

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7003035号
(P7003035)

(45)発行日 令和4年2月10日(2022.2.10)

(24)登録日 令和4年1月5日(2022.1.5)

(51)国際特許分類

F I

C 1 2 Q 1/6883(2018.01)

C 1 2 Q 1/6883

Z Z N A

C 1 2 N 15/09 (2006.01)

C 1 2 N 15/09

Z

請求項の数 17 (全83頁)

(21)出願番号	特願2018-522783(P2018-522783)	(73)特許権者	509067555
(86)(22)出願日	平成28年11月3日(2016.11.3)		アスラジェン, インコーポレイテッド
(65)公表番号	特表2019-502367(P2019-502367 A)		アメリカ合衆国 テキサス 7 8 7 4 4, オースティン, ウッドワード ストリート 2 1 5 0, スイート 1 0 0
(43)公表日	平成31年1月31日(2019.1.31)	(74)代理人	100145403
(86)国際出願番号	PCT/US2016/060389		弁理士 山尾 憲人
(87)国際公開番号	WO2017/079478	(74)代理人	100122301
(87)国際公開日	平成29年5月11日(2017.5.11)		弁理士 富田 憲史
審査請求日	令和1年10月18日(2019.10.18)	(74)代理人	100157956
(31)優先権主張番号	62/250,476		弁理士 稲井 史生
(32)優先日	平成27年11月3日(2015.11.3)	(74)代理人	100170520
(33)優先権主張国・地域又は機関	米国(US)		弁理士 笹倉 真奈美
		(72)発明者	エラン・ブラム
			アメリカ合衆国 7 8 6 1 3 テキサス州シ
			最終頁に続く

(54)【発明の名称】 リピート配列の核酸サイズ検出のための方法

(57)【特許請求の範囲】

【請求項 1】

i . リピート領域を増幅すること；

i i . キャピラリー電気泳動による高分解能断片分析を実施すること；

i i i . 増幅産物のラダーを得ること；および

i v . 内部サイズ分類標準を用いて、リピート領域をサイズ分類すること

を含む、核酸試料のリピート領域をサイズ分類する方法であって、

内部サイズ分類標準が：

チャンネルにおいて増幅産物のラダーのリピートプロファイルを同定すること；

前のピーク位置および区間値を用いて、リピートプロファイルにおいて予想されるピーク位置を決定すること、

予想されるピーク位置を含む第1のウィンドウ内にピーク位置を同定すること、および

同定されたピーク位置におけるシグナル強度が振幅基準を満たす場合、同定されたピーク位置をリピートピーク位置のセットに加えること

を含む、リピートピーク位置のセットを繰り返し、作成すること；ならびに

リピートピーク位置のセットおよび対応する断片サイズのセットを用いて、リピートピーク位置とリピート断片サイズとの間の線形関係を推定し、それによって内部サイズ分類標準を作成すること

によって作成される、方法。

【請求項 2】

(a) リピートピーク位置のセットを繰り返し、作成すると同時に、リピートピーク位置のセットの1つまたは複数のシグナル強度を用いて、振幅基準を反復して、アップデートすること；

(b) リピートピーク位置のセットを繰り返し、作成すると同時に、リピートピーク位置のセットの1つまたは複数のピーク位置を用いて、区間値を反復して、アップデートすること；

(c) リピートピーク位置のセットを繰り返し、作成する前に、リピートプロファイルの第1の部分の統計的尺度を用いて振幅基準を初期化すること；および／または

(d) リピートピーク位置のセットを繰り返し、作成する前に、リピートプロファイルの第2の部分の周期性を用いて区間値を初期化すること

をさらに含む、請求項1に記載の方法。

【請求項3】

リピート領域をサイズ分類することが、外部サイズ分類標準を使用することをさらに含み、外部サイズ分類標準が：

チャンネルにおいて増幅産物のラダーのピークを同定することであって、ピークが局所的ノイズ閾値を超える、同定すること；

断片サイズの第1セットおよび対応するピーク位置の第1のセットを用いるピーク位置と断片サイズとの間の線形関係を、チャンネルの第1の領域における同定されたピークを用いて、繰り返し、再推定することであって、繰り返し再推定が、

漸減する断片サイズを断片サイズの第1のセットへ含めること、および

漸減する断片サイズに対応するピーク位置を、対応するピーク位置の第1のセットへ含めること

を含む、繰り返し、再推定すること；

断片サイズの第2のセットおよび対応するピーク位置の第2のセットを用いる、チャンネルの第2の領域についてのピーク位置と断片サイズとの間の非線形関係を、チャンネルの第2の領域における同定されたピークを用いて、推定すること；および

線形関係と非線形関係を組み合わせて、それによって外部サイズ分類標準を作成することによって作成される、請求項1から2のいずれかに記載の方法。

【請求項4】

繰り返し再推定が、

漸減する断片サイズの1つ、およびピーク位置と断片サイズとの間の再推定された線形関係を用いて、予想されるピーク位置を決定すること；

予想されるピーク位置を含むウィンドウ内にチャンネルにおける実際のピーク位置を決定すること；ならびに

実際のピーク位置を、対応するピーク位置の第1のセットへ含めること

をさらに含む、請求項3に記載の方法。

【請求項5】

チャンネルにおいて増幅産物のラダーのリピートプロファイルを同定すること、リピートピーク位置のセットを繰り返し、作成すること、ならびにリピートピーク位置のセットおよび対応する断片サイズのセットを用いて、リピートピーク位置とリピート断片サイズとの間の線形関係を推定すること、によって、内部サイズ分類標準を作成すること；ならびに内部サイズ分類標準および外部サイズ分類標準を用いてアフィン変換を作成すること；およびアフィン変換を外部サイズ分類標準に適用して、移動度補正済みサイズ分類標準を得ること、によって、移動度補正済みサイズ分類標準を作成すること

をさらに含む、請求項3または4に記載の方法。

【請求項6】

リピート領域をサイズ分類することが、少なくとも1つの遺伝子特異的ピークを同定すること；および増幅産物のラダーを用いて少なくとも1つの遺伝子特異的ピークをサイズ分類することをさらに含む、請求項1から5のいずれかに記載の方法。

【請求項7】

少なくとも1つの遺伝子特異的ピークをサイズ分類することが：
バックグラウンドモデルを区分的にスケールリングすることによって、バックグラウンドモデルから動的閾値を作成して、動的閾値を得ること；
動的閾値を用いて、リピートプロファイルにおいて遺伝子特異的ピーク位置を決定すること；
増幅産物のラダーを用いて、リピートサイズを遺伝子特異的ピーク位置に関連づけること；
および
遺伝子特異的ピークサイズの表示をアウトプットすること
を含む、請求項6に記載の方法。

【請求項8】

バックグラウンドモデルを区分的にスケールリングすることが：
第1の断片サイズより上で、かつ第2の断片サイズより下のサイズを有する増幅産物に対応するバックグラウンドモデルの第1の領域を決定すること；
第2の断片サイズより上のサイズを有する増幅産物に対応するバックグラウンドモデルの第2の領域を決定すること；
最初のスケールリング因子から、最初のスケールリング因子より小さい第2のスケールリング因子まで変動する第1のスケールリング因子を、バックグラウンドモデルの第1の領域に掛けること；および
第2のスケールリング因子をバックグラウンドモデルの第2の領域に掛けること
を含む、請求項7に記載の方法。

【請求項9】

リピートプロファイルにおいて遺伝子特異的ピーク位置を決定することが、
リピートプロファイルにおいて第1の位置に第1のピークを同定すること、および
第1の位置におけるリピートプロファイルの第1の値が、第1の位置における動的閾値の第1の値を超えることを決定すること
を含む、請求項7または8に記載の方法。

【請求項10】

リピートプロファイルにおいて遺伝子特異的ピーク位置を決定することが、
リピートプロファイルにおいて第2の位置に第2のピークを同定することであって、第2ピークが第1のピークに隣接している、同定すること；
第2の位置におけるリピートプロファイルの第2の値が振幅基準を満たすことを決定することであって、振幅基準が第1の値に基づいている、決定すること；および
第2の位置におけるリピートプロファイルの第2の値が、第2の位置における動的閾値の第2の値を超えることを決定すること
をさらに含む、請求項9に記載の方法。

【請求項11】

リピート領域を増幅することが少なくとも2つの異なるプライマーを含み、少なくとも2つの異なるプライマーが第1のプライマーおよび第2のプライマーを含む、請求項1から10のいずれかに記載の方法。

【請求項12】

(a) 第1のプライマーがリピート領域の一部を含む；
(b) 第2のプライマーがリピート領域の外側の位置にアニールする；および／または
(c) 少なくとも2つの異なるプライマーが第3のプライマーおよび任意の第4のプライマーをさらに含む
請求項11に記載の方法。

【請求項13】

リピート領域をサイズ分類することが、
(a) リピートプロファイルにおいてピークをコールするための閾値を動的に決定すること；
(b) スライドウィンドウを用いてリピートプロファイルにおいてピークをコールするこ

10

20

30

40

50

と；

- (c) リピートプロファイルにおいて振幅閾値より下にピークを内挿すること；
- (d) リピートプロファイルにおいて推定上のピーク位置を用いてサンプリング単位から塩基対単位へマッピングする較正曲線を作成すること；
- (e) 第1のチャンネルをシグナルアーチファクトについて補正すること；
- (f) 第1の電磁的に検出可能な部分を検出するように構成された、第1のチャンネルの動的範囲を拡張すること；および／または
- (g) サイズ分類標準基準、リピートプロファイルシグナル・ノイズ比基準、リピートプロファイル混入基準、および／またはマイナーアレル感度基準の達成を決定すること、をさらに含む、請求項1から12のいずれかに記載の方法。

10

【請求項14】

第1のチャンネルをシグナルアーチファクトについて補正することが、
第1のチャンネルにおいて、可能性のある気泡位置を含むウィンドウを同定すること；
ウィンドウ内で、チャンネルについてのシグナル強度間の相関を決定すること；および
決定された相関に基づいて、ウィンドウ内のチャンネルについてのシグナル強度を置き換えること；および／または
第1のチャンネルについてのシグナル強度に基づいて漏出位置を同定すること；
第2のチャンネルについてのシグナル強度に基づいて、漏出位置を含むウィンドウを決定すること；および
第2のチャンネルについてのシグナル強度を置き換えること
を含む、請求項13に記載の方法。

20

【請求項15】

第1のチャンネルの動的範囲を拡張することが、
第1のチャンネルにおいて飽和領域を含むウィンドウを同定すること；
ウィンドウ内の第1のチャンネルについてのシグナル強度、およびウィンドウ内の第2の電磁的に検出可能な部分を検出するように構成された、第2のチャンネルについてのシグナル強度を用いて、組み合わせられたシグナル強度を決定すること；および
第1のチャンネルについてのシグナル強度を、組み合わせられたシグナル強度に置き換えること
を含む、請求項13または14に記載の方法。

30

【請求項16】

(a) サイズ分類標準基準が、内部サイズ分類標準適合度基準、外部サイズ分類標準適合度基準、および／または内部サイズ分類標準を外部サイズ分類標準と比較する一貫性基準を含む；
(b) 遺伝子特異的ピークの位置に関連づけられたリピートサイズがゼロ未満である時、リピートプロファイル混入基準に不合格である；
(c) 遺伝子特異的ピークの位置におけるリピートプロファイルの値の最大値に対するリピートプロファイルのノイズ値の比率が閾値を超える時、マイナーアレル感度基準が満たされている、
請求項13～15のいずれかに記載の方法。

【請求項17】

リピート領域の増幅のための少なくとも2つの異なるプライマー；
バッファ；ならびに
請求項1から16のいずれかに記載の方法を実施する分析ソフトウェアについての少なくとも1つのソフトウェアキー；および／または
請求項1から16のいずれかに記載の方法を実施する分析ソフトウェア
を記憶する少なくとも1つの非一過性媒体
を含む、核酸試料のリピート領域を分類するまたはジェノタイプピークサイズ分類するためのキット。

40

【発明の詳細な説明】

【技術分野】

50

【0001】

この出願は、参照により全体として本明細書に組み入れられている、2015年11月3日に出願された米国仮出願第62/250,476号の優先権の利益を主張する。

【0002】

本開示は、核酸分析のための、データ処理および計算アノテーションシステムおよびデバイス、ならびに対応する方法に関する。特に、本開示は、核酸のリピート(repeat)領域をサイズ分類する方法に関する。

【背景技術】

【0003】

ヌクレオチドリピート(例えば、ホモポリマー領域、ジヌクレオチドリピート、トリヌクレオチドリピート、ヘキサヌクレオチドリピートなど)の領域を含む遺伝子座は、ヒトまたは動物ゲノムにおいてよく見られる。豊富なGC(グアニン-シトシン)含有量を有する遺伝子座もまたよく見られ、一方、AT(アデニン-チミン)リッチな含有量を有する座が報告および研究されている。いくつかの状況において、GCもしくはA/Tリッチな領域の伸長(expansion)、またはヌクレオチドリピートの伸長は、様々な疾患状態と関連づけることができる。例えば、X染色体上に位置する脆弱X精神遅滞-1遺伝子(FMR1)の5'非翻訳領域(UTR)におけるCGGリピートの伸長は、脆弱X症候群(FXS)ならびに様々な障害および表現型に関連している。たいていの人々において、トリヌクレオチドCGGは、FMR1遺伝子の5'非翻訳領域(UTR)、「CGGリピート領域」においておよそ5~44回、リピートしている。この領域における約45個より多いCGGリピート、特に約200個より多いCGGリピートへの伸長は、FXSと関連づけられている。FXS表現型は、精神遅滞、自閉症、不安、および他の認知または行動状態が挙げられ得る(J. Mol. Diag. 10(6): 496-501 (2008))。同様に、FMR2遺伝子の5'UTRにおけるCCGトリヌクレオチドリピート領域「CCGリピート領域」の伸長は、X連鎖性知的障害、特に脆弱X症候群E(FRAXE)に関連している。FRAXEは、X連鎖性精神遅滞の一般的な型である。他の例では、リピート長多型が、疾患状態と関連づけられている。例えば、TOMM40遺伝子のイントロン6は、ポリTリピート領域を含有し、集団においてリピート長多型を示すことが報告されている(rs10524523)。TOMM40ポリTサイズは、遅発型アルツハイマー病、および高齢者における認知能力と関連していると報告されている(The Pharmacogenetics Journal 10:375-3840 (2010);および Alzheimer's and Dementia 9:132-136 (2013)参照)。さらに、C9ORF72遺伝子におけるイントロンの(G4C2)_nヘキサヌクレオチドリピート伸長は、一般的な集団においておよそ1/600の頻度で観察されており、全ての筋萎縮性側索硬化症(ALS)および前頭側頭葉型認知症(FTD)症例のおよそ10%において存在する。30個より少ないリピートは正常とみなされ、一方、病源性C9ORF72伸長は、数百~数千個のリピートを含み得る。したがって、患者ジェノタイプ(genotype)を正確に測定し、サイズ分類し、再構築するための方法は、これらを始めとする疾患の診断および処置にとって有益であり得る。

【0004】

FMR1およびFMR2におけるCGGおよびCCGリピートなどのヌクレオチドリピートを含む配列を評価するための方法には、制限酵素消化およびポリメラーゼ連鎖反応(PCR)ストラテジーが挙げられる。制限消化分析は、リピート領域のサイズの粗い測定を提供することができる。しかしながら、制限消化分析は、分解能が限定され得、短い割り込み配列(interruption)(例えば、CGGリピート領域内のAGG割り込み配列)を容易には検出せず、かつメチル化状態を決定することができない。

【0005】

PCRストラテジーは、リピート領域をサイズ分類すること、および様々なジェノタイプを再構築することにおいて、より高い精度を提供し得る。しかしながら、長いリピート配列を含む遺伝子座、またはこれらの座についてジェノタイプを再構築する能力を邪魔するGCもしくはA/Tリッチな配列を含有する遺伝子座の増幅およびシーケンシングにおい

10

20

30

40

50

て制限が存在する。例えば、FMR1におけるCGGリピートの分析のためにPCR手順を最適化するための努力が試みられており、それには、従来のPCRアッセイへの改変が挙げられる (Genome Res. 6(7): 633-8, (1996); J. Mol. Diag. 8: 544-550, (2006); および Am. J. Med. Genet. 51(4): 527-34, (1994) 参照)。最近になって、200個を超えるCGGまたはCCGリピートを有するゲノム座のより信頼性の高い増幅を可能にするPCR技術が開発されている。

【0006】

現行のワークフローストラテジーは、PCRをサイズ分解能技術、例えば、キャピラリー電気泳動と共に用いる場合が多い。キャピラリー電気泳動は、PCRに基づいたアッセイの定量的能力を向上させて、DNA産物の単一塩基分解能に至るまでの正確な分解能を可能にする。DNA産物のサイズ分類を容易にするために、典型的には、例えば、特定のキャピラリー電気泳動適用のための目的のサイズ範囲に渡る、公知のサイズの、プールされた色素標識DNA断片を用いる、外部標準キャリブレーションが用いられる。それでも、これらの標準を適用することにより達成される実現可能な能力にも関わらず、このアプローチは、いくつかの欠点を有し、それらには、以下が挙げられるが、それらに限定されない：
i. 市販されている色素標識DNAラダーの高い費用；
ii. 標準ラダーについての専用色素チャンネルの使用により引き起こされる、キャピラリー電気泳動多重化バンド幅の低減；
iii. PCR増幅分析物と標準DNAとの間の塩基組成または配列の違いによる歪んだ電気泳動移動度に起因するサイズ分類の不正確さがあり、特別注文のDNAラダーの使用を必要とし、それは、特にリピート障害PCR産物について、サイズ分類するのに用いるのに面倒または非効率的であり得ること。さらに、FMR1 PCRおよびCE産物の断片サイズ分析は、訓練されたオペレーターにより手作業で行われ、それは、大きな試料セットについては面倒であり、かつ他の点では合理化されたワークフローへ曖昧さと主観性の両方を導入し得る。

【発明の概要】

【0007】

したがって、GCまたはA/Tリッチでもあり得る、リピート領域をサイズ分類し、それらのリピート領域と関連したジェノタイプを再構築する向上した方法の必要性が当技術分野において存在する。本明細書に開示された方法は、1つまたは複数のヌクレオチドリピート領域をサイズ分類し、それらからジェノタイプを再構築するための増幅に基づいた方法における、単独での、または外部標準と組み合わせた、内部標準の作成および使用に関する。

【0008】

図面は、必ずしも一定の縮尺、または網羅的であるとは限らない。それより、一般的に、本明細書に記載された発明の原理を図示することに重きを置かれる。この明細書に組み入れられ、かつその一部を構成する添付の図面は、本開示と一致したいくつかの実施形態を例証し、かつその説明と一緒に、本開示の原理を説明する役割を果たす。

【図面の簡単な説明】

【0009】

【図1】図1は、ジェノタイプピークサイズ分類のためのシステムの例示的な高レベル表示を描く。

【図2】図2は、ゲノムリピート領域を自動的に分析するための例示的な過程を示す。

【図3】図3Aは、外部サイズ分類標準を作成するのに用いられる例示的な外部サイズ分類ラダーを描く。図3Bは、例示的な外部サイズ分類標準を描く。

【図4】図4Aは、増幅産物のラダーの例示的なチャンネルを描く。図4Bは、リピートプロファイル (repeat profile) を用いた、内部サイズ分類標準の作成を描く。

【図5】図5は、バックグラウンドモデルおよび動的閾値を用いてアノテートされた、例示的なリピートプロファイルを描く。

【図6】図6Aは、(n, n+1) ジェノタイプを示すヘテロ接合体女性についての例示

的なリピート領域を描く。図 6 B は、例示的な伸長したリピート領域を描く。

【図 7】図 7 A ~ 7 C は、異なる品質管理問題を有する試料を描く。図 7 A は、不十分な R O X ラダーを有する試料を描く。図 7 B は、不十分な P C R 増幅を有する試料を描く。図 7 C は、混入ピークを有する試料を描く。

【図 8】図 8 は、ジェノタイプピークサイズ分類のための例示的なコンピュータシステムを描く。

【図 9】図 9 は、実施例 1 に記載されたある特定の実施形態による実験設計の概略図を描く。

【図 10】図 10 は、リピートピーク局在化が遺伝子特異的産物のサイズ分類に利用される過程のグラフ表示を描く。各リピートの期待される位置は、独立型サイズ分類ラダーとしてか、または外部サイズ分類標準の断片移動度補正としてかのいずれかで、サイズ分類のための較正曲線を作成するのに用いられ得る。

10

【図 11】図 11 は、濃い灰色での 1 1 0 6 試料確証データセットにおけるジェノタイプの分布と比較した、薄い灰色での一般的な集団における F M R 1 ジェノタイプの分布を描く。

【図 12】図 12 は、内部由来のサイズ分類ラダーを用いて決定されたリピート領域長と、R O X ラダー (R O X 1 0 0 サイズラダー、A s u r a g e n P / N : 1 4 5 1 9 4) と呼ばれる外部サイズ分類標準から導かれたサイズ分類ラダーとの間の一致を描く。

【図 13】図 13 は、臨床セットにおいて評価されたメジャーアレルの手作業でのサイズ分類と自動サイズ分類との比較を描く。

20

【図 14】図 14 A および 14 B は、遺伝子特異的産物に対するアルゴリズムの分析感度を詳述する図を描く。矢印は、自動サイズ分類により生じたコールを示す。図 14 A は、追加の遺伝子特異的産物の検出を描く。図 14 B は、低量の伸長したアレルの検出を描く。

【図 15】図 15 A および 15 B は、F M R 1 ジェノタイプ範囲に渡るアッセイの分析感度を詳述する図を描く。矢印は、自動サイズ分類により生じたコールを示す。図 15 A は、正常な試料を描く。図 15 B は、マイナーアレルを有する前突然変異試料を描く。

【図 16】図 16 は、例示的なアッセイおよびソフトウェアコンポーネントの図を示す、

【図 17】図 17 は、ある特定のアッセイを描く外部ラダーの R O X プロファイルを描く。

【図 18】図 18 A および 18 B は、臨床試料の大きな (n = 1 1 0 6) セットの自動サイズ分類分析を試験した結果を描く。

30

【図 19】図 19 A および 19 B は、複数機器の R U S H インプット量試験の結果を描く。

【図 20】図 20 A および 20 B は、人工マイナーアレルインプット滴定試験の結果を描く。

【図 21】図 21 A および 21 B は、R U S H 試料滴定試験の結果を描く。

【図 22】図 22 A および 22 B は、正常なジェノタイプを有する試料についての自動サイズ分類分析の例示的な結果を描く。図 22 C および 22 D は、前突然変異ジェノタイプを有する試料についての自動サイズ分類分析の例示的な結果を描く。

【図 23】図 23 A および 23 B は、伸長型試料についての自動サイズ分類分析の例示的な結果を描く。図 23 C および 23 D は、低レベルのマイナーアレル同定およびサイズ分類の例示的な結果を描く。

40

【図 24】図 24 は、正常、前突然変異、および伸長型のジェノタイプ範囲を通してのジェノタイプの混合物を有する対照試料についての自動サイズ分類分析の例示的な結果を描く。

【図 25】図 25 A および 25 B は、9 5 % 前突然変異試料のバックグラウンドにおける 5 % 完全突然変異試料の混合物を含む対照試料についての自動サイズ分類分析の例示的な結果を描く。図 25 A は、全てのコールされたジェノタイプを含む完全な試料を描き、一方、図 25 B は、完全突然変異コールを示すズームインバージョンを描く。

【発明を実施するための形態】

【0 0 1 0】

今、本開示によるある特定の例示的实施形態が詳細に参照され、それらのある特定の例は

50

、添付の図面において図示されている。

【0011】

本明細書で用いられたセクション見出しは、系統化することをのみ目的とし、記載された主題を限定するものとして解釈されるべきではない。非限定的に、特許、特許出願、論説、書籍、および論文を含む、この出願に引用された全ての文書、または文書の部分は、任意の目的のために、全体として参照により本明細書に明確に組み入れられている。参照により組み入れられた刊行物および特許または特許出願が、本明細書に含有された本発明と矛盾する限りにおいては、本明細書が、いかなる矛盾した材料にも優先する。

【0012】

本発明を理解するのに助けるために、ある特定の用語がまず、定義される。追加の定義は、本出願を通して提供される。

【0013】

特許請求の範囲および／または本明細書において用語「を含む」と共に用いられる場合の語「1つの(a)」、「1つの(an)」、または「その(the)」の使用は、「1つ(one)」を意味し得るが、それはまた、「1つまたは複数」、「少なくとも1つ」、および「1つまたは1つより大きい」という意味とも一致する。

【0014】

この出願において、単数形の使用は、具体的に他に指定がない限り、複数形を含む。また、この出願において、「または」の使用は、他に指定がない限り、「および／または」を意味する。さらに、用語「を含んでいる(including)」、加えて、「を含む(includes)」および「を含んだ(included)」などの他の語形の使用は、限定的ではない。本明細書に記載されたいかなる範囲も、その端点、およびその端点間の全ての値を含むものと理解される。

【0015】

本明細書で用いられる場合、用語「A／トリッチな」、「A／トリッチさ」、および「リピート性A／トリッチセグメント」は、下記で定義されたホモポリマーセグメント、または $(T_nA)_m$ 、 $(AT_n)_m$ 、 $(TAn)_m$ 、または $(AnT)_m$ (n は2以上であり、 m は、リピート性A／トリッチなセグメントの長さが10残基以上である)を含むセグメントを指す。 n の値は、そのセグメントを通して一定である必要はない。したがって、リピート性A／トリッチセグメントの例には、AATAATAATAAT、AATAAATAATAAT、AATAAATAAATAAT、AATAAATAAATAATなどが挙げられる。 $(T_nA)_m$ 、 $(AT_n)_m$ 、 $(TAn)_m$ 、または $(AnT)_m$ を含むセグメントに関して、いくつかの実施形態において、 n は2から10までの範囲の値である。いくつかの実施形態において、 n は3から10までの範囲の値である。いくつかの実施形態において、 n は4から10までの範囲の値である。いくつかの実施形態において、 n は2から8までの範囲の値である。いくつかの実施形態において、 n は3から8までの範囲の値である。いくつかの実施形態において、 n は4から8までの範囲の値である。いくつかの実施形態において、 n は2から6までの範囲の値である。いくつかの実施形態において、 n は3から6までの範囲の値である。いくつかの実施形態において、 m は2から20までの範囲の値である。いくつかの実施形態において、 m は3から20までの範囲の値である。いくつかの実施形態において、 m は4から20までの範囲の値である。いくつかの実施形態において、 m は2から15までの範囲の値である。いくつかの実施形態において、 m は3から15までの範囲の値である。いくつかの実施形態において、 m は4から15までの範囲の値である。いくつかの実施形態において、 m は2から10までの範囲の値である。いくつかの実施形態において、 m は3から10までの範囲の値である。いくつかの実施形態において、 m は4から10までの範囲の値である。いくつかの実施形態において、 m は2から8までの範囲の値である。いくつかの実施形態において、 m は3から8までの範囲の値である。いくつかの実施形態において、 m は4から8までの範囲の値である。いくつかの実施形態において、リピート性A／トリッチセグメントの長さは、約10残基から約60残基までの範囲内にある。いくつかの実施形態において、リピート性A／トリッチセグメントの長さ

10

20

30

40

50

は、約10個の連続した (consecutive) 残基から約40個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約15個の連続した残基から約40個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約20個の連続した残基から約40個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約5個の連続した残基から約50個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約10個の連続した残基から約50個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約15個の連続した残基から約50個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約20個の連続した残基から約50個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約5個の連続した残基から約60個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約10個の連続した残基から約60個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約15個の連続した残基から約60個の連続した残基までの範囲内にある。いくつかの実施形態において、リピート性A/Tリッチセグメントの長さは、約20個の連続した残基から約60個の連続した残基までの範囲内にある。他に指示がない限り、リピート性A/Tリッチセグメントは、以下の段落で説明されているような割り込み配列を含み得る。いくつかの実施形態において、リピート性A/Tリッチセグメントは、割り込み配列を含まない。

【0016】

本明細書で用いられる場合、「漏出 (bleed-over)」は、蛍光標識PCR産物が十分な強度で蛍光を発し、異なって標識されたPCR産物へ論理的に割り当てられた重複蛍光周波数放出帯域に有意なシグナルを生じる時に起こる。この漏出は、異なるチャンネルについての蛍光検出器が重複スペクトル感度を有する場合に生じ得る。この漏出は、多重反応において特定のチャンネルに固有の産物を検出する過程をコンボリューション (convolute) 得る。例えば、HEXチャンネルにおけるPCR産物は、十分な強度で蛍光を発して、ROXチャンネルにおける記録されたシグナル強度に影響し得る。

【0017】

本明細書で用いられる場合、「GCリッチな」、「GCリッチさ」、および「リピート性GCリッチセグメント」は、GヌクレオチドもしくはCヌクレオチドを含む下記で定義されたホモポリマーセグメント、またはGヌクレオチドおよびCヌクレオチドのリピート性パターンを含むセグメントを指す。CGGリピート、CCGリピート、GGGGCCリピート、および任意の散在AGG割り込み配列が含まれる。グアニン残基、シトシン残基、またはそれらの類似体である、核酸またはその核酸の断片における総核酸塩基残基に占める割合またはパーセンテージは、リッチさを定義する。例えば、正確に30個のシトシン、正確に30個のグアニン、正確に1個のシトシン類似体、および正確に1個のグアニン類似体を含有する100個のヌクレオチド配列は、62%のGCリッチさを有する。いくつかの実施形態において、「GCリッチな」核酸または核酸の領域は、約50%より多いグアニン残基、シトシン残基、またはそれらの類似体を含有するものである (例えば、約50%、51%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、または99.5%より多いグアニン残基、シトシン残基、もしくはそれらの類似体、または中間の任意のパーセンテージ)。

【0018】

本明細書で用いられる場合、用語「ホモポリマーセグメント」は、連続してリピートされたG、C、A、T、またはUなどのヌクレオチドを含む核酸のセグメントを指す。

【0019】

他に指示がない限り、ホモポリマーセグメント、GCリッチなリピート、またはA/Tリッチなリピートは、さもなければ連続した、またはリピートした、一連のヌクレオチドに

において割り込み配列を含み得る。割り込み配列は、一連を構成する他のヌクレオチドとは異なる任意の数のヌクレオチドであり得る。いくつかの実施形態において、割り込み配列は、単一ヌクレオチドである。割り込み配列を含むホモポリマーセグメントの例は、第 1 の数の T 残基、その後、1 個の C 残基、その後、第 2 の数の T 残基である。

【0020】

割り込み配列を含むホモポリマーセグメントの例は、第 1 の数の U 残基、その後、1 個の C 残基、その後、第 2 の数の U 残基である。割り込み配列を含むホモポリマーセグメントの別の例は、第 1 の数の A 残基、その後、1 個の G 残基、その後、第 2 の数の A 残基である。前述の例における A、T、または U 残基の第 1 および第 2 の数は、例えば、5 ～ 10 の範囲内であり得る。いくつかの実施形態において、前述の例における A、T、または U 残基の第 1 および第 2 の数は、6 ～ 10 の範囲内である。いくつかの実施形態において、前述の例における A、T、または U 残基の第 1 および第 2 の数は、7 ～ 10 の範囲内である。いくつかの実施形態において、前述の例における A、T、または U 残基の第 1 および第 2 の数は、8 ～ 10 の範囲内である。いくつかの実施形態において、前述の例における A、T、または U 残基の第 1 および第 2 の数は、9 ～ 10 の範囲内である。あるいは、ホモポリマーセグメントは、連続した一連のヌクレオチド（割り込まれていない）を含む。

【0021】

本明細書で用いられる場合、「核酸」は、対象から単離されている、および／またはリピート領域のサイズ分類が求められている、任意の連続する (continuous) ヌクレオチド残基または類似体である。核酸は、対象から単離された遺伝子、遺伝子断片、またはゲノム領域を含み得る。本明細書で用いられる場合、「ジェノタイプ」は、核酸の核酸塩基配列であり、またはそれを含む。

【0022】

本明細書で用いられる場合、「ピーク位置」は、勾配が sign をプラスからマイナスへ変化させるシグナルのインデックスであり得る。ピーク位置の他の定義は、構想されたシステムおよび方法から逸脱することなく用いられ得る。例えば、以下の通りである：

$$f(x, c) = \Delta_1 \text{ sign}(\Delta_1 s(x, c))$$

【0023】

【数 1】

$$P(c) = \{x \mid f(x, c) = -2 \wedge s(x, c) > 50\}$$

【0024】

式中、 $s(x, c)$ は、機器チャネル c についてのインデックス x におけるシグナル強度を表し、 $f(x, c)$ は、そのシグナル強度の一階微分係数の sign の微分係数であり、 $P(c)$ は、 $f(x, c)$ が -2 に等しい、チャネル c における全てのインデックスのセットを表す。

【0025】

本明細書で用いられる場合、「ピークショルダー」は、シグナル強度が振幅閾値を超えるピーク位置に最も近いポイントであり得る。この振幅閾値は、平均値より 2 標準偏差だけ上であり得る。いくつかの実施形態において、標準偏差および平均値は、ピークショルダーを含むシグナルの区間に対して計算され得る。左ピークショルダーは、ピーク位置より低いインデックスを有するピークショルダーであり得、右ピークショルダーは、ピーク位置より高いインデックスを有するピークショルダーであり得る。他の実施形態において、ピークショルダーは、統計的分布をピークにフィッティングさせることにより同定することができ、ショルダーは、分布の末端（パーセンタイルカットオフにより定義される）における位置に割り当てられる。さらに他の実施形態において、ピークショルダーは、ピーク領域の一階微分係数を用いて同定することができ、ショルダーは、一階微分係数の絶対値が閾値より下である位置に割り当てられる。

【0026】

本明細書で用いられる場合、「リピート領域」または「ヌクレオチドリピート領域」は、1～20 核酸塩基残基長のリピート性配列（例えば、ホモポリマー、ジヌクレオチド、トリヌクレオチド、テトラヌクレオチド、ペンタヌクレオチド、ヘキサヌクレオチド配列など）を含む核酸または核酸の領域を指し、その短い配列は、2 回以上（例えば、2 回、3 回、4 回、5 回、10 回、15 回、20 回、50 回、100 回、200 回、500 回、またはそれ以上のリピート）、リピートされる。例えば、ヌクレオチドリピートは、CGG、CCG、GGGGCC などの短い配列が 2 回以上、リピートされている核酸の領域を包含する。リピート領域は、ホモポリマー、例えば、一連の A または T ヌクレオチドであり得、リピート領域は、割り込み配列またはリピート変異体を含んでもよい。核酸または核酸の領域は、リピートと GC リッチな領域の両方、またはリピートと AT リッチな領域の両方であり得る（しかし、そうである必要はない）。例えば、核酸または核酸の領域は、グアニン残基、シトシン残基、またはそれらの類似体のジマー、トリマー、テトラマー、またはヘキサヌクレオチドリピートを包含し得る。

【0027】

核酸は、1 つまたは複数の割り込み配列を含有する 1 個または複数のヌクレオチドリピート領域、A/T リッチな領域、または GC リッチな領域を包含し得る。本明細書で用いられる場合、核酸内の「割り込み配列」は、リピートパターンと一致しない、核酸内の 1 個もしくは複数の核酸塩基残基もしくは類似体の存在を指し、または GC リッチな領域において、G もしくは C（またはその類似体）以外の核酸塩基を含む。例えば、GC リッチなヌクレオチドリピート領域は、40 個の CGG リピート内に散在した 2 個の AGG 配列を有する、40 個の CGG トリヌクレオチドリピートを包含し得る。

【0028】

本明細書で用いられる場合、「シグナル強度」は、「相対蛍光単位」または RFU によって本明細書では表されるが、蛍光の他の測定単位が、構想されたシステムおよび方法から逸脱することなく、用いられ得る。

