否かを決定するための方法に関し、ここで、基準となる捕捉効率と異なる捕捉効率は、生物学的サンプル中の異常な長さを有する標的核酸の存在を示す。用語「異常な」は、「正常な」長さとの比較に基づく相対的な用語であることが認識されるだろう。いくつかの実施形態において、正常な長さは、正常（例えば、健常者または非保因者の表現型）に関連する長さである。したがって、異常な長さは、正常な長さよりも短いかまたは長い長さである。いくつかの実施形態において、異常な長さの存在は、その遺伝子座が疾患または疾患保因者表現型に関連する高いリスクを示唆する。いくつかの実施形態において、異常な長さは、その被験体が、ある疾患もしくは状態を有するか、またはある疾患もしくは状態（例えば、その遺伝子座に関連するもの）の保因者であることを示唆する。しかしながら、異常な長さの存在を検出することに関する実施形態の記載は、予想される長さまたはコントロールの長さと異なる長さの存在を検出することも支持することが認識されるだろう。

10

20

【0048】

いくつかの実施形態において、本発明の局面は、標的核酸（例えば、標的核酸内のサブターゲット（sub-target）領域）の長さを推定することに関する。いくつかの実施形態において、本発明の局面は、標的核酸に対する検出プローブのハイブリダイゼーションを可能にする条件下で複数の検出プローブと標的核酸を接触させること（ここで、各検出プローブは、標的核酸の第1の領域にハイブリダイズする第1のアームおよび標的核酸の第2の領域にハイブリダイズする第2のアームを含むポリヌクレオチドであり、ここで、第1の領域および第2の領域は、標的核酸の共通の鎖に存在し、第1の領域の5'末端と第2の領域の3'末端との間の標的のヌクレオチド配列は、サブターゲット核酸のヌクレオチド配列である）；および複数の検出プローブとハイブリダイズした複数のサブターゲット核酸を捕捉すること；および複数のサブターゲット核酸におけるあるサブターゲット核酸の出現頻度を測定すること（ここで、複数のサブターゲット核酸におけるあるサブターゲット核酸の出現頻度は、サブターゲット核酸の長さを示す）によって、標的核酸の長さを推定するための方法に関する。核酸の長さを推定するための方法は、標的核酸領域が、既知のコントロール長よりも短いか、中間であるか、または長いかを決定するために、既知の核酸の長さに対する2つ以上の基準となる効率と、標的核酸領域に対する捕捉効率を比較する工程を包含し得ることが認識されるだろう。いくつかの実施形態において、異なる既知の長さの一連の核酸を用いることにより、目的の標的核酸領域の長さを評価するための検量線が提供され得る。

30

40

【0049】

いくつかの実施形態において、欠失または挿入を有すると疑われる標的領域の捕捉効率は、その捕捉効率を、正常の捕捉効率を示す参照と比較することによって測定される。いくつかの実施形態において、捕捉効率は、基準となる捕捉効率よりも低い。いくつかの実施形態において、被験体は、標的領域に挿入を有すると同定される。いくつかの実施形態において、捕捉効率は、基準となる捕捉効率よりも高い。いくつかの実施形態において、被験体は、標的領域に欠失を有すると同定される。いくつかの実施形態において、被験体は、挿入についてヘテロ接合であると同定される。いくつかの実施形態において、被験体は、欠失についてヘテロ接合であると同定される。

【0050】

50

本明細書中に記載される方法（例えば、タイリング/スタガリング、タギング、サイズ検出および/または感度増強）のいずれかのいくつかの実施形態において、本発明の局面は、サブターゲット核酸（またはサブターゲット核酸の配列）を捕捉することに関する。いくつかの実施形態において、分子反転プローブの手法が用いられる。いくつかの実施形態において、分子反転プローブは、その5'末端に第1の標的化アームを含み、その3'末端に第2の標的化アームを含む核酸の直鎖状の一本鎖であり、ここで、その第1の標的化アームは、サブターゲット核酸の一方の末端に隣接する第1の領域と特異的にハイブリダイズすることができ、第2の標的化アームは、標的核酸の同じ鎖におけるサブターゲット核酸の他方の末端に隣接する第2の領域と特異的にハイブリダイズすることができる。いくつかの実施形態において、第1および第2の標的化アームは、約10~約100ヌクレオチド長である。いくつかの実施形態において、第1および第2の標的化アームは、約10~20、20~30、30~40または40~50ヌクレオチド長である。いくつかの実施形態において、第1および第2の標的化アームは、同じ長さを有する。いくつかの実施形態において、第1および第2の標的化アームは、異なる長さを有する。いくつかの実施形態において、プローブセットにおける第1の標的化アームと第2の標的化アームとの各対は、同じ長さを有する。したがって、一方の標的化アームのほうが長い場合、他方は、対応してより短い。これにより、いくつかの実施形態における精度管理工程が、捕捉されたすべてのプローブ/標的配列産物が多重化された複数の捕捉反応の後に同じ長さを有することを保証することが可能になる。いくつかの実施形態において、介在領域が、第1および第2の標的化アームの一方または両方の長さの任意の差異を吸収する（accommodate）ように変更される場合、プローブセットは、同じ長さを有するように設計され得る。

10

20

30

40

50

【0051】

いくつかの実施形態において、第1および第2の標的化アームのハイブリダイゼーション T_m は、同様のものである。いくつかの実施形態において、第1および第2の標的化アームのハイブリダイゼーション T_m は、互いの2~5以内である。いくつかの実施形態において、第1および第2の標的化アームのハイブリダイゼーション T_m は、同一である。いくつかの実施形態において、第1および第2の標的化アームのハイブリダイゼーション T_m は、経験的に決定される最適条件に近いが、必ずしも同一でない。

【0052】

いくつかの実施形態において、分子反転プローブの第1および第2の標的化アームは、異なる T_m を有する。例えば、第1の標的化アーム（分子反転プローブの5'末端におけるもの）の T_m は、第2の標的化アーム（分子反転プローブの3'末端におけるもの）の T_m よりも高いことがある。本発明の局面によると、理論に拘束するつもりはないが、第1の標的化アームに対する T_m が相対的に高いことにより、第1の標的化アームが、第2の標的化アームの3'末端の伸長産物によってハイブリダイゼーション後に置き換えられることを回避するかまたは防ぐのが助けられ得る。本明細書中で使用されるような標的化アームの T_m に対する参照は、相補的配列（例えば、標的化アームの配列と相補的である配列を有する標的核酸の領域）を有する核酸に対する標的化アームのハイブリダイゼーション T_m に関するものであることが認識されるだろう。また、本明細書中に記載される標的化アームの T_m は、任意の適切な方法を用いて計算され得ることも認識されるだろう。例えば、いくつかの実施形態において、ある実験方法（例えば、ゲルシフトアッセイ、ハイブリダイゼーションアッセイ、融解曲線分析法、例えば、インターカレート色素（例えば、二本鎖DNAに結合した色素など）からのシグナルレベルをモニタリングしつつ温度勾配（temperature ramp）を行うことによるSYBR色素を用いるPCR機器における融解曲線分析法）を用いることにより、1つ以上の T_m が経験的に測定され得る。いくつかの実施形態において、最適な T_m は、形成された産物の数を評価すること（例えば、複数のMIPプローブの各々について）、および最適な T_m を、すべての標的化アームに対する T_m のヒストグラムにおける中心点として決定することによって、決

定され得る。いくつかの実施形態において、予測アルゴリズムを用いることにより、 T_m が理論的に決定され得る。いくつかの実施形態において、配列がその標的とハイブリダイズするときのG/CおよびA/T塩基対の数ならびに/またはハイブリダイズ産物の長さに基づいた比較的単純な予測アルゴリズムが用いられ得る(例えば、例えば、 $64.9 + 41 * ([G + C] - 16.4) / (A + T + G + C)$ 、例えば、Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) *Nucleic Acids Res* 6: 3543 - 3557を参照のこと)。いくつかの実施形態において、より複雑なアルゴリズムを用いることにより、塩基のスタッキングのエントロピーおよびエンタルピー、イオン濃度ならびにプライマー濃度の作用が説明され得る(例えば、Santa Lucia J (1998), *Proc Natl Acad Sci USA*, 95: 1460 - 5を参照のこと)。いくつかの実施形態において、アルゴリズムは、改変パラメータ(例えば、塩基対のエントロピー/エンタルピー値に対する最近接パラメータ)を使用し得る。本発明の局面は、この点に限定されないので、任意の適当なアルゴリズムを使用してもよいことが認識されるだろう。しかしながら、種々の方法によって、同じ配列に対して異なる T_m の計算値または予測値がもたらされ得ることも認識されるだろう。したがって、いくつかの実施形態において、異なる T_m について所定の類似度または差異を有するプローブのセットを設計するとき、 T_m の決定または予測における任意の体系的な差異の負の影響を回避するように、プローブセットに対する種々の配列の T_m を決定するために同じ経験的な方法および/または理論的な方法が用いられる。

【0053】

いくつかの実施形態において、第1の標的化アームの T_m は、第2の標的化アームの T_m よりも約1、約2、約3、約4、約5 高いかまたは約5 超高いことがある。いくつかの実施形態において、複数のプローブにおける各プローブ(例えば、5~10個の異なるプローブのセットにおける各プローブ、少なくとも10個の異なるプローブのセットにおける各プローブ、10~50個の異なるプローブのセットにおける各プローブ、50~100個の異なるプローブのセットにおける各プローブ、100~500個の異なるプローブのセットにおける各プローブ、500~1,000個の異なるプローブのセットにおける各プローブ、1,000~1,500個の異なるプローブのセットにおける各プローブ、1,500~2,000個の異なるプローブのセットにおける各プローブ、2,000~3,000個の異なるプローブのセットにおける各プローブ、3,000~5,000個の異なるプローブのセットにおける各プローブ、5,000~10,000個の異なるプローブのセットにおける各プローブまたは少なくとも5,000個の異なるプローブのセットにおける各プローブ)は、独特の第1の標的化アーム(例えば、それらすべてが異なる配列を有する)および独特の第2の標的化アーム(例えば、それらすべてが異なる配列を有する)を有する。いくつかの実施形態において、プローブの少なくとも10%(例えば、プローブの少なくとも25%、25%~50%、50%~75%、75%~90%、90%~95%もしくは95%超またはすべて)について、第1の標的化アームは、第2の標的化アームの相補的配列に対する第2の標的化アームの T_m よりも高い(例えば、約1、約2、約3、約4、約5 高いかまたは約5 超高い)第1の標的化アームの相補的配列に対する T_m を有する。いくつかの実施形態において、第1の標的化アームの各々は、それらのそれぞれの相補的配列に対する同様または同一の T_m を有し、第2の標的化アームの各々は、それらのそれぞれの相補的配列に対する同様または同一の T_m を有する(そして、第1の標的化アームは、第2の標的化アームよりも高い T_m を有する)。例えば、いくつかの実施形態において、第1のアームの T_m は、約58であり得、第2のアームの T_m は、約56であり得る。いくつかの実施形態において、第1のアームの T_m は、約68であり得、第2のアームの T_m は、約65であり得る。いくつかの実施形態において、異なる標的化アームに対する T_m の類似度(例えば、1、2、3、4、5の範囲内)または同一性は、各アームに対して経験的なデータに基づくべきであるか、または各アームに対して同じ予測アルゴリズム(例えば、Wa

10

20

30

40

50

l l a c e , R . B . , S h a f f e r , J . , M u r p h y , R . F . , B o n n e r , J . , H i r o s e , T . , a n d I t a k u r a , K . (1 9 7 9) N u c l e i c A c i d s R e s 6 : 3 5 4 3 - 3 5 5 7 , S a n t a L u c i a J (1 9 9 8) , P r o c N a t l A c a d S c i U S A , 9 5 : 1 4 6 0 - 5 または他のアルゴリズム) に基づくべきであることが認識されるだろう。

【 0 0 5 4 】

いくつかの実施形態において、分子反転プローブの第1の標的化アーム(分子反転プローブの5'末端におけるもの)のT_mは、標的核酸におけるその相補的配列からの第1の標的化アームの置き換えを防ぐのに十分に安定であるように選択される。いくつかの実施形態において、第1の標的化アームのT_mは、50~55、少なくとも55、55~60、少なくとも60、60~65、少なくとも65、少なくとも70、少なくとも75 または少なくとも80 である。上で考察されたように、特定の標的化アームが経験的または理論的に決定され得ることが認識されるだろう。種々の理論モデルを用いることにより、T_mが決定され得、また、特定の配列に対して予想されるT_mは、その予測に使用されるアルゴリズムに応じて異なり得ることが認識されるだろう。いくつかの実施形態において、複数のプローブ中の各プローブ(例えば、5~10個の異なるプローブのセットにおける各プローブ、少なくとも10個の異なるプローブのセットにおける各プローブ、10~50個の異なるプローブのセットにおける各プローブ、50~100個のセット異なるプローブにおける各プローブ、100~500個の異なるプローブのセットにおける各プローブ、または少なくとも500個の異なるプローブのセットにおける各プローブ)は、異なる第1の標的化アーム(例えば、異なる配列)を有するが、異なる第1の標的化アームの各々は、標的核酸におけるその相補的配列に対する同様または同一のT_mを有する。いくつかの実施形態において、異なる標的化アームに対するT_mの類似度(例えば、1、2、3、4、5の範囲内)または同一性は、各アームに対して経験的なデータに基づくべきであるか、または各アームに対して同じ予測アルゴリズム(例えば、Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) Nucleic Acids Res 6:3543-3557, Santa Lucia J (1998), Proc Natl Acad Sci USA, 95:1460-5 または他のアルゴリズム) に基づくべきであることが認識されるだろう。

【 0 0 5 5 】

いくつかの実施形態において、サブターゲット核酸は、核酸リピートを含む。いくつかの実施形態において、核酸リピートは、ジヌクレオチドまたはトリヌクレオチドリピートである。いくつかの実施形態において、サブターゲット核酸は、核酸リピートの異常な増加または減少の非存在下において10~100コピーの核酸リピートを含む。いくつかの実施形態において、サブターゲット核酸は、核酸リピートを含む脆弱X遺伝子座の領域である。いくつかの実施形態において、一方または両方の標的化アームは、核酸リピートの領域のすぐ隣の標的核酸における領域とハイブリダイズする。いくつかの実施形態において、一方または両方の標的化アームは、任意の核酸リピートを含まない領域によって核酸リピートの領域から分離された標的核酸における領域とハイブリダイズする。いくつかの実施形態において、分子反転プローブは、捕捉されるサブターゲット核酸ならびに必要なに応じて第1の標的化アームおよび/または第2の標的化アームを配列決定するために使用され得るプライマー結合領域をさらに含む。

【 0 0 5 6 】

いくつかの実施形態において、本発明の局面は、生物学的サンプル中の複数の異なる標的核酸の長さを評価することに関する。いくつかの実施形態において、その複数の標的核酸は、複数の異なる分子反転プローブを用いて分析される。いくつかの実施形態において、異なる分子反転プローブの各々は、3'および5'末端の各々に第1の標的化アームと第2の標的化アームとの異なる対を含む。いくつかの実施形態において、異なる分子反転プローブの各々は、同じプライマー結合配列を含む。

【0057】

いくつかの実施形態において、本発明の局面は、被験体から得られた生物学的サンプル由来の核酸を分析することに関する。いくつかの実施形態において、その生物学的サンプルは、血液サンプルである。いくつかの実施形態において、その生物学的サンプルは、組織サンプル、特定の細胞集団、腫瘍サンプル、循環腫瘍細胞または環境サンプルである。いくつかの実施形態において、生物学的サンプルは、単一の細胞である。いくつかの実施形態において、核酸は、複数の異なる被験体から得られた生物学的サンプルにおいて分析される。いくつかの実施形態において、生物学的サンプル由来の核酸は、多重反応において分析される。生物学的サンプルは、サンプル中の複数の細胞から得られた複数のコピー数のゲノムを含むことが認識されるだろう。したがって、サンプルは、複数の独立したコピー数の目的の標的核酸領域を含み得、その標的核酸領域の捕捉効率は、本明細書中に記載されるようにそのサイズを評価するために使用され得る。

10

【0058】

いくつかの実施形態において、本発明の局面は、捕捉された標的核酸の量（例えば、捕捉されたサブターゲット核酸配列の量）を測定することによって核酸の捕捉効率を評価することに関する。いくつかの実施形態において、捕捉された標的核酸の量は、独立して捕捉された標的核酸分子の数（例えば、サブターゲット領域の配列を有する独立して捕捉された分子の量）を測定することによって測定される。いくつかの実施形態において、捕捉された標的核酸の量は、捕捉された核酸の基準となる量と比較される。いくつかの実施形態において、基準となる量は、基準となる核酸の独立して捕捉された分子の数を測定することによって測定される。いくつかの実施形態において、基準となる核酸は、欠失または挿入を含むと疑われない生物学的サンプル中の異なる遺伝子座の核酸である。いくつかの実施形態において、基準となる核酸は、捕捉反応に加えられる、既知サイズおよび既知量の核酸である。本明細書中に記載されるように、独立して捕捉された核酸配列の数は、核酸サンプルをプローブの調製物（例えば、本明細書中に記載されるようなMIPプローブ）と接触させることによって測定され得る。その調製物は、複数のコピー数の同じプローブを含み得、よって、複数の独立したコピー数の標的領域が、異なるプローブ分子によって捕捉されることが認識されるだろう。配列を実際に捕捉するプローブ分子の数は、捕捉された分子の量または数を、任意の適当な手法を用いて測定することによって評価され得る。この数は、サンプル中の標的分子の数と、それらの標的分子の捕捉効率の両方を反映するものであり、それは、本明細書中に記載されるような標的分子のサイズに関する。したがって、捕捉効率は、標的核酸の存在量について照らし合わせることによって、例えば、捕捉された標的分子の数または量を適切なコントロール（例えば、コントロール核酸の既知のサイズおよび量、または生物学的サンプル中に同じ量で存在するはずであり、いずれの挿入もしくは欠失も含まないと予想される、異なる遺伝子座）と比較することによって、評価され得る。他の因子（例えば、領域の配列、GC含有量、二次構造の存在など）も、特定の標的核酸領域の捕捉効率に影響し得ることが認識されるだろう。しかしながら、これらの因子は、適切なコントロール（例えば、同様の特性を有する公知の配列、同じ配列、同じ頻度で生物学的サンプル中に存在すると予想される他のゲノム配列など、またはそれらの任意の組み合わせ）を用いることによって説明され得る。

20

30

40

【0059】

いくつかの実施形態において、本発明の局面は、ある遺伝子座に対する捕捉効率が、基準となる捕捉効率と統計学的に有意に異なる場合に、その遺伝子座の1つ以上の対立遺伝子に挿入または欠失を有すると被験体を同定することに関する。

【0060】

本明細書中に記載される捕捉手法のいずれか（例えば、MIPによる捕捉手法）のために用いられるハイブリダイゼーション条件は、公知のハイブリダイゼーション緩衝液およびハイブリダイゼーション条件に基づき得ることが認識されるだろう。

【0061】

いくつかの実施形態において、本明細書中に開示される方法は、欠失または挿入の検出

50

が重要である任意の用途にとって有用である。

【0062】

いくつかの実施形態において、本発明の局面は、核酸配列の分析を、サンプリングされる核酸のタイプで異なる系統誤差を有する2つ以上の異なる核酸調製手法からの結果に基づかせることに関する。本発明によると、体系的であり、かつ単純に核酸の捕捉中または増幅中の確率的影響に起因しない、種々の配列バイアスを種々の手法は有する。したがって、核酸調製における変動を克服するために必要な過剰サンプリングの程度は、そのバイアスを克服するのに十分であること（例えば、2～5倍、5～10倍、5～15倍、15～20倍、20～30倍、30～50倍またはその中間からそれより高い倍数の過剰サンプリング）が必要である。

10

【0063】

いくつかの実施形態によると、種々の手法は、種々の特徴的なバイアスまたは系統誤差を有する。例えば、1つの手法は、サンプル分析を目的の遺伝子座における1つの特定の対立遺伝子に偏らせ得るのに対し、別の手法は、サンプル分析を同じ遺伝子座の別の対立遺伝子に偏らせ得る。したがって、同じサンプルが、配列分析にむけて核酸を調製するために使用される手法のタイプに応じて、異なると同定され得る。各手法が、目的の多型配列に対して異なる相対的な感度を有するので、これは、感度の限界を効率的に表す。

【0064】

本発明の局面によると、核酸分析の感度は、異なる核酸調製工程からの配列を組み合わせ、組み合わせられた配列情報を診断アッセイのために（例えば、被験体が目的の遺伝子座においてホモ接合であるかまたはヘテロ接合であるかに関するコールを生成するために）用いることによって、高められ得る。

20

【0065】

いくつかの実施形態において、本発明は、生物学的サンプルに対して第1の調製方法を用いて標的核酸の第1の調製物を得て、生物学的サンプルに対して第2の調製方法を用いて標的核酸の第2の調製物を得て、第1の核酸調製物と第2の核酸調製物との両方において得られた配列をアッセイして、そして第1の核酸調製物と第2の核酸調製物との両方からの配列情報を用いることにより、生物学的サンプル中の標的核酸の遺伝子型を決定することによって、核酸検出アッセイの感度を高める方法を提供し、ここで、その第1の調製方法および第2の調製方法は、異なる体系的な配列バイアスを有する。いくつかの実施形態において、第1の核酸調製物および第2の核酸調製物は、配列アッセイを行う前に組み合わせられる。いくつかの実施形態において、別個の配列アッセイが、第1の核酸調製物および第2の核酸調製物において行われ、両方のアッセイからの配列情報が、組み合わせることにより、生物学的サンプル中の標的核酸の遺伝子型が決定される。いくつかの実施形態において、第1の調製方法は、増幅ベース、ハイブリダイゼーションベース、または環状プローブベースの調製方法である。いくつかの実施形態において、第2の方法は、増幅ベース、ハイブリダイゼーションベース、または環状プローブベースの調製方法である。いくつかの実施形態において、第1および第2の方法は、異なるタイプの方法である（例えば、それらのうちの一方だけが、増幅ベース、ハイブリダイゼーションベース、または環状プローブベースの調製方法であり、他方の方法は、他の2つのタイプの方法のうちの一つである）。したがって、いくつかの実施形態において、第2の調製方法は、増幅ベース、ハイブリダイゼーションベース、または環状プローブベースの調製方法であるが、但し、第2の方法は、第1の方法と異なる。しかしながら、いくつかの実施形態において、両方の方法が、異なる方法であるならば（例えば、両方ともが、増幅ベースまたはハイブリダイゼーションベースの方法であるが、異なるタイプの増幅方法またはハイブリダイゼーション方法、例えば、異なる相対的なバイアスを有する方法であるならば）、それらの両方の方法は、同じタイプの方法であってもよい。

30

40

【0066】

増幅ベース（例えば、PCRベースまたはLCRベースなど）の調製方法において、ゲノム遺伝子座（標的核酸）は、オリゴヌクレオチドプライマー対を用いて各遺伝子座を選

50

択的に増幅するポリメラーゼ連鎖反応またはリガーゼ連鎖反応（または他の増幅方法）を用いて直接単離される。プライマーは、標的核酸とハイブリダイズするのに十分、および標的核酸の増幅をプライミングするのに十分、標的配列と相補的であることが理解されるべきである。様々な分野において公知の方法のいずれか1つが、プライマーの設計および合成のために利用され得る。プライマーの一方または両方が、標的配列と完全に相補的であり得る。縮重プライマーも使用され得る。プライマーは、標的配列と相補的でないが下流の用途を容易にする追加の核酸（例えば、制限酵素認識部位および識別配列（例えば、ソース配列）を含む）も含み得る。PCRベースの方法には、単一の標的核酸の増幅および多重増幅（複数の標的核酸の並行した増幅）が含まれ得る。