【0029】

本明細書で用いられる場合、用語「鋳型」は、核酸合成反応における延長（extension）のためにプライマーと相互作用する核酸を指す。

【0030】

I. リピート領域のサイズ分類およびジェノタイプの再構築

図 1 は、ジェノタイプピークサイズ分類のためのシステム 100 の例示的な高レベルな表示を描く。システム 100 は、PCR デバイス 110、キャピラリー電気泳動（CE）デバイス 120、アノテーションデバイス 130、および実験室情報管理システム 140 を包含し得る。PCR デバイス 110 は、当業者によく知られた PCR 機器を包含し得る。例えば、PCR デバイスは、サーマルサイクルを包含し得る。非限定的例として、ABI モデル 9700 サーマルサイクラーが用いられ得る。PCR デバイス 110 は、核酸試料のリピート領域を増幅するように構成され得る。

【0031】

CE デバイス 120 は、当業者によく知られた CE 機器を包含し得る。例えば、ABI モデル 3100、3130、3730、または 3500 CE デバイスが用いられ得る。CE デバイス 120 は、高分解能断片分析を実施するように構成され得る。いくつかの実施形態において、CE デバイスは、増幅断片をサイズによって分離するために用いられ得る。いくつかの実施形態において、CE デバイスは、増幅産物のラダーを生成するために用いられ得る。他の実施形態において、CE デバイスは、増幅断片をサイズによって分離することにより増幅産物のラダーを生成するように用いられ得る。ある特定の実施形態において、CE デバイスは、リピート領域サイズ分類情報を得るために用いられる。CE デバイス 120 は、増幅産物のこのラダーを示すアウトプットを提供するように構成され得る。いくつかの実施形態において、このアウトプットは、ファイルに 1 つまたは複数のチャネルを包含し得る。ファイルは、.FSA ファイル、または当業者に知られた類似したファイルであり得る。

【0032】

アノテーションデバイス130は、専用コンピュータデバイス、デスクトップ、ワークステーション、オールインワンシステム、コンピュータクラスター、端末、メインフレーム、モバイルコンピュータデバイス、または他のコンピュータデバイスを含み得る。アノテーションデバイス130は、スタンドアロンであり得、またはより大きいシステムの部分であり得る、サブシステムの部分であり得る。例えば、アノテーションデバイス130は、遠隔設置され、かつパブリックネットワークまたは専用プライベートネットワークを通して通信する分散サーバーを含み得る。いくつかの実施形態において、アノテーションデバイス130は、少なくとも一部、クラウドコンピューティングインフラストラクチャ上のバーチャルシステムとして、実行され得る。開示された実施形態と一致して、アノテーションデバイス130は、データおよび／またはソフトウェア命令を記憶するように構成された1つまたは複数の記憶デバイスを含み得、またはそれらと通信し得る。記憶されたデータおよび／またはソフトウェア命令は、1つまたは複数のソフトウェアプログラムを含み得る。例えば、記憶されたデータおよび／またはソフトウェア命令は、分析ソフトウェアを含み得る。アノテーションデバイス130は、開示された実施形態と一致した1つまたは複数の方法を実施するようにこの分析ソフトウェアを実行し得る。ある特定の態様において、アノテーションデバイス130は、アノテーションデバイス130から遠隔操作で、この分析ソフトウェアを実行し得る。例えば、アノテーションデバイス130は、1つまたは複数の遠隔デバイスにアクセスして、記憶された分析ソフトウェアを実行し得る。ある特定の実施形態において、アノテーションデバイス130は、分析ソフトウェアの記憶、実行、および／または実装に基づいた特定の装置またはシステムとして構成され得る。アノテーションデバイス130は、CEデバイス120および実験室情報管理システム140などのシステム100の他のコンポーネントと通信するように構成され得る。アノテーションデバイス130は、Ethernet、FireWire、USB、RS-232、SCSI、WLAN、Bluetooth、または類似したインターフェイスを用いてシステム100のこれらのコンポーネントと通信し得る。

10

20

【0033】

アノテーションデバイス130は、リピータゲノム領域をサイズ分類するように構成され得る。このサイズ分類は、自動的に実施され得る。例えば、アノテーションデバイス130は、FMR1サイズ分類結果を作成するように構成され得る。アノテーションデバイス130は、リピータゲノム領域をサイズ分類し、および／または遺伝子産物位置を同定するように、シグナル処理技術、統計的技術、および機械学習技術の組合せを用い得る。アノテーションデバイス130は、この分析の結果をアウトプットし、および／またはこの分析の工程を仲介するように構成され得る。アノテーションデバイス130は、これらの指示を、実験室情報管理システム140へ、またはディスプレイ、プリンター、記憶デバイス、もしくは別のシステムへアウトプットするように構成され得る。

30

【0034】

実験室情報管理システム140は、専用コンピュータデバイス、デスクトップ、ワークステーション、オールインワンシステム、コンピュータクラスター、端末、メインフレーム、モバイルコンピュータデバイス、または他のコンピュータデバイスを含み得る。実験室情報管理システム140は、スタンドアロンであり得、またはより大きいシステムの部分であり得る、サブシステムの部分であり得る。例えば、実験室情報管理システム140は、遠隔設置され、かつパブリックネットワークまたは専用プライベートネットワークを通して通信する分散サーバーを含み得る。いくつかの実施形態において、実験室情報管理システム140は、少なくとも一部、クラウドコンピューティングインフラストラクチャ上のバーチャルシステムとして、実行され得る。開示された実施形態と一致して、実験室情報管理システム140は、データおよび／またはソフトウェア命令を記憶するように構成された1つまたは複数の記憶デバイスを含み得、またはそれらと通信し得る。記憶されたデータおよび／またはソフトウェア命令は、1つまたは複数のソフトウェアプログラムを含み得る。実験室情報管理システム140は、開示された実施形態と一致した1つまたは

40

50

複数の方法を実施するように、記憶された1つまたは複数のソフトウェアプログラムを実行し得る。ある特定の態様において、実験室情報管理システム140は、実験室情報管理システム140から遠隔操作で、記憶された1つまたは複数のソフトウェアプログラムを実行し得る。例えば、実験室情報管理システム140は、1つまたは複数の遠隔デバイスにアクセスして、記憶された1つまたは複数のソフトウェアプログラムを実行し得る。ある特定の実施形態において、実験室情報管理システム140は、ソフトウェア命令の記憶、実行、および／または実装に基づいた特定の装置またはシステムとして構成され得る。実験室情報管理システム140は、CEデバイス120およびアノテーションデバイス130などのシステム100の他のコンポーネントと通信するように構成され得る。実験室情報管理システム140は、Ethernet、FireWire、USB、RS-232、SCSI、WLAN、Bluetooth、または類似したインターフェイスを用いてシステム100のこれらのコンポーネントと通信し得る。

10

【0035】

実験室情報管理システム140は、試料および対応するデータを管理するように構成され得る。他の態様において、実験室情報管理システム140は、ワークフローを自動化するために用いられ得る。いくつかの実施形態において、実験室情報管理システム140は、試料管理実験室情報管理システム (Sample Manager Laboratory Information Management System)、ワトソン実験室情報管理システム (Watson Laboratory Information Management System)、ノーチラス実験室情報管理システム (Nautilus Laboratory Information Management System)、または臨床実験室情報管理システム (Clinical Laboratory Information Management System) を実行するように構成され得る。当業者は、適切な実験室情報管理システムについて容易にわかるだろう。実験室情報管理システム140は、ゲノム試料に関する情報を受け取るように構成され得る。実験室情報管理システム140は、アノテーションデバイス130、記憶デバイス、または別のシステムからのこの情報を受け取り得る。実験室情報管理システム140は、この情報を記憶のために整え、かつ関連臨床医に見せるように構成され得る。

20

【0036】

当業者により認識されているように、図1に描かれたデバイスの特定の配置は、限定することを意図するものではない。例えば、システム100は、追加のデバイス、またはより少ないデバイスを含む場合がある。同様に、システム100の個々のデバイスの機能が、複数のデバイスに渡って分散され得、システム100の異なるデバイスにより実施される複数の機能が、単一のデバイスによって実施され得る。

30

【0037】

システム100は、リピート領域をサイズ分類し、適宜、ジェノタイプを再構築する以下の方法を実施するように構成され得る。いくつかの実施形態において、リピート領域のサイズ分析は、リピート領域を増幅すること、および増幅された産物を用いて、リピート領域サイズを決定することを含み得る。ある特定の実施形態において、核酸のリピート領域が増幅されて、増幅産物のラダーが得られる。増幅産物のラダーは、リピート領域をサイズ分類することにおける内部標準として用いられ得る。いくつかの実施形態において、内部標準は、任意の外部標準なしに用いられる。他の実施形態において、内部標準は、外部標準と組み合わせて用いられる。いくつかの実施形態において、外部標準は、蛍光標識DNAラダー、例えば、ROXサイズ標準であり得る。より具体的な実施形態において、ROXサイズ標準は、ROX 1000サイズラダー (Asuragen P/N: 145194) であり得る。リピート領域の増幅パターンは、リピート領域をサイズ分類するために用いられ得る。当業者は、増幅産物のラダーが、リピートプロファイル、リピートエレメント周期性、第1の増幅産物、増幅産物カウント数、および／または一定のエレメント長などの、リピート領域をサイズ分類することにおいて有用なある特定の特徴を有することを理解するであろう。ある特定の実施形態において、サイズ分類情報は、再構築される

40

50

ジェノタイプを作製するために、患者において障害を診断するために、もしくは患者の子孫の障害のリスクを診断するために、または伸長したリピート領域に関連した障害を有する患者を処置することにおいて、用いられ得る。

【0038】

様々な実施形態において、リピート領域をサイズ分類する方法は、リピート領域を含む核酸または部分を増幅して、一連の増幅断片を生成することを含む。増幅断片は、特定の断片において増幅されたリピート単位の数に対応する、異なる長さのものであり得る。他の実施形態において、リピート領域を含む核酸または部分が増幅されず、さらなる分析のために直接的に単離され、断片化される。いくつかの実施形態において、増幅断片（または増幅されていない断片）は、例えば、高分解能断片分析、例えば、遺伝子アナライザ、マイクロチップアナライザ（例えば、バイオアナライザ）、キャピラリー電気泳動、またはラダーの増幅断片を分析するための別の高分解能方法を用いる分析などのサイズ分解能技術を用いて、サイズによって分離される。例えば、キャピラリー電気泳動が用いられ得る。ある特定の実施形態において、バイオアナライザなどのマイクロチップ電気泳動が用いられ得る。様々な実施形態において、高分解能断片分析は、例えば、キャピラリー電気泳動電気泳動図において異なるリピート数の増幅産物に対応するピークを評価することにより、増幅産物のラダーを生成するために用いられる。いくつかの実施形態において、個々のリピート単位の公知の長さは、リピート単位のラダーを、ヌクレオチド塩基対（bp）長を示すラダーへ変換するために用いることができる。例えば、リピート単位を含む核酸領域の増幅およびキャピラリー電気泳動は、3ヌクレオチドの単位によって長さが異なる増幅断片のラダーを生じ、増幅産物のラダーについてのパラメータ情報を用いて、ラダーをヌクレオチド長の測定単位への変換を可能にする。いくつかの実施形態において、ラダーは、目的の核酸においてリピート領域のサイズを決定するために用いられる。いくつかの実施形態において、ラダーは、リピート領域を含む核酸の他の部分のサイズ、またはラダーを生成する同じ反応において増幅された目的の他の核酸のサイズを決定するために用いられる。いくつかの実施形態において、リピート領域サイズは、ジェノタイプを再構築するために用いられる。ある特定の実施形態において、リピート領域における任意の割り込み配列への順方向および逆方向での距離などの追加のパラメータもまた同定され（例えば、キャピラリー電気泳動電気泳動図から）、ジェノタイプを再構築するためにリピート領域サイズと共に用いられる。ある特定の実施形態において、リピート配列における割り込み配列は、増幅産物のラダーにおいて検出される。

【0039】

様々な実施形態において、リピート領域、例えば、GCリッチもしくはA/Tリッチな領域をサイズ分類し、および／もしくは特徴づけるための方法、ならびに／またはリピート領域を含むジェノタイプを再構築するための方法が本明細書で開示される。例えば、本明細書で開示される方法は、CGGリピートまたはCCGリピートを含む核酸またはその断片からリピート領域をサイズ分類するために用いることができる。本明細書で開示された方法は、ホモポリマーセグメントなどのA/Tリッチセグメントを含む核酸またはその断片からリピート領域をサイズ分類するために用いることができる。サイズ分類の方法は、リピート領域において割り込み配列を決定するための方法、加えて、サイズ分類に基づいてジェノタイプを再構築する方法（単独で、またはリピート領域における任意の割り込み配列への順方向および逆方向での距離などの追加のパラメータと組み合わせて）と共に用いることができる。

【0040】

いくつかの実施形態において、方法は、対象から単離された、FMR1もしくはFMR2遺伝子、もしくはその断片、またはFMR1もしくはFMR2の5'UTR、もしくはその断片のリピート領域をサイズ分類するために用いることができる。ある特定の実施形態において、本明細書で開示された方法は、CGGリピートパターンおよび位置、ならびにAGG割り込み配列の組織化ならびに／またはFMR1の5'UTR内のメチル化を含む、対象由来の試料においてFMR1遺伝子についてのジェノタイプを再構築するのを助けるた

10

20

30

40

50

めに用いられる。他の実施形態において、本明細書で開示された方法は、FMR1についてのように、CCGリピートパターンを含む、FMR2についてのジェノタイプを再構築するのに助けるために用いられる。さらに他の実施形態において、本明細書で開示された方法は、TOMM40のリピー領域をサイズ分類するのに助けるために用いられる。他の実施形態において、方法は、C9ORF72のリピー領域をサイズ分類するのに助けるために用いられる。

【0041】

いくつかの実施形態において、本明細書で開示された方法は、患者試料由来の核酸またはその断片のリピー領域のサイズを決定するために用いられ、核酸が、少なくとも1つのリピートまたはGCリッチもしくはA/Tリッチな領域を有し、かつ前記患者の両親の少なくとも1人由来の関連ジェノタイプが知られていない。いくつかの実施形態において、リピート領域を含む、目的の核酸または部分が患者試料から単離される。様々な単離および精製方法が知られており、用いることができる。ある特定の実施形態において、本明細書で開示された方法は、患者試料由来の例えばFMR1またはFMR2の、CCGまたはCCGリピート領域のサイズを決定するために用いられる。ある特定の実施形態において、本明細書で開示された方法は、患者試料由来のヘキサマーリピート、例えば、C9ORF72のGGGGCCリピートのサイズを決定するために用いられる。ある特定の実施形態において、本明細書で開示された方法は、患者試料由来のホモポリマーリピート、例えば、TOMM40のポリTリピート領域のサイズを決定するために用いられる。いくつかの実施形態において、患者の両親の少なくとも1人由来の、FMR1、FMR2、C9ORF72、またはTOMM40ジェノタイプなどの関連ジェノタイプが知られていない。

【0042】

ある特定の実施形態において、核酸試料のリピー領域のサイズ分類のための方法は、患者由来の試料を提供することを含み、試料が、1つまたは複数のリピート領域またはGCリッチもしくはA/Tリッチな領域を有する核酸またはその断片を含有する。いくつかの実施形態において、核酸を特徴づける情報（すなわち、「パラメータ情報」）が収集される。いくつかの実施形態において、パラメータ情報は、リピートプロファイル、リピートエレメント周期性、第1の増幅産物、増幅産物カウント数、および／または一定のエレメント長を含む、増幅産物のラダーから得られる特徴を含む。いくつかの実施形態において、リピートプロファイルは、電気泳動図において観察されるピークのパターンである。いくつかの実施形態において、増幅産物は、リピート領域を、選択されたプライマーを用いて増幅することにより産生される断片の蔓延（spread）である。いくつかの実施形態において、リピート領域の全長は、パラメータ情報から計算される。いくつかの実施形態において、追加のパラメータ情報、例えば、目的の領域のGCリッチさもしくはA/Tリッチさのパーセント、ならびに／またはリピートもしくはGCリッチもしくはA/Tリッチな領域における任意の割り込み配列への順方向および逆方向での距離に関する情報が作成される。いくつかの実施形態において、収集された情報は、自動分析を行うようにプログラミングされたプロセッサを含む装置を用いて自動的に分析される。ある特定の実施形態において、サイズ分類解答（solution）または解答ジェノタイプの精度は、それがパラメータ情報と適合することを確認するようにジェノタイプを手作業で分析することにより、または任意の他の確認アッセイ（例えば、制限酵素消化、サンガーシーケンシング、またはハイスループットシーケンシングの他の形式）を行うことにより評価することができる。いくつかの実施形態において、サイズ分類解答または解答ジェノタイプは、コンピュータ上に電子的にディスプレイもしくは記憶することができ、またはその後の診断および治療目的のために印刷することができる。

【0043】

ある特定の実施形態において、リピート領域のサイズ分類は、リピート領域において、突然変異もしくはジェノタイプを検出するために、またはリピート領域における突然変異に関連した障害もしくは障害のリスクを診断し、もしくは診断するのに助けるために用いることができ、例えば、FMR1、FMR2、C9ORF72、またはTOMM40関連の

10

20

30

40

50

突然変異、ジェノタイプ、または障害が挙げられる。

【0044】

様々な実施形態において、核酸のリピート領域を特徴づけるサイズ分類情報は、増幅および高分解能断片分析などの任意の適切な方法を用いて得ることができる。ある特定の実施形態において、サイズ分類情報（例えば、パラメータ情報のサブセット、核酸を特徴づける情報）は、増幅産物のラダーに関するリピートプロファイル、リピートエレメント周期性、第1の増幅産物、増幅産物カウント数、および／または一定のエレメント長を含む。いくつかの実施形態において、リピート領域の全長、加えてリピート領域の開始点から最初の、またはそれに続く割り込み配列までの順方向および逆方向での距離が、パラメータ情報に含まれる。いくつかの実施形態において、パラメータ情報を分析するように、およびリピート領域をサイズ分類するように、適宜、その情報からジェノタイプを再構築するようにプログラミングされたプロセッサを含む装置が提供される。ある特定の実施形態において、装置は、核酸を特徴づける情報から核酸のジェノタイプを再構築するために用いられる。いくつかの実施形態において、装置は、増幅産物のラダーにおいて各産物のサイズを評価する。いくつかの実施形態において、リピート領域の長さならびに順方向および逆方向での割り込み配列に基づいた全ての可能なジェノタイプ再構築が作製されて、全てのパラメータ情報を満たす再構築が選択される（例えば、順方向と逆方向の両方において正しい位置に割り込み配列を配置するジェノタイプ）。ある特定の実施形態において、装置は、スクリーン上にディスプレイされ、将来の使用のためにデジタル的に保存され、または記録された紙として印刷され得る、再構築されたジェノタイプの報告書を提供する。

10

20

【0045】

様々な実施形態において、核酸に関するパラメータ情報は、核酸をサイズ分類するのに適した増幅産物のラダーに関する情報を含む限り、当技術分野において公知の任意の方法を用いて得ることができる。いくつかの実施形態において、パラメータ情報は、増幅産物のラダーに関する、リピートプロファイル、リピートエレメント周期性、第1の増幅産物、増幅産物カウント数、および／または一定のエレメント長を含む。核酸部位を特異的に切断する制限酵素が、リピート領域を分析し、それにより、パラメータ情報を作成するために用いることができる。例えば、FMR1のCGGリピート区画内のAGG割り込み配列の存在は、制限酵素EciI（New England Biolabs Inc.、Ipswich、MA、USA）を用いて核酸を消化することにより検出することができる。制限酵素は、消化された産物のラダーを生成するために用いられ得、そのラダーをサイズ分類に用いることができる。他の実施形態において、必要な情報を作成するために増幅方法を用いることができる。例えば、制限消化および／またはPCR方法を、1つまたは複数のCGGまたはCCGリピート領域を決定することにおいて、患者から単離されたFMR1もしくはFMR2遺伝子またはその断片と共に用いることができる。

30

【0046】

国際公開番号WO／2014／015273に開示された方法は、全体として参照により本明細書に組み入れられており、それは、リピート領域を分析し、リピートサイズ、およびリピート領域における任意の割り込み配列までの順方向および逆方向での距離を含むパラメータ情報を得るための、その刊行物に開示された、PCRおよびキャピラリー電気泳動方法を含む。

40

【0047】

いくつかの実施形態において、リピート領域を増幅して増幅産物を生成するための適切な方法には、ポリメラーゼ連鎖反応（PCR）、リアルタイムPCR（RT-PCR）、核酸配列ベース増幅（nucleic acid sequence-base amplification）（NASBA）、リガーゼ連鎖反応、多重ライゲーション依存性プローブ増幅（multiplex ligatable probe amplification）、インバーダーテクノロジー（Third Wave）、ローリングサークル増幅、インビトロ（in vitro）転写、鎖置換増幅、転写媒介性増幅（TMA）、RNA（例えば、Eberwine）増幅、ループ媒介性等温増幅、または当業者に知られ

50

ている任意の他の方法が挙げられる。例えば、FMR1リピート領域増幅は、CGGリンカープライマーおよびヒトFMR1 PCRキット (Asuragen Inc.、Austin、USA) を用いた2段階PCRアプローチを用いて、生じさせることができる。Tassone et al., J Mol Diagn. 10(1):43-49 (2008); Chen et al., J Mol Diagn. 12(5): 589-600 (2010); Yrigollen et al., PLoS One 6(7): e21728 (2011)を参照されたい。例えば、少なくとも1つのGCリッチな領域を含む核酸が、(a) CGG、CCG、GCG、CGC、GCC、またはGGCリピートを含む第1のプライマーおよびGCリッチな領域の外側の位置にアニールする第2のプライマーを含む、少なくとも2つのPCRプライマーを提供すること；(b) 少なくとも2つの異なるプライマーを用いて核酸にPCRを実施することであって、ここで、PCRが、PCR産物のセットを生じる、実施すること；(c) PCR産物のセットを高分解能技術（キャピラリー電気泳動など）を用いて分解して、PCR産物のサイズおよび存在量の表示を生成させること；ならびに(d) PCR産物のサイズおよび存在量情報からGCリッチな領域の長さ、およびGCリッチな領域内に割り込み配列が位置しているかどうか、またはどこに位置しているかを導くことにより、分析することができる。

【0048】

様々な実施形態において、PCR増幅された核酸は、例えば、当業者によく知られたキャピラリー電気泳動 (CE) 機器などの高分解能断片アナライザ、例えば、ABIモデル3100、3130、3730、または3500 CE機器 (Applied Biosystems、Carlsbad、CA) を用いて、リピート領域サイズ分類情報を得るように分析される。増幅された核酸を電気泳動的に、または別の方法で、サイズ分類し、および／またはシーケンシングする能力がある任意の機器を含む他の道具もまた用いることができる。サイズ分類および他のパラメータ情報を収集する任意の他の方法（例えば、サンガーシーケンシングまたはハイスループットシーケンシングの他の形式）もまた用いることができる。米国特許出願公開第2010/0209970号、第2010/0243451号、および第2012/0107824号に記載されたPCR方法などの、FMR1遺伝子またはその断片を分析するための様々な技術は、リピートスペーシングにより長さが分割された増幅産物のラダーを生じる場合が多い。例えば、FMR1およびFMR2の5' UTRまたはそれらの断片においてCGGおよびCCGリピート座を特徴づけるためのリピート領域サイズ分類およびジェノタイプ再構築は、米国特許出願公開第2010/0243451号（参照により本明細書に組み入れられている）に記載された、段落【0040】～【0051】、【0056】～【0060】、【0065】～【0067】、【0089】、【0094】、および【0104】において開示されたプライマー、ポリメラーゼ、試薬、および反応条件を含む、方法を用いて、作製することができる。追加として、米国特許出願公開第2010/0209970号、第2010/0243451号、および第2012/0107824号（全体として参照により本明細書に組み入れられている）は、GCリッチな領域を分析するためのPCR方法および試薬を記載する。

【0049】

例えば、いくつかの実施形態において、FMR1およびFMR2パラメータ情報は、リピート領域の外側にアニールするプライマー、およびリピート配列、並べ換え配列 (sequence permutation)、または配列 (GCG、CCG、CGC、GCC、またはGGC) の逆相補体にアニールするプライマーを用いて作成することができる。リピート領域の外側（上流または下流）にアニールすることができるプライマーは、フォワードプライマーまたはリバースプライマーであり得る。プライマーは、リピート領域を挟む配列にアニールしてもよい。フォワードプライマーの例には、CGG TGG AGG GCC GCC TCT GAG C（配列番号1）、CAG GCG CTC AGC TCC GTT TCG GTT T（配列番号2）、CAG TCA GGC GCT CAG CTC CGT TTC G（配列番号3）、TCC GGT GGA GGC CCG CCT CTG AGC（配列番号4）、GGT TCG GCC TCA GTC AGG CGC TCA GCT CC G TTT CG（配列番号5）、GGG TTC GGC CTC AGT CAG GCG CTC AGC TCC GTT T CG（配列番号6）、GCG GGC CGG GGG TTC GGC CTC AGT CA（配列番号7）、CAG

CGG GCC GGG GGT TCG GCC TCA G (配列番号 8)、GCA GCG GGC CGG GGG TTC GGC CTC A (配列番号 9)、GGG CCG GGG GTT CGG CCT CAG TCA G (配列番号 10)、GGG GTT CGG CCT CAG TCA GGC GCT CA (配列番号 11)、GGG GTT CGG CCT CAG TCA GGC GCT CAG (配列番号 12)、GGC GCT CAG CTC CGT TTC GGT TTC ACT TCC (配列番号 13)、TCA GGC GCT CAG CTC CGT TTC GGT TTC A (配列番号 14)、CAC TTC CGG TGG AGG GCC GCC TCT GA (配列番号 15)、TTC CG G TGG AGG GCC GCC TCT GAG C (配列番号 16)、および TCA GGC GCT CAG CTC CGT TTC GGT TTC ACG GCG GCG GCG GCG GA (配列番号 44) が挙げられる。リバープライマーの例には、CGC ACT TCC ACC ACC AGC TCC TCC A (配列番号 17)、GGA GCC CGC CCC CGA GAG GTG (配列番号 18)、GGG AGC CCG CCC CCG AG A GGT (配列番号 19)、CGC ACT TCC ACC ACC AGC TCC TCC AT (配列番号 20)、CGG GAG CCC GCC CCC GAG AGG TG (配列番号 21)、CCG GGA GCC CGC CC C CGA GAG GT (配列番号 22)、CCG GGA GCC CGC CCC CGA GAG GTG (配列番号 23)、CGC CGG GAG CCC GCC CCC GAG AGG TG (配列番号 24)、GCG CCG GGA GCC CGC CCC CGA GAG GT (配列番号 25)、CGC CGG GAG CCC GCC CCC GAG A GG T (配列番号 26)、GCG CCA TTG GAG CCC CGC ACT TCC ACC A (配列番号 27)、GCG CCA TTG GAG CCC CGC ACT TCC A (配列番号 28)、AGC GCC ATT G GA GCC CCG CAC TTC C (配列番号 29)、CGC CAT TGG AGC CCC GCA CTT CCA C (配列番号 30)、TTG GAG CCC CGC ACT TCC ACC ACC A (配列番号 31)、AGC CCC GCA CTT CCA CCA CCA GCT CCT C (配列番号 32)、GAG CCC CGC ACT TCC ACC ACC AGC TCC T (配列番号 33)、CAT TGG AGC CCC GCA CTT CCA CCA CCA G (配列番号 34)、CCC GCA CTT CCA CCA CCA GCT CCT CCA TCT (配列番号 35)、TAG AAA GCG CCA TTG GAG CCC CGC ACT TCC (配列番号 36)、AAG CGC C AT TGG AGC CCC GCA CTT CC (配列番号 37)、AAG CGC CAT TGG AGC CCC GCA CTT CCC CGC CGC CGC CGC CG (配列番号 43)、および AAG CGC CAT TGG AGC CCC GCA CTT CCC CGC CGC CGC CGC CT (配列番号 45) が挙げられる。

【0050】

いくつかの実施形態において、FMR1 および FMR2 アッセイは、プライマー TCAGGC GCTCAGCTCCGTTTCGGTTTCACTTCCGGT (配列番号 38)、AGCGTCTACTGTCTCG GCACTTGCCCGCCGCGCCG (配列番号 39)、TCA GGC GCT CAG CTC CGT TTC G GT TTC A (配列番号 40)、および TCAGGCGCTCAGCTCCGTTTTCGGTTTCA CGGCGG CGGCGG (配列番号 41) を用いることができる。方法は追加として、配列番号 1 ~ 38 または 40 のいずれかの配列を含むプライマー、ならびに 3' 末端に付加された C G G またはその並べ換え型および逆相補体 (例えば、G C G、C C G、C G C、G C C、または G G C) の追加のリピートを含むプライマーを用いることを含む。いくつかの実施形態において、プライマーにおける C G G リピートまたは並べ換え型の数は、4 個または 5 個である。いくつかの実施形態において、プライマーは、12 ~ 15 ヌクレオチドまたはそれ以上に及ぶ C G G リピート (またはその並べ換え型) の配列を含有する。いくつかの実施形態において、プライマーは、3 リピートから 10 リピートまでの範囲である C G G リピート (またはその並べ換え型) の配列を含有する。プライマーは、3、4、5、6、7、8、9、または 10 リピート、ならびに適宜、1 個または 2 個の C および / または G 残基の追加の部分的リピートを含有し得る。

【0051】

いくつかの実施形態において、リピート領域または G C リッチな領域にアニールするプライマーは、割り込みエレメントを含む領域における部位に対する優先的な結合活性を有する。割り込みエレメントの部位との優先的な結合は、例えば、PCR 反応において、リピート領域または G C リッチな領域の外側に結合する逆方向の第 2 のプライマーと共にそのプライマーを用いることにより、割り込みエレメントを含む少なくとも 1 個の産物の選択的増幅を生じ得る。優先的な結合活性は、例えば、C G G および A G G エレメント、またはそれらの並べ換え型および / もしくは逆相補体を含む部位、例えば、(1) 1 個の A G

G エlement、またはAを含むA G G Elementの一部、および(2) 3個、4個、5個、または6個のC G G Element、および適宜、追加の部分的C G G Elementを含む部位に特異的であり得る。

【0052】

いくつかの実施形態において、リピート領域またはG C リッチな領域にアニールし、かつ割り込みElementを含む部位と優先的に結合するプライマーは、リピートまたはG C リッチな配列にアニールするプライマーの部分内に、またはその末端にA、T、またはU 残基を含み得る。例えば、プライマーは、C G G、C C G、G C G、C G C、G C C、またはG G C リピートのひと続きの中に、または末端にA、T、またはU を有し得、例えば、上記の配列番号44および45を参照されたい。A、T、またはU 残基は、プライマーの3'末端に存在し得る。A、T、またはU 残基がC G G、C C G、G C G、C G C、またはG C C、G G C リピートの末端に存在する場合、A、T、またはU 残基と最後の完全なC G G、C C G、G C G、C G C、G G C、またはG G C リピートとの間に部分的C G G、C C G、G C G、C G C、G C C、またはG G C リピートがあってもよいし、なくてもよい。A、T、またはU 残基の代わりに、他の天然ヌクレオチド残基と比べてT/UまたはA 残基と優先的に塩基対形成する非天然ヌクレオチド残基を用いることは可能である。同様に、C G G、C C G、G C G、C G C、G C C、またはG G C リピートを構成する1個または複数のG および/またはC 残基の代わりに、他の天然ヌクレオチド残基と比べてC またはG 残基と優先的に塩基対形成する1個または複数の非天然ヌクレオチド残基を用いることも可能である。さもなければC G G、C C G、G C G、C G C、G C C、またはG G C リピートを構成する配列(適宜、A、T、U 残基、または上記で論じられているような対応する非天然残基と共に)内の1個または複数のそのような非天然ヌクレオチド残基の存在は、C G G、C C G、G C G、C G C、G C C、またはG G C リピートの配列としての本開示の脈絡内での前記配列のアイデンティティを無効にすることはない。非天然ヌクレオチド残基は、アデニン、チミン、グアニン、シトシン、およびウラシル(それぞれ、A、T、G、C、およびU)以外の核酸塩基を含むヌクレオチド残基である。A またはT/U 残基と優先的に塩基対形成する非天然ヌクレオチド残基の例には、非限定的に、他の天然残基と比べてA またはT/U 残基と優先的に塩基対形成するT、U、またはA 残基の付加体(例えば、5-置換ウラシル類似体)、ならびに、例えば、プソイドウラシルおよびジアミノプリンなどの核酸塩基を含む残基が挙げられる。

【0053】

いくつかの実施形態において、C9ORF72パラメータ情報は、リピート領域の外側にアニールするプライマー、およびリピート配列、並べ換え配列、またはその配列の逆相補体にアニールするプライマーを用いて作成することができる。リピート領域の外側(上流または下流)にアニールすることができるプライマーは、必要に応じてフォワードプライマーまたはリバースプライマーであり得る。これらの配列は、リピート領域を挟む配列にアニールし得る。フォワードプライマーの例には、T G C G C C T C C G C C G C C G C G G G G C G C A G G C A C C G C A A C C G C A (配列番号46)が挙げられる。リバースプライマーの例には、C G C A G C C T G T A G C A A G C T C T G G A A C T C A G G A G T C G (配列番号47)、T G C G C C T C C G C C G C C G C G G G C G C A G G C A C C G C A A C C G C A C C C C G G C C C C G G C C C C G G (配列番号48)、C G C A G C C T G T A G C A A G C T C T G G A A C T C A G G A G T C G C C G G G C C G G G G C C G G G G (配列番号49)が挙げられる。

【0054】

いくつかの実施形態において、TOMM40パラメータ情報は、リピート領域の外側にアニールするプライマー、およびリピート配列、並べ換え配列、またはその配列の逆相補体にアニールするプライマーを用いて作成することができる。リピート領域の外側(上流または下流)にアニールすることができるプライマーは、フォワードプライマーまたはリバースプライマーであり得る。これらの配列は、リピート領域を挟む配列にアニールし得る

。フォワードプライマーの例には、C C A A A G C A T T G G G A T T A C T G G C（配列番号50）が挙げられる。リバースプライマーの例には、G A T T G C T T G A G C C T A G G C A T T C（配列番号51）が挙げられる。

【0055】

いくつかの実施形態において、第1のプライマーは、増幅産物のラダーを作製する場合に用いられ、割り込みエレメントを含まないリピートまたはGCリッチな領域における部位に対する優先的な結合活性を有する。割り込みエレメントの存在は、この方法の結果において、合成が割り込みエレメントを含む部位と結合した第1のプライマーの延長を含んだであろう、相対的に低レベルの産物により、シグナルを発生し得る。これらの低レベルは、電気泳動図において、より高いピークにより囲まれた、ギャップまたは低いピークのセットとして現れ得る。いくつかの実施形態において、割り込みエレメントを含むリピートまたはGCリッチな領域における部位に対する優先的な結合活性を有する第1のプライマーが提供される。割り込みエレメントの存在は、アンカード（anchored）アッセイにおいて、合成が割り込みエレメントを含む部位と結合した第1のプライマーの延長を含んだ、相対的に高レベルの産物により、シグナルを発生する。その高レベルは、電気泳動図において、より低いピークおよび／またはベースラインシグナルにより囲まれたスパイクとして現れ得る。

【0056】

パラメータ情報を作成する方法は、少なくとも2つまたは少なくとも3つの異なるプライマーを提供することを含む増幅反応に関わり得る。いくつかの実施形態において、少なくとも3つの異なるプライマーが提供され、そのプライマーの1つが、リピートまたはGCリッチな領域の外側で優先的に結合するプライマーであり、第2のプライマーは、リピートまたはGCリッチな領域内で優先的に結合し、第3のプライマーは、第1のプライマーまたは第2のプライマーのいずれかの部分列である。いくつかの実施形態において、1つのプライマーは、CGGリピートおよび5'フラップ配列を含むキメラプライマーであり、別のプライマーは、そのキメラプライマーの5'フラップ配列の配列を有する。キメラプライマーの5'フラップ配列の配列を有するプライマーは、キメラプライマーの非リピート配列全体を有し得るが、必ずしもそうとは限らないことは留意されるべきである。言い換えれば、1つのプライマーの一部または全部の配列が、別のプライマーの配列により含まれ得、例えば、キメラプライマーは5'フラップ配列を含み、別のプライマーは、5'フラップの一部または全部の配列を含み得る。いくつかの実施形態において、プライマーは、CGGリピート配列の12～15ヌクレオチドを含有する。5'フラップ配列は、CGGリピート領域に隣接した、もしくはその近くの配列に対応し得、またはそれは、CGGリピート領域内とその周辺の配列とは無関係であり得る。いくつかの実施形態において、キメラプライマーの長さは、およそ35、40、45、50、または55ヌクレオチドであり得る。いくつかの実施形態において、プライマーの1つまたは複数は、60℃から75℃までの範囲、例えば、およそ60℃、65℃、70℃、または75℃の融解温度を有する。

【0057】

いくつかの実施形態において、少なくとも3つの異なるプライマーが提供され、1つのプライマーが、別のプライマーの濃度より低い濃度で提供される。例えば、キメラプライマーは適宜、キメラプライマーの5'フラップ配列の配列を有するプライマーより低い濃度で提供される。倍数の差として表される濃度の比は、2から10,000までの範囲、またはそれ以上、例えば、10、20、50、100、200、500、1,000、2,000、5,000、または10,000（またはその間の任意の値）であり得る。そのような実施形態において、より低い濃度で存在するプライマーは、増幅反応の初期ラウンドで枯渇し得、その結果、延長は、一般的に、完全に、またはほとんど完全に、まだ存在するプライマー（最初、相対的により高い濃度で存在していた）からである。

【0058】

いくつかの実施形態において、パラメータ情報を作成する方法は、1より高いGC/AT比で、かつリピートまたはGCリッチな鋳型を含むDNAの合成を誘導する総dNTP濃

10

20

30

40

50

度で、 $dNTP$ を提供することを含む。米国特許出願公開第2010-0209970号を参照。 GC/AT 比は、約1.1、1.2、1.4、1.6、2、2.5、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、またはそれより高くあり得る。 GC/AT 比は、1.1から20の間、1.1から15の間、1.1から10の間、1.1から8の間、1から15の間、1.1から7の間、1.1から6の間、1.1から5の間、1.2から25の間、1.4から25の間、1.6から25の間、2から25の間、3から25の間、4から25の間、5から25の間、2から15の間、2.5から10の間、または4から10の間であり得る。総 $dNTP$ 濃度は、約0.4 mM、0.5 mM、0.6 mM、0.7 mM、0.8 mM、0.9 mM、1 mM、1.2 mM、1.5 mM、2 mM、または3 mMであり得る。 $dNTP$ 濃度は、0.4 mMから3 mMの間、0.5 mMから3 mMの間、0.6 mMから3 mMの間、0.7 mMから3 mMの間、0.8 mMから3 mMの間、0.9 mMから3 mMの間、1 mMから3 mMの間、0.4 mMから2 mMの間、0.4 mMから1.5 mMの間、0.4 mMから1.2 mMの間、0.4 mMから1 mMの間、0.4 mMから0.9 mMの間、0.4 mMから0.8 mMの間、0.4 mMから0.7 mMの間、0.5 mMから2 mMの間、0.5 mMから1 mMの間、または0.6 mMから0.9 mMの間であり得る。「 GC/AT 比」とは、所定の溶液または混合物における、 $dCTP$ 、 $dGTP$ 、およびそれらの全てのヌクレオチド類似体の合計の濃度の、 $dATP$ 、 $dTTP$ 、 $dUTP$ 、およびそれらの全てのヌクレオチド類似体の合計の濃度に対する比を意味する。「 $dNTP$ 」は、デオキシヌクレオチド三リン酸を表し、 $dATP$ 、 $dCTP$ 、 $dGTP$ 、 $dTTP$ 、 $dUTP$ 、およびそれらの類似体を指す。「ヌクレオチド類似体」は、天然の塩基アデニン(A)、シトシン(C)、グアニン(G)、チミン(T)、またはウラシル(U)以外の塩基部分、デオキシリボースと同一または類似した糖部分、および少なくとも1個のリン酸または複数のリン酸(二リン酸または三リン酸)部分を含む分子またはイオンである。ヌクレオチド類似体は、 $dTTP$ の類似体が一般的に $dUTP$ の類似体でもあり、その逆もまた同様であることを別として、それが、三リン酸および糖部分(その両方の構造と立体配置がポリメラーゼによる核酸二重らせんへの取り込みに適している)、ならびに塩基(その、核酸二重らせんにおける塩基対形成性および核酸二重らせんにおけるDNAポリメラーゼにより取り込みの座位が、前に列挙された5個のヌクレオチドの1個と最も類似している)を含む場合、特定のヌクレオチド、特に $dATP$ 、 $dCTP$ 、 $dGTP$ 、 $dTTP$ 、または $dUTP$ の類似体である。非限定的に「ヌクレオチド」、「塩基」、「核酸塩基」、または「残基」を含む用語と共に用いられる用語「類似体」は、あたかもそれが「ヌクレオチド」と共に用いられたかのように、同じように解釈されるべきである。

【0059】

いくつかの実施形態において、パラメータ情報を作成する方法は、PCR増幅反応についてのバッファーを提供することをさらに含み得る。バッファーは、例として、かつ非限定的に、トリス(ヒドロキシメチル)アミノメタン(Tris)、ビストリスプロパン、重炭酸塩、リン酸塩、グリシン、ヒスチジン、4-(2-ヒドロキシエチル)-1-ピペラジンエタンスルホン酸(HEPES)、3-(N-モルホリノ)プロパンスルホン酸(MOPS)、ならびにそれらの様々なコンジュゲート塩基/酸および塩を含み得る。

【0060】

いくつかの実施形態において、パラメータ情報を作成する方法は、鋳型依存性様式で $dNTP$ からDNAを合成する少なくとも1つのDNAポリメラーゼを提供することを含み得る。DNAポリメラーゼは、野生型、改変型、好熱性、キメラ、遺伝子組換え型、および/または2つ以上のポリメラーゼの混合物を含み得る。DNAポリメラーゼは、Exact Polymerase (5 PRIME GmbH)、AccuSure (商標) DNAポリメラーゼ (Bioline)、Phusion (商標) AccuPrime (商標) Pfx (Invitrogen)、Platinum Taq DNAポリメラーゼ高フイデリティ (Invitrogen)、Phire (商標) ホットスタートDNAポリメ

10

20

30

40

50

ラーゼ (New England Biolabs)、Phusion (登録商標) ホット
 スタート高フィデリティDNAポリメラーゼ (New England Biolabs)
 、JumpStart (商標) REDTaq (商標) DNAポリメラーゼ (Sigma-
 Aldrich)、PfuUltra (商標) ホットスタートDNAポリメラーゼ (St-
 ratagene)、PfuTurbo (登録商標) CxホットスタートDNAポリメラ-
 ーゼ (Stratagene)、PrimeSTAR (商標) HS DNAポリメラーゼ
 (タカラバイオ株式会社)、Extensor Hi-Fidelity PCR酵素 (A
 Bgene)、ACCUZYME (商標) DNAポリメラーゼ (Bioline)、SA
 HARA (商標) DNAポリメラーゼ (Bioline)、VELOCITY DNAポ
 リメラーゼ (Bioline)、GeneChoice (登録商標) AccuPOL (商
 標) DNAポリメラーゼ (GeneChoice, Inc.)、GeneChoice (登
 録商標) UniPOL (商標) DNAポリメラーゼ (GeneChoice, Inc.
 .)、エロンガーゼ酵素ミックス (Invitrogen)、Pfx50 (商標) DNA
 ポリメラーゼ (Invitrogen)、Phusion DNAポリメラーゼ (New
 England Biolabs)、KOD HiFi DNAポリメラーゼ (Novagen)
 、KOD XL DNAポリメラーゼ (Novagen)、Expand 20kb
 PLUS耐熱性DNAポリメラーゼ混合物 (Roche Applied Science)
 、Expand高フィデリティPLUS耐熱性DNAポリメラーゼ混合物 (Roche
 Applied Science)、Expand高フィデリティ耐熱性DNAポリメラ-
 ーゼ混合物 (Roche Applied Science)、Expand Long
 Template耐熱性DNAポリメラーゼ混合物 (Roche Applied Sci-
 ence)、Easy-A (商標) 高フィデリティPCRクローニング酵素 (Strat-
 agene)、EXL (商標) DNAポリメラーゼ (Stratagene)、Herc-
 ulase (登録商標) 増強型DNAポリメラーゼ (Stratagene)、Herc-
 ulase (登録商標) II Fusion DNAポリメラーゼ (Stratagene)
 、Kapa LongRange (商標) DNAポリメラーゼ (Kapa Biosys-
 tems)、Kapa HiFi (商標) DNAポリメラーゼ (Kapa Biosys-
 tems)、Kapa2G (商標) ロバストDNAポリメラーゼ (Kapa Biosys-
 tems)、Kapa2G (商標) ロバストホットスタートDNAポリメラーゼ (Kap-
 a Biosystems)、Kapa2G (商標) Fast DNAポリメラーゼ (Ka-
 pa Biosystems)、Kapa2G (商標) FastホットスタートDNAポ
 リメラーゼ (Kapa Biosystems)、LA TAQ DNAポリメラーゼ (タ
 カラバイオ株式会社)、Optimase DNAポリメラーゼ (Transgenom-
 ic, Inc.)、Exo-Pfu DNAポリメラーゼ (Stratagene)、HotMaster Taq
 DNAポリメラーゼ (5 PRIME GmbH)、HotTaq DNAポリメラーゼ (Abnova Corporation)、AmpliTaq
 Gold (登録商標) DNAポリメラーゼ (Applied Biosystems)、
 Bst DNAポリメラーゼ Lg Frag (New England Biolabs)
 、MasterAmp (商標) Tfl DNAポリメラーゼ (EPICENTRE Bio-
 technologies)、Red Hot DNAポリメラーゼ (ABgene)、T
 hermoprime Plus DNAポリメラーゼ (ABgene)、Taq-red
 DNAポリメラーゼ (AppliChem GmbH)、BIO-X-ACT (商標) L
 ong DNAポリメラーゼ (Bioline)、BIO-X-ACT (商標) Short
 DNAポリメラーゼ (Bioline)、Bioline HybriPol (商標) D
 NAポリメラーゼ (Bioline)、BioTherm Taq DNAポリメラーゼ (e-
 Enzyme LLC)、EU-Taq DNAポリメラーゼ (eEnzyme LLC)、Synergy Taq
 DNAポリメラーゼ (eEnzyme LLC)、Gene
 Choice (登録商標) RedPOL (商標) DNAポリメラーゼ (GeneChoi-
 ce, Inc.)、AccuPrime (商標) GC-Rich DNAポリメラーゼ (I-
 nvitrogen)、PyroPhage (登録商標) 3173 DNA Polym

10

20

30

40

50

erase、Exo Minus (Lucigen)、9 Degrees North (改変型) DNAポリメラーゼ (New England Biolabs)、Therminator DNAポリメラーゼ (New England Biolabs)、Pwo DNAポリメラーゼ (Roche Applied Science)、Paq5000 (商標) DNAポリメラーゼ (Stratagene)、YieldAce (商標) DNAポリメラーゼ (Stratagene)、e2TAK (商標) DNAポリメラーゼ (タカラバイオ株式会社)、またはP. コダカラエンシス (P. kodakaraensis)、P. フリオサス (P. furiosus)、T. ゴルゴナリウス (T. gorgonarius)、T. ジリジイ (T. zilligii)、T. リトラリス (T. littoralis) 「Vent (商標)」、P. GB-D 「Deep Vent」、T. 9 N-7、T. アグレガンス (T. aggregans)、T. バロシイ (T. barossii)、T. フミコランス (T. fumicolans)、T. セラー (T. celer)、パイロコッカス種 (Pyrococcus sp.) ST700株、T. パシフィカス (T. pacificus)、P. アビシー (P. abyssii)、T. プロファンダス (T. profundus)、T. サイクリ (T. siculi)、T. ヒドロサーマリス (T. hydrothermalis)、サーモコッカス種 (Thermococcus sp.) GE8株、T. チオレデュセンス (T. thioereducens)、P. ホリコシイ (P. horikoshii) もしくはT. オンヌリネウス (T. onnurineus) NA1、サーモコッカス種9°N-7、サーモコッカス種GIE-J、サーモコッカス種MAR-13、サーモコッカス種GB-C、サーモコッカス種GIE-H、サーマス・アクアティカス (Thermus aquaticus)、サーマス・サーモフィラス (Thermus thermophilus)、サーマス・カルドフィルス (Thermus caldophilus)、サーマス・フィリフォルミス (Thermus filiformis)、サーマス・フラバス (Thermus flavus)、サーモトガ・マリティマ (Thermotoga maritima)、バチルス・ステアロサーモフィルス (Bacillus stearothermophilus)、もしくはバチルス・カルドテナックス (Bacillus caldotenax) 由来の天然に存在するDNAポリメラーゼを含み得る。

【0061】

いくつかの実施形態において、プライマーの少なくとも1つは、放射線学的または電磁氣的に検出可能な部分を含む。放射線学的に検出可能な部分には、ベータまたはガンマ粒子などの検出可能な粒子を放射する放射性同位元素、例えば、 ^{14}C 、 ^3H 、 ^{32}P 、 ^{33}P 、 ^{35}S 、および ^{125}I が挙げられる。電磁氣的に検出可能な部分には、検出可能な様式で電磁放射と相互作用する（光吸収、発光、または両方を含む）化学的実体、例えば、発色団およびフルオロフォア、例えば、フルオレセイン、FAM、シアニン色素、ローダミン色素などが挙げられる。例示的なフルオロフォアには、FAM (商標) (フルオレセイン)、HEX (商標)、TET (商標)、JOE (商標)、VIC (登録商標)、NED (商標)、PET (登録商標)、ROX (商標)、TAMRA (商標)、およびTexas Red (登録商標) が挙げられる。

【0062】

別の例において、TOMM40またはその断片のA/Tリッチセグメント座を特徴づけるためのリピート領域サイズ分類およびジェノタイプ再構築は、参照により組み入れられている、米国仮特許出願第62/196,239号に記載された、プライマー、ポリメラーゼ、試薬、および反応条件を含む方法を用いて作製され得る。

【0063】

II. 増幅産物のラダー

様々な実施形態において、目的の核酸（またはリピート領域を含むその核酸の部分）の増幅から生成された増幅産物（本明細書において増幅断片とも呼ばれる）は、電気泳動、好ましくはキャピラリー電気泳動に供され、リピート領域のサイズは、電気泳動により生じた増幅産物のラダーを用いて決定される。いくつかの実施形態において、増幅産物のラダ

一は、それ自体、リピート領域の長さを決定するための内部標準として用いられる。内部標準は、例えば、内部サイズ分類ラダー較正を用いて、増幅産物のラダーから計算され得る。他の実施形態において、増幅産物のラダーは、核酸のサイズを決定するための内部標準として、かつ外部標準と組み合わせて、用いられる。ある特定の実施形態において、外部標準は、外部サイズ分類ラダー較正を用いて計算され得る。追加の実施形態において、内部サイズ分類ラダー較正（増幅産物のラダー）は、外部サイズ分類ラダー較正（外部標準）と組み合わせて用いられ得る。下記で詳細に記載されているように、内部標準の適合度、外部標準の適合度、および内部標準と外部標準との間の一貫性はまた、試料品質管理に用いられ得る。

【0064】

本明細書に提供された方法において、増幅産物のラダーは、リピート領域を含む核酸の領域を増幅し、かつキャピラリー電気泳動などの高分解能断片分析方法を実施することにより、得ることができる。いくつかの実施形態において、電気泳動（例えば、キャピラリー電気泳動）は、たった1個のリピート単位だけ異なる増幅産物を識別することができる（例えば、CGGリピート領域において、たった3ヌクレオチドだけ異なる増幅産物を識別することができる）。いくつかの実施形態において、リピート単位は、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20ヌクレオチド長、またはそれ以上である。いくつかの実施形態において、リピート単位は、2ヌクレオチド長である。いくつかの実施形態において、リピート単位は、3ヌクレオチド長である。いくつかの実施形態において、リピート単位は、6ヌクレオチド長である。いくつかの実施形態において、電気泳動は、3ヌクレオチドまたはそれ未満の増幅産物を識別することができる。いくつかの実施形態において、電気泳動は、1bpの増幅産物を識別することができる。いくつかの実施形態において、リピート領域は、ホモポリマーセグメントであり、電気泳動は、1ヌクレオチド長異なる増幅産物を識別する。

【0065】

様々な実施形態において、リピートプロファイルは、電気泳動により生じた増幅産物のラダーから作成される。いくつかの実施形態において、リピートプロファイルは、サイズ分類すること、例えば、リピート領域、または患者試料において評価されている核酸における目的の任意の他の領域をサイズ分類することに用いられる。一部の場合では、リピートプロファイルは、リピートシグナルの始まりを検出することを含む。さらに他の実施形態において、リピートプロファイルは、電気泳動図における開始の長さを含む。追加の実施形態において、リピートプロファイルはリピート数を含む。

【0066】

様々な実施形態において、リピートピーク同定についてのアルゴリズムは、いくつかの段階で働く。ある特定の実施形態において、リピートシグナルの始まりは、まず、機器のサンプリング周波数に基づいてリピートシグナルが開始するウィンドウについての情報を用いて、検出される。他の実施形態において、その後、分位に基づいた分析を用いて、リピートシグナルが開始し、そして終了する範囲が決定される。いくつかの実施形態において、周波数に基づいた分析を用いて、サンプリング単位におけるリピート周期性が決定される。ある特定の実施形態において、リピート周期性を用いて、リピートピークがコールされるであろうウィンドウサイズが通知される。他の実施形態において、分位に基づいたアプローチを用いて、リピートピークがコールされるべきである閾値が導かれる。いくつかの実施形態において、スライドウィンドウが、単一のリピートピークをコールするために用いられ、各ウィンドウについてのそのコールされたピークが、その範囲内で最大の大きさをもつマイナスの二階微分係数を有すると定義される。ある特定の実施形態において、リピート周期性を用いて、リピートピークがコールされるであろうウィンドウサイズが通知される場合、ピークが見出されず、またはシグナルが閾値を下回るならば、ウィンドウの中心におけるリピートピークの位置が外挿され得る。ある特定の実施形態において、ピークがコールされる時、サンプリング単位でのリピートピーク間の差に基づいたウィンドウのサイズが調整され得る。

10

20

30

40

50

【0067】

様々な実施形態において、リピートピーク同定のためのアルゴリズムは、第1に、機器のサンプリング周波数に基づいてリピートシグナルが開始するウィンドウについての情報を用いて、リピートシグナルの始まりを検出することにより、働く。ある特定の実施形態において、ピーク同定のためのアルゴリズムは、第1に、機器のサンプリング周波数に基づいてリピートシグナルが開始するウィンドウについての情報を用いてリピートシグナルの始まりを検出し、第2に、分位に基づいた分析を用いて、リピートシグナルが開始し、そして停止する範囲を決定することにより、働く。他の実施形態において、ピーク同定のためのアルゴリズムは、第1に、リピートシグナルが機器のサンプリング周波数に基づいて開始するウィンドウについての情報を用いてリピートシグナルの始まりを検出し、第2に、分位に基づいた分析を用いて、リピートシグナルが開始し、そして停止する範囲を決定し、第3に、周波数に基づいた分析を用いて、サンプリング単位でのリピート周期性を決定することにより、働く。いくつかの実施形態において、ピーク同定のためのアルゴリズムは、第1に、機器のサンプリング周波数に基づいてリピートシグナルが開始するウィンドウについての情報を用いてリピートシグナルの始まりを検出し、第2に、分位に基づいた分析を用いて、リピートシグナルが開始し、そして停止する範囲を決定し、第3に、周波数に基づいた分析を用いて、サンプリング単位でのリピート周期性を決定し、第4に、リピート周期性を用いて、リピートピークがコールされるであろうウィンドウサイズを通知することにより、働く。ある特定の実施形態において、ピーク同定のためのアルゴリズムは、第1に、機器のサンプリング周波数に基づいてリピートシグナルが開始するウィンドウについての情報を用いてリピートシグナルの始まりを検出し、第2に、分位に基づいた分析を用いて、リピートシグナルが開始し、そして停止する範囲を決定し、第3に、周波数に基づいた分析を用いて、サンプリング単位でのリピート周期性を決定し、第4に、リピート周期性を用いて、リピートピークがコールされるであろうウィンドウサイズを通知し、第5に、分位に基づいたアプローチを用いて、リピートピークがコールされるであろう閾値を導くことにより、働く。ある特定の実施形態において、ピーク同定のためのアルゴリズムは、第1に、機器のサンプリング周波数に基づいてリピートシグナルが開始するウィンドウについての情報を用いてリピートシグナルの始まりを検出し、第2に、分位に基づいた分析を用いて、リピートシグナルが開始し、そして停止する範囲を決定し、第3に、周波数に基づいた分析を用いて、サンプリング単位でのリピート周期性を決定し、第4に、リピート周期性を用いて、リピートピークがコールされるであろうウィンドウサイズを通知し、第5に、分位に基づいたアプローチを用いて、リピートピークがコールされるであろう閾値を導き、第6に、スライドウィンドウを用いて、各ウィンドウについてのコールされるピークが、その範囲内で最大の大きさをもつマイナスの二階微分係数を有するとして定義される単一のリピートピークをコールすることにより、働く。いくつかの実施形態において、ピークが見出されず、またはシグナルが、第4段階において決定された閾値を下回るならば、ウィンドウの中心としてのリピートピークの位置が外挿され得る。他の実施形態において、ピークがコールされる時、サンプリング単位でのリピートピーク間の差に基づいたウィンドウのサイズが調整され得る。

【0068】

ある特定の実施形態において、リピートエレメント周期性が決定される。他の実施形態において、リピートエレメントサイズが、フーリエ変換分析などの周波数に基づいた分析を用いて決定される。いくつかの実施形態において、方法は、機器サンプリングドメインから塩基対リピートドメインへ変換することを含む。他の実施形態において、その情報は、増幅産物のラダーにおいて各増幅産物のサイズを決定するために用いられる。いくつかの実施形態において、ラダーにおけるリピートエレメントのヌクレオチド長を決定することから導かれる情報は、リピート領域のサイズを決定するための較正曲線を作成するために用いられる。より具体的な実施形態において、リピート領域キャピラリー電気泳動は、リピート領域をサイズ分類するために増幅産物のラダーに対して正規化される。ある特定の実施形態において、リピートプロファイル全体が、較正曲線を作成するために用いられる

10

20

30

40

50

。他の実施形態において、ピークは、例えば、ピーク高さおよび／またはピーク形状に基づいて、較正曲線用を選択される。さらに他の実施形態において、ピークは、リピートプロファイル周期性の一貫性の品質に基づいて選択される。追加の実施形態において、リピートプロファイル周期性の一貫性は、例えば、エントロピーフィルタまたは周波数フィルタとして、リピートピーク位置シグナルの変換バージョンに閾値を適用することにより、決定される。

【0069】

ある特定の実施形態において、アノテーションデバイス130は、リピートプロファイル全体を用いて、サイズ分類標準を作成するように構成され得る。様々な実施形態において、アノテーションデバイス130は、ピーク特性に基づいてサイズ分類標準を作成するためにピークを選択し得る。ピーク特性は、ピーク高さおよび／またはピーク形状であり得る。様々な実施形態において、アノテーションデバイス130は、リピートプロファイル周期性の一貫性などのリピートプロファイル周期性特徴に基づいてサイズ分類標準を作成するためにピークを選択し得る。例えば、リピートプロファイル周期性の一貫性は、アノテーションデバイス130により、ピーク間の差におけるエントロピーフィルタとして、または周波数フィルタとしてのいずれかで、リピートピーク位置シグナルの変換バージョンに閾値を適用することにより決定され得る。

【0070】

ある特定の実施形態において、リピートエレメント周期性は、リピート性モチーフまたはリピート性配列の周波数を反映し、例えば、リピートエレメント周期性は、FMR1遺伝子座についての3塩基対である。リピートエレメント周期性は、遺伝子座におけるリピート長に依存して1、2、3、4、5、もしくは6、またはそれ以上であり得る。ある特定の実施形態において、第1の増幅産物は、ラダー内の増幅産物のセットにおける最短の増幅産物の長さである。ある特定の実施形態において、増幅産物カウント数は、異なる長さの産物の数である。ある特定の実施形態において、増幅数のカウントは、期待される周期性に従って訂正することができる。ある特定の実施形態において、各鋳型についての一定のエレメント長は、プライマーおよび鋳型により決定される固定化断片である。

【0071】

さらなる実施形態において、方法は、シグナルアーチファクトについて補正することを含む。そのようなシグナルアーチファクトには、混入ピーク、欠測ピーク、気泡、閃光シグナル(flare up signal)、または他の蛍光チャネルからのシグナルの漏出を挙げることができる。

【0072】

リピート領域のサイズ分類が、様々な実施形態において、サイズ分類のために、外部標準の代わりに増幅産物のラダーの使用を含むことを、当業者は容易に認識するだろう。

【0073】

様々な実施形態において、追加のパラメータ情報は、リピート領域における任意の割り込み配列への順方向および逆方向での距離など、電気泳動の電気泳動図から得られる。いくつかの実施形態において、追加のパラメータ情報と組み合わされたリピート領域サイズは、ジェノタイプを再構築するために用いられる。いくつかの実施形態において、パラメータ情報の一部を満たす全ての可能なジェノタイプを評価して、全てのパラメータ情報を満たす解答ジェノタイプを同定する機械可読媒体を実行する装置上で再構築が行われる。いくつかの実施形態において、このジェノタイプの自動再構築のための装置、方法、および機械可読媒体は、全体として参照により組み入れられている、国際公開特許番号WO／第2014／015273に記載されたものである。

【0074】

III. 自動サイズ分類分析

様々な実施形態において、リピート領域の自動サイズ分類のための方法が提供される。いくつかの実施形態において、リピート領域の自動サイズ分類は、まず、試料においてリピートプライマーピークおよび遺伝子特異的プライマーピークを同定および定義する。ある

特定の実施形態において、局所的ウィンドウにおけるピーク形状、ピーク規模、またはピーク間の距離が考慮される。他の実施形態において、標準曲線は、リピートプライマーピークの少なくとも3つを用いて作成され得る。より具体的な実施形態において、標準曲線は、リピートプライマーピークの全部を用いて作成され得る。ある特定の実施形態において、標準曲線は、全てのリピートプライマーピークを用いて、1セットの変数についての値を決定することにより、作成される。いくつかの実施形態において、変数には、例えば、リピートエレメントサイズ、最初のピークに先行する塩基対、リピートエレメントサイズを掛けた、リピートプライマーにおけるリピートエレメントの数マイナス1、ピークカウント数、および相補的 (complementary) プライミングに用いられる一定のエレメント長が挙げられ得るが、それらに限定されない。ある特定の実施形態において、標準曲線は、以下の変数についての値を決定することにより作成される：— Z = リピートエレメントサイズ (bp)、— X = 最初のピークに先行する塩基対 ($X = [RP \text{ プライマーにおけるリピートエレメントの数} - 1] * Z$)、— N = ピークカウント数 (1から N までの合計ピーク)、および Anc = 相補的プライミングに用いられる一定のエレメント長 (bp で)。より具体的な実施形態において、標準曲線は、30のCGGリピートの正常な男性を用いて作成される。いくつかの実施形態において、30のCGGリピートの男性を用いて、以下の変数が計算される：— Z = リピートエレメントサイズ (bp で) ($AmplideX \text{ FMR1}$ について：3 bp リピート)、— X = 最初のピークに先行する塩基対、 $X = [RP \text{ プライマーにおけるリピートエレメントの数} - 1]$ ($AmplideX \text{ FMR1}$ について： $X = [5 \times CGG - 1] * 3 = 12 \text{ } bp$)、— N = ピークカウント数 (1から N までの合計ピーク) (30 CGGの正常なFXS男性試料における最後のピークについて、 $N = 26$ (26個のスタッターピークが存在する))、— Anc = 相補的プライミングに用いられる一定のエレメント長 (bp で) ($AmplideX \text{ FMR1}$ 例について、一定のエレメント長は127 bp である)。いくつかの実施形態において、標準曲線がいったん作成されたならば、自動化過程は、スタッターパターンにおける各 n ピークの正確なサイズを計算する：例えば、 N ピークについてのサイズ (bp) = $[X + ZN + Anc]$ 。ある特定の実施形態において、自動化過程は、ギャップ (bp で) + X を全体のサイズに外挿することにより、リピートプロファイルへの配列割り込みを無効にする。例えば、FMR1およびAGG割り込み配列について： $= 3 \text{ } bp + 12 \text{ } bp = 15 \text{ } bp$ 。ある特定の実施形態において、自動化過程は、計算されたサイズ (上記のような) に対して、ピーク (またはそれらのタイムスタンプ) についての観察されたCEサイズをプロットすることにより、リピートプライマーピークから較正曲線を作成し、遺伝子特異的サイズ計算のために、その導かれた線形回帰関数を用いた。

【0075】

ある特定の実施形態において、サイズ分類標準は、リピートプロファイル内の断片についてのサイズを示すために用いられ得る。他の実施形態において、サイズ分類曲線は、外挿により、リピート領域の外側の断片のサイズを示すために用いられ得る。

【0076】

ある特定の実施形態において、リピートプライミング化PCR (Repeat Primed-PCR) (RP-PCR) が、リピート領域アセスメントに用いられ得る。RP-PCRは、縮重プライマーを用いて、ジェノタイプに一致した複数のリピート断片を作製し得る。RP-PCRを用いるいくつかの実施形態は、リピートアセスメントのための方法として、サイズ分類を回避して、スタッターピークの直接的カウントを用いる。この直接的カウントは、式： r (リピート領域カウント数) $a = N + X / Z$ を使用し得る。例えば、FMR1について、 $26 + 12 / 3 = 30 \text{ } CGG$ 。

【0077】

様々な実施形態において、方法は、サイズ分類分析を行うようにプログラミングされたプロセッサを含む装置 (例えば、コンピュータ) を用いて行うことができる。いくつかの実施形態において、プロセッサは、核酸に関する情報 (パラメータ情報) を受け取り、その後、その核酸について解答ジェノタイプを再構築するようにプログラミングされる。いく

10

20

30

40

50

つかの実施形態において、パラメータ情報は、リピート領域サイズ情報である。いくつかの実施形態において、パラメータ情報は、リピート領域における任意の割り込み配列への順方向および逆方向での距離をさらに含み得る。いくつかの実施形態において、リピート領域サイズ、および適宜、任意の割り込み配列への順方向および逆方向での距離が、ジェノタイプの自動再構築のために装置により用いられる。いくつかの実施形態において、装置はまた、インプット情報および／または解答ジェノタイプを表示するためのモニターを含む。いくつかの実施形態において、解答ジェノタイプは、装置に電子的に記憶され、および／またはさらなる診断的もしくは治療的使用のために印刷することができる。

【0078】

下記の実施例においてより詳細に記載されているように、方法は、いくつかの実施形態において、FMR1遺伝子においてCGGリピート領域のジェノタイプをサイズ分類するために用いることができる。FMR1の5' UTRは、1つまたは複数のCGGリピート領域を含み得、そのリピート領域のそれぞれが、その領域内に1個または複数のAGG割り込み配列を含有し得る。2個以上のAGG割り込み配列が存在する場合、これらは、一般的に、連続して存在しない（すなわち、(AGG)_n（nは2以上である）を含むCGGリピート領域を見出すことはまれである）。

【0079】

IV. 試料

目的の核酸を含有する様々な試料が、リピート領域または核酸をサイズ分類する、開示された方法において用いることができる。様々な実施形態において、試料は、ヒトまたは非ヒト動物から得られる。例えば、試料は、患者試料であり得る。「患者試料」は、患者由来の任意の生物学的検体である。試料という用語には、血液、血清、血漿、尿、脳脊髄液、涙、唾液、リンパ液、透析液、洗浄液、精液、および／または他の液体試料などの生体液、ならびに生物学的起源の細胞および組織を含むが、それらに限定されない。細胞および組織には、口腔細胞、口内洗浄収集物、または毛包を含む皮膚細胞が挙げられ得る。その用語はまた、ヒトから単離された細胞、または培養細胞、細胞上清、および細胞可溶化物を含む、それら由来の細胞を含む。それはさらに、器官または組織の培養から引き出された液体、組織生検試料、腫瘍生検試料、糞便試料、および生理学的組織から抽出された液体、ならびに固形組織、組織切片、および細胞可溶化物から分離された細胞を含む。それはまた、脳由来のものなど、死後固形組織試料も含み得る。試料という用語はまた、目的の核酸を含む、ヒトまたは非ヒト動物から得られる任意の他の細胞または非細胞の検体を含む。いくつかの実施形態において、試料は、少なくとも約80 ng未満、100 ng未満、150 ng未満、200 ng未満、500 ng未満、1,000 ng未満、1,500 ng未満、2,000 ng未満、2,500 ng未満、3,000 ng未満、4,000 ng未満、または5,000 ng未満の、目的の核酸を含有する。

【0080】

場合によっては、試料は、1つまたは複数の目的の核酸を含む。目的の核酸は、ゲノムDNAであり得る。目的のゲノムDNAまたは他の核酸は、本発明の方法に供される前に、試料の他のDNAおよび非DNA成分から分離され得る。DNA精製および分離の多くの方法が当技術分野で知られており、開示された方法と共に用いられ得る。いくつかの実施形態において、目的の核酸は、インビトロで合成された核酸を含み得る。インビトロでの核酸合成の例には、PCRなどの増幅反応、インビトロでの転写、インビトロでの逆転写、インビトロでのプライマー延長、シーケンシング反応、ホスホラミダイトに基づいた核酸合成、およびそれらの組合せが挙げられる。

【0081】

いくつかの実施形態において、試料中の目的の核酸は、リピート領域、例えば、1つまたは複数のリピート性GCリッチセグメントを含み得る。ある特定の実施形態において、試料中の目的の核酸は、FMR1および／もしくはFMR2遺伝子、もしくはその断片、またはFMR1および／もしくはFMR2の5' UTRの少なくとも部分（例えば、FMR1の5' UTRのCGGリピートまたはFMR2の5' UTRにおけるCCGリピートを含む

10

20

30

40

50

部分)を含み得る。ある特定の実施形態において、核酸のサイズは、約50bp、100bp、200bp、300bp、500bp、もしくは700bp、または1kb、1.5kb、2kb、2.5kb、3kb、4kb、5kb、7kb、もしくは10kb、またはその間の任意の値であり得る。いくつかの実施形態において、核酸のサイズは、50bpから10kbの間、100bpから10kbの間、200bpから10kbの間、300bpから10kbの間、500bpから10kbの間、700bpから10kbの間、1kbから10kbの間、1.5bpから10kbの間、2bpから10kbの間、3bpから10kbの間、50bpから7kbの間、50bpから5kbの間、50bpから4kbの間、50bpから3kbの間、50bpから2kbの間、50bpから1.5kbの間、100bpから7kbの間、200bpから5kbの間、または300bpから4kbの間であり得る。