【0067】

ハイブリダイゼーションベースの調製方法は、さらなる操作のために標的核酸を選択的に固定化する工程を包含し得る。いくつかの実施形態において10~200ヌクレオチド長であり得る1つ以上のオリゴヌクレオチド（固定化オリゴヌクレオチド）が使用され、そのオリゴヌクレオチドは、遺伝子座の標的領域にわたってハイブリダイズすることにより、その標的領域が固定化されることが理解されるべきである。いくつかの実施形態において、固定化オリゴヌクレオチドは、ハイブリダイゼーションが行われる前に固定化されるか（例えば、Roche/Nimblegen「配列捕捉」）、または例えば、ストレプトアビジンでコーティングされたマイクロビーズに結合することによってハイブリダイゼーションの後に標的核酸を選択的に固定化するために使用され得る部分（例えば、ビオチン）を含むように調製される（例えば、Agilent「SureSelect」）。

【0068】

環状化選択ベースの調製方法は、目的の各領域を共有結合的に閉じた環状分子に選択的に変換し、次いで、その環状分子は、任意の環状化されていない直鎖状核酸を除去すること（通常、例えばエキソヌクラーゼを用いた、酵素的な除去）によって単離される。目的の領域に隣接する末端を有するオリゴヌクレオチドプローブが設計される。それらのプローブは、ゲノム標的へのハイブリダイズを可能にし、酵素を用いることにより、第1に（必要に応じて）、プローブ末端間の任意のギャップを埋め、第2に、閉じたプローブをライゲートする。いくつかの実施形態において、環状化の後、任意の残りの（非標的）直鎖状核酸が除去することにより、標的核酸の単離（捕捉）が生じ得る。環状化選択ベースの調製方法は、分子反転プローブによる捕捉反応および「セクター」による捕捉反応を含む。しかしながら、本発明の局面は、この点に限定されないもので、他の手法を使用してもよい。いくつかの実施形態において、標的核酸の分子反転プローブによる捕捉は、標的核酸における多型の存在を示す。

【0069】

各調製手法のバイアスプロファイルを評価するためおよび比較するために、様々な方法が使用され得る。次世代シーケンシングを用いることにより、ある特定の調製方法から得られた単離された各標的核酸の存在量が定量的に測定され得る。この存在量は、コントロール存在量の値（例えば、標的核酸の既知の出発存在量）および/または代替の調製方法を使用することによって測定された存在量と比較され得る。例えば、標的核酸のセットは、3つの調製方法のうち1つ以上によって単離され得る；その標的核酸は、増幅法を用いてx回、ハイブリダイゼーション濃縮（hybridization enrichment）法を用いてy回、および環状化選択法を用いてz回、観察され得る。ペアワイズ相関係数（pairwise correlation coefficient）は、各存在量の値の間（例えば、xとy、xとzおよびyとzの間）についてコンピュータで計算されることにより、調製方法の対の間の核酸単離におけるバイアスが評価され得る。単離のメカニズムが各アプローチにおいて異なるので、存在量は、通常、互いに異なるものであり、大部分は互いに相関関係がない。

【0070】

いくつかの実施形態において、本発明は、生物学的サンプル中の標的核酸を代表する核酸調製物を得る方法を提供し、その方法は、生物学的サンプルに対して第1の調製方法を

10

20

30

40

50

用いて標的核酸の第1調製物を得て、生物学的サンプルに対して第2の調製方法を用いて標的核酸の第2の調製物を得て、そしてその第1の核酸調製物および第2の核酸調製物を混合することにより、その生物学的サンプル中の標的核酸を代表する混合された調製物を得ることによる、方法である。

【0071】

本明細書中に記載されるいずれかの方法のいくつかの実施形態において、標的核酸の第3の調製物が、第1の調製方法および第2の調製方法とは異なる第3の調製方法を用いて得られ、ここで、第1、第2および第3の調製方法のすべてが、異なる体系的な配列バイアスを有する。本明細書中に記載されるいずれかの方法のいくつかの実施形態において、異なる調製方法が、多重核酸分析の感度を高めるために生物学的サンプル中の複数の異なる遺伝子座に対して用いられる。いくつかの実施形態において、標的核酸は、表1から選択される遺伝子の配列を有する。

10

【0072】

しかしながら、本発明の遺伝子タイピング方法は、いくつかの工程を含み得、それらの各々は、独立して、本明細書中に記載される1つ以上の異なる調製手法を含み得ることが認識されるだろう。いくつかの実施形態において、核酸調製物は、本明細書中に記載される1つ以上の（例えば、2、3、4、5つまたはそれ以上の）異なる手法（例えば、増幅、ハイブリダイゼーションによる捕捉、環状プローブによる捕捉など、またはそれらの任意の組み合わせ）を用いて得られ得、その核酸調製物は、最初の調製物に対して用いられた手法とは独立して選択される1つ以上の異なる手法（例えば、増幅、ハイブリダイゼーションによる捕捉、環状プローブによる捕捉など、またはそれらの任意の組み合わせ）を用いて分析され得る。

20

【0073】

いくつかの実施形態において、本発明の局面は、核酸アッセイの感度を高めるための組成物、キット、デバイスおよび分析方法も提供する。本発明の局面は、遺伝子タイピング分析の信頼水準を高めるために特に有用である。しかしながら、本発明の局面は、任意の適当な核酸分析（例えば、2つ以上の配列バリエーションがサンプル中に存在するか否かを決定するように設計された核酸分析であるがこれに限定されない）において使用され得る。

【0074】

いくつかの実施形態において、本発明の局面は、複数の核酸プローブ（例えば、10～50個、50～100個、100～250個、250～500個、500～1,000個、1,000～2,000個、2,000～5,000個、5,000～7,500個、7,500～10,000個またはそれより少ないか、それより多いか、もしくは中間の数の異なるプローブ）に関する。いくつかの実施形態において、各プローブ、またはプローブの各サブセット（例えば、10～25%、25～50%、50～75%、75～90%または90～99%）は、異なる第1の標的化アームを有する。いくつかの実施形態において、各プローブ、またはプローブのサブセットの各プローブ（例えば、10～25%、25～50%、50～75%、75～90%または90～99%）は、異なる第2の標的化アームを有する。いくつかの実施形態において、第1および第2の標的化アームは、同じ存在配列によって分断される。いくつかの実施形態において、第1および第2の標的化アームは、それぞれの標的核酸（例えば、ゲノム遺伝子座）において同じ長さまたは類似の長さ（例えば、核酸の数、例えば、0～25、25～50、50～100、100～250、250～500、500～1,000、1,000～2,500またはそれより長いもしくは中間の数のヌクレオチド）によって分断される標的核酸配列と相補的である。いくつかの実施形態において、各プローブ、またはプローブのサブセット（例えば、10～25%、25～50%、50～75%、75～90%または90～99%）は、第1のプライマー結合配列を含む。いくつかの実施形態において、そのプライマー結合配列は、同じである（例えば、それは、シーケンシング反応または他の伸長反応をプライミングするために使用され得る）。いくつかの実施形態において、各プローブ、またはプローブのサブセット（例えば、10～25%、25～50%、50～75%、75～90%

30

40

50

または90～99%)は、独特の識別配列タグ(例えば、予め決められており、各プローブを識別するために使用することができる、識別配列タグ)を含む。

【0075】

いくつかの実施形態において、本明細書中に開示される方法は、感度が重要な任意の用途にとって有用である。例えば、不均一な組織サンプルにおける癌変異の検出、母体に循環する胎児DNAにおける変異の検出、および着床前遺伝子診断法において単離された細胞における変異の検出。

【0076】

本発明のいくつかの局面によると、生物学的サンプル中の核酸における多型を検出する方法が提供される。いくつかの実施形態において、その方法は、生物学的サンプルに対してある調製方法(例えば、本明細書中に開示される調製方法のいずれか)を用いて核酸調製物を得る工程、ならびにその核酸調製物に対して分子反転プローブ捕捉反応を行う工程を包含し、ここで、その核酸調製物の標的核酸の分子反転プローブによる捕捉(例えば、変異検出MIPを用いる捕捉)は、標的核酸における変異(多型)の存在を示唆し、ここで、必要に応じてその多型は、表2から選択される。

10

【0077】

本発明のいくつかの局面によると、生物学的サンプル中の核酸の遺伝子タイピングを行う方法が提供される。いくつかの実施形態において、その方法は、生物学的サンプルに対してある調製方法を用いて核酸調製物を得る工程、その核酸調製物の標的核酸をシーケンシングする工程、ならびに生物学的サンプルに対して分子反転プローブによる捕捉反応を行う工程(ここで、生物学的サンプル中の標的核酸の分子反転プローブによる捕捉は、標的核酸における多型の存在を示す)、そのシーケンシング反応および捕捉反応の結果に基づいて標的核酸の遺伝子タイピングを行う工程を包含する。

20

【0078】

本明細書中に開示される方法のいくつかの実施形態において、標的核酸は、表1から選択される遺伝子の配列を有する。

【0079】

本明細書中に記載される任意の1つ以上の実施形態は、複数の遺伝マーカーを並行して評価するために使用されることが認識されるだろう。したがって、いくつかの実施形態において、本発明の局面は、複数の異なる遺伝子座における1つ以上のマーカー(例えば、1つ以上の対立遺伝子)の存在を並行して決定することに関する。したがって、複数の遺伝性障害のリスクまたは存在が、並行して評価され得る。いくつかの実施形態において、1つ以上の遺伝性障害を有する出生児を有するリスクが評価され得る。いくつかの実施形態において、親または子の生物学的サンプル(例えば、着床前、出生前、出生時または出生後の段階におけるサンプル)に対して評価が行われ得る。いくつかの実施形態において、本開示は、患者サンプル(例えば、血液、着床前胚、絨毛または羊水サンプル)から複数の遺伝子座(例えば、表1または2から選択される複数の標的核酸)を分析するための方法を提供する。患者または被験体は、ヒトであり得る。しかしながら、本発明の局面がこの点に限定されないので、本発明の局面は、ヒトに限定されず、他の種(例えば、哺乳動物、鳥類、爬虫類、他の脊椎動物または無脊椎動物)にも適用され得る。被験体または患者は、男性(雄)または女性(雌)であり得る。いくつかの実施形態において、生殖遺伝カウンセリングに関連して、夫婦の男性メンバーおよび女性メンバー由来のサンプルが分析され得る。いくつかの実施形態において、例えば、動物育種プログラムに関連して、適合する繁殖パートナーもしくは最適な繁殖パートナーを決定するためか、または特定の形質に対する戦略を決定するため、あるいは1つ以上の疾患または状態を回避するために、複数の雄および雌の被験体由来のサンプルが分析され得る。したがって、1つ以上の本発明の実施形態から得られた情報に基づいて、生殖リスクが決定され得、そして/または生殖に関するアドバイスが提供され得る。

30

40

【0080】

しかしながら、本発明の局面は、目的の遺伝子座における1つ以上の対立遺伝子の存在

50

が医学的判断（例えば、疾患のリスクまたは検出、疾患の予後、治療の選択、治療のモニタリングなど）と関連がある任意の医学的評価に関連して使用され得ることが認識されるだろう。本発明のさらなる局面は、腫瘍組織もしくは循環腫瘍細胞における、癌を引き起こす細胞経路内もしくは処置レジメンの有効性を予測する細胞経路内の変異の検出、または環境内もしくは被験体、例えば、ヒト被験体から得られたサンプル中の病原生物の検出および同定に関連して使用され得る。

【0081】

これらおよび他の本発明の局面は、以下の説明ならびに非限定的な実施例および図面においてより詳細に説明される。

【図面の簡単な説明】

10

【0082】

【図1】図1は、タイリングプローブのレイアウトの非限定的な実施形態を图示している。

【図2】図2は、ずらされた（staggered）プローブのレイアウトの非限定的な実施形態を图示している。

【図3】図3は、交互にずらされた（alternating staggered）プローブのレイアウトの非限定的な実施形態を图示している。

【図4】図4A、BおよびCは、識別タグ配列と標的配列とを組み合わせるための様々な非限定的な方法を表している（NNNNは、識別タグ配列を表している）。

【図5】図5は、標的配列および識別タグ配列に基づいて遺伝子タイピングするための非限定的な方法を表している。

20

【図6】図6は、MIP捕捉反応のシミュレーションの非限定的な結果を表している。

【図7】図7は、シーケンシングカバレッジ（sequencing coverage）の非限定的なグラフを表している。

【図8】図8は、MIPを用いるとき、短い配列が長い配列よりも高効率で捕捉されることを图示している。

【図9】図9は、反復領域（太い波線）と、隣接する独特の配列（太い直線（straight line））との両方を含む領域のパドロック（MIP）捕捉の非限定的なスキームを图示している。

【図10】図10は、標的ギャップサイズと反復領域のリードの相対数との間の非限定的な仮説上の関係性を图示している。

30

【図11A】図11Aは、二倍体ゲノム由来のFMR1リピート領域のMIP捕捉を表している。

【図11B】図11Bは、MIP捕捉プローブおよび独特の識別タグを用いた、二倍体ゲノムにおけるFMR1リピート領域の長さの二対立遺伝子の分離のための調製方法を表している。

【図11C】図11Cは、二倍体ゲノムにおけるFMR1リピート領域の長さの分析を表している。

【図12】図12は、本発明のアルゴリズムの実施形態の概略図である。

【図13】図13は、MIP捕捉を用いた、標的1つあたりの存在量のグラフの非限定的な例を图示している；そして

40

【図14】図14は、2つのMIP捕捉反応の相関関係の非限定的なグラフを示している。

【発明を実施するための形態】

【0083】

詳細な説明

本発明の局面は、遺伝子型を評価するため、特に、被験体の1つ以上の遺伝子座の対立遺伝子の同一性（または二倍体生物の同一性）を決定するための、調製および分析の方法および組成物に関する。本発明の局面は、遺伝分析におけるあいまいさおよびエラーの種々の起源の特定に部分的に基づき、また、遺伝分析の種々の段階におけるこれらのエラー

50

およびあいまいさを回避するため、減少させるため、認識するため、および/または解決するための1つ以上のアプローチの特定に部分的に基づく。本発明の局面は、核酸評価技術の1つ以上の調製工程および/または分析工程に関連するバイアスおよび/または確率的変動に対処するための方法および組成物に関する。いくつかの実施形態において、調製方法は、遺伝分析の結果を歪めるバイアスのリスクを回避するかまたは減少させるように適合され得る。いくつかの実施形態において、分析方法は、誤解釈を引き起こし得るデータの変動(例えば、不正確なコール(例えば、被験体が実際にはヘテロ接合であるときのホモ接合のコールまたは被験体が実際にはホモ接合であるときのヘテロ接合のコール))を認識するようおよび補正するよう適合され得る。本発明の方法は、任意のタイプの変異、例えば、一塩基の変異(例えば、挿入、欠失、トランスポージョンまたはトランジションなど)、複数の塩基の挿入、欠失、重複、逆位および/もしくは他の任意の変異、またはそれらの組み合わせに対して使用され得る。

10

【0084】

いくつかの実施形態において、代表的なシーケンシングリードよりもコア配列のリピートの長さが長く、それにより、1つ以上のコア配列単位の欠失または重複が生じているか否かをもっぱら配列リードに基づいて決定することが困難である場合、さらなる手法または代替の手法を用いて、そのコア配列の複数のリピートを特徴とする遺伝子座が扱われ得る。

【0085】

いくつかの実施形態において、アッセイ結果の高い信頼度は、i)異なるバイアスを有する(例えば、異なるバイアスを有すると知られている)2つ以上の異なる調製手法および/もしくは分析手法を選択すること、ii)2つ以上の異なる手法を用いて患者サンプルを評価すること、iii)2つ以上の異なる手法からの結果を比較すること、ならびに/またはiv)それらの結果がその2つ以上の異なる手法に対して矛盾がないか否かを決定することによって得られ得る。いくつかの実施形態において、工程(iv)における決定が、それらの結果に矛盾がない(例えば、同じである)ことを示唆する場合、そのアッセイ結果の高い信頼度が得られる。他の実施形態において、工程(iv)における決定が、それらの結果に矛盾がある(例えば、それらの結果があいまいである)ことを示唆する場合、工程(i)において選択された2つ以上の異なる調製手法および/または分析手法と比べて異なるバイアスを有する(例えば、異なるバイアスを有すると知られている)1つ以上のさらなる調製手法および/または分析手法を用いることにより患者サンプルが評価され、そしてその1つ以上のさらなる調製手法および/または分析手法の結果を、工程(ii)からの結果と比較することにより、矛盾が解消される。

20

30

【0086】

いくつかの実施形態において、2つ以上の独立したサンプルが、被験体から得られ得、独立して分析され得る。いくつかの実施形態において、2つ以上の独立したサンプルは、ほぼ同時に得られる。いくつかの実施形態において、2つ以上の独立したサンプルは、複数の異なる時点において得られる。いくつかの実施形態において、2つ以上の独立したサンプルを使用することにより、確率的な測定ノイズの排除、正規化および/または定量化が促進される。2つ以上の独立したサンプルが、本明細書中に開示される方法のいずれか(例えば、ヒトまたは他の動物被験体における病原体のプロファイリング、腫瘍の進行/後退のモニタリング、循環腫瘍細胞の分析、母体循環系中の胎児細胞の分析、および環境内の病原体の分析/モニタリング/プロファイリングのための方法を含む)に関して得られ得ることが認識されるべきである。

40

【0087】

いくつかの実施形態において、本明細書中に記載される手法の1つ以上が、複数の患者サンプルを並行して評価するために単一のアッセイプロトコルにおいて組み合わせられ得る。

【0088】

本発明の局面は、複数の患者サンプルの、費用効果は高いが信頼できるハイスループッ

50

トの遺伝子タイピング（例えば、並行して行われるもの、例えば、多重反応におけるもの）にとって有用であることが認識されるだろう。いくつかの実施形態において、本発明の局面は、多重分析におけるエラーの頻度を低下させるために有用である。ある特定の実施形態は、複数の反応（例えば、複数の遺伝子座および/または複数の患者サンプル）が処理されている場合に、特に有用であり得る。例えば、10～25、25～50、50～75、75～100個またはそれ以上の遺伝子座が、並行して処理され得る任意の数の被験体サンプル（例えば、1～25、25～50、50～100、100～500、500～1,000、1,000～2,500、2,500～5,000個またはそれ以上もしくは中間の数の患者サンプル）から、各被験体に対して評価され得る。本発明の種々の実施形態は、2つ以上の標的捕捉反応および/または2つ以上の患者サンプル分析を単一の多重反応において並行して行うことを含むことが認識されるだろう。例えば、いくつかの実施形態において、複数の捕捉反応（例えば、異なる標的遺伝子座に対して異なる捕捉プローブを用いる反応）が、単一の患者サンプルに対する単一の多重反応において行われ得る。いくつかの実施形態において、複数の患者サンプルのうちの各1つから捕捉された複数の核酸が、単一の多重分析反応において組み合わせられ得る。いくつかの実施形態において、異なる被験体からのサンプルは、被験体特異的（例えば、患者特異的）タグ（例えば、独特の配列タグ）でタグ化され、各産物からの情報は、同定された被験体に対して割り当てられ得る。いくつかの実施形態において、各患者サンプルに対して用いられる異なる捕捉プローブの各々は、共通の患者特異的タグを有する。いくつかの実施形態において、捕捉プローブは、患者特異的タグを有しないが、各被験体からの捕捉された産物は、患者特異的タグで標識された1つの増幅プライマーまたは増幅プライマーの対を用いて増幅され得る。本発明の局面は、この点に限定されないもので、単一の患者サンプルからの捕捉された産物と患者特異的タグを関連づけるための他の手法を使用してもよい。本明細書中で使用される患者特異的タグは、特定のアッセイにおいて、同定された患者に割り当てられる独特のタグのことを指し得ることが認識されるだろう。同じタグが、異なる患者サンプルセット（例えば、異なる患者からのサンプル）（その各々は、それらのタグのうちの1つに割り当てられる）を用いた別個の多重分析において使用され得る。いくつかの実施形態において、1つのアッセイから次のアッセイへの汚染のリスクを低下させるため、および特定のアッセイに存在すると予想されないタグの存在に基づいて汚染の検出を可能にするために、独特のタグの種々のセットが、連続的な（例えば、交互の）多重反応において使用され得る。

10

20

30

40

50

【0089】

本発明の実施形態は、いくつかの異なる状況：生殖の状況、疾患スクリーニング、癌を有する被験体の同定、ある疾患に対して高リスクを有する被験体の同定、いくつかの因子の1つ以上に従う被験体の集団の層別化、例えば、特定の薬物に対する応答性、特定の薬物に対する有害反応（または、ゆえにリスク）を欠くかもしくは欠かないこと、および/または医学的記録のための情報の提供（例えば、1つ以上の遺伝子座におけるホモ接合性、ヘテロ接合性）のいずれかに対して使用され得る。本発明は、患者サンプルのゲノム分析に限定されないことが認識されるだろう。例えば、本発明の局面は、病原体を検出するための環境サンプルのハイスループット遺伝分析にとって有用であり得る。

【0090】

いくつかの実施形態において、本明細書中に開示される方法は、1つ以上の遺伝性障害の診断にとって有用である。いくつかの実施形態において、本明細書中に開示される方法を用いて診断され得る遺伝性障害は、Ashkenazi Jewish集団に蔓延している遺伝性障害である。いくつかの実施形態において、それらの遺伝性障害は：21-ヒドロキシラーゼ欠損の(Deficient)先天性副腎過形成；ABC8関連の高インスリン症；アルファ-サラセミア(Constant SpringおよびMR関連を含む)；アリアルスルファターゼA欠損の異染性白質ジストロフィ(Metyachromatic Leukodystrophy)；ピオチニダーゼ欠損症-ホロカルボキシラーゼ合成酵素欠損症；ブルーム症候群；カナバン病；CFTR関連障害-嚢胞性線維

症；シトルリン血症Ⅰ型；複合MMA & ホモシスチン尿症 - db1C；ジストロフィン異常症（DMD & BMD）；家族性自律神経障害；ファンコニー貧血 - FANCC；古典的ガラクトース血症：ガラクトキナーゼ欠損症（Deficiency）およびガラクトースエピメラーゼ欠損症；ゴーシェ病；GJB2関連DFNB1非症候性難聴および聴覚消失；グルタル酸血症Ⅰ型；異常ヘモグロビン症ベータ鎖障害；糖原病ⅠA型；メーブルシロップ尿症；ⅠA、ⅠB、2、3型；中鎖アシル - コエンザイムA；デヒドロゲナーゼ欠損 - MCADD；メチルマロン酸血症；ムコリピド - シスⅣ；ネマリンミオパシー；ニーマン・ピック（Niemann - Pick）A型 - 酸性スフィンゴミエリナーゼ欠損症；非ケト - シス型高グリシン血症 - グリシン脳症；オルニチントランスカルバミラーゼ欠損症；PKUフェニルアラニンヒドロキシラーゼ欠損症；プロピオン酸血症；短鎖アシル - CoAデヒドロゲナーゼ欠損症 - SCADD；スミス・レムリ・オピッツ症候群；脊髄性筋萎縮症（SMN1） - SMA；テイ・サックス - HexA欠損症；アッシャー症候群（Usher Syndrome）Ⅰ型（ⅠB型、ⅠC型、ⅠD型、ⅠF型、ⅠG型）；X連鎖精神遅滞ARX関連障害；小脳形成不全（Cerebellar Hypoplasia）および特有の（sistinctive）顔貌を伴うX連鎖精神遅滞；X連鎖精神遅滞；9、21、30、46、58、63、88、89を含む；X連鎖精神遅滞：FM1関連障害 - FRXA、脆弱X MR；X連鎖SMR：レンベニング症候群1；Zellweger Spectrum障害 - Zellweger、NALDおよび/または乳児レフサム病を含むペルオキシソームの二機能性酵素欠損症から選択される。しかしながら、これらのすべて、これらのサブセット、他の遺伝子またはそれらの組み合わせが、使用され得る。

【0091】

いくつかの局面によると、本開示は、多重診断方法に関する。いくつかの実施形態において、多重診断方法は、複数の遺伝子座（例えば、表1の遺伝子座）を並行して捕捉する工程を包含する。いくつかの実施形態において、遺伝子座は、1つ以上の多型（例えば、表2の多型）を有し、それらの遺伝子型は、疾患を引き起こす対立遺伝子に対応する。したがって、いくつかの実施形態において、本開示は、複数の遺伝性障害を並行して評価するための方法を提供する。

【0092】

いくつかの実施形態において、着床前、出生前、出生時または出生後の段階において複数の遺伝性障害を並行して診断するための方法が提供される。いくつかの実施形態において、本開示は、患者サンプル（例えば、血液、着床前胚、絨毛または羊水サンプル）から複数の遺伝子座（例えば、表1から選択される複数の標的核酸）を分析するための方法を提供する。患者または被験体は、ヒトであり得る。しかしながら、本発明の局面は、この点に限定されないので、本発明の局面は、ヒトに限定されず、他の種（例えば、哺乳動物、鳥類、爬虫類、他の脊椎動物または無脊椎動物）にも適用され得る。被験体または患者は、男性（雄）または女性（雌）であり得る。いくつかの実施形態において、生殖遺伝カウンセリングに関連して、夫婦の男性メンバーおよび女性メンバー由来のサンプルが分析され得る。いくつかの実施形態において、例えば、動物育種プログラムに関連して、適合する繁殖パートナーもしくは最適な繁殖パートナーを決定するためか、または特定の形質に対するストラテジーを決定するため、あるいは1つ以上の疾患または状態を回避するために、複数の雄および雌の被験体由来のサンプルが分析され得る。

【0093】

しかしながら、他の任意の疾患、および/または疾患もしくは障害（アレルギー、処置に対する応答性、処置および予後に対する癌腫瘍のプロファイリング、患者の感染のモニタリングおよび同定、ならびに環境内の病原体のモニタリングが挙げられるがこれらに限定されない）に対する危険因子も研究され得ることが認識されるだろう。

【0094】

1. 多重増幅反応における提示のバイアスの減少；

いくつかの実施形態において、本発明の局面は、例えば診断目的での、遺伝子座の多重

検出において、バイアスを減少させる方法および再現性を高める方法に関する。

【0095】

潜在的に複雑な混合物中の特定の核酸配列を検出するためまたは増幅するために、分子反転プローブ技術が使用される。分子反転プローブの使用は、単一ヌクレオチド多型の検出 (Hardenbolら、2005 Genome Res 15:269-75) およびエキソンの大きなセットの調製増幅 (Porrecaら、2007 Nat Methods 4:931-6, Krishnakumaraら、2008 Proc Natl Acad Sci USA 105:9296-301) について証明されている。一般に数千の標的が、数千のプローブを含む単一反応において捕捉され得るので、上記方法の主要な利点の1つは、高い多重化の程度に対して能力がある点である。しかしながら、例えば、増幅効率に関連する課題 (例えば、Turner EHら、Nat Methods 2009 Apr 6:1-2を参照のこと) のせいで、研究および診断における上記方法の実際の有用性は限定されている。

10

【0096】

本開示の局面は、多重的なゲノム捕捉方法およびシーケンシング方法における系統誤差 (バイアス)、すなわち、標的核酸の提示の大きなばらつき、および捕捉された標的核酸 (例えば、生物学的サンプルから単離された標的核酸) のプール内のヘテロ接合性対立遺伝子の不均等なサンプリングに関連する課題を克服するための有効な方法の発見に部分的に基づく。したがって、いくつかの実施形態において、本開示は、多重捕捉方法における標的核酸の検出におけるばらつきを減少させる方法を提供する。他の実施形態において、方法は、捕捉プールにおける対立遺伝子の提示を改善し、ゆえに、相違する検出結果を改善する。ある特定の実施形態において、本開示は、標的核酸 (例えば、遺伝子座) を捕捉するための調製方法を提供し、その方法は、標的核酸の重複した領域を捕捉する複数のプローブ (例えば、分子反転プローブMIP) の種々のセットの使用を含み、従来技術の方法と比べて、捕捉プール内の標的核酸のより均一な提示を達成する。他の実施形態において、方法は、例えば、診断目的での、遺伝子座の大規模な並行捕捉に関連するバイアスまたはバイアスのリスクを低下させる。他の実施形態において、複数の遺伝子座の並行した検出の再現性を高めるための (例えば、標的核酸の捕捉に対する多型の影響を減少させることによって) 方法が提供される。さらなる実施形態において、複数の遺伝子座の並行した分析に対する、プローブの合成および/またはプローブの増幅のばらつきの影響を減少させるための方法が提供される。

20

30

【0097】

いくつかの局面において、本開示は、複数の異なるプローブを含むプローブセットを提供する。本明細書中で使用されるとき、「プローブ」は、標的核酸またはその部分領域の同じ鎖に隣接する核酸と相補的である5'領域および3'領域に隣接した中央領域を有する核酸である。例示的なプローブは、分子反転プローブ (MIP) である。「標的核酸」は、遺伝子座であり得る。例示的な遺伝子座は、本明細書中の表1 (RefSeq遺伝子の欄) に開示されている。

【0098】

プローブは、代表的には、捕捉効率/増幅効率に部分的に影響すると知られているある特定の制約 (例えば、融解温度、G/C含有量など) を満たすように設計されているが (Ballら (2009) Nat Biotech 27:361-8 および Dengら (2009) Nat Biotech 27:353-60)、広く均一であるかまたは高度に再現性のある捕捉効率/増幅効率を確実にするのに十分な制約のセットは、以前に得られていない。本明細書中に開示されるように、均一性および再現性は、標的内の各塩基が2つ以上のプローブによって捕捉されるように標的1つあたり複数のプローブを設計することによって高められ得る。いくつかの実施形態において、本開示は、捕捉される標的1つあたり複数のMIPを提供し、ここで、所与の標的核酸に対して設計されるセット内の各MIPは、中央領域、ならびに標的核酸内の (少なくとも部分的に) 異なる核酸にハイブリダイズする5'領域および3'領域 (「標的化アーム」) (標的核酸の部分領域と

40

50

すぐ隣接する)を有する。よって、種々の標的化アームと充填(f i l l - i n)配列との間の効率の差は、単一の標的に対して複数のM I Pにわたって平均され得、より均一かつ再現性のある捕捉効率をもたらされる。

【0099】

いくつかの実施形態において、上記方法は、各標的に対して単一のプローブを設計する工程を包含する(標的は、一塩基ほど小さいか、またはキロベースもしくはそれより大きい連続した配列ほど大きいことがある)。

【0100】

場合によっては、1~200bpの範囲内の長さを有する分子(例えば、標的核酸またはその部分領域)を捕捉するプローブを設計することが好ましいことがある(本明細書中
10
で使用されるとき、bpとは、二本鎖核酸における塩基対のことを指すが、しかしながら、長さがbpで示される場合、塩基対とは対照的に、同じ数の塩基長を有する一本鎖核酸も本発明によって企図されることが認識されるだろう)。しかしながら、プローブの設計は、そのように限定されない。例えば、プローブは、場合によっては、最大10、20、30、40、50、60、70、80、90、100、200、300、400、500、1000bpまたはそれ以上のbpの範囲の長さを有する標的を捕捉するように設計され得る。

【0101】

捕捉分子(例えば、標的核酸またはその部分領域)の長さは、複数の考慮すべき事柄に基づいて選択されることが認識されるべきである。例えば、標的の分析が、シーケンシング、例えば、次世代の配列決定装置を用いたシーケンシングを含む場合、その標的の
20
長さは、ショットガンライブラリーの構築が必要ないように、代表的にはシーケンシングのリード長とマッチするべきである。しかしながら、本発明の局面は、この点に限定されないので、捕捉された核酸は、任意の適当なシーケンシング手法を用いて配列決定され得ることが認識されるだろう。

【0102】

いくつかの標的核酸は、1つのプローブで捕捉されるには大きすぎることも認識されるべきである。その結果として、標的全体を分析するために、標的核酸の複数の部分領域を捕捉することが必要な場合がある。

【0103】

いくつかの実施形態において、標的核酸の部分領域は、少なくとも1bpである。他の実施形態において、標的核酸の部分領域は、少なくとも10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000bpまたはそれ以上である。他の実施形態において、標的核酸の部分領域は、標的核酸長の最大10%、20%、30%、40%、50%、60%、70%、80%、90%、95%またはそれ以上のパーセントである長さを有する。

【0104】

当業者はまた、M I Pの設計において、プローブの長さとの関係性が考慮されることも認識するだろう。いくつかの実施形態において、M I Pは、対応する標的(例えば、標的核酸の部分領域、標的核酸)よりも数百塩基対(例えば、最大100、20
40
0、300、400、500、600、700、800、900、1000bpまたはそれ以上)長いように設計される。

【0105】

いくつかの実施形態において、標的核酸の部分領域の長さは、異なってよい。例えば、標的核酸が、プローブハイブリダイゼーションが不可能であるかまたは非効率的である領域を含む場合、問題のある核酸とのハイブリダイゼーションを回避するため、および完全な標的核酸を含む核酸を捕捉するために、1つ以上の異なる長さの部分領域を捕捉するプローブを使用することが必要であり得る。

【0106】

本発明の局面は、各標的核酸を増幅するために、複数のプローブ、例えば、M I Pを使
50

用することを含む。いくつかの実施形態において、所与の標的に対するプローブのセットが、標的全体にわたって「タイル張り」するように設計され得、一連のより短いサブターゲットとして標的が捕捉される。いくつかの実施形態において、所与の標的に対するプローブのセットが、その標的全体にわたって「タイル張り」するように設計される場合、そのセット内のいくつかのプローブは、隣接する非標的配列を捕捉する。あるいは、そのセットは、標的に隣接するハイブリダイゼーション領域の正確な位置を「ずらす」ように設計され得、種々の標的化アームを有する複数のプローブによって標的全体が捕捉され（および場合によっては、隣接する非標的配列が捕捉され）、タイリングの必要性がなくなる。選択される特定のアプローチは、標的セットの性質に左右され得る。例えば、小さい領域が捕捉される場合、末端がずらされたアプローチが、適切であり得る一方、より長い領域が望まれる場合は、タイリングが選択され得る。すべての場合において、病理学的遺伝子座を標的にするプローブに対するバイアス許容度の量は、所与の分子を捕捉するために使用される種々のMIPの数を変更することによって調節（「調整」）され得る。

10

【0107】

いくつかの実施形態において、「カバー率（coverage factor）」すなわち、1つの分子において1つの塩基対を捕捉するために使用されるプローブの数は、条件を指定するための重要なパラメータである。タイリングアプローチ（例えば、図1を参照のこと）を用いるかまたはずらしアプローチ（staggered approach）の1つ（例えば、図2または3を参照のこと）を用いるかに応じて、標的1つあたりのプローブの種々の数が示される。

20

【0108】

図1は、単一の標的にわたってタイリングされている10個の捕捉されたサブターゲットを示しているタイリングプローブのレイアウトの非限定的な実施形態を図示している。標的内の各位置は、3つのサブターゲットによってカバーされており、1塩基対あたりのMIPの性能は、3つのプローブに対して平均される。

【0109】

図2は、3つのMIPのセットによって捕捉されている標的を示している、ずらされたプローブのレイアウトの非限定的な実施形態を図示している。各MIPは、黒色で示されている標的全体に加えて、（場合によっては）灰色で示されている追加的な標的外の配列を捕捉し、各MIPの標的化アームは、異なる配列に当たっている。標的内の各位置は、3つのサブターゲットによってカバーされており、1塩基対あたりのMIPの性能は、3つのプローブに対して平均される。標的化アームは、示される黒色または灰色の領域のすぐ隣を捕らえる。いくつかの実施形態において、標的化アーム（図示せず）は、互いと重複しないように設計され得ることが認識されるだろう。

30

【0110】

図3は、3つのMIPのセットによって捕捉されている標的を示している、交互にずらされたプローブのレイアウトの非限定的な実施形態を図示している。各MIPは、黒色で示されている標的全体に加えて、（場合によっては）灰色で示されている追加的な標的外の配列を捕捉し、各MIPの標的化アームは、異なる配列に当たっている。標的内の各位置は、3つのサブターゲットによってカバーされており、1塩基対あたりのMIPの性能は、3つのプローブに対して平均される。標的化アームは、示される黒色または灰色の領域のすぐ隣を捕らえる。

40

【0111】

任意のレイアウトに対して、隣接するタイリングプローブまたはずらしプローブにおける標的化アームは、重複するように、重複しないように、またはプローブのサブセットだけに対して重複するように、設計され得ることが認識されるだろう。

【0112】

任意のレイアウトに対するある特定の実施形態において、約3～約10というカバー率が使用される。しかしながら、これらの方法はそのように限定されず、最大2、3、4、5、6、7、8、9、10、20またはそれ以上のカバー率が使用され得る。選択される

50

カバー率は、使用されるプローブのレイアウトに依存し得ることが認識されるべきである。例えば、タイリングアプローチでは、所望のカバー率のために、標的1つあたりのプローブの数は、代表的には、標的の長さ、サブターゲットの長さおよび隣接するサブターゲットの開始位置との間隔（刻み幅）の関数である。例えば、3という所望のカバー率の場合、20bpという開始部位の隔たりおよび60bpというサブターゲットの長さによって、200bpの標的が12個のMIPでカバーされ得る（図1）。よって、標的核酸1つあたりのプローブの数および捕捉される分子の長さを変更することによって、特定のカバー率が達成され得る。ずらしアプローチでは、一定の長さの標的核酸が、いくつかの部分領域または「超標的（super-target）」として捕捉される（それらは、標的核酸および追加の隣接する核酸を含む分子であり、それらは、様々な長さであり得る）。例えば、50bpの標的は、「ずらされた」配置（図2）または「交互にずらされた」配置（図3）において3つのプローブを用いて3というカバー率で捕捉され得る。

10

【0113】

カバー率は、検出バイアスが容認できる程度に動かされ得る。一部の例では、バイアス許容度が小さい場合、おそらくはより高いカバー率を用いて、標的核酸のより多くの部分領域を標的にすることが望ましいことがある。いくつかの実施形態において、カバー率は、最大2、3、4、5、6、7、8、9、10またはそれ以上である。

【0114】

いくつかの実施形態において、タイリングプローブのレイアウトが用いられ、標的の長さが1bpより長く、かつ刻み幅（標的の5'末端と、隣接する標的の5'末端との間の距離）が標的またはその部分領域の長さより短いとき、プローブ数 = $T / (S / C) + (C - 1)$ であるような標的の長さ（T）、サブターゲットの長さ（S）およびカバー率（C）に基づいて特定の標的に対するプローブの数をコンピュータで計算することが可能である。

20

【0115】

いくつかの局面において、本開示は、複数の分子が並行して増幅されるとき増幅効率の均一性を高める方法；増幅効率の再現性を高める方法；増幅効率に対する、標的とするプローブのばらつきとの関与を低減させる方法；所与の標的核酸に対する、プローブハイブリダイゼーション領域内の多型の影響を減少させる方法；および/またはMIPによる多重増幅が核酸シーケンシングによる分析用の調製工程として使用されるときに下流のワークフローを単純化する方法を提供する。

30

【0116】

標的に隣接する領域下の標的核酸内の多型は、ハイブリダイゼーション、ポリメラーゼによる充填および/またはライゲーションを干渉し得る。さらに、これは、ただ1つの対立遺伝子に対して生じ得、対立遺伝子の脱落をもたらし、最終的には下流のシーケンシングの正確度を低下させる。いくつかの実施形態において、任意の所与の標的を捕捉するための複数のハイブリダイゼーション部位を有するMIPのセットを用いるとき、MIPのセット内のすべての標的化アームが変異の位置をカバーするわけではないので、多型が原因の損失の確率は、実質的に低下する。

40

【0117】

多数の配列が必要とされるので、MIP捕捉反作用のプローブは、プログラム可能なマイクロアレイにおいて合成され得る。これらの方法の合成収量は少ないので、MIP増幅反作用の十分なプローブを生成するために、その後、増幅工程が必要である。多重オリゴヌクレオチド合成とプールされたものの増幅との組み合わせにより、一様でない合成エラー率および提示のバイアスがもたらされる。所与の標的に対するすべてのプローブが同じエラー率およびバイアスを有するわけではないので、各標的に対して複数のプローブを合成することによって、これらの起源による変動が平均され得る。

【0118】

本明細書中に開示される多重増幅ストラテジーは、SNPの検出におけるように分析的に用いられ得るか、またはしばしば、次世代シーケンシングもしくは他のシーケンシ

50

ング手法に対して調製的に用いられ得る。調製的な状況において、増幅反応のアウトプットは、通常、ショットガンライブラリープロトコルへのインプットであり、それはその後、シーケンシングプラットフォームへのインプットになる。次世代シーケンシングは、エキソンなどのアンプリコンよりも著しく短いリードをもたらすので、ショットガンライブラリーが部分的に必要である。本明細書中に記載されるマルチタイリング (multi-tiled) アプローチによってもたらされるバイアスの減少に加えて、タイリングは、ショットガンライブラリーの調製に対する必要性もなくす。プローブ、例えば、MIPを設計するとき、捕捉分子の長さを指定することができるので、その捕捉分子は、配列決定装置のリード長にマッチするように選択することができる。このように、リードは、エキソンに対するプローブセット内の各捕捉分子の開始位置のおかげでそのエキソンに沿って「歩く」ことができる。

10

【0119】

2. 核酸調製におけるバイアスに関連する分析のエラーの減少：

いくつかの実施形態において、本発明の局面は、バイアスを減少させ、かつ下流の定量的用途の信頼度および正確度を高める、DNAシーケンシング関連技術における調製工程に関する。

【0120】

現在、次世代 (ポロニー (pony) ベースの) シーケンシング (とりわけ、ゲノムリシーケンシング、遺伝子発現についてのRNA-seq、メチル化についての亜硫酸水素塩シーケンシングおよびImmune-seqを含む) を利用してデータを生成する多くのゲノムクスアッセイが存在する。定量的測定 (遺伝子型コールを含む) を行うために、これらの方法は、核酸の元のサンプルにおける配列の提示に対する代理として、所与のゲノム遺伝子座のシーケンシングリードの数を利用する。これらの手法の大部分は、目的のサンプルを代表するDNA分子の高複雑度ライブラリーを構築する調製工程を必要とする。これには、DNAの化学的または生化学的な処理 (例えば、亜硫酸水素塩処理)、ゲノムの特定のサブセットの捕捉 (例えば、パドロックプローブによる捕捉、溶液ハイブリダイゼーション) および様々な増幅法 (例えば、ポリメラーゼ連鎖反応、全ゲノム増幅、ローリングサークル増幅) が含まれ得る。

20

【0121】

系統誤差および偶然誤差は、ゲノム増幅法およびシーケンシングライブラリー構築法に関連する共通の問題である。例えば、ゲノムシーケンシングライブラリーは、ライブラリー構築プロセスのエラー (バイアス) の結果として、起源ゲノム由来の特定の配列の提示過剰または提示不足を含み得る。そのようなバイアスが、シーケンシングライブラリー内に存在しないゲノム由来の標的配列またはシーケンシングライブラリー内の検出不可能なゲノム由来の標的配列に生じるとき、特に問題であり得る。例えば、シーケンシングライブラリー内のゲノム由来の特定の対立遺伝子の配列 (例えば、ヘテロ接合の対立遺伝子) の提示不足は、シーケンシングライブラリーにおいて明らかなホモ接合の提示をもたらし得る。最も下流のシーケンシングライブラリーの定量化手法は、確率論的なカウントプロセスに左右されるので、これらの問題は、代表的には、統計学的に有意な判断を行うために必要な最小数の観察結果を得るのに十分なサンプリング (過剰サンプリング) によって対処されている。しかしながら、過剰サンプリングという戦略は、通常、低カウントポアソンノイズ (low-count Poisson noise) の排除に限定され、そのアプローチは、リソースをむだにし、そのような実験を行うために必要とされる費用を高める。さらに、過剰サンプリングは、そのデータに基づくある特定の結論 (例えば、診断コール) における統計学的信頼度の低下をもたらし得る。したがって、シーケンシングライブラリー調製方法におけるバイアスを克服するための新しいアプローチが必要とされている。

30

40

【0122】

本開示の局面は、ゲノムの捕捉、増幅およびシーケンシングの方法における系統誤差および偶然誤差 (バイアス) に関連する問題、すなわち、核酸の捕捉および増幅における

50

大きなばらつきならびにシーケンシングライブラリーにおけるヘテロ接合対立遺伝子の不均衡な提示に関連する問題を克服するための方法の発見に部分的に基づく。したがって、いくつかの実施形態において、本開示は、核酸の捕捉および増幅におけるばらつきを減少させる方法を提供する。他の実施形態において、その方法は、シーケンシングライブラリーにおける対立遺伝子の提示を改善し、ゆえに、相違する検出結果を改善する。ある特定の実施形態において、本開示は、個別の核酸分子を独特にタグ化する識別タグ配列の使用を含む、標的核酸（例えば、遺伝子座）を捕捉するための調製方法を提供する。いくつかの実施形態において、その識別タグ配列のおかげで、識別タグ配列と標的配列との対がシーケンシング反応において観察される頻度に基づいたバイアスの検出が可能になる。他の実施形態において、その方法は、例えば診断目的での、遺伝子座の捕捉、増幅およびシーケンシングに関連する、バイアスによって引き起こされるエラー、またはバイアスのリスクを減少させる。

10

20

30

40

50

【0123】

本発明の局面は、独立して捕捉されたおよび/または分析された（例えば、増幅の前、またはバイアスを導入し得る他のプロセスの前に）個別の標的分子と、独特の配列タグ（識別タグ配列と呼ばれる）を関連づけることに関する。これらのタグは、独立した標的分子を互いと識別するために有用であり、それによって、既知数の個別の標的分子に基づいた分析が可能になる。例えば、アッセイにおいて得られた複数の標的分子配列の各々が、異なる識別タグと関連する場合、それらの標的配列は、互いから独立していると考えられ得、遺伝子型の尤度は、この情報に基づいて決定され得る。対照的に、アッセイにおいて得られる複数の標的分子配列の各々が、同じ識別タグと関連する場合、それらの標的分子配列のおそらくすべてが、そのアッセイにおけるこの標的分子の提示過剰に起因して（例えば、偏った増幅に起因して）、同じ標的分子を起源とするものだろう。これは、各核酸が異なる識別タグに関連した状況よりも少ない情報しか提供しない。いくつかの実施形態において、独立して単離された閾値数の分子（例えば、識別タグと標的配列との独特の組み合わせ）が分析されることにより、被験体の遺伝子型が決定される。

【0124】

いくつかの実施形態において、本発明は、調製核酸のプール（ライブラリー）を含む組成物に関し、そのプール内の調製核酸の各々は、バイアスの影響を検出するためおよび減少させるため、ならびに標的核酸配列の遺伝子タイピングを行うための「識別タグ配列」を含む。本明細書中で使用されるとき、「識別タグ配列」は、核酸（調製核酸）の配列であり、それは、複数の異なる単離された核酸の状況において、独立して単離された独特の核酸を同定するものである。代表的には、識別タグ配列は、核酸調製方法の1つ以上の段階において標的核酸の起源を同定するために使用される。例えば、多重核酸捕捉反応において、識別タグ配列は、複数の独立した標的核酸捕捉事象を識別するための根拠を提供する。また、多重核酸増幅反応において、識別タグ配列は、例えば、標的核酸の複数の独立した主要なアンプリコンを識別するための根拠を提供する。よって、ある調製方法に関する単離された核酸の標的核酸配列と識別タグ配列との組み合わせ（標的：識別タグ配列）は、独立して単離された独特の標的核酸を同定するための根拠を提供する。図4A～Cは、識別タグ配列と標的配列とを組み合わせるための方法の様々な非限定的な例を表している。