10

【0082】

様々な実施形態において、試料中の目的の核酸は、ホモポリマーセグメントなどの1つまたは複数のリピート性A/Tリッチセグメントを含み得る。ある特定の実施形態において、A/Tリッチセグメントは、(i)少なくとも10個のA残基、少なくとも10個のT残基、もしくは少なくとも10個のU残基を含むホモポリマーセグメントであって、少なくとも10個のA、T、もしくはU残基が、連続し、もしくは1~3個の他のヌクレオチドによって1回、中断されている、ホモポリマーセグメント、または(ii)(T_nA)_m、(A_nT)_m、(T_nA)_m、または(A_nT)_mを含むセグメントであって、nが2以上であり、mが、リピート性A/Tリッチセグメントの長さが10残基以上であるような、セグメントである。いくつかの実施形態において、核酸鑄型は、ホモポリマーセグメントなどの1つまたは複数のリピート性A/Tリッチセグメントを含むことが知られ得る。核酸鑄型は、ホモポリマーセグメントなどの1つまたは複数のリピート性A/Tリッチセグメントを含むのではないかと疑われ得る。ある特定の実施形態において、核酸のサイズは、約50bp、100bp、200bp、300bp、500bp、もしくは700bp、または1kb、1.5kb、2kb、2.5kb、3kb、4kb、5kb、7kb、もしくは10kb、またはその間の任意の値であり得る。いくつかの実施形態において、核酸のサイズは、50bpから10kbの間、100bpから10kbの間、200bpから10kbの間、300bpから10kbの間、500bpから10kbの間、700bpから10kbの間、1kbから10kbの間、1.5bpから10kbの間、2bpから10kbの間、3bpから10kbの間、50bpから7kbの間、50bpから5kbの間、50bpから4kbの間、50bpから3kbの間、50bpから2kbの間、50bpから1.5kbの間、100bpから7kbの間、200bpから5kbの間、または300bpから4kbの間であり得る。

20

30

【0083】

様々な実施形態において、多重アッセイは、2つ以上の核酸領域の並行分析に用いることができる。いくつかの実施形態において、多重PCR反応は、核酸の少なくとも1つのリピート領域をサイズ分類するために用いることができる。ある特定の実施形態において、第1および第2の核酸領域が増幅される。特定の実施形態において、第2の核酸領域が増幅される。他の実施形態において、第1、第2、および第3の核酸領域が増幅される。他の実施形態において、第2および第3の核酸領域が増幅される。さらに他の実施形態において、第2の核酸、および適宜、第3の核酸領域が増幅される。ある特定の実施形態において、核酸領域は、ラダー増幅産物についての鑄型と異なる。いくつかの実施形態において、少なくとも1つの核酸領域、少なくとも2つの核酸領域、少なくとも3つの核酸領域、少なくとも4つの核酸領域、または少なくとも5つの核酸領域が増幅される。いくつかの実施形態において、多重アッセイは、2つ以上の遺伝子座をサイズ分類するために用いることができる。いくつかの実施形態において、多重アッセイは、3つ以上の遺伝子座をサイズ分類するために用いることができる。いくつかの実施形態において、多重アッセイは、FMR1およびFMR2の少なくとも1つのリピート領域をサイズ分類するために用いることができる。他の実施形態において、多重アッセイは、FMR1およびC9ORF

40

50

72の少なくとも1つのリピート領域をサイズ分類するために用いることができる。いくつかの実施形態において、多重アッセイは、FMR2およびC9ORF72の少なくとも1つのリピート領域をサイズ分類するために用いることができる。ある特定の実施形態において、多重アッセイは、FMR2およびC9ORF72の少なくとも1つのリピート領域をサイズ分類するために用いることができる。他の実施形態において、多重アッセイは、伸長したリピート領域に関連した障害、例えば、脊髄小脳変性症、筋緊張性ジストロフィー、またはハンチントン病について用いることができる。ある特定の実施形態において、多重アッセイは、2つ以上の蛍光標識と共に用いることができる。

【0084】

V. リピート領域サイズ分類の装置および機械可読媒体

様々な実施形態において、1つまたは複数のリピート領域のサイズ分類に用いられ、および適宜、リピート領域を含有する核酸についてのジェノタイプの再構築に用いられる装置が開示される。いくつかの実施形態において、核酸リピート領域のサイズまたは特徴に関する情報が、ジェノタイプを再構築するのに用いられる装置に提供される。いくつかの実施形態において、リピート領域はまた、割り込み配列、例えば、CGGリピート領域におけるAGG割り込み配列を含む。いくつかの実施形態において、リピート領域サイズ、ならびに任意の割り込み配列への順方向および逆方向での距離を含むパラメータ情報が、ジェノタイプを再構築するのに用いられる装置に提供される。

【0085】

様々な実施形態において、核酸試料のリピート領域をサイズ分類し、適宜、例えば、再構築されるジェノタイプを作製するための装置が開示される。いくつかの実施形態において、装置は、メモリデバイスへ通信可能に連結されたプロセッサを含む。いくつかの実施形態において、プロセッサにより実行される場合、リピート領域のサイズ分類およびジェノタイプ再構築分析をプロセッサに行わせる機械実行可能な命令が、メモリデバイスに記憶される。ある特定の実施形態において、機械実行可能な命令は、(a) リピート領域を増幅すること、(b) 高分解能断片分析を実施すること、(c) 増幅産物のラダーを得ること、および(d) 増幅産物のラダーを内部標準として用いて、リピート領域の長さを決定することをプロセッサにさせる。ある特定の実施形態において、装置はさらに、プロセッサおよびメモリデバイスへ通信可能に連結されたモニターを含み、メモリデバイスに記憶された機械実行可能な命令が、解答ジェノタイプをモニター上に表示するようにプロセッサに命令する。いくつかの実施形態において、装置はさらに、プロセッサおよびメモリデバイスへ通信可能に連結されたプリンターを含み、メモリデバイスに記憶された機械実行可能な命令が、解答ジェノタイプをプリンターで印刷するようにプロセッサに命令する。

【0086】

様々な実施形態において、リピート領域をサイズ分類するために用いられる装置は、核酸に関するパラメータ情報（例えば、増幅産物のラダーに関する、リピートプロファイル、リピートエレメント周期性、第1の増幅産物、増幅産物カウント数、および／または一定のエレメント長）のインプットを受け入れることができる。いくつかの実施形態において、装置は、パラメータ情報を用いて、増幅産物のラダーにおける各増幅産物のサイズ、および／または核酸の全体のサイズを決定するようにプログラミングされる。装置は、その結果を表示し、および／またはアーカイブに入れるようにプログラミングされ得る。いくつかの実施形態において、装置は、その結果を表示し、および／またはアーカイブに入れるための手段を含む。

【0087】

様々な実施形態において、本明細書に開示された装置は、プロセッサおよびメモリデバイスを含み、メモリデバイスが、核酸に関するパラメータ情報のインプットを受け入れて、サイズ分類分析を行うようにプロセッサに命令する機械可読な命令を含有し、そのサイズ分類分析が、標準曲線を作成するために、以下の式によって表され得る：— Z = リピートエレメントサイズ (b_p で)、— X = 最初のピークに先行する塩基対、 X = [リピートプライマーにおけるリピートエレメントの数 − 1] * Z 、— N = ピークカウント数 (

10

20

30

40

50

1 から N までの合計ピーク)、 $-A_{nc}$ = 相補的プライミングに用いられる一定のエレメント長 (bp) で)。いくつかの実施形態において、スタッター (shutter) パターンにおける各 n ピークの正確なサイズを決定するために、以下の式が用いられ得る: N ピークについてのサイズ (塩基対) = $[X + ZN + A_{nc}]$ 。結果として、この式は、リピート領域のサイズを提供する。

【0088】

いくつかの実施形態において、装置はさらに、解答ジェノタイプを表示するための手段 (例えば、ジェノタイプを視覚的に表示するモニター、ジェノタイプをデジタル形式で保存するデータ記憶媒体、および/または解答ジェノタイプをプリンター、もしくは他の電子記憶もしくはディスプレイデバイスへ送信するための接続) を含む。

10

【0089】

いくつかの実施形態において、装置はコンピュータであり、コンピュータは、プロセッサ、およびそれに記憶されたコンピュータコードを有するメモリデバイスを含み、コンピュータコードは、核酸に関するパラメータ情報のインプットを受け入れ、その後、増幅産物のラダーのリピートプロファイルを決定し、それにより、リピート領域をサイズ分類するようにプロセッサに命令する。いくつかの実施形態において、コンピュータはまた、インプット情報および/または再構築されたジェノタイプを表示するためのモニターを含む。いくつかの実施形態において、再構築されたジェノタイプは、コンピュータに電子的に記憶され、および/またはさらなる診断的もしくは治療的使用のために印刷することができる。様々な実施形態において、コンピュータは、ユーザーとの対話を可能にするためのデバイスを含む。例えば、コンピュータは、ユーザー (患者、医者、または他のヘルスケアワーカー) がパラメータ情報を入力すること、および/または再構築されたジェノタイプにアクセスし、操作することを可能にするキーボードおよび/またはポインティングデバイス (例えば、マウスまたはトラックボール) を含む得る。

20

【0090】

様々な実施形態において、リピート領域のサイズ分類、および適宜、ジェノタイプの再構築を行えという命令は、装置に機械可読媒体 (例えば、機械実行可能な命令、ソフトウェア、コンピュータコード、コンピュータプログラムなど) において記憶され得る。例えば、機械可読媒体は、C++、C#、Java、Perl、Python、Julia、R、Go、Ruby、Scala、JavaScript、または任意の他のコンピュータコードに適切なフォーマットで記憶されたコンピュータコードを含み得る。機械可読媒体は、核酸に関するパラメータ情報を用いてリピート領域のサイズ分類を行うために装置へ命令を与えることができる。様々な実施形態において、機械可読媒体上の命令は、(a) リピート領域を増幅すること、(b) 高分解能断片分析を実施すること、(c) 増幅産物のラダーを得ること、および (d) 増幅産物のラダーを内部標準として用いて、リピート領域の長さを決定することを装置に命令することができる。

30

【0091】

いくつかの実施形態において、機械可読媒体上の命令は、サイズ結果をモニター上に表示するように装置に命令する。いくつかの実施形態において、機械可読媒体上の命令は、サイズ結果をプリンターで印刷するように装置に命令する。

40

【0092】

機械可読媒体に記憶された命令は、装置によって (例えば、コンピュータ内のプロセッサによって) 用いられ得る、命令、情報、および/またはデータを提供する任意のコード、記号、または他のシグナルであり得る。いくつかの実施形態において、機械可読媒体に記憶された命令は、核酸に関するパラメータ情報を受け取り、リピート領域をサイズ分類するための分析を行い、核酸のサイズを記憶または送信するように装置に命令するプログラムをコードする。

【0093】

いくつかの実施形態において、機械可読媒体に記憶された命令は、リピート領域サイズ分類分析プログラムを実行するように装置に命令する。いくつかの実施形態において、プロ

50

グラムは、核酸のサイズを表示し、および／またはアーカイブに入れるように（例えば、サイズをモニター上に表示するように、サイズをデータ記憶媒体に保存するように、および／またはサイズをプリンターもしくは他の電子的記憶もしくはディスプレイデバイスへ送信するように）との命令を含む。

【0094】

いくつかの実施形態において、機械可読媒体に記憶された命令は、モニター上にグラフィカルディスプレイを提供するユーザーインターフェイスをさらにコードする。いくつかの実施形態において、インターフェイスは、ユーザーが核酸に関するパラメータ情報を入力するのを（例えば、ユーザーがデータファイルをアップロードするのを可能にすることにより、またはユーザーがユーザーインターフェイス上に示されたディスプレイフィールドへ情報を入力するのを可能にすることにより）可能にする。いくつかの実施形態において、ユーザーインターフェイスは、インプットデータおよび／またはサイズ結果を表示し、および／または保存するための様々な方法（例えば、データをユーザーのモニター上に表示すること、データを特定化された電子デバイスまたは電子アドレスへ送ること、データを特定の場所へ印刷し、および／または保存することによる）など、パラメータ情報を分析するための選択肢をユーザーに提供する。

【0095】

様々な実施形態において、核酸サイズは、装置へ物理的に接続された記憶媒体に（例えば、コンピュータ上のハードドライブなどの内部メモリデバイスに）データとして記憶され得、および／または装置へ通信可能に接続されている（例えば、有線またはワイヤレスインターネットまたはインターネット接続などによる）リモート記憶デバイスに記憶され得る。いくつかの実施形態において、ユーザーインターフェイスは、自動的に、サイズを特定の場所に記憶させ、サイズを印刷し、および／またはサイズを特定化された電子デバイスもしくは電子アドレスへ（例えば、核酸サイズを要請していた医療専門家のEメールアドレスへ）送るための選択肢をユーザーに提供する。

【0096】

VI. 使用方法

様々な実施形態において、上記で開示された方法は、伸長したリピート領域、例えば、GCリッチな領域もしくはA/Tリッチなリピート領域を検出するために、および／または患者において障害を診断するために、もしくは患者の子孫において障害のリスクを診断するために、用いることができる。いくつかの実施形態において、方法は、例えば、（1）患者から試料を得ること、（2）CGGもしくはCCGリピートを含む領域またはリピート性A/Tリッチセグメントなどの1つまたは複数のリピート領域を有する核酸を試料から単離すること、（3）1つまたは複数のリピート領域を有する核酸の領域を増幅すること、（4）キャピラリー電気泳動を実施すること、（5）増幅産物のラダーを得ること、（6）増幅産物のラダーを内部標準として用いて、リピート領域の長さを決定することを含む、リピート領域に関連した遺伝的障害を診断し、リスクを診断し、または処置するために用いることができる。いくつかの実施形態において、方法は、リピート領域において任意の割り込み配列を検出すること、および割り込み配列への順方向および逆方向での距離を決定することをさらに含み得る。いくつかの実施形態において、リピート領域サイズ、および適宜、任意の割り込み配列への順方向および逆方向での距離は、ジェノタイプを再構築するために用いられる。いくつかの実施形態において、リピート領域サイズ、および適宜、再構築されたジェノタイプは、伸長したリピート領域を検出するために、または伸長したリピート領域、例えば、GCリッチな領域、もしくはホモポリマーセグメントなどのA/Tリッチな領域に関連した遺伝的障害を診断するために用いられる。いくつかの実施形態において、リピート領域の長さおよび／または再構築されたジェノタイプは、患者または患者の子孫において遺伝的障害のリスクを予測するために用いられる。いくつかの実施形態において、リピート領域の長さおよび／または再構築されたジェノタイプは、患者の子孫において遺伝

10

20

30

40

50

的障害のリスクを検出するために用いられる。ある特定の実施形態において、方法は、リピート領域の長さおよび／または再構築されたジェノタイプに基づいて適切な処置決定（例えば、妊娠カウンセリングおよび／または不妊治療を提供すること）を行うことを含む。いくつかの実施形態において、方法は、リピート領域の長さおよび／または再構築されたジェノタイプに基づいて遺伝的障害を有すると同定された患者に適切な処置を施すことを含む。

【0097】

例えば、方法は、伸長したFMR1またはFMR2アレルに関連した障害を診断し、および／またはそのリスクを予測し、および／またはそれに関する処置決定を行うために、FMR1またはFMR2核酸またはその断片を患者試料から単離すること、CGGまたはCCGリッチなリピート領域を増幅すること、キャピラリー電気泳動を実施して、増幅産物のラダーを得ること、および増幅産物のラダーを用いて、CGGまたはCCGリッチなリピート領域サイズを決定することを含み得る。例えば、200より多いCGGもしくはCCGリピートのサイズが、患者において脆弱X症候群もしくは脆弱X（FRAXE）精神遅滞を検出するために用いることができ、または35～45より多いCGGもしくはCCGリピートのサイズの範囲が、患者の子孫において脆弱X症候群もしくは脆弱X（FRAXE）精神遅滞のリスクを検出するために用いることができる。いくつかの実施形態において、方法はまた、CGGまたはCCGリッチなリピート領域において任意の割り込み配列への順方向および逆方向での距離を検出すること、リピートサイズおよび割り込み情報を用いてFMR1またはFMR2アレルについてのジェノタイプを再構築すること、および再構築されたジェノタイプを用いて、伸長したFMR1またはFMR2アレルに関連した障害を検出することを含む。

【0098】

多数の遺伝子およびゲノム領域は、GCリッチなまたはA/Tリッチな領域を含む、リピート領域を含み、そのリピート領域は遺伝的障害に関連しており、そのことは、それらを有望な診断的および治療的標的にさせている。したがって、様々な実施形態において、本明細書に開示されたリピート領域をサイズ分類する方法は、これらの遺伝子座について用いることができ、関連した遺伝的障害について、診断し、予知し、処置し、および／または処置決定を導くために用いることができる。いくつかの実施形態において、本明細書に開示されたリピート領域をサイズ分類する方法は、FMR1またはFMR2遺伝子を分析するために用いることができる。いくつかの実施形態において、これらの方法は、FMR1の5'UTRにおけるCGGリピート領域およびFMR2の5'UTRにおけるCCGリピート領域の長さに関連している、FXS、FRAXE、FXTAS、FXPOI、およびドーパミン応答性パーキンソニズムの診断を助けることができる。例えば、5'UTRにおける約45より多いCGGリピートを有する再構築されたFMR1ジェノタイプ、および特に、約200より多いCGGリピートを有するジェノタイプは、FXSおよび関連した障害を診断するために、加えて、患者の子孫においてその障害のリスクを診断するために用いることができる。

【0099】

さらなる実施形態において、リピート領域をサイズ分類する方法は、伸長したリピート領域の、他の障害に関連したジェノタイプを検出するために用いられ得、例えば、脊髄小脳変性症1型、脊髄小脳変性症2型、脊髄小脳変性症3型、脊髄小脳変性症6型、脊髄小脳変性症7型、脊髄小脳変性症8型、フリードライヒ運動失調症、進行性ミオクロノスτένかん、筋緊張性ジストロフィーI、筋緊張性ジストロフィーII、ハンチントン病、筋萎縮性側索硬化症、歯状核赤核淡蒼球ルイ体萎縮症、脊髄小脳変性症、筋萎縮性側索硬化症（ALS）、前頭側頭型認知症（FTD）、およびアルツハイマー病である。これらの状態に関連した遺伝子座は当技術分野において公知であり、それらには、非限定的に、SCA1、SCA2、SCA3、CACNA1A、SCA7、SCA8、X25、CSTB、C9ORF72、DMPK、ZNF9、HTT、AR、ATN1、ATXN1-3、ATXN7、ATXN10、CACNA1A、SCA8、PPP2R2B、CNBP、TB

P、およびTOMM40が挙げられる。例えば、Nat Genet. 1996 May;13(1):105-8; Nat Genet. 1996 May;13(1):109-13を参照。これらの座におけるGCリッチな領域および／またはリピート領域の過伸長および／または過剰メチル化は疾患に関連しており、本明細書に開示された方法を用いたこれらの突然変異および伸長の検出は、検出された状態について、処置の一部として、または処置を導くために用いることができる。表1は、本明細書に開示された方法と共に用いることができる遺伝子座の例、およびそれらの座におけるリピート領域と疾患ジェノタイプまたは表現型との間の関係を示す。ある特定の実施形態において、方法は、リピート領域アレル内の20より多い、30より多い、35より多い、40より多い、50より多い、100より多い、110より多い、または200より多いリピートのリピート長を検出する。

【0100】

【表1】

表1:本明細書に開示された方法と共に用いられ得る遺伝子座、およびそれらの座におけるリピート領域と疾患遺伝子型または表現型との間の関係

疾患	遺伝子	リピート数		リピート位置	リピート変異体
		正常	突然変異体		
脆弱X症候群	FMR1	(CGG)<45	(CGG)>200	5'-UTR	AGG
脆弱X(FRAXE)精神遅滞	FMR2	(CCG)<35	(CCG)>200	5'-UTR	CTG
筋緊張性ジストロフィー	DMPK	(CTG)<35	(CTG)>50	3'-UTR	CCG, CTC
脊髄小脳変性症8型	SCA8	(CTG)<40	(CTG)>110	アンチセンスRNA	CCG, CTA, CTC, CCA または CTT
フリードライヒ運動失調症	X25	(GAA)<35	(GAA)>100	イントロン1	GGA, GAG
球脊髄性筋萎縮症	AR	(CAG)<30	(CAG)>40	翻訳領域	
ハンチントン病	IT15	(CAG)<40	(CAG)>40	翻訳領域	
歯状核赤核淡蒼球ルイ体萎縮症	DRPLA	(CAG)<35	(CAG)>50	翻訳領域	
脊髄小脳変性症1型	SCA1	(CAG)<40	(CAG)>40	翻訳領域	CAT
脊髄小脳変性症2型	SCA2	(CAG)<30	(CAG)>35	翻訳領域	CAA
脊髄小脳変性症3型	SCA3	(CAG)<40	(CAG)>40	翻訳領域	
脊髄小脳変性症6型	CACNA1A	(CAG)<20	(CAG)>20	翻訳領域	
脊髄小脳変性症7型	SCA7	(CAG)<40	(CAG)>40	翻訳領域	正常アレルは割り込みをもたない
進行性ミオクロースてんかん型	CSTB	(C ₄ GC ₄ GCG (配列番号10))<3	(C ₄ GC ₄ GCG (配列番号10))>50	プロモーター	
アルツハイマー病	TOMM40				
筋萎縮性側索硬化症および前頭側頭型認知症	C9ORF72	(GGGGCC)<30	(GGGGCC)>30		
筋緊張性ジストロフィーII型	CNBP (ZNF9)	(CCTG)<26	(CCTG)>75	イントロン1	

【0101】

例えば、リピート領域をサイズ分類し、および／またはジェノタイプを再構築することは

、脊髄小脳変性症1型および2型などのSCA1またはSCA2の障害に関連したジェノタイプを検出するために用いることができ、その障害は、それらのCAGリピート領域の伸長に関連している。例えば、リピート領域をサイズ分類することは、SCA1またはSCA2遺伝子における1個または複数のCAGリピートの全長、加えて、CAGリピートにおけるCATまたはCAA割り込み配列への順方向および逆方向での距離に関する情報を提供することができる。1個または複数のCAGリピートの全長、および任意の割り込み配列への順方向または逆方向かのいずれかでの距離を用いてリピート領域をサイズ分類することは、SCA1またはSCA2遺伝子について1セットの可能性のあるジェノタイプを生じさせるために適用することができる。リピート領域をサイズ分類することは、突然変異もしくはジェノタイプを検出するために、またはSCA1もしくはSCA2関連突然変異、ジェノタイプ、もしくは障害を診断し、もしくは診断するのを助けるために、加えて、その障害を処置し、および処置決定を導くために用いることができる。

10

【0102】

他の実施形態において、リピート領域をサイズ分類し、および／またはジェノタイプを再構築する方法は、リピート性A／Tリッチセグメントに関連した障害などの伸長したリピート領域に関連した他の障害を検出するために用いられ得る。いくつかの実施形態において、その障害は、神経変性疾患である。いくつかの実施形態において、神経変性疾患はアルツハイマー病である。アルツハイマー病は、遅発型アルツハイマー病であり得る。リピート性A／Tリッチセグメントに関連した他の遺伝子座は当技術分野において公知であり、例えば、遺伝子TOMM40である。いくつかの実施形態において、アセスメントされているリピート座位はTOMM40のイントロン6の全部または一部である。いくつかの実施形態において、アセスメントされているTOMM40遺伝子のイントロン6の部分は、ポリTリピート多型(re 10524523)を含有する。

20

【0103】

自動サイズ分類分析

図2は、ゲノムリピート領域を自動的に分析するための例示的な過程を示す。この過程は、シグナル前処理工程201、サイズ分類標準を作成する工程203、および遺伝子産物のサイズ分類工程205を含み得る。当業者に認識されているように、構想された実施形態から逸脱することなく、追加の工程が実施され得、工程が除去され得、工程の順序が変わり得る。アノテーションデバイス130は、開示された実施形態と一致して、これらの工程を実施するように構成され得る。

30

【0104】

工程201において、アノテーションデバイス130は、生データを受け取って、前処理するように構成され得る。生データは、CEデバイス120から、または別のデバイスから受け取られ得る。例えば、アノテーションデバイス130は、生データを記憶デバイスから引き出すように構成され得る。生データは、CE実験の各チャネルについてのシグナル強度を記憶するデータファイルに受け入れられ得る。これらのデータファイルは、シグナル強度をJSONに基づいたフォーマットで記憶し得る。

【0105】

いくつかの実施形態において、PCRアッセイは、Genetic Analyzer機器(3130/3500/3700)のApplied Biosystemsファミリーと共に実行され得、その機器の全ては、Applied Biosystemsにより維持される独自仕様のフォーマットでデータをエクスポートする。このフォーマットは、Fragment Sequence Analysis (FSA) フォーマットと呼ばれ、一連の独自仕様に従ってコード化された、キャピラリー電気泳動(CE)実験からの蛍光データを含有する。アノテーションデバイス130は、このファイルフォーマットに直接的にアクセスするように構成され得る。例えば、アノテーションデバイス130は、プログラムアクセスおよび操作のために、そのファイルにおける情報をデコードして、jsonに基づいたフォーマットへ編成するように設計されたパーサーを用い得る。このパーサーは、Bio::Trace::ABIF (フリーソフトウェアとして認可されている

40

50

と呼ばれるperlプログラミング言語についてのオープンソースモジュールを用い得る。このパーサーは、異なるGenetic Analyzer機器（3130／3500／3700）に渡って実行された1000個より多い試料に関して確認されており、GeneMapper（FSAフォーマットにアクセスするための現行標準）を通して見られる未処理の蛍光データと正確に一致していることが示されている。パーサーのアウトプットは、CE実験の複数チャンネルについてのシグナル強度を記憶し得るデータファイルを含み得る。これらのデータファイルは、シグナル強度をJSONに基づいたフォーマットで記憶し得る。

【0106】

データファイルの少なくとも1つのチャンネルは、増幅産物のラダーに対応し得る。このチャンネルは、リピートプロファイルを含み得る。いくつかの実施形態において、データファイルの別のチャンネルは、公知のサイズを有する産物のラダーに対応し得る。例えば、データファイルのこの他のチャンネルは、ROXラダーなどの外部ラダーに対応し得る。

【0107】

いくつかの実施形態において、アノテーションデバイス130は、リピートプロファイルが存在するデータファイルにおいて、目的の領域を検出するように構成され得る。アノテーションデバイス130は、周波数に基づいた分析を用いてリピートプロファイルの周期性を動的に決定し得る。アノテーションデバイス130は、リピートプロファイルにおけるリピートピークをコールするための閾値を動的に決定し得る。閾値は、振幅閾値であり得る。アノテーションデバイス130は、スライドウィンドウを用いて、リピートプロファイルにおけるリピートピークをコールし得る。アノテーションデバイス130は、閾値より下のリピートプロファイルにおけるリピートピークを内挿し得る。この内挿は、アノテーションデバイス130により作成されたサイズ分類標準の精度を向上させ得る。アノテーションデバイス130は、リピートプロファイルにおける推定上の開始ピーク位置およびその後のピーク位置を用いて、サンプリング単位から塩基対単位へマッピングするサイズ分類標準を作成し得る。

【0108】

前処理は、パースされたデータを正規化して、異なる構成で実行された試料に渡る差を調整し得る。例えば、PCRアッセイは、データ解釈のためにCEに基づいた読み取りを用い、CE機器に体系的に存在するシグナルアーチファクトによるコンボリューションを受けやすくなり得る。パースされたデータの各チャンネルは、下流データ処理を単純化し、かつそのロバスト性を増加させるために、アノテーションデバイス130によりフィルタにかけられ得る。いくつかの態様において、ローパスフィルタまたはバンドパスフィルタが、データを平滑化するためにチャンネルの1つまたは複数に適用され得る。ローパスフィルタは、Butterworth、Savitzky-Golay、移動平均、または他の類似したフィルタであり得る。CEデバイスにおける各チャンネルについてのベースライン蛍光値が、デバイスの耐用年限を通して継続的に較正されなければならないため、各チャンネルは、実験室に渡る不適切な機器較正および／または機器構成の変動性を無効にするために、アノテーションデバイス130により正規化され得る。アノテーションデバイス130は、チャンネルからの値を減算することによりチャンネルを再較正するように構成され得る。この値は、チャンネルのシグナル強度の統計値であり得る。例えば、その値は、チャンネルのシグナル強度の10パーセンタイルであり得る。10パーセンタイルは、シグナル強度において一般的に遭遇されるシャープなマイナスのゆらぎによって影響されることなく、シグナルにおけるより低い値をロバストに表し得る。下記の方程式において、 $s(x, c)$ は、機器チャンネルcについての位置xにおけるシグナル強度を表すものとする：

$$b(c) = Q_{10}(s(x, c))$$

$$s_{norm}(x, c) = s(x, c) - b(c)$$

【0109】

アノテーションデバイス130は、シグナル前処理中の気泡または混入物質などのアーチファクトから生じるアーチファクトを除去するように構成され得る。CE実験中のキャピ

ラリー管に存在する気泡は、シグナル強度に大きなスパイクを生じ得る。これらのスパイクは、遺伝子特異的産物またはROXチャンネルサイズ分類ピークとして誤って解釈されて、間違った結果を生じ得る。しかしながら、気泡からの蛍光は、チャンネルの全部に、同じような程度の大きさで影響し、アノテーションデバイス130が気泡アーチファクトを同定し、かつ除去することを可能にする。

【0110】

第1の工程において、アノテーションデバイス130は、パースされたデータにおいて全チャンネルに渡る50 RFUを超える全てのピークの位置を見出すように構成され得る。アノテーションデバイス130は、ピークインデックスの交差が複数チャンネルに渡って起こる時、候補気泡アーチファクト位置を決定するように構成され得る。例えば、チャンネル（{FAM, HEX, NED, ROX}）におけるピークインデックスの交差は以下：
 $C = P(FAM) \cap P(HEX) \cap P(NED) \cap P(ROX)$
 のように表され得る。

【0111】

アノテーションデバイス130は、気泡アーチファクトが各候補位置に存在するかどうかを決定するように構成され得る。第1の工程として、アノテーションデバイス130は、チャンネルにおいて、可能性のある気泡位置を含むウィンドウを同定するように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、候補位置におけるシグナル強度ピークの左ショルダーおよび右ショルダー $[h_{il}, h_{ir}]$ を決定するように構成され得る。以下の方程式において、 $S(i, c)$ は、チャンネル c についての左ショルダーと右ショルダーとの間のシグナル強度の関数であり得、 i は、候補ピーク位置を表し得る。

$$S(i, c) = s([h_{il}, h_{ir}], c)$$

【0112】

第2の工程として、アノテーションデバイス130は、ウィンドウ内の複数チャンネルに渡るシグナル強度間の相関を決定するように構成され得る。この相関は、チャンネルに渡ってシグナル強度を比較する、ペアワイズ順位相関有意検定または任意の他の類似性尺度であり得る：

【0113】

【数2】

$$PC(i) = \{p_{rank}(S(i, X), S(i, Y)) \mid X, Y \in CH \circ CH\}$$

【0114】

ここで、検定される機器チャンネルのセットはCHであり、 $PC(i)$ は、候補ピーク位置 i についてのペアワイズ順位相関値のセットである。

【0115】

アノテーションデバイス130は、順位相関有意検定が、全てのペアワイズ比較に渡って有意閾値より低い有意値を生じる場合、気泡アーチファクトが候補位置に存在すると決定するように構成され得る。

【0116】

【数3】

$$B = \{i \mid i, \max PC(i) < T\}$$

【0117】

ここで、 B は、気泡アーチファクトを有する候補位置を示し、 T は、有意閾値である。 T は0.0001から0.01の間であり得、例えば、0.005の有意閾値は、独立した訓練データセットを用いて実験的に検証されている。

【0118】

アノテーションデバイス130は、気泡アーチファクトを置き換えるように構成され得る。いくつかの態様において、ウィンドウ内のチャンネルについてのシグナル強度は、アノテーションデバイス130によって、シミュレートされたノイズに置き換えられ得る。シミュレートされたノイズは、気泡を囲む領域についてのシグナル強度を用いてアノテーションデバイス130によって決定される平均および標準偏差を有する、ガウスノイズであり得る：

$$bkg(i, c) = s([h_{il} - d, h_{il}], c) \cup s([h_{ir}, h_{ir} + d], c)$$

$$\mu(i, c) = \text{平均 } bkg(i, c)$$

$$\sigma(i, c) = \text{標準偏差 } bkg(i, c)$$

$$s([h_{il}, h_{ir}], c) \sim N(\mu(i, c), \sigma(i, c))$$

【0119】

ここで、 $bkg(i, c)$ は、気泡を囲む領域についての値のセットである。この例において、領域は、左ピークショルダーおよび右ピークショルダーから d 位置単位、延長する。非限定的例として、 d は5から50の間であり得る。

【0120】

アノテーションデバイス130は、データファイルにおいてチャンネルのダイナミックレンジを延長するように構成され得る。いくつかの実施形態において、チャンネルは、最初の電磁的に検出可能な部分を検出するように構成され得る。例えば、チャンネルは、FAMチャンネルであり得る。アノテーションデバイス130は、シグナル飽和の領域に渡ってピーク形状を外挿することによりチャンネルのダイナミックレンジを延長し得る。飽和は、電磁的に検出可能な部分が機器RFUセンサーの収集限界より高い発光で蛍光を発する場合に起こり得、その結果として、ピーク形状に関する情報の損失を生じる。しかしながら、収集についての波長スペクトルは、チャンネルに渡っての漏出を許容するため、飽和した領域についてのピーク形状は、同様の波長での蛍光を捕獲するチャンネルから外挿することができる。

【0121】

第1の工程において、アノテーションデバイス130は、チャンネルにおいてウィンドウを同定するように構成され得る。ウィンドウは、チャンネルの飽和した領域を含み得る。例えば、アノテーションデバイス130は、シグナル強度が振幅閾値を超えるチャンネルの領域を決定し得る。この振幅閾値は、実験的に導かれ得、機器特異的であり得る。下記の方程式において、 $s(x, c)$ は、機器チャンネル c についてのインデックス x におけるシグナル強度を表すものとし、 L は、飽和判定基準を満たす全ての位置インデックスのセットを表すものとし、 T は、ある機器特異的閾値であるとする：

【0122】

【数4】

$$L = \{x \mid s(x, c) > T\}$$

【0123】

いくつかの態様において、 c は、FAMチャンネルを含み得る。様々な態様において、 T は、1000 RFUから40000 RFUの間であり得、 T は、飽和が起こっているRFUレベルを記載する。アノテーションデバイス130は、 L におけるインデックスでのシグナル強度を改変するように構成され得る。例えば、アノテーションデバイス130は、データファイルにおける、チャンネルについてのウィンドウ内のシグナル強度、および1つまたは複数の他のチャンネルについてのウィンドウ内のシグナル強度を用いた、組み合わされたシグナル強度を決定し得る。これらの他のチャンネルは、他の電磁的に検出可能な部分を検出するように構成され得る。例えば、これらの他のチャンネルは、NEDチャンネルまたはHEXチャンネルであり得る。シグナル強度は、線形的に、または非線形的に結合され得る。いくつかの実施形態において、組み合わされたシグナル強度を決定することは、第1

10

20

30

40

50

のチャンネルにおける計算されたピークの形状を外挿することを含む。例えば、組み合わされたシグナル強度は、チャンネルについてのウィンドウ内のシグナル強度の線形結合を含み得る。組み合わされたシグナル強度は、オフセットをさらに含み得る。非限定的例として、アノテーションデバイス 130 は、NED チャンネルからの RFU 値を FAM チャンネルへ結合するように構成され得る：

【0124】

【数5】

$$s_{ext}(x, c) = \begin{cases} s(x, FAM) + s(x, NED), & x \in L \\ s(x, FAM), & \text{そうではない場合} \end{cases}$$

10

【0125】

前述の例に示されているように、アノテーションデバイス 130 は、ウィンドウ内のチャンネルについてのシグナル強度を、組み合わされたシグナル強度に置き換えるように構成され得る。

【0126】

アノテーションデバイス 130 は、開示された実施形態と一致して、工程 203 においてサイズ分類標準を作成するように構成され得る。アノテーションデバイス 130 は、このサイズ分類標準を用いて、シグナルにおける位置単位（POP7 ゲルにおいて移動した距離と相似）から塩基対サイズへ変換するように構成され得る。サイズ分類標準は、非一過性メモリに記憶されたデータまたは命令を含み得る。下で記載されているように、アノテーションデバイス 130 は、内部サイズ分類標準および外部サイズ分類標準の少なくとも 1 つを用いて、この全部のサイズ分類標準を作成するように構成され得る。いくつかの実施形態において、アノテーションデバイス 130 は、サイズ分類標準をディスプレイ、プリンター、システム 100 の別のコンポーネント（例えば、実験室情報管理システム 140）、または別のシステムにアウトプットするように構成され得る。

20

【0127】

アノテーションデバイス 130 は、開示された実施形態と一致して、工程 205 においてジェノタイプピークを同定し、かつサイズ分類するように構成され得る。この工程において、アノテーションデバイス 130 は、アッセイにおける様々なプライマーセットの、測定されたシグナルへの異なる効果を補正するバックグラウンドモデルを作成し得る。このバックグラウンドモデルは、測定されたシグナルにおいて少なくとも 1 つの遺伝子特異的産物ピークを同定するために、アノテーションデバイス 130 によって用いられ得る。

30

【0128】

アノテーションデバイス 130 は、サイズ分類標準を用いて、リピート領域の少なくとも 1 つおよび少なくとも 1 つの遺伝子特異的産物ピークをサイズ分類するように構成され得る。アノテーションデバイス 130 は、工程 203 において作成されたサイズ分類標準を用い得る。このサイズ分類標準は、内部サイズ分類標準、外部サイズ分類、または内部サイズ分類標準と、内部サイズ分類標準または別のサイズ分類標準との両方から作成された組合せ型サイズ分類標準であり得る。例えば、アノテーションデバイス 130 はまた、記憶デバイスから引き出された、システム 100 の別のコンポーネントから受け入れられた、または別のシステムから受け入れられたサイズ分類標準を用いるように構成され得る。いくつかの実施形態において、アノテーションデバイス 130 は、少なくとも 1 つの遺伝子特異的産物ピークの指示、および／またはリピート領域のサイズを、ディスプレイ、プリンター、システム 100 の別のコンポーネント（例えば、実験室情報管理システム 140）、または別のシステムにアウトプットするように構成され得る。このアウトプットは、最初のゲノム試料を提供する患者のジェノタイプの指示を含み得る。

40

【0129】

図 3A は、外部サイズ分類標準を作成するのに用いられる例示的な外部サイズ分類ラダーを描く。上で記載されているように、アノテーションデバイス 130 は、工程 203 にお

50

いてサイズ分類標準を作成するように形成され得る。C E 実験における断片サイズ分類のための現行のゴールドスタンダードは、外部から加えられた公知のサイズの色素標識分子の使用を必要とし、その分子は、P C R アッセイに用いられる他の電磁的に検出可能な部分により生じた周波数スペクトルの外側のバンドに蛍光ピークを生じる（例えば、A m p l i d e X（登録商標）F M R 1 P C R 産物）。これらの蛍光ピークは、アッセイにより生成された標的産物とは無関係に同定され得る。いくつかの態様において、これらのピークは、F S A シグナルにおける位置（サンプリング単位での）を断片サイズ（塩基対での）に関係づける外部サイズ分類標準を作成するために、アノテーションデバイス 1 3 0 によって用いられ得る。アノテーションデバイス 1 3 0 は、R O X 蛍光ピークを自動的に同定し、かつ標識し、同時に、さもなければ結果として誤標識されたピークを生じるだろう（例えば、G e n e M a p p e r に基づいたワークフローまたは同様のソフトウェアを用いる場合）アーチファクトを検出するように構成され得る。アノテーションデバイス 1 3 0 により用いられるラベリングシステムおよび方法は、将来的なアッセイに用いられるように構想される任意のサイズ分類ラダー（例えば、R O X 1 0 0 0 または R O X 2 0 0）へ延長され得る。

【0130】

第 1 の工程において、アノテーションデバイス 1 3 0 は、データファイルにおいてチャンネルで生じたアーチファクトを同定するように構成され得る。このチャンネルは、外部から加えられた、公知のサイズの色素標識サイズ分類ラダー分子に関連づけられ得る。アーチファクトは「漏出」アーチファクトであり得る。

【0131】

P C R 産物に対応するチャンネルにおけるシグナル強度が振幅閾値を超える場合、その P C R 産物は、他のチャンネルに影響するのに十分な強度で蛍光を発している可能性がある。したがって、アノテーションデバイス 1 3 0 は、P C R 産物に対応するチャンネルにおけるシグナル強度に基づいて可能性のある漏出位置をまず、同定するように構成され得る。これらの位置は、機器特異的閾値を超えるシグナル強度を有し得る。この機器特異的閾値は、実験的に決定され得る。それは、他の方法で推定され得る。下記の方程式において、 $s(x, c)$ は、機器チャンネル c についてのインデックス x におけるシグナルを表すものとし、 $T(\text{機器})$ は、機器特異的閾値であるとし、 B は c における漏出位置インデックスのセットを表すものとする：

$$f(x, c) = \Delta_1 \cdot \text{sgn}(\Delta_1 \cdot s(x, c))$$

【0132】

【数 6】

$$B = \{x \mid f(x, c) = -2 \wedge s(x, c) > T(\text{機器})\}$$

【0133】

【数 7】

$$T(\text{機器}) = \begin{cases} 20000, & \text{機器} \\ 8000, & \text{機器} \end{cases} = \begin{cases} 35XX, 37XX \\ 31XX \end{cases}$$

【0134】

いくつかの態様において、 c は F A M チャンネルであり得る。上記で列挙された R F U 値および機器は例示であり、限定することを意図されない。類似した値が同じ機器について用いられ得、追加の値が、類似した機器について決定され得る。

【0135】

第 2 の工程において、アノテーションデバイス 1 3 0 は、第 1 のチャンネルから第 2 のチャンネルへの漏出の程度を決定するように構成され得る。いくつかの実施形態において、アノテーションデバイス 1 3 0 は、漏出位置を含むウィンドウを決定し得る。これらのウィン

ドウは、第1のチャンネルにおけるシグナル強度に基づいてアノテーションデバイス130により決定され得る。例えば、アノテーションデバイス130は、漏出位置を囲む左のピークショルダー位置と右のピークショルダー位置との間の領域としてウィンドウを定義し得る。左のピークショルダー位置および右のピークショルダー位置は、ピークの左側の領域および右側の領域においてノイズプロファイルのアセスメントし、その後、そのノイズプロファイルからのパラメータを用いて、ピークシグナルがノイズから有意に偏向する閾値を決定することにより、アノテーションデバイス130によって決定され得る。いくつかの実施形態において、ノイズプロファイルは、ガウス分布に従うと仮定され、ピークショルダーは、シグナルが平均値より上へ2標準偏差分、偏向する点として標識される。当業者により認識されているように、他の振幅閾値またはノイズ分布モデルが用いられ得る。いくつかの態様において、左のノイズプロファイルおよび右のノイズプロファイルは、独立してパラメータ化され得る。下記の方程式は、漏出位置 i についてのこの過程の非限定的例を記載する：

$$bkg_l(i) = s([i - 60, i - 30], c)$$

$$bkg_r(i) = s([i + 30, i + 60], c)$$

$$\mu_l(i) = \text{平均 } bkg_l(i), \mu_r(i) = \text{平均 } bkg_r(i)$$

$$\sigma_l(i) = \text{標準偏差 } bkg_l(i), \sigma_r(i) = \text{標準偏差 } bkg_r(i)$$

【0136】

【数8】

$$h_{ll} = \min(\{x, x \in [i - d, i] \wedge s(x, c) > \mu_l(i) + 2\sigma_l(i)\})$$

【0137】

【数9】

$$h_{lr} = \max(\{x, x \in [i, i + d] \wedge s(x, c) > \mu_r(i) + 2\sigma_r(i)\})$$

【0138】

ここで、 c は第1のチャンネルであり、FAMチャンネルまたは別のチャンネルであり得、 d は30～100サンプリング単位の間であり得る。

【0139】

第3の工程において、アノテーションデバイス130は、左のピークショルダーと右のピークショルダーとの間の領域に渡ってノイズをシミュレートするように構成され得る。このシミュレートされたノイズは、第2のチャンネルにおけるシグナルバックグラウンドを模倣するようにアノテーションデバイス130により選択されたパラメータを有するガウスであり得る。下記の方程式は、漏出位置 i についてのこの過程の非限定的例を記載する：

$$bkg(i) = s([h_{il} - d_2, h_{il}], ROX) \cup s([h_{ir}, h_{ir} + d_2], c_2)$$

$$\mu(i) = \text{平均 } bkg(i)$$

$$\sigma(i) = \text{標準偏差 } bkg(i)$$

$$s([h_{il}, h_{ir}], c_2) \sim N(\mu(i), \sigma(i))$$

【0140】

ここで、 c_2 は、第2のチャンネルであり、ROXチャンネルまたは別のチャンネルであり得、 d_2 は、5～50サンプリング単位の間であり得る。第4の工程において、アノテーションデバイス130は、他のチャンネルについてのウィンドウ内のシグナル強度を、シミュレートされたノイズに置き換えるように構成され得る。

【0141】

図3Aは、サンプリング単位の範囲に渡ってシグナル強度301(RFUでの)を含み得る、漏れ込み位置除去のアウトプットである、チャンネル300を描く。チャンネル300は、ROX PCR産物に対応し得る。示されているように、シグナル強度301は、実際

のピークおよびアーチファクト（例えば、アーチファクト 315 および 317）を含み得る。アノテーションデバイス 130 は、チャンネル 300 において、ピークが局所的ノイズ閾値を超える時、ピークを同定するように構成され得る。この同定は、偽陽性ピークのコールである可能性が高いピークアーチファクトを除去することを含み得る。

【0142】

第1の工程において、アノテーションデバイス 130 は、チャンネルにおいて、スライドウィンドウを用いて計算された振幅閾値を超える、可能性のあるピークを決定するように構成され得る。このスライドウィンドウは、アノテーションデバイス 130 により、シグナル強度 301 に渡って実行され得、250～750 サンプル単位幅の間であり得る。例えば、スライドウィンドウは、500 サンプル単位幅であり得る。アノテーションデバイス 130 は、ウィンドウ内のシグナル強度 301 の、平均および標準偏差などの統計値を決定し得る。アノテーションデバイス 130 は、その決定された統計値に基づいて、可能性のあるピークを、振幅閾値を超えるものとして決定し得る。例えば、アノテーションデバイス 130 は、平均ノイズレベルより上へ 3 標準偏差分を超えるピークを同定し得る。

【0143】

第2の工程において、アノテーションデバイス 130 は、「ショルダー」アーチファクトによって引き起こされる偽陽性ピークを同定するように構成され得る。偽陽性ピークを同定するために、アノテーションデバイス 130 は、近くのピークのうちの最大のもののみを選択し得る。例えば、アノテーションデバイス 130 は、可能性のあるピークを含む区間内でピーク高さを比較し得る。アノテーションデバイス 130 は、その可能性のあるピークが、その区間における別の可能性のあるピークより小さいと決定し得る。したがって、アノテーションデバイス 130 は、その可能性のあるピークが偽陽性ピークであると決定し得、この可能性のあるピークを、次の分析のために、そのチャンネルにおけるピークから排除し得る。その区間は、25～75 サンプル単位幅の間であり得る。

【0144】

アノテーションデバイス 130 は、開示された実施形態と一致して、外部サイズ分類標準を作成するために、チャンネルにおけるピークを断片サイズに関連づけるように構成され得る。いくつかの実施形態において、アノテーションデバイス 130 は、断片サイズとの関連づけのために、チャンネルにおいて最も可能性が高いピークを選ぶ繰り返しアプローチを用い得る。例えば、アノテーションデバイス 130 は、断片サイズの第1のセットおよび対応するピーク位置の第1のセットを用いたピーク位置と断片サイズとの間の線形関係を、チャンネルの第1の領域における同定されたピークを用いて、繰り返し、再推定し得る。このアプローチは、最初の条件を選択するために、より大きい断片サイズと関連づけられた低ノイズプロファイルを利用する。当業者により認識されているように、アノテーションデバイス 130 は、代替として、データに適用されるモデルに従って、残差を最小化する最適化ルーチンを用いて、ピークを断片サイズに関連づけるように構成され得る。この最適化ルーチンは、モデルに対する適合度を定量化する判定基準に基づいて、モデルにおいて検討されるピークを繰り返し、含め、および除去することができる。

【0145】

第1の工程において、アノテーションデバイス 130 は、チャンネルにおいて、所定の塩基対長より大きい期待断片サイズを、最も遠い（キャピラリーにおける距離で）ピークに自動的に関連づけるように構成され得る。例えば、最後のピーク位置（例えば、ピーク位置 321）は、最大の期待断片サイズ（例えば、断片サイズ 323）と自動的に関連づけられ得る。次に最も大きいピークは、次に最も大きい期待断片サイズに関連づけられ得る。所定の塩基対長は、およそ 500 塩基対であり得る。

【0146】

第2の工程において、アノテーションデバイス 130 は、塩基対の所定の範囲内の標識されたピークにモデルをフィッティングする（fit）ように構成され得る。当業者により認識されているように、塩基対によって表された値はまた、リピート数によっても表され

10

20

30

40

50

得、塩基対による値の表現は、限定的であることを意図されない。例えば、線形サイズ分類ラダーは、1次最小二乗回帰を、所定の範囲内での標識されたピークにフィッティングすることにより、アノテーションデバイス130により作成され得る。所定の範囲は、350～550塩基対の間から始まり得、650～750塩基対の間で終わり得る。モデルは、アノテーションデバイス130が、サンプリング単位を塩基対長および／またはリピート数へ変換するのを可能にし得る。

【0147】

第3の工程において、アノテーションデバイス130は、チャンネル300の第1の領域310における同定されたピークを用いて、このモデルを繰り返し、再推定するように構成され得る。いくつかの実施形態において、サンプリング単位と塩基対との間の関係は、第1の領域310に渡って線形であり得る。チャンネルのこの領域は、塩基対の所定の数より少ないものに対応するピークを含み得る。アノテーションデバイス130は、断片サイズの第1のセットおよび対応するピーク位置の第1のセットを用いた、ピーク位置と断片サイズとの間の線形関係を繰り返し、再推定するように構成され得る。アノテーションデバイス130は、より大きい断片サイズからより小さい断片サイズへ進み、線形関係を再推定するために用いられる断片サイズのセットおよび対応するピーク位置のセットへ、漸減する断片サイズを含め得る。例えば、断片サイズ313が次に最も大きい断片サイズである時、アノテーションデバイス130は、断片サイズ313および対応するピーク位置311を断片サイズの第1のセットおよび対応するピーク位置の第1のセットへ加えるように構成され得る。その後、アノテーションデバイス130は、第1の領域310における、断片サイズ313とより大きい断片サイズ、および第1の領域310における、ピーク位置311とより大きいピーク位置を用いて、ピーク位置と断片サイズとの間の線形関係を再推定し得る。

【0148】

各繰り返しにおいて、いくつかの実施形態では、アノテーションデバイス130は、最新の線形関係を用いて、次の断片サイズに対応するピーク位置を決定するように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、次の断片サイズおよび最新の線形関係を用いて、予想されるピーク位置を決定し得る。この次の断片サイズは、漸減する断片サイズのうちの1つであり得、最新の線形関係は、上記で論じられた再推定された線形関係の1つであり得る。その後、アノテーションデバイス130は、その予想されるピーク位置を含むウィンドウを決定し得る。このウィンドウは、予想されるピーク位置の両側に5～50サンプリング単位の間を含み得る。アノテーションデバイス130は、予想されるピーク位置を含むウィンドウ内にチャンネルにおける実際のピーク位置を決定し得る。この実際のピーク位置は、シグナル強度301の微分係数の`sign`の微分係数を用いて、上記のように決定され得る。アノテーションデバイス130は、その実際のピーク位置を、対応するピーク位置の第1のセットへ含むように構成され得る。このように、実際のピーク位置が変化するため、最新の線形関係が適切にアップデートされ得る。さらに、この方法は、データにおいてアーチファクトを同定するための追加の方法を提供し得る。例えば、アーチファクト317は、ウィンドウ（例えば、ウィンドウ319）内になく、そのため、アノテーションデバイス130が、線形関係を再推定する時、このアーチファクト317をスキップし得る。したがって、線形関係の推定は向上し得る。

【0149】

このように、アノテーションデバイス130は、ゲルのよりノイズが多い領域においてピーク関連づけの精度を増加させるように、新しいデータ点との線形関係を継続してアップデートし得る。この繰り返しアプローチは、不適切なサイズ分類ラダーパラメータ化に寄与し得るROXチャンネルにおけるシグナルアーチファクトを無視することにおいて、現行の方法（GeneMapper、GeneMarker）より特異的であることが示されており、プライマードイマーピークをROX断片ピークとして取り違えることに対してロバストでもある。

【0150】

アノテーションデバイス130は、第2の領域320について、ピーク位置と断片サイズとの間の非線形関係を決定するように構成され得る。いくつかの実施形態において、この非線形関係は、1次、2次、もしくは3次スプラインモデルなどのスプラインモデル、または2次もしくは3次多項式モデルを含み得る。アノテーションデバイス130は、断片サイズ323およびピーク位置321などの、第2の領域320における断片サイズのセットおよび対応する同定されたピーク位置のセットを用いて、非線形関係を決定し得る。いくつかの態様において、図3Aに示されているように、第1の領域310および第2の領域320は重複し得ない。例えば、第2の領域320のより低い境界は、第1の領域310のより高い境界に等しくあり得る。第2の領域320は、650〜750塩基対より大きい断片を含み得る。

10

【0151】

図3Bに示されているように、アノテーションデバイス130は、第1の領域330についての線形関係331と第2の領域340についての非線形関係341を組み合わせることにより、外部サイズ分類標準350を作成するように構成され得る。外部サイズ分類標準350を作成するために、アノテーションデバイス130は、線形関係331および非線形関係341を再サンプリングすることにより追加の点を生じさせ得る。外部サイズ分類標準350を作成するために、一変量スプラインモデルが、これらの追加の点にフィッティングされる。いくつかの実施形態において、追加の点は、外部サイズ分類標準に沿って均一な間隔で、例えば、一定数の塩基対だけ異なって、置かれ得る。追加の点の数は、断片サイズの最初の数より実質的に大きくあり得る。例えば、断片サイズの最初の数より2〜10倍多い追加の点が作成され得る。例えば、40〜200個の間の追加の点が用いられ得る。

20

【0152】

いくつかの代替の実施形態において、アノテーションデバイス130は、最適性基準を満たすサイズ分類標準を推定するように構成され得る。いくつかの態様において、アノテーションデバイス130は、データに適用されるモデルに従って、残差を最小化する最適化ルーチンを用いて、ピークを断片サイズに関連づけ得る。この最適化ルーチンは、モデルに対する適合度を定量化する判定基準に基づいて、モデルにおいて検討されるピークを繰り返し、含め、および除去することができる。

【0153】

例えば、アノテーションデバイス130は、チャンネルにおいて、可能性のあるピークの2つ以上のサブセットを作成するように構成され得る。この作成は、決定論的であり得、または少なくとも部分的にランダムであり得る。非限定的例として、各サブセットは、第1のピークを含み得る。アノテーションデバイス130は、各サブセットについてサイズ分類標準を決定するように構成され得る。これらのサイズ分類標準は、線形関係、あるいはスプラインモデルまたは2次もしくは3次多項式モデルなどの非線形関係を含み得る。いくつかの態様において、アノテーションデバイス130は、比較点を作成するために、サイズ分類標準を再サンプリングするように構成され得る。アノテーションデバイス130は、再サンプリングされたサイズ分類標準と参照モデルとの間の比較に基づいて、再サンプリングされたサイズ分類標準の少なくとも一部について費用関数を計算するように構成され得る。参照モデルは、期待されるサイズ分類標準を含み得る。当業者により認識されているように、アーチファクトを含むサブセットを用いて作成されたサイズ分類標準は、参照モデルと大いに異なり得る。これらのモデルについての費用関数の値は、ほとんどまたは全くアーチファクトを含まない対象についての費用関数の値より大きい可能性が高い。費用関数は、L1ノルム、L2ノルム、または当業者に公知の他の費用関数を含み得る。したがって、アノテーションデバイス130は、リピート領域の少なくとも1つおよび少なくとも1つの遺伝子特異的ピークをサイズ分類する場合、費用関数を最小化するサイズ分類標準を用いるように構成され得る。この方法は、有利には、可能性のあるピークにおいてアーチファクトの同定を必要とせず、そのようなアーチファクトが、より高い費用をもたらし得るからである。当業者により認識されているように、この方法は、内部サイ

30

40

50

ズ分類標準または外部サイズ分類標準を推定するために用いられ得る。

【0154】

図4Aは、開示された実施形態と一致した、データファイルの例示的なチャンネルを描く。場合によっては、上記の外部サイズ分類標準は、外部から加えられた公知のサイズの色素標識PCR産物とゲノム領域のPCR断片との間の組成の違いのために、不正確である可能性がある。例えば、キャピラリー電気泳動におけるROX断片移動度は、FMR1断片移動度と異なり、ROX断片のヌクレオチドのバランスがとれた性質と比較して、FMR1断片のGCリッチな性質による。これらの不正確のために、アノテーションデバイス130は、内部サイズ分類標準を作成するように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、内部サイズ分類標準および外部サイズ分類標準を用いて移動度が補正されたサイズ分類標準を作成するように構成され得る。外部サイズ分類標準は、図3Aおよび3Bに関して上で記載されているように、ROXチャンネルから導かれ得る。アノテーションデバイス130による内部サイズ分類標準の作成は、データファイルのチャンネルにおいてリピートプロファイルを同定し、リピートピーク位置とリピート断片サイズとの間の線形関係を推定することを含み得る。

10

【0155】

図4Aに示されているように、データファイルのチャンネルは、リピートプロファイル410を含み得る。リピートプロファイルは、最小の検出されたPCR断片ピークから始まり、最大の検出されたPCR断片ピークまたは遺伝子産物ピークで終わるチャンネルの部分を含み得る。例えば、リピートプロファイル410は、2000サンプリング単位後から始まり得、4500～5000サンプリング単位の間で終わり得る。図4Aに描かれているように、リピートプロファイル410は、漸増的に大きくなる断片に対応するピークのリピートの連続を表示し得る。ゲノム試料に依存して、リピートプロファイル410はまた、4500サンプリング単位あたりのピークなど、1つまたは複数の遺伝子産物ピークを表示し得る。

20

【0156】

アノテーションデバイス130は、データファイルのチャンネルにおいてリピートプロファイル410を同定するように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、ROXラダーなどの外部サイズ分類標準を用いて、リピートプロファイル410の始まりの位置を予想し、近似し得る。アノテーションデバイス130が、品質管理基準（下記）を満たす外部サイズ分類標準を作成することができない場合、または内部モデルのみを用いる場合、アノテーションデバイス130は、以下の過程を用いて、リピートプロファイル410の始まりを決定し得る。

30

【0157】

第1の工程において、アノテーションデバイス130は、リピートプロファイルのおよその開始位置を決定するように構成され得る。アノテーションデバイス130は、第1のウィンドウサイズWを用いてチャンネルの総和変換（summation transformation）を実施し得る：

【0158】

【数10】

$$t(i) = \sum_{i_x}^{i_x+W} s(i_x, c), i_x = (Wx | x = 1, 2, 3 \dots)$$

40

【0159】

このように、変換されたデータ、 $t(i)$ は、重複していない第1ウィンドウとして、チャンネルについての第1のウィンドウ内のシグナル強度の和を含み得る。当業者により認識されているように、アノテーションデバイス130は、追加として、または代替として、チャンネルをローパスフィルタにかけ得る。チャンネルcは、FAMチャンネルであり得、第1のウィンドウサイズWは50～1000サンプリング単位の間であり得る。変換されたシグナルにおける最大のピーク（プライマーダイマー増幅事象により引き起こされる可能性

50

がある) について、アノテーションデバイス 130 は、上記で提供されたピークショルダーの議論に従って、少なくとも最右のピークショルダーを見出し得る。

【0160】

t(i) を用いて、最右のピークショルダーの位置を決定した後、アノテーションデバイス 130 は、チャンネルの第 2 のウィンドウ内のシグナル強度を周波数ドメインへ変換し、第 2 ウィンドウ内のシグナル強度の優位周波数が周波数基準を満たす時を決定するように構成され得る。例えば、アノテーションデバイス 130 は、この位置から始まる、第 2 のウィンドウ内のチャンネルの優位周波数を計算し得る。第 2 のウィンドウは、100～200 サンプル単位の間であり得る。アノテーションデバイス 130 は、シグナルの優位周波数と所定の周波数との間の差が、実験的に導かれた差の基準を満たす、最初の第 2 のウィンドウとしてリピートプロファイルのおよその始まりを決定し得る。

10

【0161】

第 2 の工程において、アノテーションデバイス 130 は、リピートプロファイルの厳密な開始位置を決定するように構成され得る。アノテーションデバイス 130 は、第 3 のウィンドウ内のシグナル強度の統計的尺度を用いて、この厳密な開始位置を決定し得る。例えば、正確なリピート開始部位は、第 3 のウィンドウ内のシグナル強度の所定のパーセンタイルより大きい最初の位置として、アノテーションデバイス 130 により決定され得る。いくつかの実施形態において、第 3 のウィンドウは、およその開始位置で始まり得る。下記の方程式において、a はシグナル開始部位のおよその位置を表すものとし、c はチャンネルとする：

20

$w = [a, a + 1000]$

【0162】

【数 11】

開始点 = $\min(\{x, x \in w \wedge s(x, c) > Q_{85}(s(w, c))\})$

【0163】

この例において、統計的尺度は、85 パーセンタイルであるが、そのパーセンタイルは、例えば、70～99 パーセンタイルの間であり得る。同様に、第 3 のウィンドウの幅は 1000 サンプル単位であるが、第 3 のウィンドウは 50～5000 サンプル単位幅の間であり得る。平均および標準偏差などの他の統計的尺度から導かれた振幅閾値もまた用いられ得る。チャンネル c は F A M チャンネルであり得る。

30

【0164】

第 2 の工程において、アノテーションデバイス 130 は、リピートプロファイルの終了位置を決定するように構成され得る。いくつかの実施形態において、アノテーションデバイス 130 は、リピートプロファイルの終了位置の終点 (end) を決定するためにチャンネルをフィルタにかけ得る。例えば、アノテーションデバイス 130 は、第 4 のウィンドウを用いてチャンネルに渡ってパーセンタイルフィルタを適用し得る。その変換が適用された後、シグナル終了位置は、振幅閾値を超える最後の変換された領域として、アノテーションデバイス 130 により選択され得る：

40

$t(i) = Q_{90}(s([i_x, i_x + 100], c))$, $i_x = (100x \mid x = 1, 2, 3 \dots)$

【0165】

【数 12】

終点 = $100 * \max(\{i, t(i) > 100\})$

【0166】

このように、変換されたデータ、t(i) は、第 4 のウィンドウに渡る、チャンネル c のパーセンタイルフィルタにかけられた値を含み得る。この例において、パーセンタイルは、

50

90パーセンタイルであるが、パーセンタイル値は、70パーセンタイルから99パーセンタイルまでの範囲であり得る。同様に、第4のウィンドウの幅は100サンプリング単位であるが、第3のウィンドウは50～5000サンプリング単位幅の間であり得る。平均および標準偏差などの他の統計的尺度から導かれた振幅閾値もまた用いられ得る。チャンネルcはFAMチャンネルであり得る。ここで、終点は、リピートプロファイル410における最終値のインデックスであり得る。振幅閾値は、機器特異的であり得、実験的に導かれ得る。ここで、その値は100RFUであるが、この値は、限定的であることを意図されない。

【0167】

図4Bは、開示された実施形態と一致した、リピートプロファイルを用いた内部サイズ分類標準の作成を描く。アノテーションデバイス130は、リピートプライマーに起因する、チャンネルにおける振幅ピークを同定するように構成され得る。アノテーションデバイス130は、これらのピークを、期待される断片サイズに関連づけて、内部サイズ分類標準を作成し得る。いくつかの態様において、下記でより詳細に論じられているように、アノテーションデバイス130は、リピートプロファイルの周期性から導かれたウィンドウを用いてリピートピークを繰り返し、コールし得る。アノテーションデバイス130は、周期性がリピートプロファイルにおいてシフトするにつれて、このウィンドウを調整し得、リピートピークが抑圧されている所に（例えば、AGG割り込み部位において）ピーク位置を内挿し得る。

【0168】

第1の工程において、シグナルの開始位置および終了位置が同定された後、アノテーションデバイス130は、最初の区間値および最初の振幅閾値を決定するように構成され得る。上で記載されているように、アノテーションデバイス130は、周波数に基づいた分析を用いてリピートプロファイルの周期性を動的に決定し得る。例えば、アノテーションデバイス130は、リピートプロファイルの最初の部分にフーリエ変換を実施して、リピートプロファイルの優位周波数を同定し得る。最初の部分は、開始位置で始まり得る。アノテーションデバイス130は、優位周波数の逆数を用いて、予想されるピーク位置を決定するための最初の区間値を計算し得る：

【0169】

【数13】

$$f_{rp} = \mathcal{F}\{s([\text{開始点}, \text{開始点} + 1000], c)\}$$

【0170】

【数14】

$$d_{rp} = \frac{1}{f_{rp}}$$

【0171】

ここで、 f_{rp} は、最初の部分に渡るチャンネルcの優位周波数である。この例において、最初の部分は、1000サンプリング単位幅であるが、最初の部分は、500～5000サンプリング単位幅の間であり得る。チャンネルcはFAMチャンネルであり得る。

【0172】

アノテーションデバイス130は、リピートプロファイルにおいて、リピートピークをコールするための閾値を動的に決定し得る。その閾値は、振幅閾値であり得る。例えば、アノテーションデバイス130は、リピートプロファイルの最初の部分に渡って計算された統計的尺度を用いて、リピート位置を同定するための最初の振幅閾値を決定するように構成され得る。最初の部分は開始位置で始まり得、統計的尺度はパーセンタイルであり得る：
 $t_{rp} = Q_{25}(s([\text{開始点}, \text{開始点} + 2000]), c)$

【0173】

ここで、 t_{rp} は、最初の振幅閾値である。この例において、25パーセンタイルが統計的尺度であるが、パーセンタイルは5パーセンタイル～50パーセンタイルの間の範囲であり得る。同様に、最初の部分は、2000サンプリング単位幅であるが、100～5000単位幅であり得る。チャンネルcはFAMチャンネルであり得る。アノテーションデバイス130は、平均および標準偏差などの他の統計的尺度を用いて、最初の振幅閾値を決定し得る。

【0174】

第3の工程において、アノテーションデバイス130は、リピートピーク位置のセットを繰り返し、作成するように構成され得る。様々な態様において、アノテーションデバイス130は、繰り返しのごとに、リピートプロファイル410において、予想されるピーク位置425を決定し得る。予想されるピーク位置425は、前のピーク位置421および区間値423に依存し得る。いくつかの態様において、アノテーションデバイス130は、スライドウィンドウを用いて、リピートプロファイルにおいてリピートピークをコールし得る。いくつかの実施形態において、区間値423は、最初の区間値から導かれ得る。アノテーションデバイス130は、予想されるピーク位置425を含むウィンドウ427内でリピートピーク位置を同定し得る。いくつかの実施形態において、ピークを選択するためのウィンドウは、区間値423の25～100%であり得る。ある特定の実施形態において、ウィンドウは、区間値423の50～100%の間、または80～100%の間である。アノテーションデバイス130は、ウィンドウ427内の最大ピークの位置を決定するように構成され得る。この同定されたピーク位置におけるシグナル強度が振幅閾値429を超える場合、この実際のピーク431が、リピートピーク位置のセットに加えられる。いくつかの実施形態において、振幅閾値429は、最初の振幅閾値から導かれ得る。次の繰り返しにおいて、前のピーク位置が、この繰り返し中に同定された実際のピーク431であり得る。

【0175】

いくつかの実施形態において、アノテーションデバイス130は、振幅閾値429および区間値423の1つまたは複数をアップデートするように構成され得る。例えば、振幅閾値429は、2つ以上の前の繰り返しについての実際のピークにおけるシグナル強度の平均であり得る。同様に、区間値423は、2つ以上の前の繰り返しについての、実際のピークと前のピークとの間の位置との差の平均であり得る。いくつかの実施形態において、これらの平均は、3～50個の以前の繰り返しに関するものであり得る。このように、アノテーションデバイス130は、リピートプロファイルの経過に渡ってのリピートピークの周期性および振幅におけるシフトに適応し得る。

【0176】

【数15】

$$x_{next} = \operatorname{argmax}_i \left(s \left(\left[i + \frac{d_{rp}}{2}, i + \frac{3}{2} d_{rp} \right] \right) \right)$$

【0177】

【数16】

$$p_{next} = \begin{cases} x_{next}, & s(x_{next}, c) > t_{rp} \\ i + d_{rp}, & \text{そうではない場合} \end{cases}$$

【0178】

ここで、 x_{next} は、そのウィンドウ内の最大のピークの位置であり、このピークは、チャンネルc（FAMチャンネルであり得る）についてのシグナル強度が振幅閾値を超える場合、リピートピーク位置のセットに加えられる。いくつかの実施形態において、アノテ

ションデバイス 130 は、閾値より下のリピートプロファイルにおけるリピートピークを内挿し得る。例えば、シグナル強度が振幅閾値を超えない時、予想されるピーク位置がリピートピーク位置のセットに加えられる。このように、アノテーションデバイス 130 は、リピートピーク振幅が縮小する領域においてピークを内挿し得る。この内挿は、アノテーションデバイス 130 により作成されるサイズ分類標準の精度を向上させ得る。

【0179】

第 4 の工程において、アノテーションデバイス 130 は、リピートピーク位置とリピート断片サイズとの間の関係を記載する内部サイズ分類標準を作成するように構成され得る。いくつかの実施形態において、アノテーションデバイス 130 は、リピートプロファイルにおける推定される開始ピーク位置および次のピーク位置を用いて、サンプリング単位から塩基対単位へマッピングする較正曲線（すなわち、サイズ分類標準）を作成し得る。例えば、アノテーションデバイス 130 は、リピートピーク位置のセットおよび対応する断片サイズのセットを用いて内部サイズ分類標準を作成し得る。アノテーションデバイス 130 は、最初のリピートピーク位置を最小の断片サイズと関連づけ、2 番目のリピートピーク位置を次に最も小さい断片サイズと関連づけるなどにより、リピートピーク位置と断片サイズとの間の対応を確立し得る。いくつかの実施形態において、各追加のリピートピーク位置は、前の断片サイズより大きい 1 つの追加のリピートである断片サイズと関連づけられ得る。アノテーションデバイス 130 は、対応する断片サイズのセットをリピートピーク位置のセットに対して回帰推定をすることにより関係を作成し得る。