【0125】

識別タグが、当該分野で公知のいくつかの異なる方法のうちのいずれか1つを用いて合成され得ることは、当業者に明らかだろう。例えば、識別タグは、ランダムなヌクレオチド付加によって合成され得る。識別タグ配列は、代表的には、あらかじめ定義された長さであり、それは、調製反応（例えば、増幅ベースの反応、環状化選択ベースの反応、例えば、MIP反応）において独特の標的：識別タグ配列を生成する尤度を制御するために選択される。識別タグ配列は、最大5、最大6、最大7、最大8、最大9、最大10、最大11、最大12、最大13、最大14、最大15、最大16、最大17、最大18、最大19、最大20、最大21、最大22、最大23、最大24、最大25またはそれ以上の

ヌクレオチド長であり得る。遺伝子タイピングの目的のために、単離された核酸が、標的核酸配列と識別タグ配列との独特の組み合わせを含み、標的核酸配列と識別タグ配列との閾値数の独特の組み合わせの観察結果 (o b s e r v a n c e) が、その遺伝子型におけるある特定の統計学的信頼度を提供する場合、その単離された核酸は、独立して単離されたと同定される。

【 0 1 2 6 】

ライブラリー調製プロセスにおいて、各核酸分子は、独特の識別タグ配列が目的の標的核酸配列 (ライブラリーを調製するための核酸配列、例えば、多型配列) とともに配列決定されることを可能にする配置において、その識別タグ配列でタグ化され得る。特定の工程において識別タグ配列を含む核酸を組み込むことによって、そのプロトコルの後の工程においてバイアスの検出および補正が可能になる。

10

【 0 1 2 7 】

独特の識別タグ配列の大きなライブラリーは、規定の長さの縮重ランダム配列ポリヌクレオチドを用いることによって作製され得る。それらのポリヌクレオチドの識別タグ配列は、シーケンシングの最後の段階において解読され得る。識別タグ配列の観察結果を用いることにより、ライブラリーの最後のシーケンシングの読み出しにおいてバイアスが検出され得、補正され得る。例えば、ランダムに生成され得る識別タグ配列の可能な総数は、 4^N であり、ここで、 N は、識別タグ配列の長さである。よって、識別タグ配列の長さは、独特の識別タグ配列を有する M I P の集団のサイズが、M I P 捕捉産物のライブラリーを生成するのに十分であるように調整され得る (ここで、標的核酸配列と識別タグ配列との独立した同一の組み合わせは稀である) ことが理解されるべきである。本明細書中で使用されるとき、標的核酸配列と識別タグ配列との組み合わせは、「標的：識別タグ配列」とも呼ばれ得る。

20

【 0 1 2 8 】

シーケンシングプロセスの最後の読み出しにおいて、各リードは、追加的な独特の識別タグ配列を有し得る。いくつかの実施形態において、識別タグ配列が、ライブラリー内にランダムに分配されているとき、その独特の識別タグ配列のすべてが、ほぼ同じ回数で観察される。したがって、識別タグ配列の出現回数は、ポアソン分布に従い得る。

【 0 1 2 9 】

いくつかの実施形態において、調製核酸 (例えば、増幅された M I P 捕捉産物) のプールにおける標的：識別タグ配列の過剰提示は、調製プロセスにおけるバイアス (例えば、増幅プロセスにおけるバイアス) を示唆する。例えば、統計学的に過剰提示される標的：識別タグ配列の組み合わせは、そのプロトコルの、M I P への識別タグ配列の組み込みと M I P 捕捉産物の実際のシーケンシングとの間の 1 つ以上の工程におけるバイアスを示唆する。

30

【 0 1 3 0 】

所与の標的：識別タグ配列のリードの数は、起源サンプル中に存在する標的配列の量を示唆し得る (代理として働き得る)。いくつかの実施形態において、起源サンプルにおける配列の出現回数は、目的の数である。例えば、本明細書中に開示される方法を用いるとき、M I P のプール内の識別タグ配列の出現率は、あらかじめ決められ得る (例えば、すべての識別タグ配列に対して同じであり得る)。したがって、増幅後およびシーケンシング後の識別タグ配列の出現率の変化は、そのプロトコルにおけるバイアスを示唆し得る。バイアスが補正されることにより、例えば診断目的のために、元の M I P プールの組成の正確な提示が提供され得る。

40

【 0 1 3 1 】

いくつかの局面によると、調製核酸分子 (例えば、独特の識別タグ配列を有するライブラリー内の各核酸である M I P) のライブラリーは、そのライブラリー内の核酸分子の数が、そのライブラリーを用いて捕捉される標的核酸分子の予定数よりも著しく多くなるように構築され得る。このことにより、調製方法の産物が、独特の標的：識別タグ配列だけを含むことが確実に成り ; 例えば、M I P 反応において、捕捉工程は、M I P ライブラリ

50

ー内の独特の識別タグ配列の集団全体をより少なくサンプリングし得る。例えば、 $1 \mu\text{g}$ のゲノムDNAを利用した実験は、約 $\sim 150,000$ コピーの二倍体ゲノムを含み得る。ライブラリー内の各MIPが、ランダムに生成された12-merの識別タグ配列を含むMIPライブラリー（約160万個の独特の識別タグ配列が存在し得る）の場合、1ゲノムコピーあたり100を超える独特の識別タグ配列が存在し得る。ライブラリー内の各MIPが、ランダムに生成された15-merの識別タグ配列を含むMIPライブラリー（約10億個の独特の識別タグ配列が存在し得る）の場合、1ゲノムコピーあたり7000を超える独特の識別タグ配列が存在し得る。ゆえに、同じ識別タグ配列が複数回組み込まれる確率は、途方もなく小さい。よって、識別タグ配列の長さは、MIP捕捉反応における標的配列の量、および標的：識別タグ配列の組み合わせが独立して複数出現すること

10

【0132】

図5は、標的配列および識別タグ配列に基づいて遺伝子タイピングを行うための非限定的な方法を表している。標的配列および識別タグ配列のシーケンシングリードは、二倍体の遺伝子型コールを行うためにつぶされる。図6は、MIP捕捉反応のシミュレーションの非限定的な結果を表しており、ここで、MIPプローブ（各々、15ヌクレオチドの識別タグ配列を有する）は、10000コピーの標的配列（例えば、ゲノム等価物）と組み合わせられている。このシミュレートされた反応では、同じ識別タグ配列を有する1コピー以上の標的配列を捕捉する確率は、0.05である。Y軸は、観察結果の数を反映している。X軸は、標的：識別タグ組み合わせの独立した出現数を反映している。図7は、シーケンシングカバレッジの非限定的なグラフを表しており、これは、対立遺伝子が十分な深度でサンプリングされることを確実にするのを助け得る（例えば、1000個の標的を仮定したときの対立遺伝子1つあたり $10\times$ または $20\times$ 最小サンプリング）。この非限定的な例では、X軸は、標的1つあたりの必要とされる総カバレッジであり、Y軸は、所与の総カバレッジが各対立遺伝子に対して少なくとも $10\times$ または $20\times$ カバレッジをもたらす確率である。

20

【0133】

MIPライブラリー調製プロセスの一部として、アダプターが、目的の分子の末端にライゲートされ得ることを当業者は認識するだろう。アダプターは、しばしば、PCRプライマー部位（増幅用またはエマルジョンPCR用）および/またはシーケンシングプライマー部位を含む。さらに、例えば、一緒に混合され得る個別のサンプル（例えば、患者サンプル）を一義的に同定するために、バーコードが含まれ得る（例えば、USPTO Publication Number US2007/0020640A1（McCluskeyら）を参照のこと）。

30

【0134】

ランダムな識別タグ配列の実際の組み込みは、当該分野で公知の様々な方法によって行われ得る。例えば、識別タグ配列を含む核酸が、ライゲーションによって組み込まれ得る。これは、識別タグ配列を有する分子を任意の平滑末端核酸にライゲートすることができるので、融通の利く方法である。上記シーケンシングプライマーが識別タグ配列と標的配列の両方を配列決定するように、それらのシーケンシングプライマーを後で組み込まなければならない。あるいは、ただ1回のライゲーションだけを行えばよいように、シーケンシングアダプターが、その3'末端の（縮重塩基として）ランダムな識別タグ配列とともに合成され得る。別の方法は、プライマー構造が、共通のアダプター配列の後にランダムな識別タグ配列、その後PCRプライミング配列で（5'から3'への順序で）配置されるように、識別タグ配列をPCRプライマーに組み込む方法である。識別タグ配列およびアダプター配列（シーケンシングプライマー部位を含み得る）は、タグとして組み込まれる。識別タグ配列を組み込む別の方法は、遺伝子捕捉反応を行う前に、パドロックプローブ中に識別タグ配列を合成することである。識別タグ配列は、標的化アームに対して3'に組み込まれるが、そのプロトコルの下流で使用され得る増幅プライマーに対しては5'に組み込まれる。識別タグ配列を組み込む別の方法は、遺伝子特異的プライ

40

50

マーまたはポリ - d T 逆転写プライマーにおけるタグとしてである。これにより、識別タグ配列を c D N A レベルで直接組み込むことが可能になる。

【 0 1 3 5 】

いくつかの実施形態において、組み込み工程における識別タグ配列の分布は、均一であると仮定され得る。この場合は、そのプロトコルの任意の部分におけるバイアスは、この分布の均一性を変化させ得、それは、シーケンシングの後に観察され得る。このことにより、最終的なアウトプットが多くの分子の並行したシーケンシングである任意の調製プロセスにおいて識別タグ配列を使用することが認められる。

【 0 1 3 6 】

識別タグ配列は、チップ上で並行して合成されるとき、複数のプローブ（例えば、M I P）に組み込まれ得、その組み込まれるヌクレオチドの縮重は、複数のプローブにおけるほぼ均一な分布を確実にするのに十分である。独特の識別タグ配列のプールの増幅自体が最初のプール内にバイアスを導入し得ることが認識されるべきである。しかしながら、ほとんどの実際の場合では、合成（例えば、カラム合成、チップベースの合成によるものなど）のスケールが十分大きいので識別タグ配列の最初のプールの増幅は必要でない。独特の識別タグ配列のプールに対する増幅工程または選択工程を回避することによって、潜在的なバイアスが最小にされ得る。

【 0 1 3 7 】

識別タグ配列の使用の1つの例は、ゲノムリシーケンシングにおけるものである。ほとんどの次世代シーケンシング機器の未処理の (r a w) 正確度が比較的低いことを考慮すると、目的のゲノム遺伝子座を過剰サンプリングすることは非常に重要である。さらに、すべての遺伝子座には2つの対立遺伝子が存在するので、サンプルがホモ接合であるかまたはヘテロ接合であるかを十分な程度の統計学的信頼度で決定するのに十分な回数、両方の対立遺伝子が観察されることを確実にするのに十分、サンプリングすることが重要である。実際は、シーケンシングを行うことにより、起源サンプル中の分子の組成物がサンプリングされる。しかしながら、複数のリードが、所与の遺伝子座に対して集められた後では、バイアス（例えば、P C R 増幅工程によって引き起こされるバイアス）に起因して、それらのリードの大部分が単一の起源分子から得られたものである可能性がある。これは、観察された標的配列の集団を歪め得、遺伝子型コールの結果に影響し得る。例えば、ヘテロ接合である遺伝子座が、ホモ接合とコールされる可能性がある。なぜなら、その遺伝子座の多くの観察結果のうち第2の対立遺伝子の観察結果がほんのわずかにしか存在しないからである。しかしながら、提示過剰の対立遺伝子は、提示過剰の識別タグ配列を有するとも見られ得る（すなわち、提示過剰の識別タグ配列を有する配列のすべてが同じ単一分子に由来し得る）ので、識別タグ配列に対する情報が入手可能である場合、この状況は、防がれ得る。ゆえに、それらの配列および識別タグ配列の対応する分布が、遺伝子型をコールするアルゴリズムへの追加のインプットとして使用されることにより、遺伝子型コールの正確度および信頼度が著しく改善され得る。

【 0 1 3 8 】

いくつかの局面において、本開示は、遺伝子座または遺伝子座の一部（例えば、表1の遺伝子座）である複数の標的配列を分析するための方法を提供する。それらの遺伝子座は、1つ以上の多型（例えば、S N P）における遺伝子型を得るためにシーケンシングすることによって分析され得る。例示的な多型は、表2に開示される。当業者は、他の多型も当該分野で公知であり、例えば、Entrez Single Nucleotide Polymorphism データベースを照会することによって、例えば、表1の遺伝子IDを用いて検索することによって、同定され得ることを認識するだろう。

【 0 1 3 9 】

10

20

30

40

【表 1 - 1】

表1: 標的核酸

遺伝子名	遺伝子ID	説明	遺伝子別名	OMIM	RefSeq 遺伝子	染色体 マップ上の 位置
CYP21A 2	1589	シトクロムP450, ファミリー21, サブファミ リーA, ポリペプチド2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP結合カセット, サブフ ァミリーC(CFTR/MRP)、メンバー8	HI; SUR; HHF1; MRP8; PHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	アルファ サラセミア/精神遅滞症 候群 X連鎖(RAD54ホモログ , <i>S. cerevisiae</i>)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1- q21.1
ARSA	410	アールスルファターゼA	MLD; ARSA	607574	NG_009260.1	22q13.31- qter; 22q13.33
PSAP	5660	プロサホシン	GLBA; SAPI; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21- q22
BTD	686	ビオチニダーゼ	BTD	609019	NG_008019.1	3p25

10

20

【 0 1 4 0 】

【表 1 - 2】

遺伝子名	遺伝子ID	説明	遺伝子別名	OMIM	RefSeq 遺伝子	染色体 マップ上の 位置
HLCS	3141	ホロカルボキシラーゼ合成酵素(ヒオチン(プロプリオニルコエンザイムA-カルボキシラーゼ(ATP加水分解))リカーゼ)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	ブルーム症候群, RecQヘリカーゼ様	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	アスパルトアシラーゼ (カナパン病)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter- p13
CFTR	1080	嚢胞性線維症 膜貫通型 コンダクタンズ制御因子 (ATP結合カセット サブファミリーC, メンバ ー7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.1 2	7q31.2
ASS1	445	アルギニノコハク酸 合成酵素1	ASS; CTLNI; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	メチルマロン酸尿症 (コ バラミン欠乏症)cb1C型 、ホモシスチン尿症を伴う	cb1C; FLJ25671; DKFZp564I122; RP11- 291L19.3; MMACHC	609831	NC_000001.9	1p34.1
IKBKAP	8518	B細胞における銅ペ 軽鎖ポリペプチド遺伝子 エンハンサーのインヒビタ ー、キナーゼ複合体関連 タンパク質	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	ファンコニー貧血, 相補群C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	ガラクトキナーゼ1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	ガラクトース-1-リン酸ウリ リルトランスフェラーゼ	GALT	606999	NC_000009.1 0	9p13
GALE	2582	UDP-ガラクトース-4 -エピメラーゼ	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	グルコシダーゼ, ベータ; 酸(グルコシルセラミダー ゼを含む)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	ギャップ結合タンパク質, ベータ2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11- q12
GCDH	2639	グルタリル-コエンザイ ムA デヒドロゲナーゼ	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	グルコース-6-ホスファターゼ, 触媒サブユニット	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	ヘモグロビン, ベータ	CD1131-C; ベータ-グロビン; HBB	141900	NG_000007.3	11p15.5

10

20

30

40

【 0 1 4 1 】

【表 1 - 3】

遺伝子名	遺伝子ID	説明	遺伝子別名	OMIM	RefSeq 遺伝子	染色体マップ上の位置
BCKDH A	593	分枝鎖ケト酸デヒドロゲナーゼE1, アルファポリペプチド	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDH B	594	分枝鎖ケト酸デヒドロゲナーゼE1, ベータポリペプチド	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	ジヒドロリポアミド分枝鎖トランスアシラーゼE2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	ジヒドロリポアミドデヒドロゲナーゼ	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	アシル-CoA脱ヒドロゲナーゼ, C-4~C-12直鎖	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	メチルマロン酸尿症(コバラミン欠乏症)cblA型	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	メチルマロン酸性尿症(コバラミン欠乏症)cblB型	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	メチルマロニルCoA脱ヒドロゲナーゼ	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	ムコリピン1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	アクチン, アルファ1, 骨格筋	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	トロポミオシン3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hsep30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	トロポニンT 1型(骨格, 遅筋型)	ANM; TNT; STNT; TNNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	ネブリン	NEM2; NEB177D; FLJ11505; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	スフィンゴミエリンホスホジエステラーゼ1, 酸リソソーム	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	グリシンデヒドロゲナーゼ(脱炭酸)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

10

20

30

40

【 0 1 4 2 】

【表 1 - 4】

遺伝子名	遺伝子ID	説明	遺伝子別名	OMIM	RefSeq 遺伝子	染色体 マップ上の 位置
GCSH	2653	グリシン開裂系タンパク質H(アミノメチルキャリア)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	アミノメチルトランスフェラーゼ	GCE; NKH; GCST; AMT	238310	NC_000003.1 0	3p21.2- p21.1
OTC	5009	オルニチンカルバモイルトランスフェラーゼ	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	フェニルアラニンヒドロキシラーゼ	PH; PKU; PKUI; PAH	612349	NG_008690.1	12q22- q24.2
DHPR	5860	キノイドジヒドロプテリジントラクターゼ	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-ヒルポイルテトラヒドロプテリンシンターゼ	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3- q23.3
PCCA	5095	プロピオニルコエンザイムAカルボキシラーゼ, アルファポリペプチド	PCCA	232000	NG_008768.1	13q32
PCCB	5096	プロピオニルコエンザイムAカルボキシラーゼ, ベータポリペプチド	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	アシル-CoA脱ヒドロゲナーゼ, C-2 ~C-3短鎖	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22- qter
DHCR7	1717	7-デヒドロコレステロールレダクターゼ	SLOS; DHCR7	602858	NC_000011.8	11q13.2- q13.5
SMN1	6606	運動ニューロン1の生存 テロメア	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	ヘキササミニダーゼA (アルファポリペプチド)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23- q24
MYO7A	4647	ミオシンVIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	1008 3	アッシャー症候群1C(常 染色体劣性, 重篤)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ- 73; NY-CO-37; NY-CO- 38; ush1cpst; PDZ-73/NY- CO-38; USH1C	605242	NC_000011.8	11p15.1- p14
CDH23	6407 2	カドヘリン様23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10q21- q22
PCDH15	6521 7	プロトカドヘリン15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1

10

20

30

40

【 0 1 4 3 】

【表 1 - 5】

遺伝子名	遺伝子ID	説明	遺伝子別名	OMIM	RefSeq 遺伝子	染色体マップ上の位置
SANS	124590	アッシャー症候群1G (常染色体劣性)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	アリストレス(aristales)関連ホメオボックス	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	オリゴフリン(oligophrenin)1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	リジン(K)特異的デメチラーゼ5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJホモログ1(E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	溶質キャリアファミリー6(神経伝達物質トランスポーター、クレアチン)、メンバー8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	ディスク(discs)、大ホモログ3(Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAPI02; SAPI02; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	テトラスパニン7	A15; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	ジンクフィンガータンパク質41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	アシル-CoA合成酵素長鎖ファミリーメンバー4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBPI	10084	ポリグルタミン結合タンパク質1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBPI	300463	NC_000023.9	Xp11.23
PEX1	5189	ペルオキシソーム生合成因子1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	ペルオキシソーム膜タンパク質3、35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	ペルオキシソーム生合成因子6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	ペルオキシソーム生合成因子10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	ペルオキシソーム生合成因子12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	ペルオキシソーム生合成因子5	PXRI; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	ペルオキシソーム生合成因子26	FLJ20695; PEX26MIT; Pex26pMIT; PEX26	608666	NG_008339.1	22q11.21

表 2 に列挙される変異は、いくつかの疾患関連遺伝子において立証されている多型である (CFTR は、嚢胞性線維症において変異しており、GBA は、ゴーシェ病において変異しており、ASPA は、カナパン病において変異しており、HEXA は、テイ・サックス病において変異している)。これらの多型は、いくつかのタイプ：挿入/欠失の長さが 3 bp の倍数でない場合にフレームシフトを引き起こし得る (およびゆえに通常、タンバ

10

20

30

40

50

ク質機能を妨害し得る)挿入/欠失多型、およびタンパク質のアミノ酸配列を変更し得、場合によっては、終止コドンの導入による完全な不活性化を引き起こし得る置換である。

【 0 1 4 4 】

【 表 2 - 1 】

表2:多型の非限定的な例

遺伝子名	遺伝子ID	SNP ID	変異	配列番号
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAAA[A/G]G GGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[- /T]AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAAGGG[A/G]C GGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]C CATAGCATTTTGTAACCTAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]C ACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAATAATAGTGTATGTACCC[C/G]GC TTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GA GGCCAGAGAGATTAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TA GATAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAGTT[A/T]G AATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]AT ATATTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]C GCCTGGCCTGATGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CC ATCAGGAAGGGTGTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAACTTTTTTTCAGGTG[- /AGA]AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAAGATTATATAAAGG[A/C]TA TTGCTTTTGAATCACAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]T TCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAAA[- /A]GTTTGAATGTTTTCTTGCATTCAGA	16

10

20

30

【 0 1 4 5 】

【表 2 - 2】

CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATA[- /TGTATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTT ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTFACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTGCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACCT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAAAA[- /A]GGAAGAAGGAATTAATAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGATATATATATATAT[A/T]TA TATATTTTTTTTTCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACCTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATAT[A/G]TA TATATATATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTATTTTCTCATATTATTT[- /T]CAGTGGCTTTTTCTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACACTACA[A/G]GT GATTTCTTTTTCTTATTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTTGTGTCTTCT[- /CTTT]AACACATCTTTTCTAGCAGAATTCA	41

10

20

30

40

【表 2 - 4】

GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgag[A/T]ggcgcgatctcggtcac cgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGAGTCCGGGCGGG[A/G]C CTGGATTGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAA[TTTTTTTT]- /G/T]gtttgagacagagcttgcctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcgtggcgggtgc[C/T]tgtaatcccacgtacttgg gaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagagag[- /AG]gagCCGGCGGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGCCGG[C/G/T] GAAATCTTATTTACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGT[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctct[- /TC]gctcgtctctctctctctctctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGTCCGTGAATGTTTGTACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGTATTTTTTC[A/G]GA CTTCTCTGGTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCTGGCCAGGAG[C/T]GG TGGCTACGCCTGTAATCCAGC	91
ASPA	443	rs62071299	TTTAAAAATGCTGATTCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93

10

20

30

40

【表 2 - 5】

ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[- /TGTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACCTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[- /GTACTT]TTTTTCCAAATTGAAGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTGTTTTGTTTTGTTTTT[- /GTTTTGTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[- /TGACT]TGTACTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[- /C]TGATGGCAGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[- /C]ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[- /T]GGTGTGCCGCACCCATTAACCTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTT[- /T]GAGACAGAGTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGAAATCTTAAGCTTTTATTTGG[- /G]TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[- /A]CCATGTGCAGGTTGTTACATATGT	114
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[- /A]GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[- /G]AAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCAGGTTCAAGAGA[- /A]TTCTCCTGCCTCAGCCTCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACCTCCATGGTGGATTGTT	118

10

20

30

40

【表 2 - 6】

ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAA[- /A]TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[- /G]AGCTGTCCTCTTGCCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[- /CA]ACTCTGAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[- /T]AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACCTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[- /C]CCCTGACCCTCCTCTATAGCATTA	124
ASPA	443	rs34015272	ATAAATGATCATCATTACAGTAGGG[- /G]TTTTGTTTGTGTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTT[- /T]CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAATAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAACCATTATT[A/G]CT AACCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTATTAAAGCAA[A/G]GG ATTCCGGCAAAACAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAACA	133
HEXA	3073	rs62022858	ATCTGCTCTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTGTTGATTTTTTTTTTTT[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTACTTTTCTTTATTGACTT[C/T]CCC TTATTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAAAA[- /A]GGAAGTAGCAGCAACAGCTTGGAAA	139
HEXA	3073	rs60920713	GTTGCCAGGGTTGAGTGCAGAGGCA[C/T]AT CTTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[- /TTTT]GAGACAGAGTCTTGCTGTGTCAACC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTAAAAAAAAAAAAAAAAAAAA[- /A]TGCAGTCTAATGGCAGAATTAGACT	144

10

20

30

40

【 0 1 5 0 】

【表 2 - 7】

HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTGA[A/G]ACA GGGTCTCTGTGTGCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCAGTGTGTACCTC[- /CCTC]TGTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTTGGGGGGGGGG[- /G]TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAAA[- /AGAAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAAGAAAAAAAAAAAAAAAAAAAA[- /A]GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[- /GT]CTGTATCCCCACTTCCAGAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[- /C]TTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCAGTGTGA[- /CCTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAAA[- /A]AAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[- /T]TTTTCTTCAATCCAACCAAGATATAT	164
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAAG[- /A]GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGCAGTTAAAAGAAAACAAA[- /C]CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCAGGACCACA[- /C]GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[- /T]TTTTTTTTTCCATAAACAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTAACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169

10

20

30

40

【 0 1 5 1 】

【表 2 - 8】

HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[- /ATG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[- /A]AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[- /TG]TCTGTATCCCCACTTCCAGAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[- /G]GGAAATAAACATCTTATTGAATTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[- /T]TTCAGGTTACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GG GGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[- /T]TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GA TTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTACACTTC[C/T]GCT GTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]AC CTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CC ATTACCTGCCACTCTCTAGCATC	180

10

20

3. 核酸標的の長さの評価：

いくつかの実施形態において、本発明の局面は、核酸リピート配列を含む領域における核酸の欠失または挿入を検出するための方法に関する。

【0152】

核酸リピート配列を含むゲノム領域は、リピート配列の数の増幅または縮小（例えば、1単位以上の反復配列の挿入または欠失）に起因する遺伝的不安定性の部位であることが多い。多数のリピート配列を含むゲノム領域の長さの不安定性は、いくつかの遺伝性および非遺伝性の疾患および状態に関連している。

【0153】

例えば、「脆弱X症候群」、すなわちマーチン・ベル症候群は、徴候が重篤から軽度につながる一連の特徴的な物理的、知的、情緒的および行動的な特色をもたらす遺伝的症候群である。この症候群は、X染色体上の単一のトリヌクレオチド遺伝子配列（CGG）の拡大に関連するものであり、正常な神経発生に必要なFMR-1タンパク質が発現できなくなる。脆弱X症候群には、繰り返しCGG配列の長さに関して一般に認められている形態が4つある；正常（29～31個のCGGリピート）、前変異（55～200個のCGGリピート）、完全変異（200個を超えるCGGリピート）および中間またはグレーゾンの対立遺伝子（40～60個のリピート）。

30

【0154】

他の例としては、1つ以上のマイクロサテライト遺伝子座（例えば、BAT-25および/またはBAT-26）における核酸リピートのゲノムコピー数の増加または減少を含むマイクロサテライト不安定性（MSI）に関連している癌が挙げられる。現在、特定の遺伝子座における核酸リピート配列の数を決定するためおよび核酸の挿入または欠失の存在を同定するためのシーケンシングベースのアッセイが多く存在する。しかしながら、そのような手法は、ある領域の全長を配列決定しないことがあるハイスループット多重分析では有用でない。

40

【0155】

対照的に、いくつかの実施形態において、本発明の局面は、遺伝子座を配列決定する必要なく（または、遺伝子座全体を配列決定する必要なく）、そのゲノム遺伝子座における挿入または欠失の存在を検出することに関する。本発明の局面は、高レベルのリピート配

50

列を含む核酸領域における挿入または欠失を検出するために特に有用である。ある遺伝子座におけるリピート配列の存在は、ある集団において、その遺伝子座におけるリピート配列の1つ以上の挿入または欠失に起因する比較的高レベルの多型に関連することが多い。それらの多型は、疾患または疾患に対する素因に関連し得る（例えば、ある特定の多型の対立遺伝子は、疾患または状態に関連する劣性の対立遺伝子である）。しかしながら、その遺伝子座における正確な配列およびリピートの数を決定するためにシーケンシング手法を用いるとき、リピート配列の存在は、しばしば、遺伝子座の分析を複雑にし、エラーのリスクを高める。

【0156】

いくつかの実施形態において、本発明の局面は、核酸捕捉法（例えば、分子反転プローブ技術に基づく核酸配列の捕捉法）を用いて、挿入または欠失を含むと疑われる（例えば、リピート配列の存在に起因して）遺伝子座の一部の捕捉頻度を評価することによって、遺伝子座のサイズを決定することに関する。本発明の局面によると、異なる生物学的サンプル（例えば、異なる被験体由来のサンプル）における目的の遺伝子座に対する捕捉効率の統計学的に有意な差は、それらのサンプルにおける相対的な長さが異なることを示唆する。その長さの差は、その遺伝子座の一方または両方の対立遺伝子におけるものであり得ることが認識されるだろう。したがって、本発明の局面は、生物学的サンプルがその多型に対してヘテロ接合において調べられているのかまたはホモ接合において調べられているのかに関係なく、多型を同定するために使用され得る。本発明の局面によると、挿入または欠失を有する1つ以上の遺伝子座を含む被験体は、適切なコントロール（例えば、既知の核酸サイズに対する捕捉効率、生物学的サンプル中の挿入または欠失を含むと疑われない他の領域に対する捕捉効率）もしくは所定の参照捕捉効率またはそれらの任意の組み合わせを用いて、1つ以上の生物学的サンプルから得られた核酸についての捕捉効率を分析することによって同定され得る。しかしながら、本発明の局面は、コントロールの性質または存在によって限定されないことが認識されるだろう。いくつかの実施形態において、捕捉効率の統計学的に有意な変動が検出される場合、被験体は、その遺伝子座における挿入または欠失に関連する疾患または状態に対してリスクがあると同定され得る。いくつかの実施形態において、挿入または欠失の正確な性質を決定するため、および被験体が1つ以上の挿入または欠失に対してヘテロ接合であるかまたはホモ接合であるかを決定するために、被験体は、より詳細に分析され得る。例えば、遺伝子座の増幅（例えば、PCR）産物のゲル電気泳動もしくはサザンブロッティング、またはそれらの任意の組み合わせが、遺伝子座の長さを実証する直交性アプローチとして用いられ得る。いくつかの実施形態において、挿入および欠失の数およびタイプを同定するために、遺伝子座のより網羅的かつ詳細な配列分析が行われ得る。しかしながら、本発明の局面にしたがって異常な長さを有すると同定された遺伝子座をさらに分析するために、他の手法も用いられ得る。

【0157】

したがって、本発明の局面は、目的のゲノム領域における異常な核酸の長さを検出することに関する。いくつかの実施形態において、本発明は、反復エレメントなどのアクセスしづらいゲノム領域のサイズを推定することを目標とする。しかしながら、本発明の方法は、正確な長さを推定することを要求しないことが認識されるだろう。いくつかの実施形態では、異常な長さを有する1つ以上の対立遺伝子が目的の遺伝子座に存在することを決定することで十分である（例えば、異常な捕捉効率の検出に基づいて）。

【0158】

非限定的な例において、本発明の局面を例証するために、トリヌクレオチドリピート（遺伝子型）のサイズが症状（表現型）と関係がある脆弱Xが用いられ得る。しかしながら、脆弱Xは、非限定的な例であり、同様の分析が、他の遺伝子座に対して行われ得る（例えば、独立して、または多重分析において同時に）ことが認識されるだろう。

【0159】

分子反転プローブ（MIP）の使用は、単一ヌクレオチド多型の検出（Hardenbolら、2005 Genome Res 15:269-75）およびエキソンの大き

10

20

30

40

50

なセットの調製的増幅 (Porrecaら、2007 Nat Methods 4:931-6, Krishnakumaraら、2008 Proc Natl Acad Sci USA 105:9296-301) について証明されている。どちらの場合においても、増幅される遺伝子座の上流および下流とハイブリダイズする末端(「標的化アーム」)を有するオリゴヌクレオチドプローブが設計される。

【0160】

いくつかの実施形態において、本発明の局面は、プローブ捕捉効率に対する長さの影響をアッセイ(例えば、ハイスループットアッセイおよび/または多重アッセイ)に用いることにより、評価される領域全体のシーケンシングを必要とせずに配列の長さを決定することができるという認識に基づく。これは、サイズが変化しやすいリピート領域にとって特に有用である。Dengら、Nature Biotech. 27:353-60 (Dengらの補足図1Gを参照のこと)から再現された図8に図示されるように、MIPを用いるとき、短い配列が長い配列よりも高効率で捕捉されることが図示されている。統計パッケージRおよびそのエフェクト(effect)モジュールをこの分析のために用いた。線形モデルを使用し、個別の因子は独立していると仮定した。破線は、95%信頼区間を表している。より短い標的配列が、長い標的配列よりも高い効率で捕捉された($p < 2 \times 10^{-16}$)。しかしながら、配列の長さの体系的分析にこの差別的な捕捉効率を使用することは、以前は認識されていなかった。

10

【0161】

いくつかの実施形態において、プローブハイブリダイゼーションの後に、ポリメラーゼ充填反応およびライゲーション反応を行うことにより、ハイブリダイズされたプローブが、所望の標的を含む共有結合的に閉じた環状分子に変換される。PCRまたはローリングサークル増幅に加えて、環状化されていない材料のエキソヌクレアーゼ消化を行うことにより、出発核酸プールから環状の標的が単離され、増幅される。この方法の主要な利点のうちの一つは、高度の多重化に対する能力であるので、一般に、数千の標的が、数千のプローブを含む単一反応において捕捉される。

20

【0162】

本発明の局面によると、反復領域は、独特の非反復配列に取り囲まれており、その非反復配列を使用することにより、例えば、PCRまたはパドロック(MIP)ベースの方法を用いて、リピート含有領域を増幅することができる。

30

【0163】

反復領域に加えて、プローブ(例えば、MIPまたはパドロックプローブ)は、ゲノム(または標的プール)において一義的に同定されるのに十分な少なくとも一つの配列を含むように設計され得る。図9に図示されるように、プローブが環状化され、増幅された後、独特の配列が同定されて、反復領域の「代表」として働き得るように、アンプリコンの末端が配列決定され得る。図9は、反復領域(太い波線)と隣接する独特の配列(太い直線)の両方を含む領域のパドロック(MIP)捕捉の非限定的なスキームを図示している。これらのプローブの領域は、領域「1」および「3」として示される標的化アームとともに示されている。シーケンシングプライマー結合部位であり得るかまたはシーケンシングプライマー結合部位を含む介在領域は、「2」と示されている。パドロックが環状化され、増幅された後、末端を配列決定することにより、目的的反復領域に相当する独特の配列の配列を得ることができる。捕捉効率は、全体的に見て、標的の長さとの逆の相関関係があるが、種々のプローブ配列は、独特の特性を有し得る。ゆえに、およそ0~150bp、150~600bpおよびそれ以上の反復サイズ(これらはそれぞれ、脆弱X症候群の正常、前変異および完全変異を表す)を識別するのに十分感度の高い最適なプローブが選択されるように、複数のプローブが、設計され得、試験され得る。しかしながら、ある範囲のリピート領域サイズの差(例えば、約3~30塩基、約30~60塩基、約60~90塩基、約90~120塩基、約120~150塩基、約150~300塩基、約300~600塩基、約600~900塩基という長さの差、または任意の中間もしくはそれより長い長さの差)を識別するように、他のプローブのサイズおよび配列が設計され得

40

50

、必要に応じて、最適化され得ることが認識されるだろう。長さの差は、サイズの増加またはサイズの減少であり得ることが認識されるだろう。

【0164】

いくつかの実施形態において、予想外の捕捉頻度の最初の決定は、サイズの差の存在を示す。いくつかの実施形態において、捕捉頻度の増加は、欠失を示唆する。いくつかの実施形態において、捕捉頻度の減少は、挿入を示唆する。しかしながら、捕捉プローブの特定の配列パラメータおよび相対的サイズ、標的領域ならびに欠失または挿入に応じて、捕捉頻度の変化は、標的領域の長さの増加または減少に関連し得ることが認識されるだろう。いくつかの実施形態において、その変化の正確な性質は、本明細書中に記載されるような1つ以上の追加の手法を用いて測定され得る。

10

【0165】

したがって、いくつかの局面において、MIPプローブは、2つのハイブリダイゼーション配列または標的化アーム（一方は直鎖状プローブの各端に位置する）を含む直鎖状の核酸鎖を含み、ここで、ハイブリダイゼーション配列の各々は、標的核酸の同じ鎖における別個の配列に相補的であり、標的核酸におけるこれらの配列は、目的の標的核酸配列の2つの末端に隣接する。ハイブリダイゼーションの際、プローブの5'末端と3'末端の両方が同じ鎖にハイブリダイズすることにより、標的領域に隣接する領域を分断するという意味において、プローブの2つの末端は互いに対して反転している（例えば、図9に図示されるように）ことが認識されるだろう。

20

【0166】

いくつかの実施形態において、ハイブリダイゼーション配列は、約10～100ヌクレオチド長、例えば、約10～30、約30～60、約60～90または約20、約30、約40もしくは約50ヌクレオチド長である。しかしながら、用途に応じて、他の長さを使用してもよい。いくつかの実施形態において、プローブの両方の標的化アームのハイブリダイゼーションTmは、同様であるように設計されるかまたは選択される。いくつかの実施形態において、異なる標的領域を捕捉するために設計された複数のプローブの標的化アームのハイブリダイゼーションTmは、同様であるように選択されるかまたは設計され、その結果、それらは多重反応において共に使用することができる。したがって、充填前のMIPプローブの代表的なサイズは、約60～80ヌクレオチド長である。しかしながら、標的化アームのサイズおよびMIPプローブに存在する他の任意の配列（例えば、プライマー結合配列またはタグ配列）のサイズに応じて、他のサイズを使用してもよい。いくつかの実施形態において、MIPプローブは、配列依存的な二次構造を回避するように設計される。いくつかの実施形態において、MIPプローブは、標的化アームが、公知の多型の領域と重複しないように設計される。いくつかの実施形態において、脆弱X遺伝子座のリピート領域を捕捉するために使用され得る標的化アームは、捕捉される鎖に応じて、以下の配列を有し得るかまたはこれらの配列と相補的な配列を有し得る。

30

左：CTCCGTTTCGGTTTCACCTTC（配列番号181）

右：ATCTTCTCTTTCAGCCCTGCT（配列番号182）。

【0167】

これらの標的化アームを用いて捕捉される代表的なサイズは、約100ヌクレオチド長（例えば、トリヌクレオチドリピートの約30リピート）である。

40

【0168】

いくつかの実施形態において、反復領域の「代表」に対して得られるリードの数は、得られるリードの総数に依存するので、そのリードの数は、標的の長さを推定する情報価値がない。これを克服するために、集団の中で多型が存在しないかまたは多型が最小である他の「コントロール」領域を標的にする1つ以上のプローブを含めることが有用である。捕捉効率が体系的に一貫するので（例えば、図9を参照のこと）、反復領域の「代表」に対して得られたリードと、コントロール領域に対して得られたリードとの比は、規定数のリピートを有するDNAを用いて調整され得る。最終的には、その比は、図10に図示されるように、リピートの長さの基準として働き得る。図10は、「コントロール」領域の

50

リードに対するリピートの「代表」のリードの比によって測定される、標的のギャップサイズと反復領域のリードの相対数との非限定的な仮定的関係性を図示している。y軸の単位は、任意である。

【0169】

いくつかの実施形態において、同様のサイズ範囲を有する標的をより上手く見分けるために、ショットガンライブラリーを作製することによって（例えば、捕捉された配列（例えば、MIPプローブを用いて捕捉された配列）からショットガンライブラリーを作製することによって）反復領域全体が配列決定され得る。リピートが長いほど、リピートの短いリードが多く得られ得る。ゆえに、標的の長さは、「反復」リードの相対数に対して2倍寄与し得、それにより、区別している標的のよりよい分離をもたらされ得る。いくつかの実施形態において、任意の所与のリピートからのリードの数は、存在するリピートの数の一次関数であり得るといことが期待される。しかしながら、いくつかの実施形態において、ポアソンサンプリングによって誘導される広がりは、考慮される必要があり得、いくつかの実施形態において、その分離を限定するのに十分に大きい場合がある。

10

【0170】

二倍体サンプル由来の両方の対立遺伝子の長さの正確な測定が望まれるとき、さらなる操作が必要になり得る。これは、測定される捕捉効率が、実際にはその2つの対立遺伝子の平均効率であり得るからである。各対立遺伝子に対する別個の測定を効率的に達成するために、後に続く（個別のゲノム遺伝子座からの）個別の捕捉事象の効率化を可能にするバーコード（例えば、配列タグ）が使用され得る。図11A~Cは、上記のアプローチを示している。反応物中に同じ識別タグ配列を有する2つの任意のMIPが存在する確率が低くなるように、所与の遺伝子座に対して、それらの骨格に多数の識別タグのうちの一つを含むMIPを合成する。MIP捕捉をそのサンプルに対して行う；その反応は、より短い標的の長さに偏り、ゆえに、その反応産物は、「長い」環よりも「短い」環を多く含み得る。各環は、独特の識別タグ配列を有するはずである。次いで、直鎖RC A（LRCA）をそれらの環に対して行う。LRCA反応において、環は、それら自体の長い直鎖状のコンカテマーに変換される。そのコンカテマーが「一定の」長さ（そのポリメラーゼの処理能力/エラー率に基づいて）に達したら、所与の環に対するLRCA反応を停止する。ゆえに、より小さい環から得られたコンカテマーは、より多いコピー数の識別タグを含み得、より大きい環から得られたコンカテマーは、より少ないコピー数の識別タグを含み得る。各識別タグ配列の数は、例えば、次世代シーケンシングによって、計数される。出現数を識別タグIDに対してプロットすると、データは当然、二倍体サンプルにおける2つの対立遺伝子の長さを反映する2つの群にクラスター化する。ゆえに、既知の標準を用いて長さの絶対較正が行われた後、対立遺伝子の長さは、このグラフから直接読み取ることができる。いくつかの実施形態において、シーケンシング法（例えば、次世代シーケンシング法）を用いることにより、1つ以上の捕捉された標的（例えば、またはそのアンプリコン）の一部が配列決定され、それらの配列を用いることにより、存在する種々のバーコードの数が計数される。したがって、いくつかの実施形態において、本発明の局面は、高度に多重化されたqPCR反応に関する。

20

30

【0171】

挿入もしくは欠失またはリピート配列が疾患または状態に関連し得る遺伝子座の他の非限定的な例が、表3および4に提供される。これらの遺伝子座のいずれか1つ以上における異常な長さの存在は、本発明の局面にしたがって評価され得ることが認識されるだろう。いくつかの実施形態において、これらの遺伝子座または他の遺伝子座の2つ以上は、同じ反応条件下において生物学的サンプル中の異なる標的核酸とハイブリダイズするように設計された種々のプローブを用いる単一の多重反応において評価され得る。

40

【0172】

【表 3】

表 3.

ポリグルタミン(PolyQ)病

タイプ	遺伝子	正常/野生型	病原性
DRPLA (歯状核赤核淡蒼球ルイ体萎縮症)	ATN1または DRPLA	6 - 35	49 - 88
HD (ハンチントン病)	HTT (ハンチンチン)	10 - 35	35+
SBMA (球脊髄性筋萎縮症または ケネディ病)	X染色体上のアンドロ ゲンレセプター	9 - 36	38 - 62
SCA1 (脊髄小脳失調症1型)	ATXN1	6 - 35	49 - 88
SCA2 (脊髄小脳失調症2型)	ATXN2	14 - 32	33 - 77
SCA3 (脊髄小脳失調症3型また はマンヤド・ジョセフ病)	ATXN3	12 - 40	55 - 86
SCA6 (脊髄小脳失調症6型)	CACNA1A	4 - 18	21 - 30
SCA7 (脊髄小脳失調症7型)	ATXN7	7 - 17	38 - 120
SCA17 (脊髄小脳失調症17型)	TBP	25 - 42	47 - 63

10

20

【 0 1 7 3 】

【表 4】

表 4.

非ポリグルタミン病

タイプ	遺伝子	コドン	正常/野生型	病理性
FRAXA(脆弱X症候群)	FMR1, X染色体上	CGG	6 - 53	230+
FXTAS(脆弱X関連振戦/運動失調症候群)	FMR1, X染色体上	CGG	6 - 53	55-200
FRAXE(脆弱XE精神遅滞)	AFF2 または FMR2, X染色体上	GCC	6 - 35	200+
FRDA(フリートライヒ運動失調)	FXN または X25, (フラタキシン)	GAA	7 - 34	100+
DM(筋緊張性ジストロフィ)	DMPK	CTG	5 - 37	50+
SCA8(脊髄小脳運動失調8型)	OSCA または SCA8	CTG	16 - 37	110 - 250
SCA12(脊髄小脳運動失調12型)	PPP2R2B または SCA12	CAG 5'末端上	7 - 28	66 - 78

10

20

以下の実施例は、本発明の局面および実施形態を例証するものであり、限定または制限を意図しない。本明細書を検討すれば、本発明の多くのバリエーションが当業者に明らかになるだろう。本発明の全範囲は、等価物の全範囲に加えて請求項、および上記のようなバリエーションに加えて明細書を参照することにより、決定されるべきである。

30

【0174】

4. 検出感度の上昇:

いくつかの実施形態において、本発明の局面は、核酸検出アッセイの感度を高めるための方法に関する。

【0175】

現在、次世代(ポロニーベースの)シーケンシング(とりわけ、ゲノムリシーケンシング、遺伝子発現についてのRNA-seq、メチル化についての亜硫酸水素塩シーケンシングおよびImmune-seqを含む)を利用してデータを生成する多くのゲノミクスアッセイが存在する。定量的測定(遺伝子型コールを含む)を行うために、これらの方法は、核酸の元のサンプルにおけるその配列の提示に対する代理として所与のゲノム遺伝子座のシーケンシングリードの数を利用する。これらの手法の大部分は、目的のサンプルを代表するDNA分子の高複雑度ライブラリーを構築する調製工程を必要とする。現在のアッセイは、いくつかの代替の核酸調製法(例えば、増幅、例えば、PCRベースの増幅;配列特異的捕捉、例えば、固定化された捕捉プローブを用いるもの;または環状化されたプローブへの標的捕捉の後の配列分析工程)のうちの1つを使用する。核酸単離手法および配列分析手法の予測不能性(確率的性質)に関連するエラーを減少させるために、現在の方法は、元の核酸サンプル中に存在するすべての配列が最終的な配列データにおいて提示される尤度を高めるために、標的核酸調製物の過剰サンプリングを伴う。例えば、ゲノムシーケンシングライブラリーは、そのライブラリーの構築プロセスにおける

40

50

確率的変動の結果として、ソース核酸サンプル（例えば、ゲノム調製物）由来の特定の配列の提示過剰または提示不足を含み得る。そのような変動は、シーケンシングライブラリー内に存在しないゲノム由来の標的配列またはシーケンシングライブラリー内の検出不可能なゲノム由来の標的配列に生じるとき、特に問題であり得る。例えば、シーケンシングライブラリー内のゲノム由来の特定の対立遺伝子の配列（例えば、ヘテロ接合の対立遺伝子）の提示不足は、シーケンシングライブラリーにおいて明らかなホモ接合の提示をもたらし得る。

【0176】

対照的に、本発明の局面は、標的核酸を単に過剰サンプリングするのではなく、核酸配列の分析を、サンプリングされる核酸のタイプで異なる系統誤差を有する2つ以上の異なる核酸調製手法からの結果に基づかせることに関する。いくつかの実施形態によると、異なる手法は、体系的でありかつ単純に核酸の捕捉中または増幅中の確率的影響に起因しない異なる配列バイアスを有する。したがって、いくつかの実施形態において、核酸調製における変動を克服するために必要な過剰サンプリングの程度は、バイアスを克服するのに十分である必要がある。いくつかの実施形態において、本発明は、異なるバイアスを有する2つ以上の異なる核酸調製手法から得られた核酸および/または配列の結果を組み合わせることによって、過剰サンプリングの必要性を低下させる方法を提供する。