【0180】

図 2 に関して上で記載されているように、アノテーションデバイスは、内部サイズ分類標準、外部サイズ分類モード、または組み合わされたサイズ分類標準を用いるように構成され得る。例えば、アノテーションデバイス 130 は、内部サイズ分類標準および外部サイズ分類標準を用いて、移動度補正済みサイズ分類標準（mobility corrected sizing standard）を作成するように構成され得る。アノテーションデバイス 130 は、内部サイズ分類標準および外部サイズ分類標準を用いるアフィン変換を作成することにより、移動度補正済みサイズ分類標準を作成し得る。このアフィン変換は、外部サイズ分類ラダーの線形成分と非線形成分の両方が移動度補正済みサイズ分類標準に寄与することを保証し得る。アフィン変換は、以下の方程式により記載され得る：
 $L_{rp}(x)$ は内部サイズ分類標準を表すものとし、 $L_{ROX}(x)$ は外部モデルを表すものとし、 $L_{NL}(x)$ は、外部モデルの非線形領域についての一変量スプラインモデルを表すものとする：

$$L_{rp}(x) = m_{rp}x + b_{rp}$$

【0181】

【数 17】

$$L_{ROX}(x) = \begin{cases} m_{ROX}x + b_{ROX}, & x \in \text{線形範囲} \\ L_{NL}(x), & \text{そうではない場合} \end{cases}$$

【0182】

【数 18】

$$L_{final}(x) = L_{ROX}(x) \left(\frac{m_{rp}}{m_{ROX}} \right) + b_{rp} - b_{rox} \left(\frac{m_{rp}}{m_{ROX}} \right)$$

【0183】

この例において、移動度補正済みサイズ分類標準の感度は、外部サイズ分類標準 $L_{ROX}(x)$ の感度と内部サイズ分類標準 $L_{rp}(x)$ の感度の比に依存し、一方、移動度補正済みサイズ分類標準のオフセットは、内部サイズ分類標準のオフセットおよび外部サイズ分類標準のオフセット、ならびに外部サイズ分類標準の感度と内部サイズ分類標準の感度

の比に依存する。アノテーションデバイス130は、アフィン変換を外部サイズ分類標準に適用して、移動度補正済みサイズ分類標準を得るように構成され得る。例えば、アノテーションデバイス130は、上記の方程式を用いて、移動度補正済みサイズ分類標準を計算し得る。

【0184】

図11に関して上で記載されているように、アノテーションデバイス130は、ジェノタイプピークを同定し、かつサイズ分類するように構成され得る。チャンネルは、リピートセグメントと遺伝子特異的増幅の両方を呈示し得、それゆえに、アノテーションデバイス130は、ジェノタイプピークを同定する前に、シグナル強度のこれらの2つの成分を分離し得る。追加として、アノテーションデバイス130はまた、下でより詳細に記載されているように、異常なジェノタイプピークを同定するように構成され得る。

10

【0185】

第1の工程において、アノテーションデバイス130は、リピート増幅事象のシグナル寄与を遺伝子特異的増幅事象のシグナル寄与から分離するための、バックグラウンドモデルを作成するように構成され得る。バックグラウンドモデルの生成は、FMR1リピート領域におけるAGG割り込み配列から生じるギャップなどの、リピート領域における割り込みにより生じるリピートプロファイルにおけるギャップに対処し得る。バックグラウンドモデルの生成はまた、リピートプロファイルのリピートピーク成分から偏向するが、デコンボリューションするための周波数に基づいたフィルタリングアプローチを可能にするだろう特性をもたない、遺伝子特異的産物ピークに対処し得る。

20

【0186】

図5は、バックグラウンドモデル520および動的閾値530を用いてアノテートされた例示的なリピートプロファイル510を描く。いくつかの実施形態において、バックグラウンドモデルは、所定のウィンドウ内のリピートプロファイルの大きさに依存し得る。リピートプロファイルにおけるギャップがリピート領域における配列割り込み570から生じる場合、バックグラウンドモデルは、その割り込み配列に近位の局所的リピートピークの大きさに依存し得る。遺伝子特異的産物ピーク540、伸長ピーク550、およびモザイク事象(mosaicism)ピーク560の間のバックグラウンドモデルもまた、割り込み配列に近位の局所的リピートピークの大きさに依存し得る。

【0187】

30

第1の工程において、アノテーションデバイス130は、リピートプロファイル510をフィルタにかけることによりバックグラウンドモデル520を作成するように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、インプットデータのスライドウィンドウの少なくとも1つの統計的尺度に基づいて値をアウトプットするように構成されたフィルタを用いて、リピートプロファイル510をフィルタにかけ得る。例えば、フィルタは、ウィンドウ内のリピートプロファイル510の中央値の和および四分位範囲をアウトプットし得る。ウィンドウは、3〜30リピートの幅の間、例えば、11リピートの幅であり得る。このデザインは、アノテーションデバイス130が、AGG割り込み配列および遺伝子特異的産物により引き起こされるリピートシグナルにおける大きなゆらぎを拒否すると同時に、ウィンドウにおいてリピートピークの高さを捕捉することを可能にし得る。当業者により認識されているように、中央値平滑化フィルタおよび線形フィルタ(例えば、Butterworthフィルタ)などの他のフィルタ型が代わりに用いられ得る。

40

【0188】

第2の工程において、アノテーションデバイス130は、バックグラウンドモデルの高周波成分を軽減するように構成されたフィルタを用いて、バックグラウンドモデル520をさらにフィルタにかけるように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、Savitzky-Golayフィルタを用いて、バックグラウンドモデル520の高周波成分を軽減し得る。Savitzky-Golayフィルタは、3〜30の幅の間、例えば、7リピートの幅であり得る。Savitzky-Golay

50

y フィルタは、リピートプロファイル動力学と調和するように整えられ、他のフィルタデザインのアウトプットで目に見えるピークおよびトラフを妨げ得る。

【0189】

第3の工程において、アノテーションデバイス130は、上記の方法に従って、生じたバックグラウンドモデル520においてピークショルダーを決定するように構成され得る。いくつかの実施形態において、アノテーションデバイス130は、ピークショルダー内のシグナル強度を、ピークショルダー値の間に線形的に内挿された値に置き換え得る。

【0190】

アノテーションデバイス130は、バックグラウンドモデルから動的閾値を生じるように構成され得る。動的閾値は、遺伝子特異的産物ピークを同定するためにアノテーションデバイス130により用いられ得る。動的閾値を生じるようにバックグラウンドモデルを動的にスケールリングすることにより、アノテーションデバイス130は、より低い断片サイズ範囲において特異性を増加させ得、一方、より高い断片サイズ範囲において感度を増加させ得る。いくつかの実施形態において、アノテーションデバイス130は、1番目の断片サイズより上で、かつ2番目の断片サイズより下のサイズを有する、増幅産物に対応するバックグラウンドモデルの第1の領域を決定するように構成され得る。アノテーションデバイス130はまた、2番目の断片サイズより上のサイズを有する、増幅産物に対応するバックグラウンドモデルの第2の領域を決定し得る。アノテーションデバイス130は、最初のスケールリング因子からその最初のスケールリング因子より低い第2のスケールリング因子までを変動する第1のスケールリング因子を、バックグラウンドモデルの第1の領域に掛け得る。例えば、第1のスケールリング因子は、最初のスケールリング因子から第2のスケールリング因子までを線形的に変動し得る。アノテーションデバイス130はまた、バックグラウンドモデルの第2の領域に第2のスケールリング因子を掛け得る。

【0191】

下記の方程式において、 M_{BG} はバックグラウンドモデル520を表すものとし、 M_G は動的閾値を表すものとし、 r_1 は第1の断片サイズに対応するインデックスを表すものとし、 r_2 は第2の断片サイズに対応するインデックスを表すものとする：

【0192】

【数19】

$$M_G(x) = \begin{cases} 3 M_{BG}, & x < r_0 \\ M_{BG} \left(\frac{3(x - r_1) + 1.5(r_2 - x)}{r_2 - r_1} \right), & r_1 < x < r_2 \\ 1.5 M_{BG}, & x > r_2 \end{cases}$$

【0193】

この例において、第1のスケールリング因子は3であり、第2のスケールリング因子は1.5であるが、これらの値は限定的であることを意図されない。第1のスケールリング因子は、1.25～10の間を変動し得、第2のスケールリング因子は1.25～10の間を変動し得る。いくつかの態様において、第1の断片サイズは、0リピート、または多くも20リピートに対応し得る。様々な態様において、第1の断片サイズは、70～190リピートの間、例えば、120リピートに対応し得る。この例において、アノテーションデバイス130は、0～120リピートの間の領域において、3から1.5へ減少する区分的スケールリング因子をバックグラウンドモデル520に適用し得、その後、120リピート後、1.5での一定のままにして、動的閾値530を得る。

【0194】

アノテーションデバイス130は、動的閾値530を用いてジェノタイプピークセットを決定するように構成され得る。アノテーションデバイス130は、上記のような、シグナル強度の微分係数の $sign$ の微分係数に基づいて、リピートプロファイルにおいて、可

能性のある位置を同定し得る。可能性のあるピーク位置におけるリピートプロファイルの値がその可能性のあるピーク位置における動的閾値の値を超える時、アノテーションデバイス 130 は、ジェノタイプピークセットにその可能性のあるピーク位置を含み得る。

【0195】

アノテーションデバイス 130 は、サイズ分類ラダーおよび遺伝子特異的産物ピーク位置を用いて、リピートサイズを遺伝子特異的産物ピークに関連づけるように構成され得る。アノテーションデバイス 130 により用いられるサイズ分類標準は、内部サイズ分類標準、外部サイズ分類標準、および移動度補正済みサイズ分類標準のうちの 1 つを含み得る。いくつかの態様において、アノテーションデバイス 130 は、以下の方程式：

【0196】

【数 20】

$$S_G = \left\{ \frac{L_{\text{サイズ分類}(i)} - S_p}{S_r}, \forall i \in G \right\}$$

を用いて、リピートサイズを産物ピーク位置に関連づけ得る。

【0197】

この例において、L サイズ分類 (x) はサイズ分類標準であるとし、S_p は遺伝子特異的産物プライマーの塩基対でのサイズであるとし、S_r はリピートの塩基対でのサイズであるとし、G を、動的閾値を用いて同定されたジェノタイプピークセットであるとする。非限定的例として、S_p は 20 ~ 1000 塩基対の間の範囲、例えば、240 塩基対であり得るが、他の値も可能である。同様に、S_r は 3 塩基対であり得るが、他の値も可能である。これらのおよび他のパラメータにおける遺伝子特異的またはリピート特異的変動は、記載された構成の範囲内である。

【0198】

アノテーションデバイス 130 は、ジェノタイプピークセットを決定した後、正常な遺伝子特異的産物ピークとして存在しない遺伝子特異的産物ピーク（すなわち、ホモ接合体女性 (female) 試料、n / n + 1 ジェノタイプ、伸長型試料）を解明するように構成され得る。ホモ接合体女性ピークは、提供された性情報を用いて解明し得る。例えば、単独でコールされる遺伝子特異的産物ピークは、女性試料についてのホモ接合体ジェノタイプへと解明され得る。

【0199】

図 6 A に示されているように、アノテーションデバイス 130 は、近位ジェノタイプ (n / n + 1) を有するゲノム試料について遺伝子特異的産物ピーク位置を決定するように構成され得る。そのような試料を同定することは、困難な問題であり得、そのような試料は、女性試料の 10 % を構成する。アノテーションデバイス 130 は、第 1 のピーク 631 の側の第 2 のピーク 633 を同定することによりそのような試料を同定し得る。例えば、アノテーションデバイス 130 は、リピートプロファイル 610 a における第 2 のピーク 633 を第 1 のピーク 631 に隣接した第 2 の位置で同定し得る。第 2 のピーク 633 は、第 1 のピーク 631 と大きさが類似し得る。例えば、アノテーションデバイス 130 は、第 2 の位置におけるリピートプロファイルの第 2 の値が、第 1 の値に基づいた振幅基準を満たすこと決定し得る。例えば、第 2 の位置におけるリピートプロファイルの第 2 の値は、第 1 の値の 70 % ~ 90 % の間である振幅閾値を超え得る。いくつかの実施形態において、第 1 のピーク 631 と第 2 のピーク 633 の両方が、動的閾値 620 を超え得る。

【0200】

図 6 B に示されているように、アノテーションデバイス 130 は、ジェノタイプピークが同定されないが、リピートプロファイル 610 b が伸長を示す時、試料を、伸長型試料と標識するように構成され得る。当業者により認識されているように、リピート領域が FM R 1 である場合、そのような試料は、200 リピートを通り過ぎて遠くへ伸長するリピートプロファイルを有する試料として存在し得る。

【0201】

任意の複雑なPCRに基づいたワークフローのように、産物は、オペレータのエラーかまたは機器のエラーのいずれかの結果として増幅できないことが時々あり得る。これは試料を解釈不可能にし得るため、アノテーションデバイス130は、再分析のために試料にフラグを付けるように構成され得る。アノテーションデバイス130は、ユーザーが結果を間違っ

10

【0202】

図7A～7Cは、異なる品質管理基準を有する試料を描く。図7Aは、不十分なROXラダーを有する試料を描く。この試料は、サイズ分類標準基準に不合格であり得る。サイズ分類標準基準は、サイズ分類標準が正しく導かれていること、およびそれが、内部キャリブ

20

【0203】

図7Bは、十分にPCR増幅されていない試料を描く。この試料はリピートプロファイルシグナル・ノイズ比基準に不合格であり得る。この基準は、ユーザーが、十分に増幅されていない試料についての結果を解釈することを防ぐ。十分に増幅されていない試料は、処理中のアルゴリズムの仮定を乱し得、間違っ

30

$L_n = Q_{75}(s([I_{rp}(1) - 200, I_{rp}(1) - 50], c))$

【0204】

【数21】

$$L_{rp} = \text{中央値}(\{s(l_{rp}(i), c), i = (1, 2, 3, 4, 5)\})$$

40

【0205】

50

【数 2 2】

$$QC_{SM} = \begin{cases} \text{合格, } L_{rp} - T_{ma} > 150 \\ \text{不合格、そうではない場合} \end{cases}$$

【0 2 0 6】

これらの例示的な方程式において、チャンネル c は F A M チャンネルであり得る。この例において、75 パーセンタイルが、リピートプロファイルのウィンドウ表示の部分に関して、アノテーションデバイス 130 により計算され得る。他のパーセンタイル、例えば、60 ~ 95 パーセンタイル、または平均および標準偏差などの他の統計的尺度が用いられ得ることを当業者は認識しているだろう。同様に、この限定的例において、リピートプロファイルのウィンドウ表示の部分は、開始位置の 200 サンプル単位前から、開始位置の 50 単位前まで伸長する。しかし、チャンネルの他のウィンドウもまた用いられ得る。

【0 2 0 7】

図 7 C は、混入ピークを有する試料を示す。この試料は、リピートプロファイル混入基準に不合格であり得る。リピートプロファイル混入基準は、不適切なジェノタイプが報告されることに寄与する可能性がある、不適切な試料調製に関係した、オフターゲット増幅または増幅アーチファクトの事例を同定するために用いられ得る。この Q C についての不合格基準は、遺伝子特異的産物ピークが、遺伝子特異的プライマーにより生じる可能性があり得ない範囲において同定されることとして定義され得る。例えば、遺伝子特異的産物ピークについて導かれたリピート数が、0 個未満のリピート（すなわち、場合によっては、240 bp 未満）であるならば、試料は、混入を有するとフラグを付けられ得る。

【0 2 0 8】

【数 2 3】

$$QC_{CT} = \begin{cases} \text{合格, } \min(S_G) > 0 \\ \text{不合格、そうではない場合} \end{cases}$$

【0 2 0 9】

マイナーアレル感度基準は、機器のバックグラウンドノイズと試料における最大のジェノタイプピークとの間の比に関して、アッセイについて可能であるレベルより下にマイナーアレルコーリング閾値を設定しないようにユーザーに警告し得る。ノイズレベル（上記で説明されているような）と最大ジェノタイプピークとの間の比率がマイナーアレル頻度を超えるならば、そのレベルでのマイナーアレルは正確に同定することができず、その試料は、ユーザーがより厳格に解釈するべきである「リスクがある」Q C とフラグを付けられる。下記の方程式において、G はジェノタイプピーク位置のセットを表すものとし、c はチャンネルを表すものとし、T_{ma} はユーザーによって特定化されたマイナーアレル閾値を表すものとし、L_n はマイナーアレル感度基準のために計算されたシグナルのバックグラウンドノイズレベルを表すものとする：

【0 2 1 0】

【数 2 4】

$$P_{max} = \max\{s(i, c), \forall i \in G\}$$

【0 2 1 1】

【数 2 5】

$$QC_{MA} = \begin{cases} \text{合格, } \frac{L_n}{P_{max}} > T_{ma} \\ \text{不合格、そうではない場合} \end{cases}$$

10

20

30

40

50

【0212】

図8は、ジェノタイプピークサイズ分類のための例示的なコンピュータシステムを描く。いくつかの実施形態において、コンピューティングシステム800は、プロセッサ801、メモリ803、ディスプレイ805、I/Oインターフェイス807、およびネットワークアダプタ809を含む。これらのユニットは、バス811により、またはワイヤレスでそれぞれの他のものと通信し得る。図8に示されたコンポーネントは、単一のデバイスまたは複数のデバイスにおいて存在し得る。

【0213】

開示された実施形態と一致して、プロセッサ801は、マイクロプロセッサ、中央処理装置(CPU)、図形処理装置(GPU)、または類似したデバイスであり得る。メモリ803は、コンピュータハードディスク、ランダムアクセスメモリ(RAM)、リムーバブル記憶装置、またはリモートコンピュータ記憶装置などの非一過性機器を含有する非一過性メモリを含み得る。いくつかの態様において、メモリ803は、ソフトウェアプログラムを記憶するように構成され得る。いくつかの態様において、プロセッサ801は、開示されたシステムおよび方法の運用を実施するようにコンピューティングシステム800を構成し得る、メモリ803に記憶された非一過性命令および／またはプログラムを実行するように構成され得る。様々な態様において、当業者により認識されているように、プロセッサ801は、開示されたシステムおよび方法の運用を実施するように、リモートメモリに記憶された非一過性命令および／またはプログラムを実行するように構成され得る。ディスプレイ805は、視覚的アウトプットを提供する任意のデバイスであり得、例えば、コンピュータモニター、LCDスクリーンなどである。I/Oインターフェイス807は、キーボード、マウス、トラックボール、オーディオインプットデバイス、タッチスクリーン、赤外線インプットインターフェイス、または類似したデバイスなどの、コンピューティングシステム800のユーザーからコンピュータシステム800へ情報を伝達するための手段を含み得る。ネットワークアダプタ809は、コンピューティングシステム800が外部ネットワークと情報を交換することを可能にするための手段を含み得る。例えば、ネットワークアダプタ809は、ワイヤレスワイドエリアネットワーク(WWAN)アダプタ、ブルートゥースモジュール、近距離無線通信モジュール、またはローカルエリアネットワーク(LAN)アダプタを含み得る。

【0214】

本明細書に開示されたアッセイを行うための試薬、分析ソフトウェア、および高分子のキットもまた本明細書に提供される。ある特定の実施形態において、試料においてリピート領域をジェノタイプピングするためのキットは、本明細書に記載されているような、リピート領域の増幅のための1つまたは複数のプライマー、バッファー、および分析ソフトウェアまたはソフトウェアキーを含む。他の実施形態において、ジェノタイプピークサイズ分類のためのキットは、リピート領域の増幅のための1つまたは複数のプライマー、バッファー、ならびに分析ソフトウェアを記憶する非一過性媒体および／またはソフトウェアキーを含む。ある特定の態様において、キットは、本明細書に記載されているような遺伝子座のリピート領域の一部に同一または相補的であるプライマーを含む。他の態様において、キットは、増幅プライマーセットまたは複数の増幅プライマーセットをさらに含み、そのプライマーの少なくとも1つが、上に記載されているような遺伝子座のリピート領域の一部に同一または相補的である配列を含む。本明細書に記載されているような分析ソフトウェアは、CD-ROMまたは他のデータ記憶デバイスなどの非一過性コンピュータ可読媒体に記憶されたデータおよび／または命令を含み得る。用語「ソフトウェアキー」は、本明細書に記載された分析ソフトウェアのダウンロードもしくはそれへのアクセスを可能にするように構成された、ソフトウェアライセンスキー、暗号化キー、URL、URL、および／またはパスワードを指す。この「ソフトウェアキー」は、紙、カード用紙、ステッカー、または類似した媒体などの非一過性媒体上に表示され得る；またはCD-ROMもしくは他のデータ記憶デバイスなどの非一過性コンピュータ可読媒体に記憶され得る(例えば、「readme」ファイルに位置する)。キットは、具体的には、例えば、この

10

20

30

40

50

明細書の特許請求の範囲 1 ～ 8 3 のいずれか一項に記載のデータ処理および計算方法を実施する分析ソフトウェアを含むことを企図する。

【0215】

キットはさらに、本明細書に記載されたアッセイを行うための酵素を含んでもよく、その酵素には、DNAポリメラーゼまたは逆転写酵素などのポリメラーゼが挙げられるが、それらに限定されない。ある特定の態様において、キットは、外部サイズ分類ラダーを含む。サイズ分類ラダーは、ROXラダー、または本明細書に記載されているようなサイズ分類ラダーであり得る。ある特定の態様において、キットは、陽性対照試料、例えば、鋳型対照試料またはプールされた細胞株対照試料を含む。

【0216】

キットはまた、プライマー、dNTP、ポリメラーゼ、および／またはバッファーを含む、リピーター領域を増幅するための試薬を含み得る。そのようなキットは、反応バッファー、増幅バッファー、および／またはポリメラーゼバッファーなどの1つもしくは複数のバッファー、DNA試料を調製するための化合物、ならびに、例えば、プローブまたは標識などの増幅産物を単離および／または検出するためのコンポーネントを含み得る。

【0217】

いくつかの実施形態において、本発明のキットは、以下（上記で論じられている方法、試薬、および組成物と一貫した）の1つまたは複数を含む：試料精製のためのコンポーネント、カオトロピック剤を有する溶解バッファーを含む；ガラス繊維フィルタまたはカラム；溶出バッファー；洗浄バッファー、アルコール溶液；およびヌクレアーゼ阻害剤。キットのコンポーネントは、例えば、水性媒体中か、または凍結乾燥の形態かのいずれかでパッケージングされ得、適切な容器において提供される。キットのコンポーネントは、乾燥粉末として提供され得る。試薬および／またはコンポーネントが乾燥粉末として提供される場合、その粉末は、適切な溶媒の添加によって再構成することができる。その溶媒もまた別の容器において提供されることが構想される。容器には、一般的に、溶媒が入れられ、適宜、等分されている、少なくとも1つのバイアル、試験管、フラスコ、ボトル、シリンジ、および／または他の容器もしくは等価物が挙げられる。キットのコンポーネントが1つおよび／または複数の溶液中で提供される場合、その溶液は、水溶液であり、無菌水溶液が特に好ましい。キットはまた、無菌の薬学的に許容されるバッファーおよび／または他の溶媒を含有するための第2の容器または等価物を含み得る。

【0218】

そのようなキットはまた、核酸分解から保護する試薬などの、DNAまたはRNAを保存または維持するコンポーネントも含み得る。そのようなコンポーネントは、ヌクレアーゼフリーもしくはRNアーゼフリーであり得、または例えばRNアーゼから、保護し得る。本明細書に記載された組成物または試薬のいずれかが、キット内のコンポーネントであり得る。追加の材料には、反応に用いられ得る、適切な反応容器、バリア組成物、増幅および／もしくはPCRのための反応混合物（バッファー、およびdNTPなどの試薬を含む）、ヌクレアーゼフリーもしくはRNアーゼフリーの水、RNアーゼ阻害剤、ならびに／または任意の追加のバッファー、化合物、補助因子、イオン性成分、タンパク質、酵素、ポリマーなどが挙げられ得る。

【実施例】

【0219】

以下の実施例は、例証する役割を果たし、決して、本開示を限定するものではない。

【実施例1】

【0220】

リピーター領域の増幅

CGGリピーター領域のサイズ分析についてのPCRに基づいたワークフローは、AmplideX（商標）FMR1 PCRアッセイ（Asuragenカタログ番号49402；米国特許出願公開第2010／0209970号）を用いて達成され得る。図9および16参照。

【実施例 2】

【0221】

内部標準を用いて核酸のリピート領域をサイズ分類するためのワークフロー

FMR1 遺伝子内の CGG リッチな領域に沿ってのプライミング事象に関連したリピートプロファイルを利用するサイズ分類方法を、開発した。キャピラリー電気泳動プロット「リピートピーク」における各ピークが、隣接したプライミング事象、すなわち、長さが 1 個余分のリピートの増幅産物に対応するため、ピークのヌクレオチドでの長さは、最初のリピートピークのサイズ（塩基対での）（プライマー配列長を考慮に入れる）を仮定し、その後、最初のピーク後の各リピートピークがその前のピークより 3 塩基対長いと仮定することにより推定することができる。この情報は、fragment sequence analysis (FSA) シグナル（サンプリング単位での）での位置を断片サイズ（塩基対での）に関係づける較正曲線を作成するために用いられる。

10

【0222】

詳細には、リピートピーク同定のためのアルゴリズムが、いくつかの段階で働く。第 1 に、リピートシグナルの始まりは、リピートシグナルが機器のサンプリング周波数に基づいて開始するウィンドウについての情報を用いて検出される。第 2 に、分位に基づいた分析を用いて、リピートシグナルが開始し、そして終了する範囲を決定する。第 3 に、周波数に基づいた分析を用いて、サンプリング単位におけるリピート周期性を決定する。第 4 に、リピート周期性を用いて、リピートピークがコールされるウィンドウサイズを通知する。第 5 に、分位に基づいたアプローチを用いて、リピートピークがコールされるはずである閾値を導く。第 6 に、スライドウィンドウを用いて、単一リピートピークをコールし、各ウィンドウについてのコールされたピークが、その範囲における最大の大きさをもつ、マイナスの二階微分係数を有すると定義される。ピークが見出されず、またはシグナルが、第 4 の段階において決定された閾値より下にある場合には、ウィンドウの中央としてのリピートピークの位置が外挿され得る。ピークがコールされる時、サンプリング単位でのリピートピーク間の差に基づいたウィンドウのサイズが調整され得る。図 10 は、遺伝子特異的産物をサイズ分類するためにピーク位置が利用される過程のグラフを描く。各リピートピークの期待される位置を用いて、サイズ分類のための較正曲線を作成する。図 10 は、カスタムプライマーの FMR2 セット (p/n 49541) と組み合わせた Amplicon X 試薬を用いて作成された。

20

30

【0223】

リピートピークが同定された後、ソフトウェアは、リピートピークについての期待される断片長（ヌクレオチドでの）に対する全てのピークインデックス（サンプリング単位での）の 3 次スプライン補間を用いて、サイズ分類のためのモデルを作成し得る。いくつかの実施形態において、サイズ分類標準は、区分的であり得、1 次多項式フィットをゲルの線形領域に、および一変量スプラインフィットをゲルの非線形領域に適用し得る。

【実施例 3】

【0224】

メジャーアレルの自動的遺伝子特異的産物同定

このアッセイについて FMR1 ジェノタイプを自動的に作製するために開発されたアルゴリズムは、サイズ分類方法と共に用いることができる。そのアルゴリズムは、遺伝子特異的産物を同定するのに大きさに基づいたアプローチを採用し、特定化された閾値に至るまで全てのアレルピークを標識する。このアルゴリズムについてのワークフローは、0 の一階微分係数およびマイナスの二階微分係数を有する、ピーク様の形状をもつ領域を同定し、相対蛍光単位 (RFU) の大きさによって領域をランク付けすることを含む。任意の工程として、性をインプットとして提供し、各性に必要とされるピークの数に戻し、アレルの数（すなわち、ホモ接合体女性試料について）を自動的に決定することを含み得る。完全突然変異分析について、コールされるリピートに合格した存在するリピートピークをアセスメントするために、伸長についてのリピートプロファイルが、分位に基づいたアプローチを用いて分析される。伸長が発生する場合には、伸長したアレルが報告される。遺伝

40

50

子特異的リピートジェノタイプは、AmplideXソフトウェアにおいて、内部的に導かれたラダーを用いて決定される。

【実施例4】

【0225】

マイナーアレルの自動的遺伝子特異的産物同定

メジャーアレルジェノタイピングに加えて、マイナーアレルカットオフの研究室特異的定義を可能にするマイナーアレル検出のための過程を確立した。マイナーアレルおよびモザイク現象は、典型的には、その現象が臨床関連カテゴリーで起こる場合のみ臨床関連であるため、アルゴリズムを、前突然変異から完全突然変異の範囲においてマイナーアレルを探索するように特異的に設計した。具体的には、マイナーアレル検出のためのアルゴリズムは、以下の工程をとり得る：（１）アルゴリズムのメジャーアレルジェノタイピング相

10

で検出されたピークのランク付けリストから、任意の過剰なピークが54CGGリピート長より長いかどうかを決定する、および（２）それらのピークについて、シグナルにおいて同定された最大の遺伝子特異的産物のユーザー定義の閾値パーセンテージより高いRFU大きさをどれが有するか（もしあれば）を決定する。閾値は、現在、10%をデフォルトにしているが、アルゴリズムへのインプットとして特定化することができる。

【実施例5】

【0226】

リピート領域をサイズ分類するためのワークフローを用いることの応用

上記で論じられたアルゴリズムの性能を、FMR1ジェノタイプ全範囲に及ぶ、500個のランダムに選択され、かつ以前にアノテートされた臨床試料の包括的セットに対して評価した。QC基準を合格した試料を、前のセクションに記載された方法を用いてジェノタイピングし、期待されたピークサイズと観察されたピークサイズとの差（リピート単位での）として精度を測定した。加えて、ROXラダー、例えば、ROX 1000サイズラダー、P/N: 145194（外部参照標準）と内部サイズ分類方法との間の一致を、試料コホートに渡る相関分析を用いて評価した。マイナーアレル検出能もまた、オペレータが同定したモザイク現象を有する7つの試料の選択されたコホートを用いて試験した。これについて、アルゴリズムを、5%マイナーアレル感度まで下げてコールするようにパラメータ化し、結果を、手作業でサイズ分類されたマイナーアレルと比較した。最後に、アルゴリズムを、分析感度および伸長にフラグと付けるためのリピートプロファイルの利用

20

30

を評価するために、5%感度対照（Asuragen、P/N 145303）を用いて試験した。この感度対照は、95%の短い女性正常および前突然変異アレル（CGG=30、56）および5%の伸長アレル（>200 CGG）で構成された。

【0227】

図11は、一般的な集団におけるFMR1ジェノタイプの発表された分布（Tassone et al., 2012）と共に、本研究で試験された患者FMR1ジェノタイプの分布を示す。その試験セットにおける患者の分布は、臨床的関連性がより高いジェノタイプについてのアルゴリズムの感度をアセスメントするために、より大きいサイズ範囲においてより高密度であった。全体的に見て、500個の試料のうちの472個がシグナル強度についてのQC閾値を合格し、前のセクションに記載された方法論を用いて、ジェノタイピングされた。埋め込まれたQC工程が28個の不合格試料（5.6%）を正確に同定し、それらは、目視検査で、完全なシグナルドロップアウト（9個の試料）かまたはリピートピーク高さの有意な欠損（19個の試料）のいずれかを有することが確認された。

40

【0228】

加えて、リピートピークサイズ分類方法が、前に用いられた方法論（ROXラダーを用いる）と、 $R^2 > 0.95$ で相関することが見出され、AmplideX FMR1ワークフローの一部としての外部校正コンポーネント（ROXラダーおよび対照試料）の必要性を最小限にすることを示唆した。図12は、ROXサイズ分類方法と内部サイズ分類方法との間の一致の例を描く。

【0229】

10

20

30

40

50

埋め込まれたQCを合格する472個の試料について、アルゴリズムは、成功裏に、全てのメジャー（非モザイク）アレル（855個）を同定し、200リピートより多いアレルを完全突然変異としてフラグを付けた。200リピートより多くを有するアレルは、内部サイズ分類方法を用いて、それらの以前に報告されたサイズ（手分析を用いて独立して決定された）の ± 1 CGG内で正確に位置づけられた。表2は、臨床的突然変異カテゴリーに関するジェノタイピング精度を詳述する。全体的に見て、2個を除いて全ての試料が、それらの期待されたカテゴリーに正しく同定され、2個の、間違って標識された試料が、それらの以前報告されたサイズの ± 1 CGGを有するとジェノタイピングされた。図13は、ジェノタイピング精度をより詳細に示し、コホートにおける全ての非モザイクピークについての期待されたFMR1ジェノタイプ対観察されたFMR1ジェノタイプを描く。図13は、実現可能性研究における全てのメジャーアレル（855個）の手作業サイズ分類と自動サイズ分類との間の相関を描く。図13は、自動FMR1ジェノタイピングワークフローが、手作業割当に基づいたワークフローとの一致した結果を生じたことを示す。

【0230】

【表2】

表2:臨床的突然変異カテゴリーに関するジェノタイピング精度。

	自動分析			
	正常 (0~44)	中間 (45~54)	前突然変異 (55~200)	完全突然変異 (>200)
手分析 正常(0~44)	162	*6	0	0
中間(45~54)	0	295	*9	0
前突然変異 (55~200)	0	*1	527	0
完全突然変異 (>200)	0	0	0	35

*手作業でアノテートされたサイズの ± 1 CGG内

【0231】

加えて、マイナーアレル同定のための過程により、7個の手作業でアノテートされたマイナーアレルのセットから全てのマイナーアレルが、それらの手作業で導かれたサイズの ± 1 CGG内に正しく同定された。図14Aは、マイナーアレル検出能を詳述する図を示す。その図における全てのマイナーアレルは、FMR1ソフトウェアにより自動的に検出され、手分析と一致して標識された。その図はまた、ユーザー特異的マイナーアレル検出のために機能的に開発された感度調整を描く。

【0232】

最後に、感度対照試料は、正確にジェノタイピングされ、伸長したアレルとしてフラグを付けられた。図14Bは、遺伝子特異的産物のアルゴリズムのラベリングを描く。

【0233】

図15Aおよび15Bは、FMR1ジェノタイプ範囲に渡るアルゴリズムの分析感度を詳述する図を示す。矢印は、自動サイズ分類により作製されたコールを示す。図8は、AmplideX PCR/CE FMR1レポーターを描き、例えば、FMR1遺伝子のCGGリピート領域を標的にするキャピラリー電気泳動に基づいたアッセイの効率および一貫

性を向上させる、高度に正確な自動FMR1分析エンジンおよびソフトウェアインターフェイスが設計された。このソフトウェアは、AmplideX（登録商標）FMR1 PCR試薬を用いて処理された1000個より多い臨床試料に関して試験され、手作業でのジェノタイピングとの100%一致を実証した。そのソフトウェアはまた、少量の遺伝子特異的産物を検出する高い感度を実証した。このソフトウェアは、FMR1アッセイワークフローについての分析時間を100倍より高く、向上させる可能性が高く、CEプロファイル曖昧性を解決することにおいてオペレータ間の一貫性を向上させる可能性をもつ。

【実施例6】

【0234】

不合格の試料の検出

シグナルの始まりに向けて起こり得るシグナルアーチファクト（例えば、小さい最初のピーク、AGG割り込み配列）に対してロバストであるように、分位に基づいた分析を用いることによりシグナルドロップアウトを検出する、再分析のために試料にフラグを付ける自動化ストラテジーを開発した。そのアルゴリズムは、RFU大きさの95分位（シグナル）およびRFU大きさの5分位（バックグラウンド）を計算することにより働き、それらの値の間の差が200RFUより下にある試料を不合格にする。試料を不合格にするための200RFU閾値は、正常なインプット量から12.4pgゲノムDNAインプット量まで下がって滴定された1セットの対照試料を用いて実験的に決定された。その閾値は、正しいジェノタイプコールをもたらす最後のインプット量を考慮することにより決定され、それらの試料についてシグナルとノイズの間の平均分離点として計算された。シグナルドロップアウトを同定することに加えて、より高いサイズ範囲においてROXラダーピークの欠如を認識することにより、時期尚早に停止された実行を同定するためのアルゴリズムを開発した。同定されたROXピークの数期待より低い場合には、アルゴリズムは、不完全なデータを有するとして試料にフラグを付ける。これは、臨床的関連に関して間違っ