【0177】

本発明によると、異なる手法は、異なる特徴的なバイアスまたは系統誤差を有する。例えば、1つの手法は、サンプル分析を目的の遺伝子座における1つの特定の対立遺伝子に偏らせ得るのに対し、別の手法は、サンプル分析を同じ遺伝子座の別の対立遺伝子に偏らせ得る。したがって、同じサンプルが、配列分析にむけて核酸を調製するために使用される手法のタイプに応じて、異なると同定され得る。各手法が、目的の多型配列に対して異なる相対的な感度を有するので、これは、感度の問題を効率的に表す。

【0178】

本発明の局面によると、核酸分析の感度は、異なる核酸調製工程からの配列を組み合わせ、組み合わせられた配列情報を診断アッセイのために（例えば、被験体が目的の遺伝子座においてホモ接合であるかまたはヘテロ接合であるかに関するコールを生成するために）用いることによって、高められ得る。

【0179】

現在、DNAシーケンシングが変異を検出する能力は、上流のサンプル単離方法（例えば、増幅、固定化濃縮、環状化捕捉などによる方法）が目的の遺伝子座を、信頼性をもって単離する能力によって限定されている。二倍体ゲノム（例えば、分子診断用シーケンシングのために提供されるヒトサンプル）についてヘテロ接合のベースコールを生成したい場合、いくつかの実施形態では、その単離方法がほぼ均一または完全に均一（目的の遺伝子座に対してヘテロ接合体またはホモ接合体と明確に「コール」されるのに少なくとも十分均一）の量の2つの配列決定された対立遺伝子を生成することが重要である。

【0180】

サンプル調製方法は、3つのクラス：1）単一またはいくつかの標的の増幅（例えば、一重（uniplex）PCR、「多重」PCR）、2）多標的（multi-target）ハイブリダイゼーション濃縮（例えば、Agilent SureSelect「ハイブリッド」捕捉 [Gnirkeら、2009, Nature methods 27:182-9]、Roche/Nimblegen「配列捕捉」 [Hodgesら、2007, Nature genetics 39:1522-7]、および3）多標的環状化選択（例えば、分子反転プローブまたはパドロックプローブ [Porrecaら、2007, Nature methods 4:931-6、Turnerら、2009, Nature methods 6:315-6]、「セクター」 [Dahlら、2005, Nucleic acids research 33:e71]）に分けられ得る。これらの方法の各々は、インプットの存在量の分布を適切に表さない単離された産物のプールを生じ得る。例えば、ヘテロ接合の位置における2つの対立遺伝子は、インプット

10

20

30

40

50

の50:50という比から下流のシーケンシングにおいてベースコールを見落とす比に歪められ得る。例えば、その比が50:50から10:90に歪められ、かつそのサンプルが10×平均カバレッジに配列決定される場合、2つの対立遺伝子のうちの1つが10個のシーケンシングリードにおいて1回も観察されない確率が高い。これにより、ヘテロ接合の位置がホモ接合に変換されることによってシーケンシング方法の感度が低下し得る(もしかすると「変異」対立遺伝子は、観察されない対立遺伝子であるかもしれない)。いくつかの実施形態において、歪められた比は、不均一な腫瘍組織に存在する変異を検出する感度を低下させる特定の問題である。例えば、不均一なサンプルにおいて分析される細胞のわずか10%しか、ヘテロ接合性の変異を有していない場合、その変異は、配列リードの50%ではなく5%に存在すると予想され得る。このシナリオでは、ロバストな高感度の検出の必要性がなおも重大であり得る。

10

【0181】

本明細書中に開示される方法は、ある特定のクラスの単離方法が種々の様式のバイアスを有するという発見に部分的に基づく。本開示は、1つのサンプルに対して複数の単離方法の組み合わせ(例えば、本明細書中に開示されるクラスのうちの少なくとも2つから1つ以上)を用いることによって下流のシーケンシングの感度を高めるための方法を提供する。これは、疾患関連変異を「見落とす」可能性を最小にするために高感度が要求される分子診断法において特に重要である。例えば、環状化選択後のシーケンシングに対して 1×10^{-3} という名目上の偽陰性のエラー率およびハイブリダイゼーション濃縮後のシーケンシングに対して 1×10^{-3} という偽陰性のエラー率が与えられるとき、そのサンプルに対して両方の手法を行うことによって、 1×10^{-6} という最終的な偽陰性の割合が達成され得る(各方法における失敗は完全に独立していると仮定する)。単一の完全浸透の変異対立遺伝子によって引き起こされる保因者の頻度が0.1の劣性の疾患の場合、見落とされる保因者診断の数は、試験された100万人の患者あたり1000人から試験された100万人の患者あたり1人にまで減少し得る。さらに、この試験が、出生前保因者スクリーニングにおいて使用される場合、一方の親において保因者であるというコールを見落とす結果として生まれる罹患小児の数は、出生児100万人あたり25人から10億人あたり25人にまで減少し得る。

20

【0182】

さらに、本開示は、疾患関連対立遺伝子を含む領域におけるシーケンシングカバレッジを効率的に上げる調製方法の組み合わせを提供する。ヘテロ接合体のエラー率は、50:50の対立遺伝子提示からの偏差と、次世代DNAシーケンシングの場合の平均存在量からの偏差との両方に広く関係しているので(単離されたより少ない標的は、一方または両方の対立遺伝子においてより少なくサンプリングされる可能性が高い)、これらの領域におけるカバレッジを選択的に上げることにより、感度も選択的に上がり得る。さらに、特定の公知の疾患関連変異の存在または非存在を検出するMIPを用いることにより、感度を選択的に高めることができる。いくつかの実施形態において、これらのMIPは、最も3'側の領域が、予想される変異と相補的であり、かつ0bpまたはそれ以上のbpの充填長を有する、標的化アームを有し得る。よって、変異が存在する場合のみ、MIPは形成し、その存在は、シーケンシングによって検出され得る。

30

40

【0183】

さらに、本明細書中に開示されるアルゴリズムを用いることにより、様々なレベルのストリンジェンシーで、所与の位置が任意の公知の疾患関連対立遺伝子を有するか否かに応じた塩基同一性の決定が行われ得る。ストリンジェンシーは、コンセンサスベースコールを生成するために必要な観察される変異リードの最小数を減少させることによって、そのような位置において低下され得る。これは、特異性の低下を犠牲にして変異対立遺伝子の検出に対する感度を効率的に高め得る。

【0184】

本発明の実施形態は、例えば、標的核酸における多型の存在を検出するために、MIPと、ハイブリダイゼーション濃縮と、必要に応じて一般的な疾患に関連する公知の特定の

50

遺伝子座を標的にした別のMIPとを組み合わせる。非限定的な例が、例えば、標的核酸における多型の存在を検出するために、MIPと、ハイブリダイゼーション濃縮と、必要に応じて一般的な疾患に関連する公知の特定の遺伝子座を標的にした別のMIPとを用いた概略図を図示している図12に例証されている。

【0185】

図13および14は、MIPベースの捕捉に対する種々の捕捉効率を図示している。図13は、MIP捕捉による標的1つあたりの存在量のグラフを示している。このグラフでは、ここに多く存在しない標的は、両方の対立遺伝子を適切にサンプリングするためにシークエンシングにおいて十分な深度でカバーされる可能性が低いので、バイアスが、ヘテロ接合体のエラー率を大きく動かす。これは、Turnerら、2009, Nature methods 6:315-6からのものである。ハイブリダイゼーション濃縮は、定性的に類似の存在量の分布をもたらすが、所与の標的の存在量は、おそらく、2つの方法の間で相関しない。図14は、Ballら、2009, Nature biotechnology 27:361-8からの2つのMIP捕捉反応の相関関係のグラフを示している。各点は、複製物1および複製物2における標的の存在量を表している。ピアソン相関 $r = 0.956$ 。これは、MIP捕捉が、標的を特定の存在量に再現性よく偏らせることを示唆する。ハイブリダイゼーション濃縮は、同様に1つの捕捉から次の捕捉へと相関する。

10

【0186】

本発明の局面によると、そのようなバイアスは、複数の被験体サンプル中の複数の遺伝子座を調べるアッセイにおいて種々の捕捉手法および/または分析手法を体系的に組み合わせることによって、検出され得るかまたは克服され得る。

20

【0187】

したがって、本明細書中に記載される実施形態のいずれか（例えば、タイリング/スタガリング、タギング、サイズ検出、感度増強アルゴリズムまたはそれらの任意の組み合わせ）において、本発明の局面は、ゲノム核酸を調製すること、および/またはそれらを1つ以上の異なるプローブ（例えば、捕捉プローブ、ハイブリダイゼーションプローブ、MIPなど）と接触させることを含むことが認識されるだろう。いくつかの実施形態において、1被験体あたり使用されるゲノム核酸の量は、1ng~10マイクログラム（例えば、500ng~5マイクログラム）の範囲である。しかしながら、それよりも多い量または少ない量（例えば、1ng未満、10マイクログラム超、10~50マイクログラム、50~100マイクログラムまたはそれ以上）も使用され得る。いくつかの実施形態において、目的の各遺伝子座について、1アッセイあたり使用されるプローブの量は、特定の用途に対して最適化され得る。いくつかの実施形態において、プローブとゲノム等価物（例えば、半数体または二倍体のゲノム等価物、例えば、核酸標的または目的の遺伝子座の各対立遺伝子または両方の対立遺伝子に対して）との比（モル比、例えば、濃度比として測定される比）は、1/100、1/10、1/1、10/1、100/1、1000/1とさまざまである。しかしながら、それより低い比、高い比または中間の比を使用してもよい。

30

【0188】

いくつかの実施形態において、各反応に使用される標的核酸およびプローブの量は、正規化されることにより、濃度または比が異なることによって引き起こされる、観察される任意の差が回避される。いくつかの実施形態において、ゲノムDNAおよびプローブを正規化するために、ゲノムDNAの濃度が、標準的な分光光度計または蛍光を用いて（例えば、蛍光インターカレート色素を用いて）読まれる。プローブ濃度は、実験によって、またはプローブ製造者が明記する情報を用いて、決定され得る。

40

【0189】

同様に、いったん遺伝子座が捕捉されると（例えば、MIPもしくは他のプローブ上に、または別の形態で）、その遺伝子座は、1つ以上のプライマーが関わる反応において増幅され得、そして/または配列決定され得る。各反応に対して加えられるプライマーの量

50

は、 $0.1 \text{ pmol} \sim 1 \text{ nmol}$ 、 $0.15 \text{ pmol} \sim 1.5 \text{ nmol}$ の範囲（例えば、およそ 1.5 pmol ）であり得る。しかしながら、他の量（例えば、それより少ない量、多い量または中間の量）を使用してもよい。

【0190】

いくつかの実施形態において、1つ以上の介在配列（例えば、MIP捕捉プローブ上の第1標的化アームと第2の標的化アームとの間の配列）、識別配列もしくはタグ配列、または標的配列（例えば、ゲノム標的配列）にハイブリダイズするように設計されていない他のプローブ配列は、標的配列または生物学的サンプル中に存在し得る他の配列（例えば、他のゲノム配列）に対する過度の相補性を回避するように（クロス-ハイブリダイゼーションを回避するように）設計されるべきであることが認識されるだろう。例えば、これらの配列は、任意のゲノム配列と十分な数のミスマッチ（例えば、30塩基のうち少なくとも5、10、15またはそれ以上のミスマッチ）を有するように、またはハイブリダイゼーション反応温度よりも低い（例えば、少なくとも5、10、15、20 またはそれ以上低い） T_m （例えば、ミスマッチ T_m ）を有するように、設計され得る。

10

【0191】

分析される核酸がDNA（例えば、ゲノムDNA）である場合、本明細書中で使用される標的化アームは、目的の遺伝子座のいずれかの鎖にハイブリダイズする（例えば、相補的である）ように設計され得ることが認識されるだろう。しかしながら、MIPプローブにおいては、一方の標的化アームに対して選択されるどちらかの鎖が他方の標的化アームに対しても用いられ得る。しかしながら、RNA分析においては、標的化アームは、転写されたRNAにハイブリダイズするように設計されるべきであることが認識されるだろう。標的配列を「捕捉する」と本明細書中で言及されるMIPプローブは、実際のところ、実際の標的分子の捕捉によってではなく鋳型に基づいた合成によって標的配列を捕捉することも認識されるだろう（例えば、アームが、標的分子にハイブリダイズする最初の段階以外、または標的分子が変性されるまでもしくは別途除去されるまで、伸長されたMIP産物に標的分子が結合したままであり得るという意味において）。

20

【0192】

いくつかの実施形態において、標的化アームは、1つの対立遺伝子または変異（例えば、SNPまたは他の多型、変異など）と相補的な配列を含み得、その結果、プローブは、その対立遺伝子または変異を有する標的核酸を優先的にハイブリダイズする（および捕捉する）ことが認識されるだろう。しかしながら、多くの実施形態において、各標的化アームは、評価されている集団の被験体において多型でない配列にハイブリダイズする（例えば、相補的である）ように設計される。これにより、すべての対立遺伝子に対して標的配列が捕捉されることおよび/または配列決定されることが可能になり、被験体間（例えば、1つ以上の遺伝子座に対するヘテロ接合またはホモ接合のコール間）の差は、本明細書中に記載されるような配列情報および/または頻度に基づき得る。

30

【0193】

配列タグ（バーコードとも呼ばれる）は、プローブ内またはプローブのファミリー内の他の位置に出現しないという点において独特であるように設計され得、また、それらは、標的にされている配列内にも出現しないことが認識されるだろう。ゆえに、それらを用いることにより、他の特徴を有する（例えば、特定の被験体および/または特定の遺伝子座に対して）特定のプローブを一義的に同定することができる（例えば、シーケンシングまたはハイブリダイゼーション特性によって）。

40

【0194】

いくつかの実施形態において、プローブもしくはプローブの領域または他の核酸は、ある特定の配列または配列の特徴（例えば、長さ、他の特性など）を含む（comprising）かまたは含む（including）と本明細書中に記載されることも認識されるだろう。しかしながら、いくつかの実施形態では、それらの実施形態のいずれかの文脈において、プローブもしくはプローブの領域または他の核酸のいずれもが、それらの領域（例えば、アーム、中央領域、タグ、プライマー部位など、またはそれらの任意の組み合

50

わせ)からなるか、またはそれらの配列からなるか、または本明細書中に記載されるような1つ以上の特徴(例えば、長さまたは他の特性など)からなる特徴を有する配列を有する(例えば、タイリングプローブもしくははずらされたプローブ、タグ化されたプローブ、長さの検出、感度増強アルゴリズムまたはそれらの任意の組み合わせについて)ことが認識されるだろう。

【0195】

本明細書中で設計されるかまたは使用されるプローブ、プライマーおよび他の核酸は、合成のもの、天然のものまたはそれらの組み合わせであり得ることが認識されるだろう。したがって、本明細書中で使用されるとき、用語「核酸」とは、複数の連結されたヌクレオチド(すなわち、ピリミジン(例えば、シトシン(C)、チミジン(T)またはウラシル(U))またはプリン(例えば、アデニン(A)またはグアニン(G))である交換可能な有機塩基に連結された糖(例えば、リボースまたはデオキシリボース)を含む分子)のことを指す。「核酸」および「核酸分子」は、交換可能に使用され得、それらは、オリゴリボヌクレオチドならびにオリゴデオキシリボヌクレオチドのことを指す。これらの用語は、ポリヌクレオチド(すなわち、ポリヌクレオチドからリン酸を引いたもの)および他の任意の有機塩基を含む核酸も含むものとする。それらの有機塩基には、アデニン、ウラシル、グアニン、チミン、シトシンおよびイノシンが含まれる。別段述べられない限り、核酸は、一本鎖または二本鎖であり得る。その核酸は、天然に存在するものまたは天然に存在しないものであり得る。核酸は、自然源から得ることができ、または核酸合成装置を用いて合成され得る(すなわち、合成のものであり得る)。核酸の回収および単離は、当該分野において通例のとおり行われ、適当な方法は、標準的な分子生物学の教科書に見られる(例えば、Maniatis' Handbook of Molecular Biologyを参照のこと)。核酸は、DNAまたはRNA(例えば、ゲノムDNA、ミトコンドリアDNA、mRNA、cDNA、rRNA、miRNAまたはそれらの組み合わせ)であり得る。天然に存在しない核酸(例えば、細菌人工染色体(BAC)および酵母人工染色体(YAC))も使用され得る。

10

20

【0196】

本発明は、核酸誘導体の使用も企図する。本明細書中に記載されるように、ある特定の核酸誘導体の使用は、特にそれらがヌクレアーゼを含み得る生物学的サンプルに曝露されたときの、それらの消化を防ぐことによって、本発明の核酸の安定性を高め得る。本明細書中で使用されるとき、核酸誘導体は、天然に存在しない核酸またはその単位である。核酸誘導体は、天然に存在しないエレメント(例えば、天然に存在しないヌクレオチドおよび天然に存在しない骨格結合(backbone linkages))を含み得る。

30

【0197】

核酸誘導体は、骨格修飾(例えば、ホスホロチオエート結合、ホスホジエステル修飾された核酸、ホスホロチオレート修飾、ホスホジエステル核酸とホスホロチオエート核酸との組み合わせ、メチルホスホネート、アルキルホスホネート、リン酸エステル、アルキルホスホノチオエート、ホスホルアミデート、カルバメート、カーボネート、リン酸トリエステル、アセトアミデート、カルボキシメチルエステル、メチルホスホロチオエート、ホスホロジチオエート、p-エトキシおよびそれらの組み合わせであるがこれらに限定されない)を含み得る。核酸の骨格の組成は、均一または不均一であり得る。

40

【0198】

核酸誘導体は、糖および/または塩基における置換または修飾を含み得る。例えば、核酸誘導体は、3'位にヒドロキシル基以外および5'位にリン酸基以外の低分子量の有機基に共有結合した骨格の糖(例えば、2'-O-アルキル化されたりボース基)を有する核酸を含む。核酸誘導体は、アラビノースなどの非リボース糖を含み得る。核酸誘導体は、置換プリンおよび置換ピリミジン(例えば、C-5プロピンで修飾された塩基、5-メチルシトシン、2-アミノプリン、2-アミノ-6-クロロプリン、2,6-ジアミノプリン、ヒポキサンチン、2-チオウラシルおよびシュードイソシトシン(pseudocytosine))を含み得る。いくつかの実施形態において、置換は、糖/塩基

50

、塩基に結合した基（ビオチン、蛍光基（フルオレセイン、シアニン、ローダミンなど）、化学的に反応性の基（カルボキシル、NH₂、チオールなどを含む）を含む）またはそれらの任意の組み合わせにおける1つ以上の置換/修飾を含み得る。

【0199】

核酸は、ペプチド核酸（PNA）、ロックト（locked）核酸（LNA）、DNA、RNAまたはそれらの共核酸（co-nucleic acids）（例えば、DNA-LNA共核酸）であり得る。PNAは、グリシンアミノ窒素およびメチレンカルボニルリンカーを介してヌクレオチド塩基に連結された2-アミノエチルグリシン残基で置き換えられたリン酸骨格を有するDNAアナログである。PNAは、ワトソン-クリック塩基対形成によってDNA標的とRNA標的の両方に結合し得、その際、場合によっては、DNAまたはRNAに基づくオリゴヌクレオチドを用いて形成され得るハイブリッドよりも強いハイブリッドを形成し得る。

10

【0200】

PNAは、ペプチド結合によって接続されるモノマーから合成される（Nielsen, P. E. 著、Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)）。それらのモノマーは、標準的な固相ペプチド合成法を用いて作製され得る。PNAの化学および合成によって、PNAの設計においてアミノ酸およびポリペプチド配列を含めることが可能になる。例えば、リジン残基を用いることにより、PNA骨格に正電荷が導入され得る。アミノ酸側鎖の修飾のために利用可能なすべての化学的アプローチが、PNAに対して直接、適用可能である。いくつかのタイプのPNAのデザインが存在し、これらには、一本鎖PNA（ssPNA）、bisPNAおよび偽相補（pseudocomplementary）PNA（pcPNA）が含まれる。

20

【0201】

PNA/DNA複合体の構造は、特定のPNAおよびその配列に依存する。ssPNAは、好ましくは、逆平行の向きで（すなわち、ssPNAのN末端がssDNAの3'末端と整列するように）、かつワトソン-クリック対形成によって、一本鎖DNA（ssDNA）に結合する。PNAはまた、フーグスティーン塩基対形成によってDNAに結合し得、それにより、二本鎖DNA（dsDNA）と三重鎖を形成し得る（Wittung, P. 著、Biochemistry 36: 7973 (1997)）。

30

【0202】

ロックト核酸（LNA）は、改変されたRNAヌクレオチドである。LNAは、DNAとハイブリッドを形成し、そのハイブリッドは、PNA/DNAハイブリッドと少なくとも同程度に安定である（Braasch, D. A. 著、Chem & Biol. 8 (1): 1-7 (2001)）。それゆえ、LNAは、PNA分子であり得るように使用され得る。LNAの結合効率は、正電荷をLNAに付加することによって、いくつかの実施形態において高まり得る。LNAは、本質的に高い結合親和性を有すると報告されている。

【0203】

市販の核酸合成装置および標準的なホスホルアミダイト化学を用いることにより、LNAが作製される。それゆえ、混合性のLNA/DNA配列の作製は、混合性のPNA/ペプチド配列の作製と同程度に簡便である。LNAモノマーの安定化の効果は、相加効果ではない。LNAモノマーは、隣接するデオキシヌクレオチドの糖環のコンフォメーションに影響し、その隣接するデオキシヌクレオチドの糖環をより安定な配置に変える（Nielsen, P. E. 著、Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)）。また、配列内のLNA残基がより少ないことにより、合成の正確度は劇的に改善される。核酸を結合体化するための生化学的なアプローチのほとんどが、LNA/DNA構築物に適用可能である。

40

【0204】

50

これらおよび他の本発明の局面は、以下の非限定的な例によって例証される。

【実施例】

【0205】

以下の実施例は、本発明の非限定的な実施形態を例証するものである。

【0206】

実施例1：ヒト標的エクソン用の捕捉プローブセットの設計

すべての標的が、部分的に重複するサブターゲットのセットとして捕捉される。例えば、タイリングアプローチにおいて、200bpの標的エクソンは、12個のサブターゲット（各々60bp長）のセットとして捕捉され得る（図1）。各サブターゲットは、他の2または3つの標的と部分的に重複するように選択される。

10

【0207】

いくつかの実施形態において、すべてのプローブが、3つの領域：1)サブターゲットからすぐ上流とハイブリダイズする配列を含む20bpの「標的化アーム」、2)増幅プライミング部位の対として使用される配列を含む30bpの「定常領域」、および3)サブターゲットからすぐ下流とハイブリダイズする配列を含む第2の20bpの「標的化アーム」から構成される。標的化アームの配列は、セット内の捕捉プローブごとに異なるが、定常領域の配列は、そのセット内のすべてのプローブについて同じであることから、すべての捕捉された標的を単一のプライマーセットで増幅することが可能である。標的化アームの配列は、20bp配列の任意の所与の対が標的ゲノム内で独特であるように（望まれない部位の偽の捕捉を防止するように）設計されるべきである。さらに、融解温度は、ハイブリダイゼーション効率が一定温度（例えば、60）においてすべてのプローブについて均一になるように、そのセット内のすべてのプローブに対して一致させるべきである。ゲノム標的と塩基対形成する能力を損ない得る強い二次構造を形成しないことを確実にするために、標的化アームの配列は、計算的にスクリーニングされるべきである。

20

【0208】

ヒトゲノムサンプルへの捕捉プローブのハイブリダイズ

ハイブリダイゼーション反応物を組み立てる：

- ・ 1.0 μ l の捕捉プローブ混合物（約 2.5 pmol）
- ・ 2.0 μ l の 10 \times Ampligase 緩衝液（Epicentre）
- ・ 6.0 μ l の 500 ng / μ l ヒトゲノム DNA（約 16.7 fmol）
- ・ 11 μ l の dH₂O。

30

【0209】

サーマルサイクラーにおいて、ゲノムDNAを変性するために95にて5分間の加熱反応を行い、次いで、60にて冷却する。60にて40時間インキュベートする。

【0210】

ハイブリダイズされたプローブを、サブターゲットを含む共有結合的に閉じた環状産物に変換する

充填/ライゲーション反応混合物を調製する：

- ・ 0.25 μ l の 2 mM dNTP 混合物（Invitrogen）
- ・ 2.5 μ l の 10 \times Ampligase 緩衝液（Epicentre）
- ・ 5.0 μ l の 5 U / μ l Taq Stoffel フラグメント（Applied Biosystems）
- ・ 12.5 μ l の 5 U / μ l Ampligase（Epicentre）
- ・ 4.75 μ l の dH₂O。

40

【0211】

1.0 μ l のこの混合物をハイブリダイズされたプローブ反応物に加え、60にて10時間インキュベートする。

【0212】

環状化されたプローブ/サブターゲット産物を未反応のプローブおよびゲノムDNAから精製する

50

エキソヌクレアーゼ反応混合物を調製する：

- ・ 21 μ l の充填 / ライゲーション反応産物
- ・ 2.0 μ l の 10 x エキソヌクレアーゼ I 緩衝液 (New England Biolabs)
- ・ 2.0 μ l の 20 U / μ l エキソヌクレアーゼ I (New England Biolabs)
- ・ 2.0 μ l の 100 U / μ l エキソヌクレアーゼ III (New England Biolabs)。

【0213】

37 で 60 分間インキュベート (Incubate) し、次いで、80 で 15 分間インキュベートすることによって熱失活させる。貯蔵のために、直ちに 4 に冷却する。

10

【0214】

プローブの「定常領域」に特異的なプライマーを用いて PCR によって環状材料を増幅する

PCR 混合物を調製する：

- ・ 5.0 μ l の 10 x Accuprime 反応緩衝液 (Invitrogen)
- ・ 1.5 μ l の 10 μ M CP-2-FA (5' - GCACGATCCGACGGTAGTGT - 3') (配列番号 183)
- ・ 1.5 μ l の 10 μ M CP-2-RA (5' - CCGTAATCGGGAAGCTGAAG - 3') (配列番号 184)
- ・ 0.4 μ l の 25 mM dNTP 混合物 (Invitrogen)
- ・ 2.0 μ l の熱失活したエキソヌクレアーゼ反応混合物
- ・ 1.5 μ l の 10 x Sybr Green (Invitrogen)
- ・ 0.4 μ l の 2.5 U / μ l Accuprime Pfx ポリメラーゼ (Invitrogen)
- ・ 37.7 μ l の dH₂O。

20

【0215】

以下のプロトコルに従ってリアルタイムサーマルサイクラーにおいて熱サイクル反応を行うが、増幅の収量がプラトーに達する前にサイクル反応を停止する (通常、8 ~ 12 サイクル)：

30

1. 95 で 5 分間
2. 95 で 30 秒間
3. 58 で 60 秒間
4. 72 で 60 秒間
5. 2 に戻り、さらに N 回。

【0216】

分析にむけて、ショットガン次世代シーケンシングライブラリーを調製する。

- ・ ゲル抽出によって、所望のアンプリコン集団を非特異的な増幅産物から精製する。
- ・ アンプリコンを、剪断に適した高分子量産物にコンカテマー化 (Concatemerize) する。
- ・ 噴霧器、BioRuptor、HydroShear、Covaris または類似の装置を用いて機械的に剪断する。DNA は、数百塩基対の長さのフラグメントに剪断されるだろう。
- ・ 使用されるシーケンシングプラットフォームによる増幅に必要なアダプターをライゲートする。必要であれば、ライゲートされた産物をライゲートされていない産物およびアダプターから精製する。

40

【0217】

実施例 2：エキソン標的セットの MIP 捕捉反応においてバイアスを検出するためおよび補正するための識別タグ配列の使用

検出 / 補正を行う際の第 1 工程は、所与のサンプルについてどれだけ多くの識別タグ配

50

列が必要であるかを決定することである。この実施例では、1000個のエキソンに対応する1000個のゲノム標的を捕捉した。識別タグ配列は、プローブの一部であるので、1番最初のプロトコル工程から生じるバイアスを測定/報告する。また、識別タグ配列は、骨格に位置するので、別個のプライミング部位から容易に配列決定され得、ゆえに、標的の配列決定に対する達成可能な総リード長に影響を及ぼさない。MIPプローブは、かなり多数の販売業者(例えば、IDT)によって、標準的なカラムベースのオリゴヌクレオチド合成法を用いて合成され、識別タグ配列は、その骨格内の「縮重」位置として導入される。各縮重位置は、合成される識別タグ配列の総数を4倍増加させるので、10ntの縮重領域は、約 1×10^6 種という識別タグ配列の複雑度を意味する。

【0218】

捕捉プローブをヒトゲノムサンプルにハイブリダイズする
ハイブリダイゼーション反応物を組み立てる：

- ・ 1.0 μ l の捕捉プローブ混合物 (約 2.5 pmol)
- ・ 2.0 μ l の 10x Ampligase 緩衝液 (Epicentre)
- ・ 6.0 μ l の 500 ng/ μ l ヒトゲノムDNA (約 16.7 fmol)
- ・ 11 μ l の dH₂O。

【0219】

サーマルサイクラーにおいて、ゲノムDNAを変性するために95℃にて5分間の加熱反応を行い、次いで、60℃にて冷却する。60℃にて40時間インキュベートする。

【0220】

ハイブリダイズされたプローブを、サブターゲットを含む共有結合的に閉じた環状産物に変換する。

【0221】

充填/ライゲーション反応混合物を調製する：

- ・ 0.25 μ l の 2mM dNTP 混合物 (Invitrogen)
- ・ 2.5 μ l の 10x Ampligase 緩衝液 (Epicentre)
- ・ 5.0 μ l の 5 U/ μ l Taq Stoffel フラグメント (Applied Biosystems)
- ・ 12.5 μ l の 5 U/ μ l Ampligase (Epicentre)
- ・ 4.75 μ l の dH₂O。

【0222】

1.0 μ l のこの混合物をハイブリダイズされたプローブ反応物に加え、60℃で10時間インキュベートする。

【0223】

環状化されたプローブ/サブターゲット産物を未反応のプローブおよびゲノムDNAから精製する

エキソヌクレアーゼ反応混合物を調製する：

- ・ 21 μ l の充填/ライゲーション反応産物
- ・ 2.0 μ l の 10x エキソヌクレアーゼ I 緩衝液 (New England Biolabs)
- ・ 2.0 μ l の 20 U/ μ l エキソヌクレアーゼ I (New England Biolabs)
- ・ 2.0 μ l の 100 U/ μ l エキソヌクレアーゼ III (New England Biolabs)。

【0224】

37℃で60時間インキュベートし、次いで、80℃で15時間インキュベートすることによって熱失活させる。貯蔵のために、直ちに4℃にて冷却する。

【0225】

プローブの「定常領域」に特異的なプライマーを用いてPCRによって環状材料を増幅する

10

20

30

40

50

PCR混合物を調製する：

- ・ 5.0 μ l の 10 \times Accuprime 反応緩衝液 (Invitrogen)
- ・ 1.5 μ l の 10 μ M CP-2-FA (5' - GCACGATCCGACGGTAGTGT - 3') (配列番号 183)
- ・ 1.5 μ l の 10 μ M CP-2-RA (5' - CCGTAATCGGGAAGCTGAAAG - 3') (配列番号 184)
- ・ 0.4 μ l の 25 mM dNTP 混合物 (Invitrogen)
- ・ 2.0 μ l の 熱失活したエキソヌクラーゼ反応混合物
- ・ 1.5 μ l の 10 \times Sybr Green (Invitrogen)
- ・ 0.4 μ l の 2.5 U/ μ l Accuprime Pfx ポリメラーゼ (Invitrogen)
- ・ 37.7 μ l の dH₂O。

10

【0226】

以下のプロトコルに従ってリアルタイムサーマルサイクラーにおいて熱サイクル反応を行うが、増幅の収量がプラトーに達する前にサイクル反応を停止する (通常、8 ~ 12 サイクル)：

- 6.95 で 5 分間
- 7.95 で 30 秒間
- 8.58 で 60 秒間
- 9.72 で 60 秒間
- 10.2 に戻り、さらに N 回。

20

【0227】

分析にむけて、ショットガン次世代シーケンシングライブラリーを調製する。

- ・ ゲル抽出によって、所望のアンプリコン集団を非特異的な増幅産物から精製する。
- ・ アンプリコンを、切断に適した高分子量産物にコンカテマー化する。
- ・ 噴霧器、BioRuptor、HydroShear、Covaris または類似の装置を用いて機械的に切断する。DNA は、数百塩基対の長さ (length) のフラグメントに切断されるだろう。
- ・ 使用されるシーケンシングプラットフォームによる増幅に必要なアダプターをライゲートする。必要であれば、ライゲートされた産物をライゲートされていない産物およびアダプターから精製する。

30

【0228】

製造者の説明書 (例えば、Illumina、ABI など) に従ってライブラリーのシーケンシングを行い、標的配列と識別タグ配列の両方を読み取る

識別タグ配列の存在量を定量化することによって検出される任意のバイアスについて補正することによってデータを分析する

リードデータから標的：識別タグの存在量の表を構築する、例えば：

【0229】

【表 5】

標的 ID	識別タグ配列 ID	カウント
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

40

任意の特定の標的：識別タグのマッピングは、偶然に 2 回以上起きないので、すべての「カウント」の記載は、「1」であるはずであり、ゆえに、その 2 回以上のマッピングは、バイアスがサンプル調製プロセスのどこかに存在する場合にのみ観察される。2 回以上観察された任意の標的：識別タグの組み合わせについては、そのようなリードのすべてが

50

、コンセンサスベースコールが決定される前に単一の読み取りに「つぶされる」。これによって、コンセンサスベースコールの正確度に対するバイアスの影響が取り消される。図5は、リピート標的：識別タグの組み合わせがつぶされている二倍体遺伝子型のコールを生成するための方法を表している。

【0230】

実施例3：MIP捕捉反应用の識別タグ配列の設計

ある標的のセットに対して、ある特定の識別タグ配列がある特定の標的配列との組み合わせにおいて2回以上偶然に観察されないと信頼できる（いくつかの統計的限界内で）ために必要な識別タグ配列の数を決定した。ある特定の長さの識別タグ配列に対する独特の識別タグ配列の総数は、4（識別タグ配列のヌクレオチドの長さ）と決定される。識別タグ配列を有するMIPプローブを用いる分子反転プローブ捕捉反応に対しては、捕捉反応を行い、同じ識別タグ配列を有する1以上のコピー数の標的配列を捕捉する確率は： $p = 1 - [N! / (N - M)!] / [N^M]$ として計算され、式中、Nは、存在し得る独特の識別タグ配列の総数であり、Mは、捕捉反応における標的配列コピーの数である。よって、識別タグ配列の長さを変化させることによって、同じ識別タグ配列を有する1以上のコピー数の標的配列を捕捉する確率が所定の確率の値に設定されたMIP捕捉反応を行うことが可能である。

10

【0231】

例えば、15ヌクレオチド長の識別タグ配列に対しては、1, 073, 741, 824個の識別タグ配列が存在し得る。各々が15ヌクレオチドの識別タグ配列を有するMIPプローブが10000コピーの標的配列（例えば、ゲノム等価物）と組み合わせられるMIP捕捉反応において、同じ識別タグ配列を有する1以上のコピー数の標的配列を捕捉する確率は、0.05である。この例では、MIP反応は、複数コピーが同じ識別タグ配列でタグ化された標的をほとんど生成しない（通常0個であるが、たまに1個またはそれ以上）。図6は、15ヌクレオチドの識別タグ配列を有する100000個の捕捉反応および10000個の標的配列に対するシミュレーションの結果を示している。

20

【0232】

実施例4：シーケンシングカバレッジの関数としての、複数の独立した遺伝子座において正確なベースコールを行うのに十分なシーケンシングリードを得るための確率の評価

30

シーケンシングカバレッジの必要条件を決定するために、モンテカルロシミュレーションを行った。このシミュレーションは、所与の遺伝子座（標的）の10000ゲノムコピー（半数が母系対立遺伝子であり、半数が父系対立遺伝子である）を仮定する。このシミュレーションは、さらに、MIP反応について1%の捕捉効率を仮定する。このシミュレーションは、ある捕捉混合物から100回、非復元抽出することにより、100個の捕捉産物のセットを得る。次いで、このシミュレーションは、その100個の捕捉産物のセットから復元抽出することにより（偏りのない増幅と仮定して）、母系または父系から「リード」を生成する。標本抽出されるリードの数は、カバレッジに依存する。次いで、高精度のベースコールを行うために必要な母系と父系の両方からの独立したリードの数（10または20リードと仮定される）を測定した。このプロセスを各カバレッジレベルについて1000回繰り返し、両方の親からの十分なリードが首尾よく得られる回数の割合を測定した。成功したベースコールを得るはずの1000個の独立した遺伝子座を有すると仮定して、この割合を1000乗し、プロットした（図7を参照のこと）。結果は、> 0.95の確率で各対立遺伝子を $\geq 10 \times$ 捕捉するためにはおよそ50×カバレッジが必要であるということを示している。

40

【0233】

実施例5：「標的」遺伝子座および「コントロール」遺伝子座のMIP捕捉

いくつかの実施形態において、標的遺伝子座の捕捉効率を正確に定量化するために、様々な長さの標的遺伝子座に対する代理として働く経験的に示されている少なくとも3セットのコントロール遺伝子座を並行して捕捉する。例えば、標的遺伝子座が、50~10

50

00bpの長さを有すると予想される場合、50、250および1000bpの長さを有するコントロール遺伝子座のセットが捕捉され得（例えば、1セットあたり20個の遺伝子座が外れ値からの妥当な保護をもたらすはずである）、それらの存在量がシーケンシングによってデジタル方式で測定され得る。これらの遺伝子座は、サンプル間の効率の変動および同じサンプルの複数のランにおける効率の変動が最小であることが観察されるように（ゆえに、「一定の効率」であるように）、選択されるべきである。これらは、存在量 - 対 - 長さの曲線の形状を定義する「参照」点として働き得る。次いで、標的の長さを決定することは、単に、検量線上の適切な点から長さを「読む」ことである。

【0234】

いくつかの実施形態において、この方法からの標的の長さの推定における統計学的信頼度は、主として3つの因子：1) 検量線を作成するために使用される存在量データの再現性/変動；2) 「コントロール」データポイントに対する回帰の適合度；3) 測定されている標的遺伝子座に対する存在量データの再現性によって動かされる。1)および2)における統計的限度は、アッセイを構築する際に測定されているので、あらかじめ判明している。さらに、適切な集団のサンプリングおよび技術的な再現性の測定がアッセイの構築に含まれているはずであるので、3)における統計的限度も、通常あらかじめ判明している。これらの3つの測定を、標的存在量の任意の所与の実験に基づく測定に対する単一のP値に統合するために標準的な統計学的方法が用いられるべきである。

【0235】

いくつかの実施形態において、校正の観察結果のセットおよびそのデータに対する線形回帰の当てはめが与えられる場合、回帰を用いることにより、長さが不明の標的遺伝子座のn回の観察結果に対して長さの値を予測することができる。まず、長さの推定値の信頼区間に対する許容可能な範囲を選択する。例えば、脆弱Xの「正常」症例(87~93bp)を「前変異」(165~600bp)の潜在的な症例と識別する場合、目標は、93bpを165bpと識別するのに十分な精度で長さを測定することである。n回の観察結果が回帰直線(regressed line)に対する方程式に代入されたときにコンピュータで計算される反応の予測値は、任意精度を有し得る。しかしながら、例えば、95%信頼水準が望まれる場合、95%信頼区間は、「正常」の長さの範囲と「前変異」の長さの範囲の両方と重複しない程度に十分短くなければならない。例を続けると、校正データからの回帰に基づいてn=400のMIP観察結果から190という長さが計算される場合、95%信頼区間は、 $190 + / - 20$ bpであり、そのサンプルが95%の確実性で「前変異」の長さを表すと結論づけられ得る。逆に、校正データがそれほどロバストでない場合、その回帰のエラー推定値はより高いことがあり、予測される反応値に対してより広い信頼区間をもたらされる。いくつかの実施形態において、95%CIが、n=400から $190 + / - 100$ bpと計算される場合、予測される反応値が「正常」の長さに対応するかまたは「前変異」の長さに対応するかを決定できない可能性がある。

【0236】

いくつかの実施形態において、予測される反応に対する信頼区間は、以下のとおり計算される：

反応に対する推定値

【0237】

【化1】



は、その反応の平均値に対する推定値と同一である：

【0238】

10

20

30

40

【化2】

$$\hat{y} = b_0 + b_1 x^*$$

。予測値に対する信頼区間は、

【0239】

【化3】

$$\hat{y} \pm t^* s_{\hat{y}}$$

によって与えられ、ここで、

【0240】

【化4】

$$\hat{y}$$

は、 x^* に対応する当てはめ値である。 t^* 値は、 $t(n-2)$ 分布に対する大きい方の $(1-C)/2$ 棄却値である。

【0241】

いくつかの実施形態において、目的の遺伝子座を分析するための手法は、以下の工程を含み得る。

【0242】

ハイブリダイズされたプローブを、サブターゲットを含む共有結合的に閉じた環状産物に変換する。

【0243】

充填/ライゲーション反応混合物を調製する：

- ・ 0.25 μ l の 2 mM dNTP 混合物 (Invitrogen)
- ・ 2.5 μ l の 10 x Ampligase 緩衝液 (Epicentre)
- ・ 5.0 μ l の 5 U/ μ l Taq Stoffel フラグメント (Applied Biosystems)
- ・ 12.5 μ l の 5 U/ μ l Ampligase (Epicentre)
- ・ 4.75 μ l の dH₂O。

【0244】

1.0 μ l のこの混合物をハイブリダイズされたプローブ反応物に加え、60 で 10 時間インキュベートする。

【0245】

環状化されたプローブ/サブターゲット産物を未反応のプローブおよびゲノム DNA から精製する

エキソヌクラーゼ反応混合物を調製する：

- ・ 21 μ l の充填/ライゲーション反応産物
- ・ 2.0 μ l の 10 x エキソヌクラーゼ I 緩衝液 (New England Biolabs)
- ・ 2.0 μ l の 20 U/ μ l エキソヌクラーゼ I (New England Biolabs)
- ・ 2.0 μ l の 100 U/ μ l エキソヌクラーゼ III (New England Biolabs)。

【0246】

37 で 60 分間インキュベートし、次いで、80 で 15 分間インキュベートすることによって熱失活させる。貯蔵のために、直ちに 4 に冷却する。

【0247】

10

20

30

40

50

プローブの「定常領域」に特異的なプライマーを用いてPCRによって環状材料を増幅する

PCR混合物を調製する：

- ・ 5.0 μ l の 10 \times Accuprime 反応緩衝液 (Invitrogen)
- ・ 1.5 μ l の 10 μ M CP-2-FA-Ilmn (プラットフォーム特異的増幅配列 + 「環定常領域」特異的配列)
- ・ 1.5 μ l の 10 μ M CP-2-RA-Ilmn (プラットフォーム特異的増幅配列 + 「環定常領域」特異的配列)
- ・ 0.4 μ l の 25 mM dNTP 混合物 (Invitrogen)
- ・ 2.0 μ l の 熱失活したエキソヌクラーゼ反応混合物
- ・ 1.5 μ l の 10 \times Sybr Green (Invitrogen)
- ・ 0.4 μ l の 2.5 U/ μ l Accuprime Pfx ポリメラーゼ (Invitrogen)
- ・ 37.7 μ l の dH₂O。

10

【0248】

以下のプロトコルに従ってリアルタイムサーマルサイクラーにおいて熱サイクル反応を行うが、増幅の収量がプラトーに達する前にサイクル反応を停止する（通常、8～12サイクル）：

- 11.95 で 5 分間
- 12.95 で 30 秒間
- 13.58 で 60 秒間
- 14.72 で 60 秒間
- 15.2 に 戻り、さらに N 回。

20

【0249】

製造者の指示書（例えば、Illumina、ABI）に従ってデジタル定量化用のサンプルにおいてシーケンシング（例えば、次世代シーケンシング）を行う。

【0250】

実施例 6：エキソン標的核酸セットのMIP捕捉反応

MIPプローブは、かなり多数の販売業者（例えば、IDT）によって、標準的なカラムベースのオリゴヌクレオチド合成法を用いて合成される。

30

【0251】

捕捉プローブをヒトゲノムサンプルにハイブリダイズする

ハイブリダイゼーション反応物を組み立てる：

- ・ 1.0 μ l の 捕捉プローブ混合物（約 2.5 pmol）
- ・ 2.0 μ l の 10 \times Ampligase 緩衝液 (Epicentre)
- ・ 6.0 μ l の 500 ng/ μ l ヒトゲノムDNA（約 16.7 fmol）
- ・ 11 μ l の dH₂O。

【0252】

サーマルサイクラーにおいて、ゲノムDNAを変性するために95にて5分間の加熱反応を行い、次いで、60にて冷却する。60にて40時間インキュベートする。

40

【0253】

ハイブリダイズされたプローブを、標的核酸を含む共有結合的に閉じた環状産物に変換する。

【0254】

充填/ライゲーション反応混合物を調製する：

- ・ 0.25 μ l の 2 mM dNTP 混合物 (Invitrogen)
- ・ 2.5 μ l の 10 \times Ampligase 緩衝液 (Epicentre)
- ・ 5.0 μ l の 5 U/ μ l Taq Stoffel フラグメント (Applied Biosystems)
- ・ 12.5 μ l の 5 U/ μ l Ampligase (Epicentre)

50

・ 4.75 μ l の dH₂O 。

【0255】

1.0 μ l のこの混合物をハイブリダイズされたプローブ反応物に加え、60 で10時間インキュベートする。

【0256】

環状化されたプローブ/標的核酸産物を未反応のプローブおよびゲノムDNAから精製する

エキソヌクラーゼ反応混合物を調製する：

・ 21 μ l の充填/ライゲーション反応産物

・ 2.0 μ l の10xエキソヌクラーゼI緩衝液 (New England Biolabs) 10

・ 2.0 μ l の20U/ μ lエキソヌクラーゼI (New England Biolabs)

・ 2.0 μ l の100U/ μ lエキソヌクラーゼIII (New England Biolabs) 。

【0257】

37 で60分間インキュベートし、次いで、80 で15分間インキュベートすることによって熱失活させる。貯蔵のために、直ちに4 に冷却する。

【0258】

プローブの「定常領域」に特異的なプライマーを用いてPCRによって環状材料を増幅する 20

PCR混合物を調製する：

・ 5.0 μ l の10xAccuprime反応緩衝液 (Invitrogen)

・ 1.5 μ l の10 μ M CP-2-FA (5' - GCACGATCCGACGGTAGTGT - 3') (配列番号183)

・ 1.5 μ l の10 μ M CP-2-RA (5' - CCGTAATCGGGAAGCTGAAG - 3') (配列番号184)

・ 0.4 μ l の25mM dNTP混合物 (Invitrogen)

・ 2.0 μ l の熱失活したエキソヌクラーゼ反応混合物

・ 1.5 μ l の10xSybrGreen (Invitrogen) 30

・ 0.4 μ l の2.5U/ μ l Accuprime Pfxポリメラーゼ (Invitrogen)

・ 37.7 μ l の dH₂O 。

【0259】

以下のプロトコルに従ってリアルタイムサーマルサイクラーにおいて熱サイクル反応を行うが、増幅の収量がプラトーに達する前にサイクル反応を停止する (通常、8~12サイクル)：