【実施例7】

【0235】

誤標識されたROXピークの同定

シグナル完全性についての自動チェックと共に、少数のサイズ分類エラーに寄与する可能性があり得るROXチャンネルにおける誤標識問題を同定するためのアルゴリズムを開発した。どのようにして誤標識問題が間違っ

【実施例8】

【0236】

FMR1サイズ分類分析

8.1. ワークフロー概観

FMR1 CGG分析についての現ワークフローは、ジェノタイプがキャピラリー電気泳動データから解釈されている点を通して合理化されている。手作業での解釈は、大量試験にとってかなりのボトルネックであり、自動アルゴリズムは、全過程を大幅に向上させる可能性がある。この研究において、高度に正確なFMR1サイズ分類結果を生じるFMR1分析の自動解決法を開発した。そのアルゴリズムの構成要素の概観は、このセクション

内で詳細に記載されている。高レベルにおいて、アルゴリズムはいくつかの段階で働く。アルゴリズムの第1の段階は、`fragment sequence analysis` ファイルから生データを抽出し、異なる構成で実行された試料に渡る差を正規化するために前処理を実施する。第2の段階は、各試料の内部のコンポーネントを用いて、シグナルにおける位置単位（POP7ゲルにおいて移動した距離に相似）から塩基対サイズへ変換するためにサイズ分類ラダーをパラメータ化することを含む。アルゴリズムの最終段階は、アッセイにおいて異なるプライマーセットからの増幅をデコンボリューションするためのモデルをパラメータ化すること、およびそのモデルを用いて、ジェノタイプピークを同定することを含む。

【0237】

8.1.1. `Fragment Sequence Analysis (FSA)` ファイルパース

`AmplideX`（登録商標）`FMR1 PCR`アッセイは、`Genetic Analyzer`機器（3130/3500/3700）の`Applied Biosystems`ファミリーと共に実行されるように設計され、その機器の全ては、`Applied Biosystems`により維持される独自仕様のフォーマットでデータをエクスポートする。このフォーマットは、`Fragment Sequence Analysis`フォーマットと呼ばれ、「`Applied Biosystems Genetic Analysis Data File Format, 2009`」に記載され、かつ全体として参照により本明細書に組み入れられた、一連の独自仕様に従ってコード化された、キャピラリー電気泳動実験からの蛍光データを含む。このファイルフォーマットに直接的にアクセスするために、情報をデコードして、プログラムでアクセスし、かつ操作することが容易である`JSON`に基づいたフォーマットへ編成するための、特定化されたパーサーを設計した。このパーサーは、`Bio::Trace::ABIF`と呼ばれるperlプログラミング言語についてのオープンソースモジュールを大量に利用する。パースソフトウェアは、異なる`Genetic Analyzer`機器（3130/3500/3700）に渡って実行された1000個より多い試料に関して確認されており、`GeneMapper (Fragment Sequence Analysis)`フォーマットにアクセスするための現行標準）を通して見られる未処理の蛍光データと正確に一致していることが示されている。

【0238】

8.1.2. `FMR1` サイズ分類分析内の前処理

8.1.2.1. シグナル平滑化

`AmplideX`（登録商標）`FMR1 PCR`アッセイは、データ解釈のためにキャピラリー電気泳動に基づいた読み取りに頼るため、そのアッセイは、`CE`機器に体系的に存在するシグナルアーチファクトによるコンボリューションを受けやすかった。処理における第1の工程として、データを平滑化するために、チャンネルのそれぞれに`Savitzky-Golay`フィルタを適用した。これは、下流処理におけるアルゴリズムにより作成される仮定が単純化されるのを可能にし、また処理後の運用においてロバスト性を増加させた。

【0239】

8.1.2.2. ベースライン正規化

平滑化後、次に、不適切な機器較正を無効にするために各チャンネルを正規化した。各チャンネルを正規化するために、シグナルから、シグナルにおける`RFU`値の10パーセンタイルを引き算した。10パーセンタイルは、実験的に選択され、それが、一般的に遭遇されるシャープなマイナスのゆらぎによって影響されることなく、シグナルにおけるより低い値をロバストに表したからである。代替の適切な値も同様に、シャープなマイナスのゆらぎによって影響されることなく、シグナルにおけるより低い値を表すだろう。

【0240】

8.1.2.3. 気泡混入除去

`AmplideX`（登録商標）`FMR1` アルゴリズムへの前処理工程として、気泡アーチ

10

20

30

40

50

ファクトを同定し、除去した。そのような気泡は、間違っただけの結果を生じる形で遺伝子特異的産物またはROXチャンネルサイズ分類ピークとして解釈され得る、シグナル強度において大きなスパイクを生じ得る。しかし、CE実行中の気泡キャピラリーチューブの存在は、同じような程度の大きさでチャンネルの全部に影響する蛍光をもたらした。このノイズピークが複数チャンネルで存在することを利用することにより、気泡アーチファクトを同定し、かつ除去した。

【0241】

気泡が見出された部位のそれぞれについて、ピークショルダー間にガウスノイズをシミュレートすることにより、気泡を除去した。そのノイズについての平均および標準偏差を、気泡を囲む領域から決定した。

【0242】

8. 1. 2. 4. シグナル飽和解決

別の前処理工程は、シグナル飽和がFAMチャンネルにおいて起こった領域に関してピーク形状を外挿することを含んだ。産物が、機器RFUセンサーの収集限界より大きい蛍光で蛍光を発した時に飽和が起こり、ピーク形状に関する情報の損失をもたらす。しかしながら、収集についての波長スペクトルは、チャンネルに渡る漏出を許容するため、飽和領域についてのピーク形状は、同様の波長での蛍光を捕獲するチャンネルから外挿することができる。同定された飽和領域において、NEDチャンネルからのRFU値を、HEXチャンネルへ加えた。

【0243】

8. 1. 3. 自動サイズ分類ラダー校正

キャピラリー電気泳動実験における断片サイズ分類のための現在のゴールドスタンダードは、外部から加えられる、公知のサイズの色素標識PCR産物の使用を必要とし、その産物は、AmplideX（登録商標）FMR1 PCR産物により生じた周波数スペクトルの外側のバンドに蛍光ピークを生じる。これらの蛍光ピークは、アッセイにより生成された標的産物とは無関係に同定することができ、fragment sequence analysisシグナルにおける位置（サンプリング単位での）を断片サイズ（塩基対サイズでの）に関係づける校正曲線を作成するために用いられる。これらの蛍光ピークを同定する過程は、GeneMapperソフトウェアにおいて自動的に操縦されるが、それは、誤標識されたピークを訂正するために手作業での検査を必要とする場合が多く、そのことは、AmplideX（登録商標）分析を実施するのに必要とされる時間を有意に増加させ得る。これを改善するために、GeneMapperに基づいたワークフローを妨げる誤標識現象に対してロバストであるようなやり方で、ROX蛍光ピークを自動的に同定し、かつ標識するアルゴリズムを開発した。加えて、そのアルゴリズムは、将来的なアッセイ開発に用いられる任意のサイズ分類ラダー（ROX 1000、ROX 200）へ容易に延長するように開発された。

【0244】

8. 1. 3. 1. 漏出アーチファクト除去

分析に関わる第1の段階は、期待されたサイズの、ROX断片ピークへの割当をコンボリューションし得る、HEXチャンネルからの漏出アーチファクトを除去することを含んだ。これらの漏出アーチファクトを検出するために、アルゴリズムは、HEXチャンネルにおいて、実験的に導かれた機器特異的閾値より上の位置を同定し、その後、その領域に渡って、シグナルバックグラウンドを模倣するガウスノイズをシミュレートした。

【0245】

8. 1. 3. 2. ROX断片ピークコーリング

分析に関わる第2の段階は、1セットの候補ROX断片ピークをコールし、偽陽性ピークコールである可能性が高いピークアーチファクトを除去することを含んだ。ROX断片ピークの候補セットを同定するために、幅が500位置単位である、データに渡るスライドフィルタを実行した。各ウィンドウについて：（a）その範囲内でのシグナルの平均および標準偏差が採用され、（b）平均ノイズレベル（CE機器についてのノイズプロファイ

10

20

30

40

50

ルは、ガウス分布に従うと仮定される)より上へ3標準偏差分を超えるピークがコールされた。これらの候補ピークを検出した後、「ショルダー」アーチファクトにより引き起こされる近位の偽陽性ピークは、選択されたピークの周囲のウィンドウにおいて最大ピークを選択することにより、解決された。

【0246】

8.1.3.3. ROX断片ピーク関連づけ

分析の第3の段階は、標識される断片ピークに関連づけられる最も可能性が高いピークを選ぶ繰り返しアプローチを用いることを含んだ。このアプローチは、最初の条件を選択するために、より大きい断片サイズに関連づけられた低ノイズプロファイルを利用した。要するに、500bpより大きい全ての期待されるROX断片サイズが、候補ピークの最も遠い(キャピラリーにおける距離による)セットと自動的に関連づけられた。その後、500bp~700bpの間の全ての標識されるROXピークのための線形サイズ分類ラダーが、1次最小二乗回帰を用いて、フィッティングされ、次(500bp未満)の断片ピークの位置を予想するために用いられた。その後、その予想された位置に最も近い候補ピークが、その断片サイズで標識され、線形サイズ分類ラダーは、そのデータ点を含むように、再びフィッティングされた。アルゴリズムは、繰り返し、この形式で、候補ピークをROX断片サイズで標識し、ゲルのノイズがより高い領域においてピーク関連づけの精度を増加させるやり方で、新しいデータ点に関して、継続して訓練する。この繰り返しアプローチは、不適切なサイズ分類ラダーのパラメータ化に寄与し得るROXチャンネルにおけるシグナルアーチファクトを無視することにおいて、現行方法(GeneMapper、GeneMarker)より特異的であることが示されており、プライマーダイマーピークをROX断片ピークと誤解することに対してもロバストである。

【0247】

8.1.3.4. サイズ分類ラダーのパラメータ化

その後、区分的サイズ分類標準が、最終のサイズ分類ラダーを用いて作成された。区分的モデルは、650bpより下のラダーピークについて線形モデルを用い、650bpより上のラダーピークについて一変量スプラインモデル(局所的サザン方法に類似(Analytical Biochemistry 100(2):319-323 (1979)参照))を用いた。最終の平滑化サイズ分類ラダーをパラメータ化するために、区分的モデルを用いて、位置単位の、断片サイズへの関連づけと共に、100個の等間隔での点を再サンプリングした。これらの再サンプリングされた点は、ラダーの平滑化最終バージョンを作成するために、一変量スプラインモデルを用いてこれらのデータにフィッティングされた。

【0248】

8.1.4. サイズ分類ラダー移動度補正

キャピラリー電気泳動におけるROX断片移動度は、FMR1断片移動度とは異なり、ROX断片のヌクレオチドバランスのとれた性質に対してFMR1断片のGCリッチな性質のためである。これらの違いを無効にするために、移動度補正因子を、FMR1リピートシグナルからパラメータ化し、ROXチャンネルから導かれたサイズ分類ラダーに適用した。この過程に関与する以下のいくつかの工程があった: 1) リピートシグナルの開始位置および終了位置の同定、2) 全てのリピート断片ピークのラベリング、3) 移動度補正の適用。

【0249】

8.1.4.1. FMR1シグナルウィンドウのパラメータ化

リピートピークを標識することに先行することとして、シグナルの目的の領域が決定された。ROXラダーを用いて、リピートプロファイルの始まりについてのおよその位置を予想した。ROXラダーのパラメータ化が失敗するか、または生じた適合度が、ROX QC基準(セクション8.1.6参照)に合わなかったならば、以下の工程が用いられた。

【0250】

200位置単位のウィンドウサイズでのデータへの総和変換が用いられた。その変換されたシグナルにおける最大ピーク(プライマーダイマー増幅事象により引き起こされる)に

ついて、ピークショルダーが同定された。最も右側のショルダーから、100位置単位ウィンドウ内のシグナルの優位周波数を計算する変換が適用された。シグナルの優位周波数が実験的に導かれた許容範囲内にある、最初のウィンドウの位置が、リピートプロファイルのおよその始まりとして用いられた。

【0251】

いったんリピート領域のおよその位置が同定されたならば、そのおよそのリピート開始部位から1000位置単位ウィンドウ内のシグナルの85パーセンタイルより大きい最初の位置を同定することにより、正確なリピート開始部位が決定された。

【0252】

リピート領域の終了位置は、100位置単位のウィンドウサイズで、シグナルに渡って90パーセンタイルフィルタを適用する変換を用いて同定された。その変換が適用された後、シグナル終了位置は、実験的に導かれた機器特異的閾値より上に入る最後の変換された領域として選択された。

【0253】

8. 1. 4. 2. リピートプライマーピークのコーリング

シグナルの開始位置および終了位置が同定された後、リピートプライマーセットにより生じたシグナルにおいて全ての増幅ピークを決定することにより分析が進行した。その後、これらのピークは、予想されるプライミング事象がもたらすサイズに関連づけられ、線形ラダー適合度が生成された。高レベルにおいて、リピートピークは、リピートピークを繰り返しコールするシグナルの周期性から導かれたウィンドウを用いてコールされ、シグナルにおいて周期性がシフトするにつれてウィンドウが調整され、リピート性ピークが抑圧されている所（すなわち、AGG割り込み部位）にピーク位置が内挿された。

【0254】

開始部位から1000位置単位ウィンドウにおけるフーリエ変換を用いて、リピートプロファイルの周期性を同定した。ピーク間の期待される開始距離は、周期性の逆数を用いて計算された。開始部位から2000位置単位ウィンドウの25パーセンタイルを用いて、ピーク位置を内挿するための閾値を導いた。

【0255】

リピートプロファイルの長さ（上記で決定された開始位置と終了位置の間）を決定するために、およそのピーク位置を含有するウィンドウを予想し、その範囲において最大ピークを選択することにより、リピート閾値より上にある次のピークコールのピークが繰り返し、選ばれた。ピークを選択するためのウィンドウは、ピーク間の距離のサイズ（シグナルの周期性により決定された場合）の $1/2$ として計算された。

【0256】

8. 1. 4. 3. サイズ分類ラダー移動度補正

リピートプロファイルについての全てのピークが同定された後、それらは、期待される断片サイズで標識され、線形サイズ分類ラダーを作成するために用いられた。この線形サイズ分類ラダーは、ROXラダーがアフィン変換を用いてマッピングされ得る、サイズ補正ラダーとしての役割を果たした。アフィン変換によるマッピングは、サイズ分類ラダーの線形成分と非線形成分の両方が移動度補正を適用されることを保証した。

【0257】

8. 1. 5. ジェノタイプピーク同定およびサイズ分類

サイズ分類分析の重要な部分は、ジェノタイプピーク同定およびサイズ分類であった。シグナルは、リピートセグメント増幅と遺伝子特異的増幅の両方を呈示するため、ジェノタイプピークの同定は、簡単ではなく、シグナルのこれらの2つの成分をデコンボリューションするための工程を含む。追加として、正常なジェノタイプピークとして存在しない困難なジェノタイプは、その過程を通して同定された。

【0258】

8. 1. 5. 1. リピートプロファイルバックグラウンド推定

遺伝子特異的増幅事象を同定することにおける第1の工程は、遺伝子特異的増幅事象のシ

10

20

30

40

50

グナル寄与からリピート増幅事象のシグナル寄与をデコンボリューションするためのバックグラウンドモデルをパラメータ化することを含んだ。この過程を困難にさせるシグナルアーチファクトの中で、最も有意なものは以下であった：（１）リピートプロファイルの大きさを有意に縮小させる FMR 1 リピート領域における A G G 割り込みであり、それは、バックグラウンドモデルを作成する間、無視されなければならないプロファイルにおける「ギャップ」を生じる；および（２）シグナルのリピート成分から有意に偏向するが、デコンボリューションへの周波数に基づいたフィルタリングアプローチを可能にするだろう特性を欠く遺伝子特異的産物ピーク。

【0259】

デコンボリューション過程は、所定のウィンドウ内のリピートシグナルの高さとして、リピートプロファイルの「バックグラウンド」をモデル化した。A G G 割り込みについて、「バックグラウンド」は、A G G 割り込みに近位の局所的リピートピークのレベルであった。遺伝子特異的産物ピークについて、「バックグラウンド」は同様に、シグナルにおける遺伝子特異的偏向に近位の局所的リピートピークのレベルであった。

【0260】

分析は、バックグラウンドモデルを作成するために以下の工程を含んだ。リピートシグナルにおける全てのピークについて、11リピートのウィンドウサイズに渡ってデータの中央値および四分位範囲をフィルタが加えた。このフィルタは、ウィンドウにおいてリピートピークの高さを捕獲するように設計されが、A G G 割り込みおよび遺伝子特異的産物により引き起こされるリピートシグナルにおける大きなゆらぎに対してロバストであった。生じたシグナルについて、データを平滑化するために7リピートのウィンドウサイズをもつ S a v i t z k y - G o l a y フィルタが用いられた。生じたシグナルにおける任意の「ピーク」が、ピークショルダーを通して線形的に内挿された。

【0261】

8. 1. 5. 2. 動的閾値を用いるジェノタイプピーク同定

より低いサイズ範囲におけるコーリングをより特異的にし、より高いサイズ範囲におけるコーリングをより高感度にする動的スケーリングアプローチを適用することにより、上記のデコンボリューションモデルから動的閾値が導かれた。高レベルにおいて、スケーリングアプローチは、0～120リピートの間の領域において3から1.5へ減少させ、その後、120リピート後においては1.5の一定のままである区分的スケール因子をデコンボリューションモデルに適用した。ピーク様形状を有する領域を同定するための上記の方法に従い、この閾値を用いてジェノタイプピークセットが決定された。遺伝子特異的産物ピークは、前のセクションにおいて導かれたサイズ分類ラダーを用いて、リピートサイズへ変換され、遺伝子特異的増幅産物からの非リピート成分の公知の断片サイズ（この場合、240bp）を用いてリピート数へ変換された。

【0262】

8. 1. 5. 3. 困難なジェノタイプの解明

ジェノタイプピークセット G を決定することにおける最初の合格後、正常な遺伝子特異的産物ピークとして存在しない困難なジェノタイプ（すなわち、ホモ接合体女性試料、 $n/n+1$ ジェノタイプ、伸長型試料）が解明された。ホモ接合体女性ピークは、提供された性情報を用いて解明され、単独でコールされたピークが、女性試料についてのホモ接合体ジェノタイプへと解明された。正常範囲内の近位のジェノタイプ（ $n/n+1$ ）を有する試料は、ジェノタイプピークの次のリピートピークを用いることにより解明された。リピートピークがジェノタイプピークに隣接して存在し、かつそのリピートピークのシグナル強度が、隣接したジェノタイプピークの高さの90%以内であった場合、そのリピートピークもまた、ジェノタイプピークとして標識された。最後に、試料についてジェノタイプピークが同定されなかったが、リピートプロファイルが伸長を示した場合、その試料は伸長型試料として標識された。これは通常、遺伝子特異的産物ピークを欠き、リピートプロファイルが200リピートを超過して十分伸長した男性（male）試料について起こった。

【0263】

8. 1. 6. 自動の埋め込まれた品質管理

結果の誤解釈を防ぐためにいくつかの品質管理尺度が用いられた。品質管理尺度の2つのカテゴリーがあった。品質管理尺度の第1のカテゴリー（サイズ分類ラダーQC、シグナル大きさQC、および混入QC）に不合格となる試料について、ジェノタイプコールは生じなかった。品質管理尺度の第2のカテゴリー（マイナーアレル感度QC）に不合格となる試料についてのジェノタイプコールは、より大きい懷疑を以て解釈された。この第2のカテゴリーは、ユーザーが、ジェノタイプコールを、それらのデータにより信頼性をもって支持されるものではないと判断することから守るはずである。

【0264】

8. 1. 6. 1. サイズ分類ラダーQC

サイズ分類ラダーQCは、サイズ分類ラダーが、内部キャリブレーションに対して、正しく導かれ、期待値とマッチすることを検証した。この尺度は、3つの異なる基準を組み合わせしており、その基準のそれぞれが、試料が合格するために満たされなければならない。第1に、ROXラダーピークに対するROXラダー適合度についての決定係数（R²）が0.98より高くなければならない。第2に、内部ラダーピークに対する内部ラダー適合度についての決定係数が0.98より高くなければならない。第3に、適合を通して等間隔での点について、内部ラダー適合度に対するROXラダー適合度についての決定係数が0.98より高くなければならない。これらの決定係数閾値は、間違ったサイズ分類を生じた試料と正しいサイズ分類を生じた試料との間を正確に識別するレベルを選択することにより、独立した訓練セットから実験的に決定された。

【0265】

8. 1. 6. 2. シグナル大きさQC

シグナル大きさQCは、試料が十分な増幅を受けたことを検証した。十分に増幅されていない試料は、処理中アルゴリズムの仮定を乱し、間違ったまたは偽陰性のジェノタイプが報告され／見逃されることをもたらす可能性があり得る。高レベルにおいて、アルゴリズムは、リピートプロファイルの開始に近位の装置のノイズレベルに対して、リピートプロファイルの開始についての十分なシグナルーノイズ比が存在することを検証する。このQCについてのSNR閾値は、間違ったサイズ分類を生じた試料と、正しいサイズ分類を生じた試料との間を正確に識別するレベルを選択することにより、独立した訓練セットから実験的に決定された。

【0266】

8. 1. 6. 3. 混入QC

混入QCは、試料がオフターゲット増幅を受けず、または間違ったジェノタイプ報告に寄与し得る不適切な試料調製に関係した増幅アーチファクトを含まないことを検証した。遺伝子特異的プライマーを用いて生じ得ない範囲において遺伝子特異的産物ピークが同定された時、試料はこのQCに不合格であった。例えば、遺伝子特異的産物ピークについて導かれたリピート数が0未満のリピート（または等価的に240bp未満）であった時、試料は混入を有するとフラグを付けられた。

【0267】

8. 1. 6. 4. マイナーアレル感度QC

混入QCは、試料が、選ばれたマイナーアレルコーリング閾値について十分低いバックグラウンドノイズを有することを検証した。このQCは、機器のバックグラウンドノイズと試料における最大のジェノタイプピークとの比率に依存する。シグナルにおけるノイズレベルと最大のジェノタイプピークとの比率がマイナーアレル頻度を超える時、その頻度におけるマイナーアレルは正確に同定することができず、その試料は、ユーザーがより厳格に解釈するべきである「リスクがある」QCというフラグを付けられる。

【0268】

8. 2. FMR1サイズ分類性能

FMR1サイズ分類分析についての性能を、複数の機器に渡っていくつかの大きなコホートについて試験した。これらの研究のそれぞれについて、アルゴリズムは、アッセイガイ

10

20

30

40

50

ドラインに従って、正しく、期待されたQC不合格にフラグを付け、QC合格試料の100%をサイズ分類した。サイズ分類についてのアッセイガイドラインは、ジェノタイプ<70リピートについて+/-1リピート、ジェノタイプ<120リピートについて+/-3リピート、およびジェノタイプ≥120リピートについてそのリピート数の+/-5%として定義された。加えて、モザイクピークを有するとオペレータにより同定された、いくつかの試料について、低レベルのモザイクピークが検出された。QCを不合格となった試料は、訓練された手作業でのオペレータにより不合格状態に値するとして独立して検証された。この研究における真のデータは、GeneMapperソフトウェアによる手作業でのサイズ分類により、訓練されたオペレータによって作成された。さらに、本研究で試験されたジェノタイプの分布は、正常な集団において期待されるジェノタイプとはほんのわずかに異なった。この研究におけるジェノタイプの分布は、ジェノタイピング精度がより大きな臨床的影響を生じる症例についてアルゴリズムをストレス試験するために、中間、前突然変異、および完全突然変異範囲において臨床的に関連するアレルを意図的に強化させた。

【0269】

8.2.1. Sally Nolan 性能研究

Sally Nolan 試料セットを、3500 CE 機器において、通常のインプットおよび条件を用いて作製した。これは、アルゴリズムを外部実験室で試験するために設計され、インプット量は、アッセイの通常の使用方法に従う。合計1040個の試料がこの研究において評価され、QCを合格した試料についてジェノタイプの100%がアッセイガイドラインに従って正確にサイズ分類された。図18Aは、本研究における試料のジェノタイプ分布を示す。図18Bは、アッセイ結果について生じた自動でのジェノタイプと手作業でのジェノタイプの比較を示し、表3は、手作業でのカテゴリー別コールと自動でのカテゴリー別コールの比較を詳述する。

【0270】

【表3】

表3:Sally Nolan 試料セットについてのカテゴリー分類能。臨床カテゴリーの手作業での割当と自動割当の間の全ての不一致は、独立したオペレータにより、突然変異境界に近いが、アッセイについてのサイズ分類ガイドライン内にあると解明されたか、または真のセットにおいて本来ラベルされない低レベルのモザイクピークを有すると解明されたかのいずれかであった。

	手分析			
	正常 (0~44)	中間 (45~54)	前突然変異 (55~200)	伸長 (>200)
正常(0~44)	147	0	0	0
中間(45~54)	10	237	0	0
前突然変異 (55~200)	0	16	476	0
伸長 (>200)	0	0	1	18

【0271】

8.2.2. 複数機器のRUSHインプット量研究

この性能研究について、アッセイのオペレータエラーに対するロバスト性を試験するために、試料のRUSHセット（ジェノタイプおよび試料特徴の多様な分布についてアッセイ

を試験することに一般的に用いられる)を、異なるインプット量で流した。加えて、各インプットレベルを、3つの異なる機器において流し、アッセイの複数機器能力を試験した。インプットレベルは、 $100\text{ ng}/\mu\text{L}$ 、 $20\text{ ng}/\mu\text{L}$ 、 $4\text{ ng}/\mu\text{L}$ 、および $0.8\text{ ng}/\mu\text{L}$ であり、それらは、上限および下限においてアッセイについての通常のインプット量($20\text{ ng}/\mu\text{L}$)に及ぶ。合計31個の試料をこの研究において評価し、QCを合格した試料についてのジェノタイプの100%が、アッセイガイドラインに従って正確にサイズ分類された。図19Aは、本研究における試料のジェノタイプ分布を示す。図19Bは、アッセイ結果について生じた自動でのジェノタイプと手作業でのジェノタイプの比較を示し、表4は、手作業でのカテゴリー別コールと自動でのカテゴリー別コールの比較を詳述する。

【0272】

【表4】

表4:研究についてのカテゴリー分類能表。本研究における各固有の試料/インプット量を、異なる機器(3130、3730、3500)に渡って3回、流した。

自動分析	手分析			
	正常 (0~44)	中間 (45~54)	前突然変異 (55~200)	伸長 (>200)
	正常(0~44)	30	0	0
	中間(45~54)	10	0	0
	前突然変異 (55~200)	0	0	33
	伸長 (>200)	0	0	37

【0273】

8.2.3. 人工マイナーアレルインプット滴定研究

この分析は、ジェノタイプの全スペクトルに渡って低レベルのマイナーアレルにフラグを付けることができる。この能力を実証するために、研究は、前突然変異試料のバックグラウンドにおいて低レベルのマイナーアレル(30リピートおよび56リピート)の存在をシミュレートした。76、96、および119におけるマイナーアレルを、異なるインプットレベルで非依存的に、前突然変異バックグラウンド($20\text{ ng}/\mu\text{L}$)へ混合した。この実験におけるマイナーアレルについてのインプットレベルのスペクトルは、 $20\text{ ng}/\mu\text{L}$ 、 $10\text{ ng}/\mu\text{L}$ 、 $5\text{ ng}/\mu\text{L}$ 、 $2.5\text{ ng}/\mu\text{L}$ 、および $1\text{ ng}/\mu\text{L}$ を含んだ。試料において前突然変異ピークをサイズ分類すること、および混合されたモザイクピークをサイズ分類することにおける分析能をアセスメントした。合計40個の試料をこの研究において評価し、QCを合格した試料についてのジェノタイプ(マイナーアレルを含む)の100%が、アッセイガイドラインに従って正確にサイズ分類された。図20Aは、本研究における試料のジェノタイプ分布を示す。図20Bは、アッセイ結果について生じた自動でのジェノタイプと手作業でのジェノタイプとの間の比較を示し、表5は、手作業でのカテゴリー別コールと自動でのカテゴリー別コールとの間の比較を詳述する。

【0274】

10

20

30

40

50

【表 5】

表 5:人工マイナーアレル研究についてのカテゴリー分類能表。各試料は並べ換えマイナーアレルでスパイクされたため、全てのジェノタイプは並べ換えジェノタイプである。

自動分析	手分析				
	正常 (0～44)	中間 (45～54)	前突然変異 (55～200)	伸長 (＞200)	
	正常(0～44)	0	0	0	0
	中間(45～54)	0	0	0	0
	前突然変異 (55～200)	0	0	38	0
	伸長 (＞200)	0	0	0	0

10

【 0 2 7 5 】

20

8. 2. 4. R U S H 試料滴定研究

この性能研究について、試料の R U S H セットは、低い試料インプットレベルにおけるアルゴリズムのロバスト性をストレス試験するために、通常のインプット量から5回の2倍段階希釈を受けた。この実験において試験されたインプットレベルは、100% (20 ng/μl)、50% (10 ng/μl)、25% (5 ng/μl)、12.5% (2.5 ng/μl)、6.2% (1.25 ng/μl)、および3.1% (0.75 ng/μl) を含む。合計66個の試料をこの研究において評価し、QCを合格した試料についてのジェノタイプの100%が、アッセイガイドラインに従って正確にサイズ分類された。図21Aは、本研究における試料のジェノタイプ分布を示す。図21Bは、アッセイ結果について生じた自動でのジェノタイプと手作業でのジェノタイプとの間の比較を示し、表6は、手作業でのカテゴリー別コールと自動でのカテゴリー別コールとの間の比較を詳述する。

30

【 0 2 7 6 】

40

50

【表 6】

表 6: 滴定研究についてのカテゴリー分類能表。

自動分析	手分析			
	正常 (0~44)	中間 (45~54)	前突然変異 (55~200)	伸長 (>200)
	正常(0~44)	15	0	0
	中間(45~54)	0	0	0
	前突然変異 (55~200)	0	0	22
	伸長 (>200)	0	0	22

【0277】

8.2.5. QC 不合格モードシミュレーション研究

本研究において ROX QC 不合格モードを試験するために、試料ジェノタイプの範囲に渡って ROX 不合格の 2 つの異なる型を有する試料セットを作製した。第 1 の不合格モードは、標識された ROX 断片を含まない RUSH 試料セットの CE 分析を含んだ。第 2 の不合格モードは、ROX 400 ラダーを含む (ROX 1000 はこのアッセイに必要とされる) RUST 試料セットの CE 分析を含んだ。合計で、ROX を含まない 13 個の試料および ROX 400 を含む 12 個の試料を分析し、サイズ分類ラダー QC 状態に関するアルゴリズムにより適切に不合格にされた。

【0278】

8.2.6. 結果応答時間の改善

アノテーションデバイス 130 は、FMR1 結果を生じるための応答時間を大幅に改善する。1000 個の試料のコホートについて、手作業でのオペレータは、1 試料あたり 1 分間、必要とすると仮定されて、コホート全体を処理するのに所要 16.6 時間となった。対照的に、アノテーションデバイス 130 は、コホート全体についての結果を 1 分 24 秒間で生じ (2 つのコアを使用する機械で)、結果を得るまでの時間を 700 分の 1 未満への増加を示した。

【0279】

8.2.7. 分析能

アッセイの分析感度をさらに例証するために、図 22A~25B は、アルゴリズムが正しく無効にし、手作業でのオペレータに従ったサイズ分類を生じることができる、様々な異なるジェノタイプおよびまれな場合の範囲を描く。例えば、図 22A および 22B は、正常なジェノタイプを有する試料についての自動サイズ分類分析の結果を描く。図 22C および 22D は、並べ換えジェノタイプを有する試料についての自動サイズ分類分析の結果を描く。図 23A および 23B は、伸長型試料についての自動サイズ分類分析の結果を描く。図 23C および 23D は、低レベルのマイナーアレル同定およびサイズ分類を描く。図 24 は、正常、並べ換え、および伸長型のジェノタイプ範囲に渡るジェノタイプの混合物を有する対照試料についての自動サイズ分類分析の結果を描く。図 25A および 25B は、95% 並べ換え試料のバックグラウンドにおける 5% 完全突然変異試料の混合物を含む対照試料についての自動サイズ分類分析の結果を描く。図 25A は、全てのコールされたジェノタイプを含む完全な試料を描き、一方、図 25B は、完全突然変異コールを示すズームインバージョンを描く。

【0280】

8.3. 考察

AmplideX（登録商標）FMR1 PCRインシリコCGG断片サイズ分析ツールを開発した。この新規なツールは、臨床的に関連したFMR1ジェノタイプの全範囲の迅速かつ正確な同定およびサイズ分類を可能にし、大量試料処理および自動データ分析を支援する。

【0281】

上記の実施例において要約された研究は、いくつかの重要な特徴をもつ、FMR1 PCR断片サイズ分析のための高性能アノテーションデバイスを実証している。これらの中には、（1）FMR1 CGGリピートサイズの全部の臨床的に関連したスペクトルについての正確なジェノタイピング；（2）低レベルのモザイク現象（1%まで下がった）を正確に同定し、かつサイズ分類する能力；（3）Genetic Analyzer（3130、3500、3730）のABIファミリーに関する複数機器の適合性；（4）キャピラリー電気泳動機器により生じるシグナルアーチファクト（気泡、不適切な較正、漏出アーチファクト、シグナル飽和、および収集ノイズ）に対するロバスト性；（5）試料の手作業での処理と比較した場合の有意に低減した分析時間（500分の1未満）、ならびに／または（6）不十分な増幅、混入アーチファクト、低品質のROXラダー、もしくは期待されたアレル検出能対試料から推測されるアレル検出能における不一致からユーザーを守るための自動QC分析および試料フラグ付けがある。

【0282】

上記の実施例は、本開示を例証することを意図され、決して限定することを意図するものではない。開示されたデバイスおよび方法の他の実施形態は、本明細書を考慮すること、ならびに本明細書に開示されたデバイスおよび方法の実施から当業者に明らかであろう。

【0283】

前述の開示された実施形態は、例証のみを目的として提示されている。本開示は、網羅的ではなく、主張された主題を、開示された正味の実施形態に限定しない。改変およびバリエーションが上記の教示に照らせば可能であり、または本発明を実施することから獲得され得ることを当業者は前述の記載から理解しているだろう。いくつかの態様において、開示された実施形態と一致した方法は、開示された方法工程を除外し得、または方法工程の開示された順序もしくは方法工程間の開示された分離度合を変更し得る。例えば、方法工程は、同じまたは類似した目的を達成するために、必要に応じて、省かれ、繰り返され、または組み合わせられ得る。様々な態様において、非一過性コンピュータ可読媒体は、開示された実施形態と一致した方法を実施するための命令を記憶し得る。これらの命令は、開示された方法工程を除外し得、または方法工程の開示された順序もしくは方法工程間の開示された分離度合を変更し得る。例えば、非一過性コンピュータ可読媒体は、同じまたは類似した目的を達成するために、方法工程を、必要に応じて、省き、繰り返し、または組み合わせる、開示された実施形態と一致した方法を実施するための命令を記憶し得る。ある特定の態様において、システムは必ずしも、あらゆる開示された部分を含むことを必要とせず、他の開示されていない部分を含み得る。例えば、システムは、同じまたは類似した目的を達成するために、部分を、必要に応じて、省き、繰り返し、または組み合わせ得る。したがって、主張される主題は、開示された実施形態に限定されず、代わりに、添付された特許請求の範囲によって、それらの等価物の完全な範囲を鑑みて、定義される。