16.95 で5分間

17.95 で30秒間

18.58 で60秒間 40

19.72 で60秒間

20.2に戻り、さらにN回 。

【0260】

分析にむけて、ショットガン次世代シーケンシングライブラリーを調製する

・ ゲル抽出によって、所望のアンプリコン集団を非特異的な増幅産物から精製する。

・ アンプリコンを、剪断に適した高分子量産物にコンカテマー化する。

・ 噴霧器、BioRuptor、HydroShear、Covarisまたは類似の装置を用いて機械的に剪断する。DNAは、数百塩基対の長さのフラグメントに剪断されるだろう。

・ 使用されるシーケンシングプラットフォームによる増幅に必要なアダプターをライゲ 50

ートする。必要であれば、ライゲートされた産物をライゲートされていない産物およびアダプターから精製する。

【0261】

製造者の説明書（例えば、Illumina、ABIなど）に従ってライブラリーのシーケンシングを行い、標的配列を読み取ることにより、標的核酸の存在量を測定する。

【0262】

実施例7：1000個の標的のセットの遺伝子タイピングを行うためのMIP、ハイブリダイゼーションおよび変異検出MIPの使用

MIP、ハイブリダイゼーションおよび変異検出MIPを用いることにより、1000個の標的のセットの遺伝子タイピングを行う。このプロトコルによって、50個の特定の公知の点変異のうちのいずれかの検出が可能になる。

10

【0263】

まず、MIP、ハイブリダイゼーションおよび変異検出MIPの別個の反応を生物学的サンプルにおいて行う。MIP捕捉反応は、本質的にはTurnerら、2009, Nature methods 6:315-6に記載されているように行う。MIPのセットは、そのセット内の各プローブが1000個の標的のうちの1つと隣接するように設計される。別個に、ハイブリダイゼーション濃縮反応は、Agilent SureSelect手順を用いて行われる。選択の前に、Illuminaの「Fragment Library」キットおよびプロトコルを用いて、濃縮されるゲノムDNAをショットガンシーケンシングライブラリーに変換する。Agilentのウェブインターフェースを用いることにより、標的核酸にハイブリダイズし得るプローブのセットを設計する。別個に、変異（例えば、特定の多型）が存在する場合のみMIPを形成し得るプローブのセットを設計する（変異検出MIP）。各変異検出MIPは、単一の公知の変異に特異的な最も3'側の塩基の同一性を有する。この変異検出MIPのセットを用いた反応を行うことにより、任意の変異対立遺伝子の存在が選択的に検出される。

20

【0264】

3つすべての反応が行われたら、2つのMIP反応物を、単一のチューブに混合し（例えば、変異検出の感度をさらに高めるために潜在的に等モルでない比で）、1つのサンプルとして次世代DNAシーケンシング機器にかける。そのハイブリダイゼーション濃縮反応物を別個のサンプルとして次世代DNAシーケンシング機器にかける。各位置における総カバレッジ、その総カバレッジでの各リードの起源、各個別のリードの質のスコア、およびその領域と重複する変異特異的MIPから得られた任意のリードの存在（または非存在）を評価することによって標的セット内の各位置においてコンセンサス二倍体遺伝子型を形成するソフトウェアアルゴリズムによって、各「サンプル」からのリードを統合する。

30

【0265】

前出の実施例は、非限定的なものであり、本発明の局面は、当業者（one or ordinary skill in the art）にとって利用可能な代替の手法および/またはプロトコルを用いて、本明細書中に記載されるように実行され得ることが認識されるだろう。

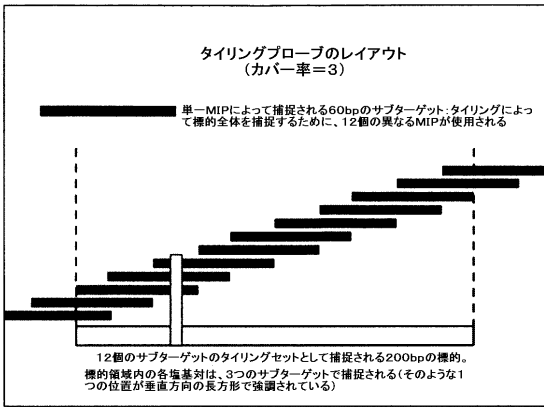
40

【0266】

前述の説明および実施例に詳しく記載されているもの以外の方法が実施され得ることは明らかである。上記の教示に鑑みて、本開示の数多くの改変およびバリエーションが存在し得、ゆえに、それらは請求項の範囲内である。本開示の各局面の好ましい特徴は、必要な変更を加えた他の各局面と同様のものである。本明細書中で言及される特許、特許出願、学術論文または他の開示を含む書面は、それらの全体が本明細書によって参考として援用される。矛盾が生じる場合には、明らかな誤りの場合を除いて本願の開示が支配する。

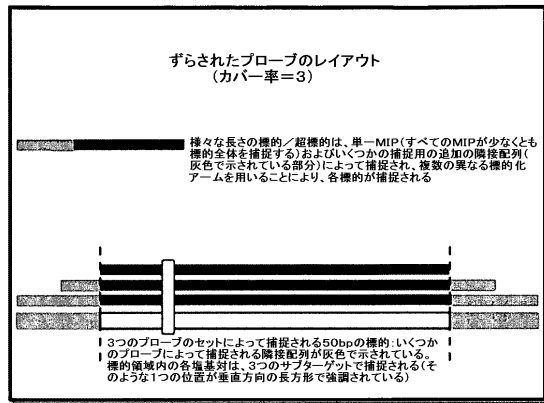
【 図 1 】

Figure 1



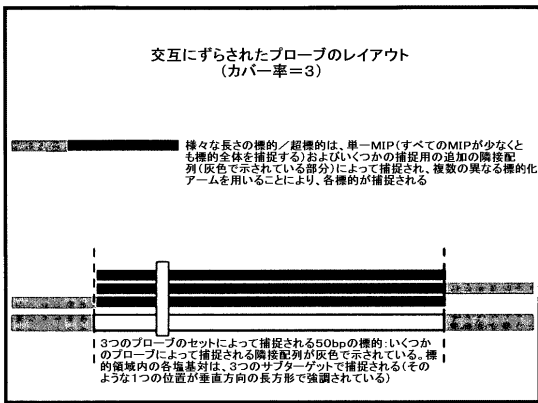
【 図 2 】

Figure 2



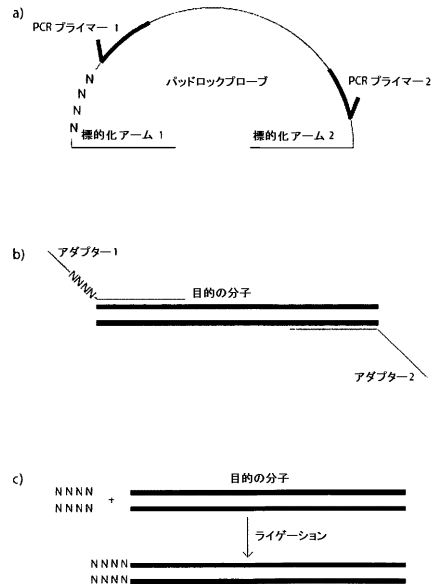
【 図 3 】

Figure 3



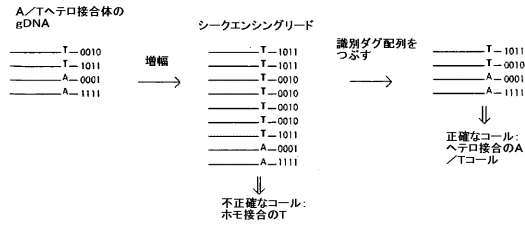
【 図 4 】

FIGURE 4



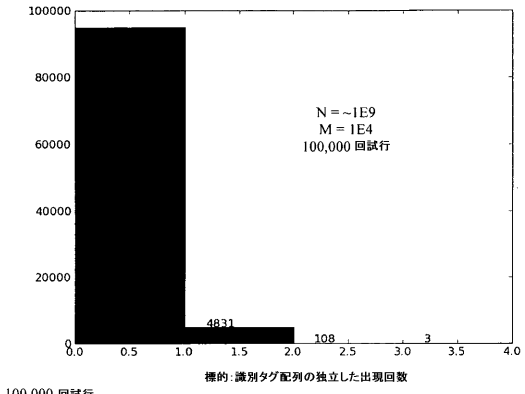
【 図 5 】

FIGURE 5



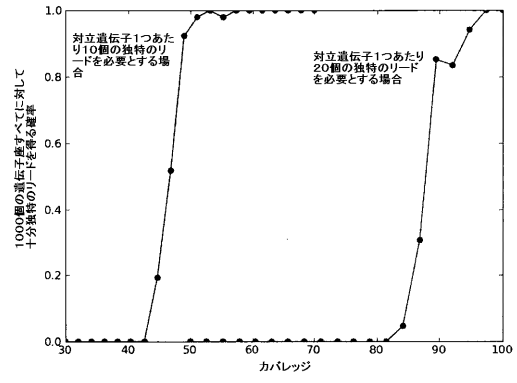
【 図 6 】

FIGURE 6



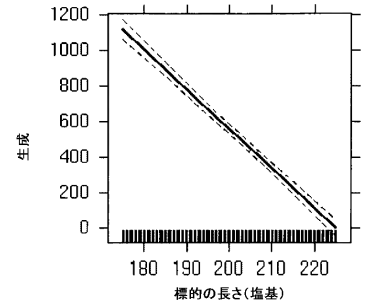
【 図 7 】

FIGURE 7



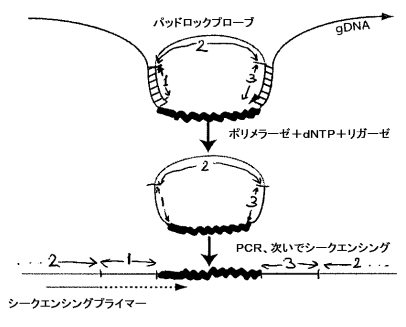
【 図 8 】

Figure 8



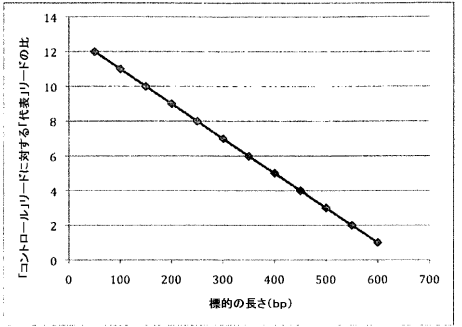
【 図 9 】

Figure 9



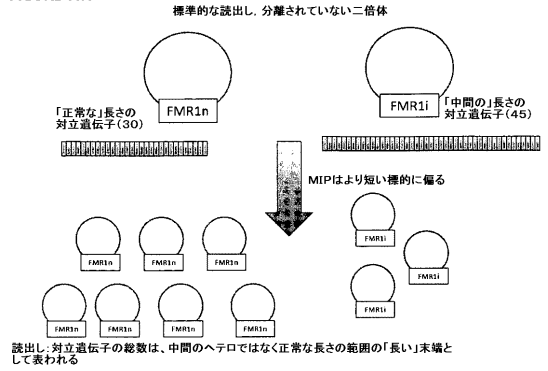
【 図 10 】

Figure 10



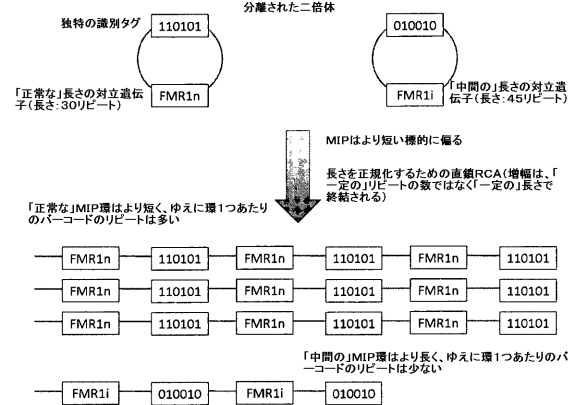
【 図 11 A 】

FIGURE 11A



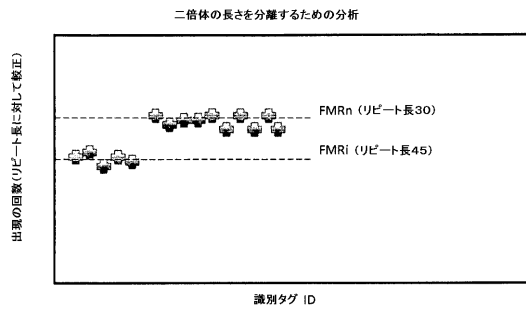
【 図 11 B 】

FIGURE 11B



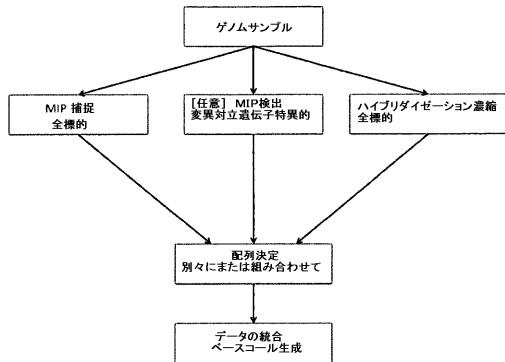
【 図 1 1 C 】

FIGURE 11C



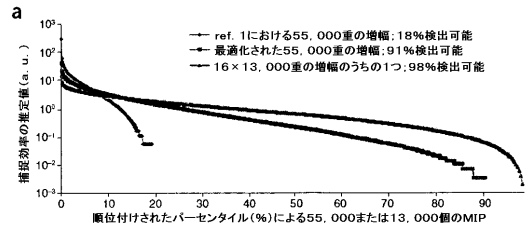
【 図 1 2 】

FIGURE 12



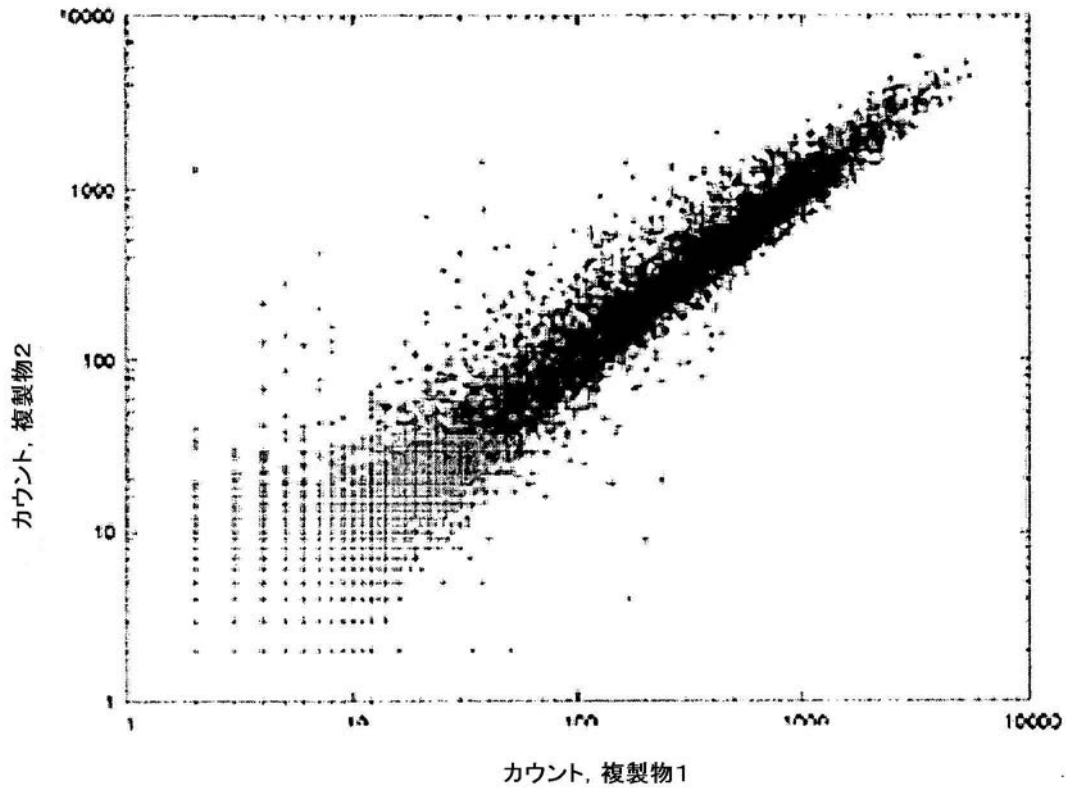
【 図 1 3 】

FIGURE 13



【 図 1 4 】

FIGURE 14



【 配列表 】

[2012525147000001.xml](#)

【 手続補正書 】

【 提出日 】 平成24年1月10日 (2012.1.10)

【 手続補正 1 】

【 補正対象書類名 】 明細書

【 補正対象項目名 】 配列表

【 補正方法 】 追加

【 補正の内容 】

【 配列表 】

[2012525147000001.app](#)

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/01293

A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - G01N 33/48 (2010.01) USPC - 436/94 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) USPC- 436/94 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) PubWEST, Google Scholar Search Terms:molecular inversion probe, multiplex, overlap, tile, stack, probe, DNA, bias, reduce, circular, padlock, genotyping, probe set		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y — A	Hardenbol et al., 'Multiplexed genotyping with sequence-tagged molecular Inversion probes' Nature Biotechnology, 2003, Vol. 21, No. 6, pages 673-678; abstract, Fig. 2; pg 673, col 2, para 2pg 674, col 1, para 2-3; pg 674, col 2, para 1-3; pg 677, col 1, para 3	66 — 1-8, 10-14 ----- 17-18, 57
X — Y	US 2003/0203370 A1 (YAKHINI et al.) 30 October 2003 (30.10.2003) abstract, para [0026]	9 ----- 10-14
Y	Mockler et al., 'Applications of DNA tiling arrays for whole-genome analysis', Genomics, 2005, Vol. 85, pages 1-15; Fig. 2; pg 2, col 1, para 4 to pg 2, col 2, para 1	1-8
Y	US 2008/0280955 A1 (MCCAMISH et al), 13 Nov. 2008 (13.11.2008), Table 9, para [0258], claims 1, 15	6
Y	WO 2007/107717 A1 (CHAMBERLAIN et al) 27 Sep. 2007 (27.09.2007), Tables 3, 4; pg 4, ln 16-33	8
&	US 2010/0129896 A1 (KNAPP) 27 May 2010 (27.05.2010)	57
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/>		
* Special categories of cited documents:		
"A"	document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier application or patent but published on or after the international filing date	"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
"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)	"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
"O"	document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search 15 Oct. 2010 (15.10.2010)	Date of mailing of the international search report 28 OCT 2010	
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/01293

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.:
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:

3. Claims Nos.: 15, 16, 19, 22-56, 64, 65, 67
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:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: claims 1-8, and 57, directed to a method of analyzing a plurality of gene loci, comprising: contacting each of a plurality of target nucleic acids with a probe set, wherein the probe set comprises a plurality of probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids in the same strand of one of a plurality of subregions in the target nucleic acid, wherein the subregions of the target are different.

- Please see extra sheet for continuation -

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.:
1-14, 17, 18, 57 and 66
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/US 10/01293

Continuation of Box III: Unity of Invention

Group II: claims 9-14, 17, 18, and 66, directed to a method of genotyping a subject, comprising determining in the sequence at least a threshold number of independently isolated nucleic acids that comprise target nucleic acid sequences and differentiator tag sequences, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences that form unique combinations, and the target nucleic acid sequence is the sequence of a genomic locus of a subject.

Group III: claims 20-21, directed to a method of determining whether a target nucleic acid has abnormal length, comprising evaluating the capture efficiency of a target nucleic acid in a sample from a subject.

Group IV: claims 58-63, directed to a method of increasing the sensitivity of a nucleic acid detection assay comprising utilizing two different preparation methods on parallel samples and assaying the sequences of the samples, using the sequence information from both determinations to evaluate the genotype of a target nucleic acid.

The inventions listed as Groups I - IV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the Group I claims is a method of analyzing a plurality of gene loci, comprising: contacting each of a plurality of target nucleic acids with a probe set, wherein the probe set comprises a plurality of probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids in the same strand of one of a plurality of subregions in the target nucleic acid, wherein the subregions of the target are different. The special technical feature of the Group II claims is a method of genotyping a subject, comprising determining in the sequence at least a threshold number of independently isolated nucleic acids that comprise target nucleic acid sequences and differentiator tag sequences, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences that form unique combinations, and the target nucleic acid sequence is the sequence of a genomic locus of a subject. The special technical feature of the Group III claims is a method of determining whether a target nucleic acid has abnormal length, comprising evaluating the capture efficiency of a target nucleic acid in a sample from a subject. The special technical feature of the Group IV claims is a method of increasing the sensitivity of a nucleic acid detection assay comprising utilizing two different preparation methods on parallel samples and assaying the sequences of the samples, using the sequence information from both determinations to evaluate the genotype of a target nucleic acid.

The only common technical element shared by the above groups is that they are related to genotyping of nucleic acids in samples. This common technical element does not represent an improvement over the prior art of US 2001/0046673 A1 to French et al. (see abstract; para [0012]- wherein length of the target was abnormal due to insertion or deletion; para [0077] - allele specific hybridization using luminescence detection; para [0107] - DNA chips generally attempt to detect SNPs by measuring the extent of hybridization of DNA targets to DNA probes immobilized on the chip surface). Therefore, the inventions of Groups I - IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.

フロントページの続き

(31)優先権主張番号 61/182,089

(32)優先日 平成21年5月28日(2009.5.28)

(33)優先権主張国 米国(US)

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

(72)発明者 ボレカ, グレゴリー

アメリカ合衆国 ニュージャージー 08226, オーシャン シティ, ガーデン パーク
ウェイ 911

(72)発明者 レーザーソン, ユリ

アメリカ合衆国 マサチューセッツ 02215, ボストン, ビーコン ストリート 645
ナンバー 215

(72)発明者 リー, チン ピリー

アメリカ合衆国 マサチューセッツ 02155, メドフォード, ステーション ランディン
グ 100, ユニット 204

(72)発明者 ワスマン, イー. ロバート

アメリカ合衆国 マサチューセッツ 01945, マーブルヘッド, ウエスト ショア ドラ
イブ 90

Fターム(参考) 4B024 AA11 CA04 CA09 CA20 HA14

4B063 QA01 QQ42 QR08 QR55 QR62 QS25 QS34

专利名称(译)	用于评估遗传标记的方法和组合物		
公开(公告)号	JP2012525147A	公开(公告)日	2012-10-22
申请号	JP2012508490	申请日	2010-04-30
[标]申请(专利权)人(译)	良好的开端遗传学		
申请(专利权)人(译)	良好的开端遗传学公司		
[标]发明人	ポレカグレゴリー レーザーソンユリ リーチンビリー ワスマンイーロバート		
发明人	ポレカ, グレゴリー レーザーソン, ユリ リー, チン ビリー ワスマン, イー. ロバート		
IPC分类号	C12Q1/68 G01N33/53 C12N15/09		
CPC分类号	C12Q1/6827		
FI分类号	C12Q1/68.ZNA.A G01N33/53.M C12N15/00.A		
F-TERM分类号	4B024/AA11 4B024/CA04 4B024/CA09 4B024/CA20 4B024/HA14 4B063/QA01 4B063/QQ42 4B063/QR08 4B063/QR55 4B063/QR62 4B063/QS25 4B063/QS34		
代理人(译)	夏木森下		
优先权	61/179358 2009-05-18 US 61/178923 2009-05-15 US 61/174470 2009-04-30 US 61/182089 2009-05-28 US		
其他公开文献	JP2012525147A5		
外部链接	Espacenet		

摘要(译)

本发明的方面涉及可用于减少偏倚和增加遗传基因座的多重分析的再现性的方法和组合物。在一些配置中，预定的制备步骤和/或核酸序列分析技术用于多个样品中的多个遗传基因座的多重分析。

Figure 1