10

20

30

40

50

【図面】
【図 1】

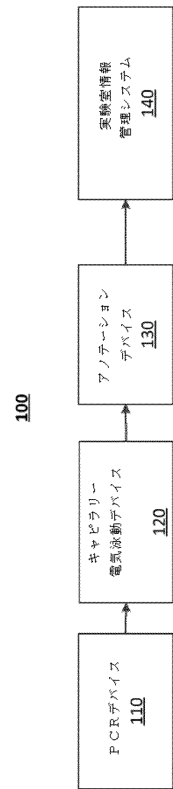


Fig. 1

【図 2】

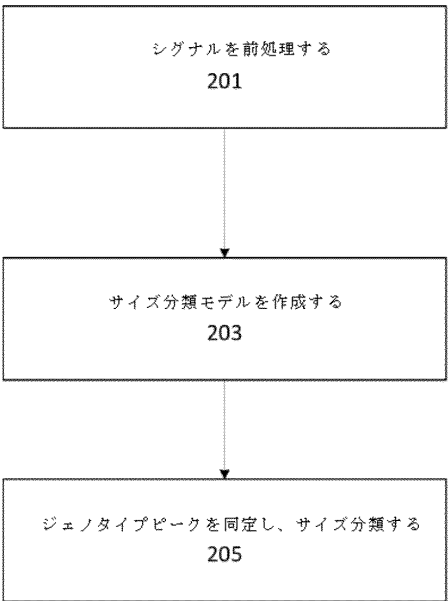


Fig. 2

【図 3 A】

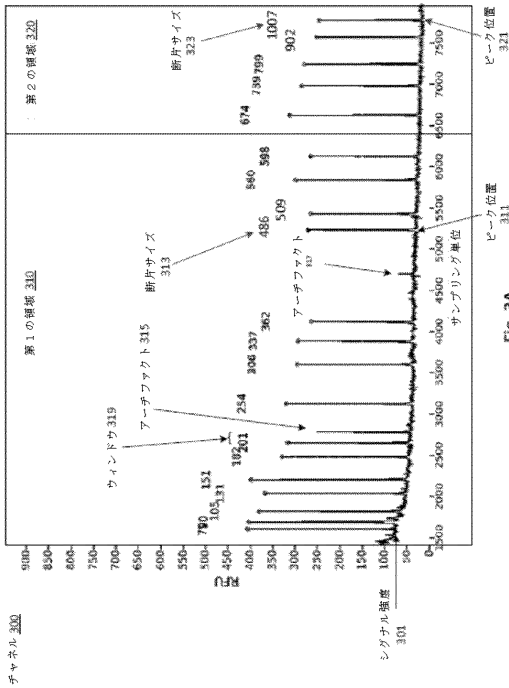


Fig. 3A

【図 3 B】

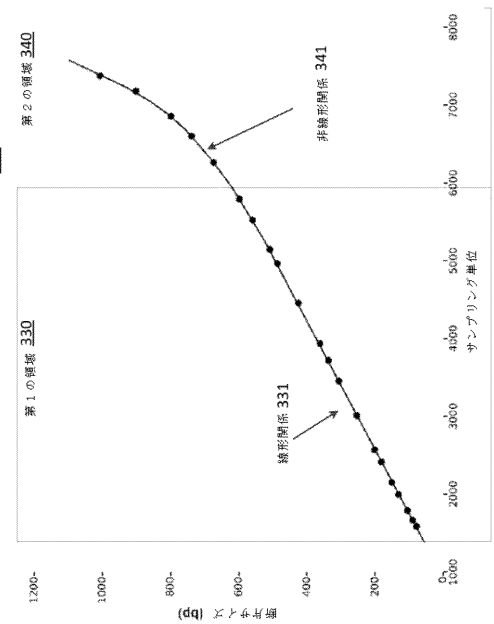


Fig. 3B

【図 4 A】

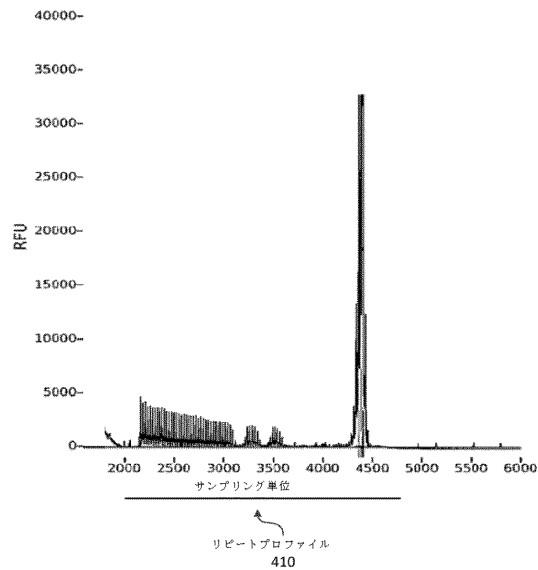


Fig. 4A

【図 4 B】

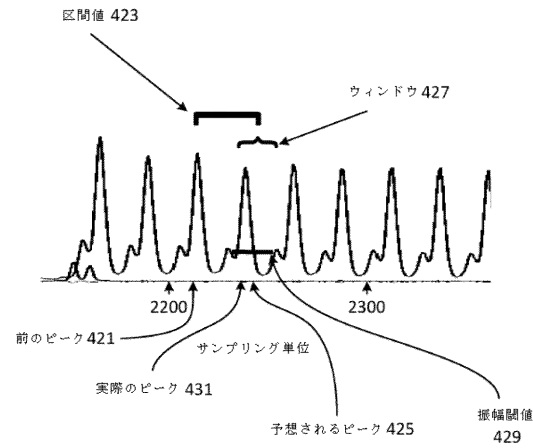


Fig. 4B

【図 5】

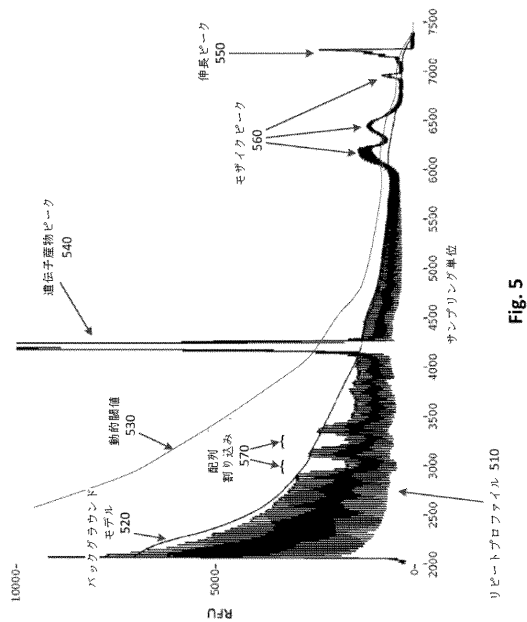


Fig. 5

【図 6 A】

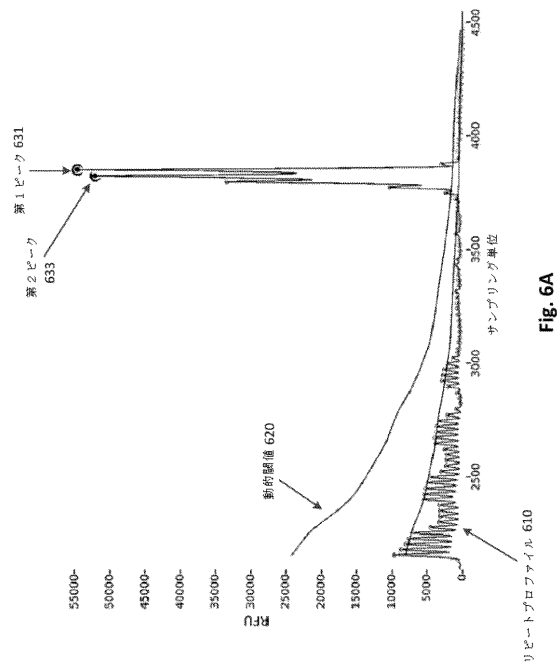


Fig. 6A

10

20

30

40

50

【図 6 B】

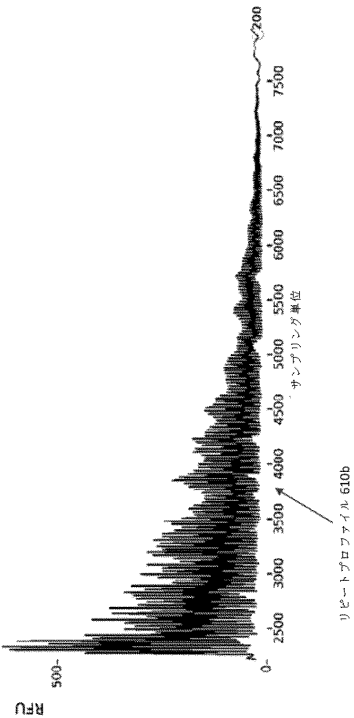


Fig. 6B

【図 7 A】

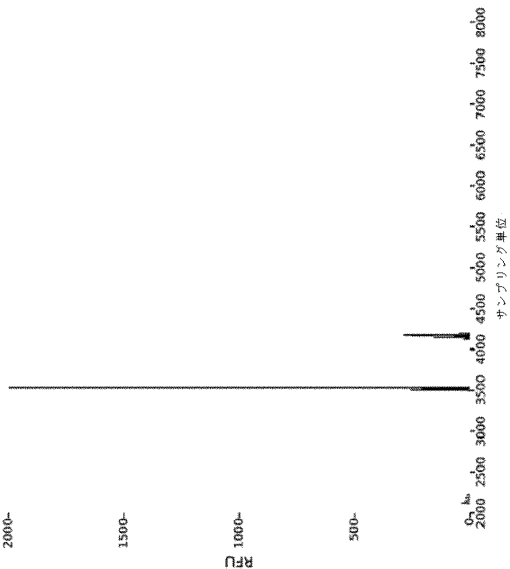


Fig. 7A

【図 7 B】

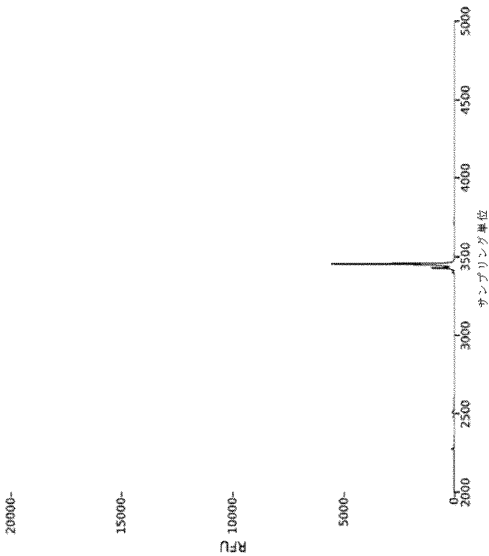


Fig. 7B

【図 7 C】

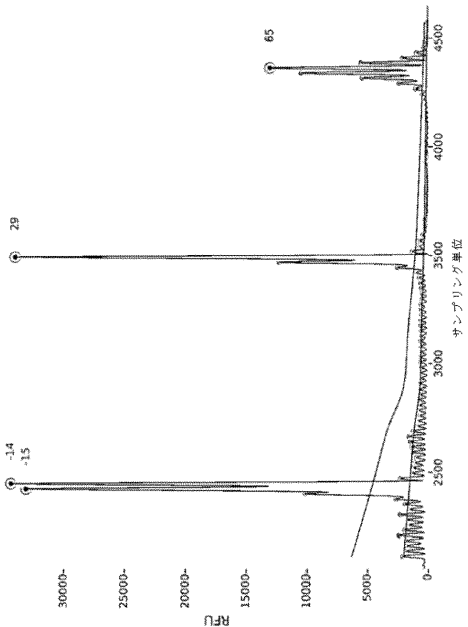


Fig. 7C

10

20

30

40

50

【図 8】

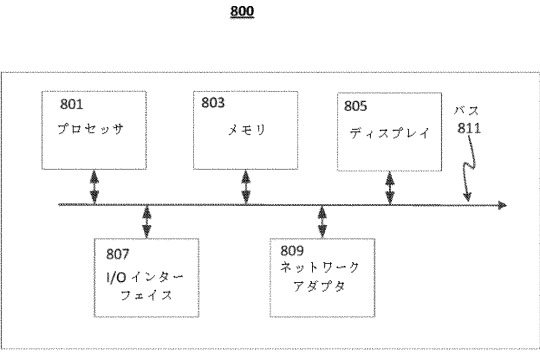


Fig. 8

【図 9】

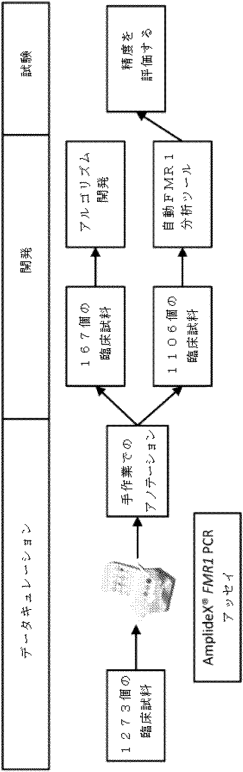


Fig. 9

【図 10】

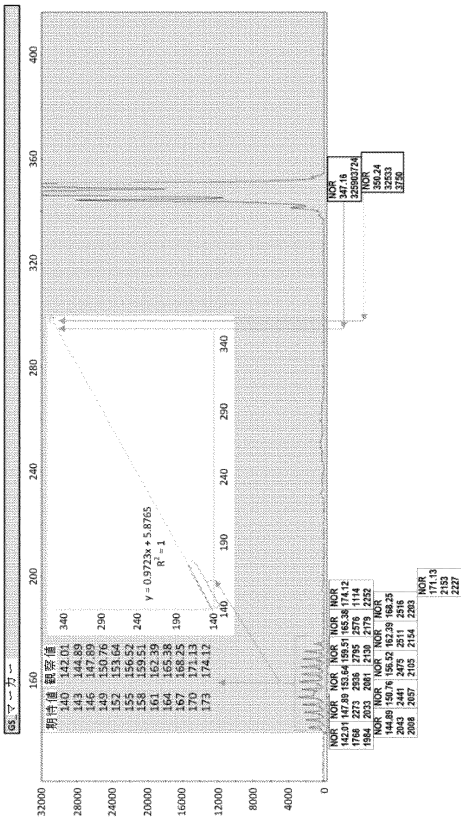


Fig. 10

【図 11】

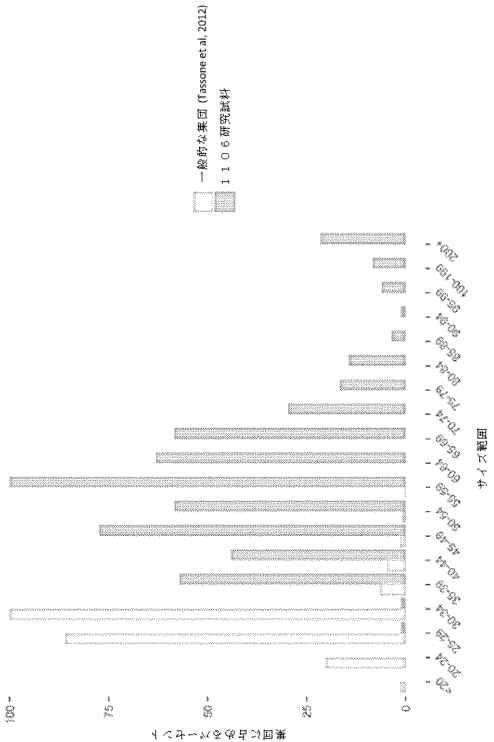


Fig. 11

【図 1 2】

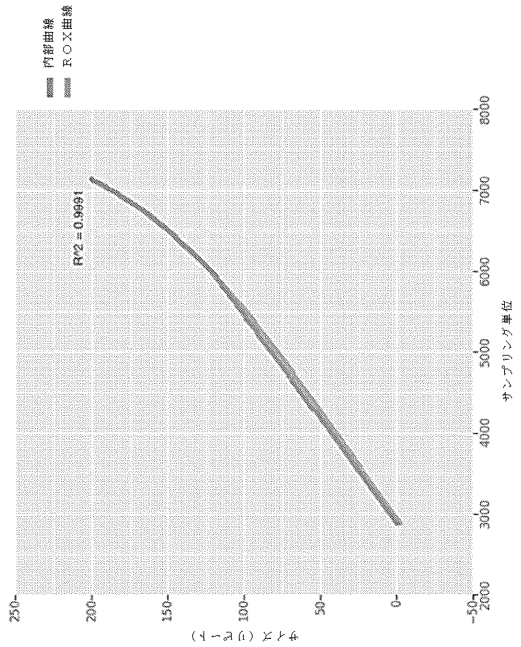


Fig. 12

【図 1 3】

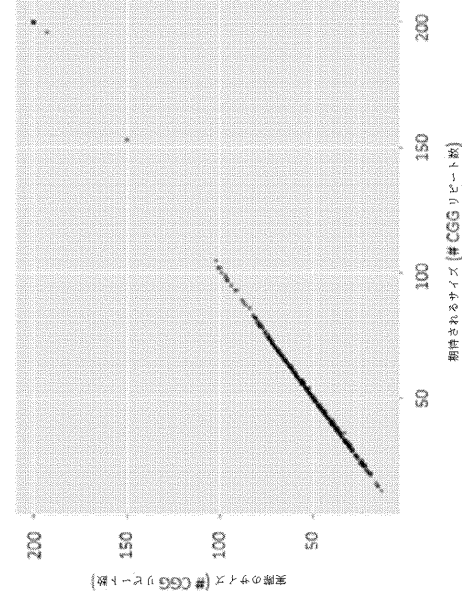


Fig. 13

【図 1 4 A】

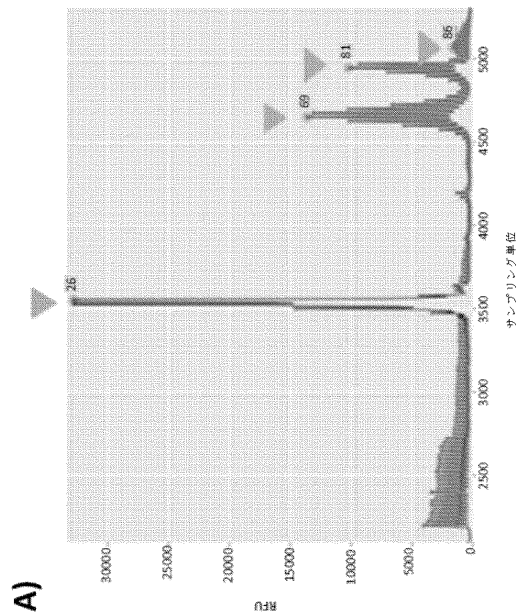


Fig. 14A

【図 1 4 B】

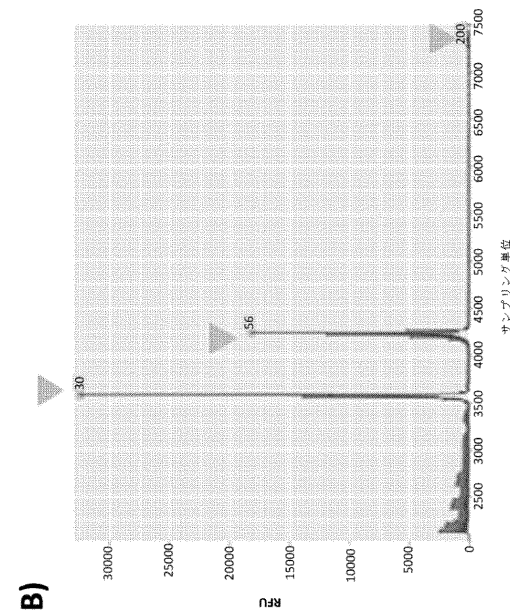


Fig. 14B

10

20

30

40

50

【図 15 A】

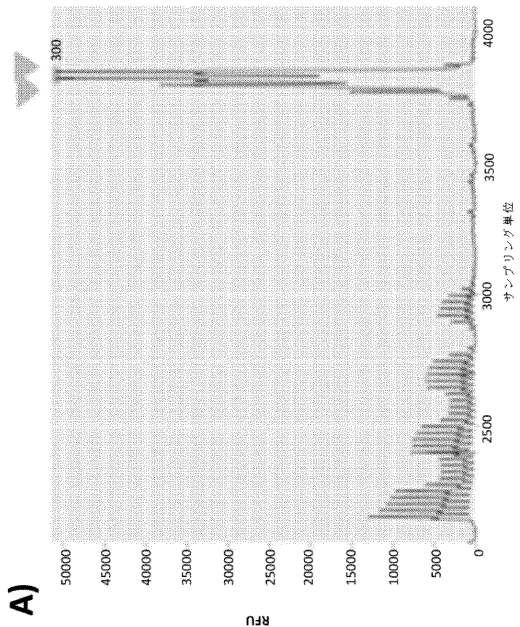


Fig. 15A

【図 15 B】

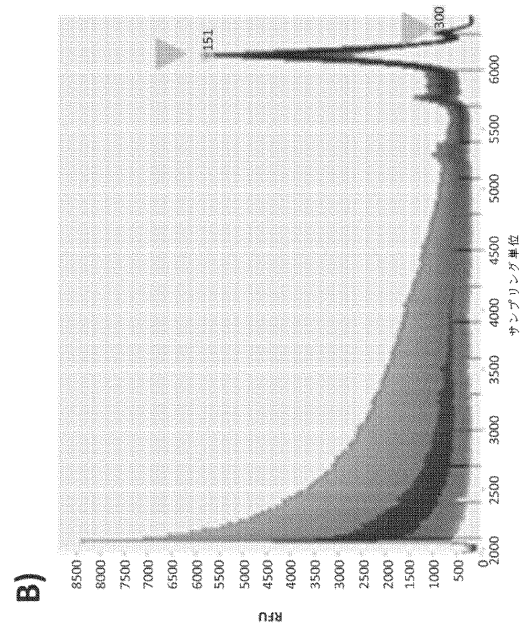


Fig. 15B

【図 16】

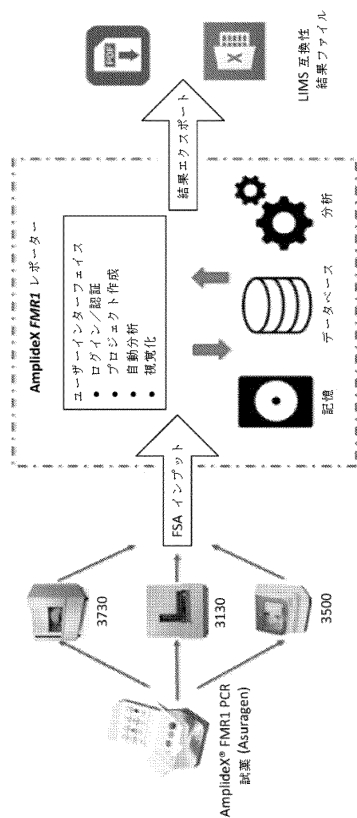


Fig. 16

【図 17】

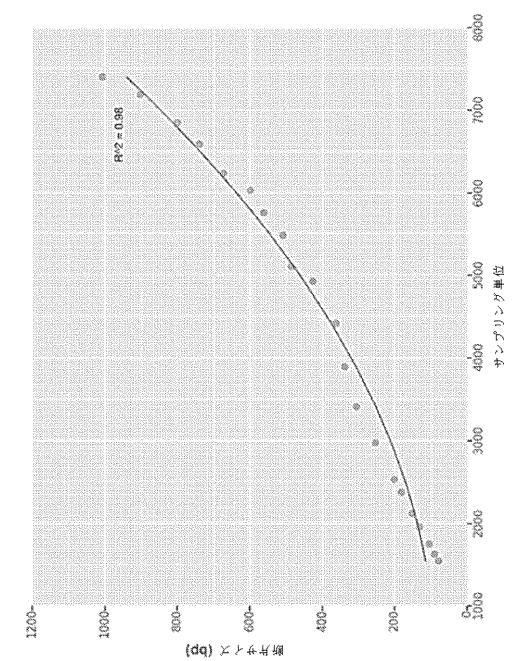


Fig. 17

10

20

30

40

50

【図 18 A】

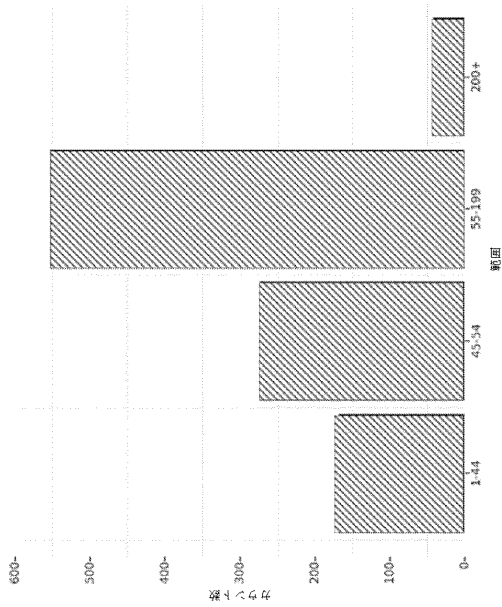


Fig. 18A

【図 18 B】

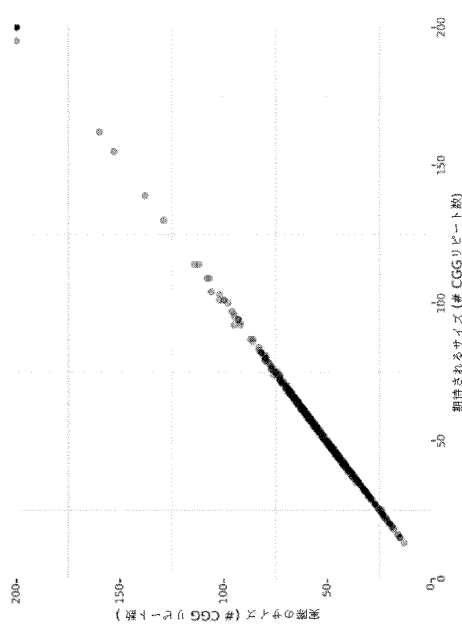


Fig. 18B

【図 19 A】

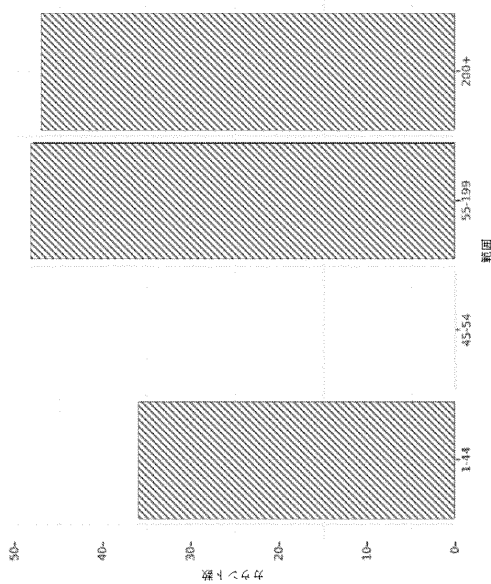


Fig. 19A

【図 19 B】

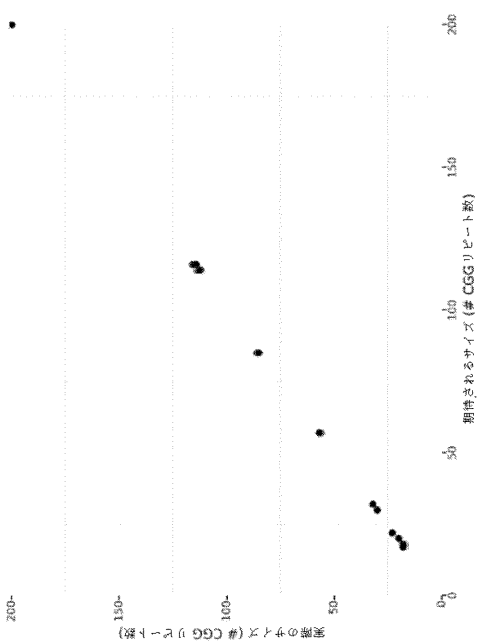


Fig. 19B

10

20

30

40

50

【図 20 A】

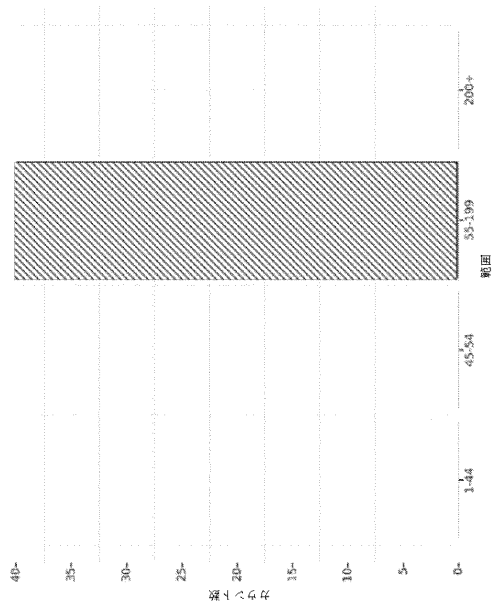


Fig. 20A

【図 20 B】

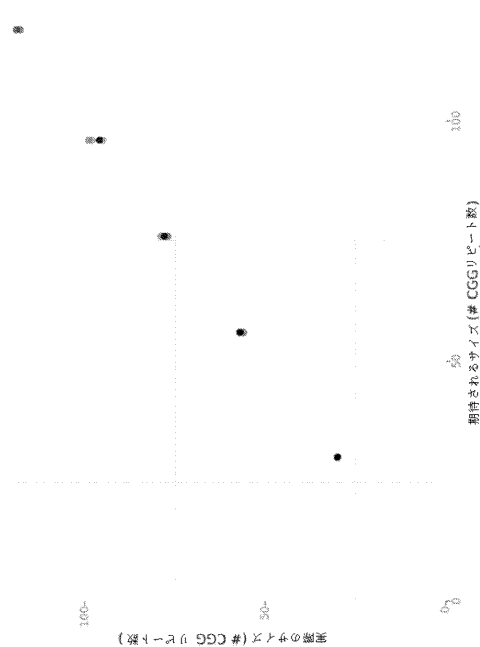


Fig. 20B

【図 21 A】

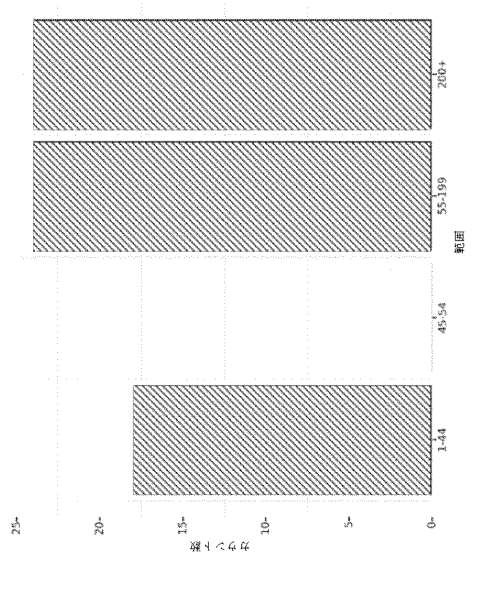


Fig. 21A

【図 21 B】

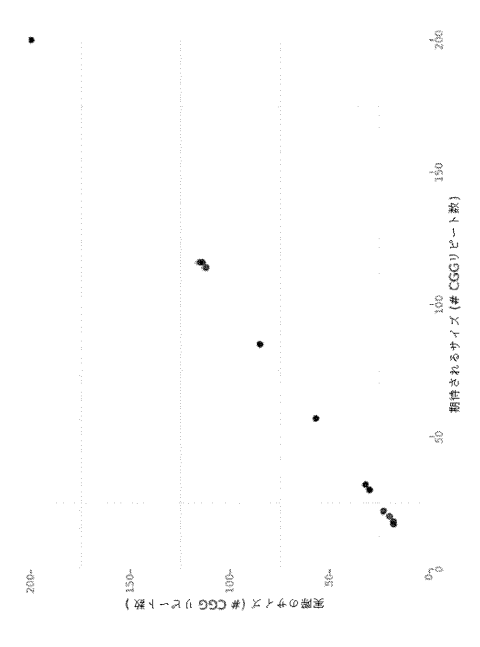


Fig. 21B

10

20

30

40

50

【図 2 2 A】

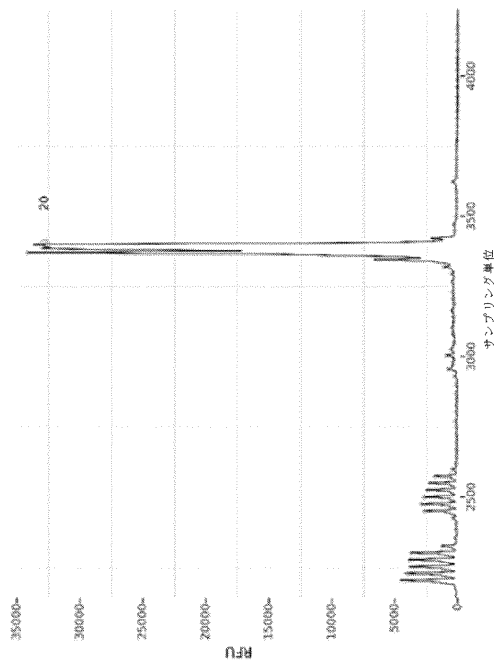


Fig. 22A

【図 2 2 B】

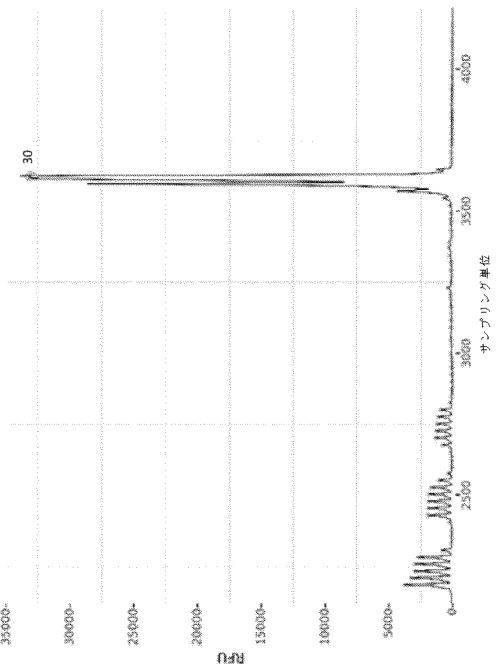


Fig. 22B

【図 2 2 C】

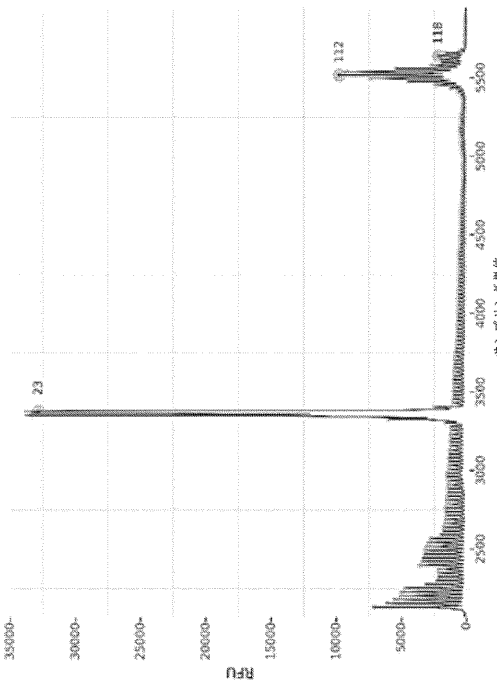


Fig. 22C

【図 2 2 D】

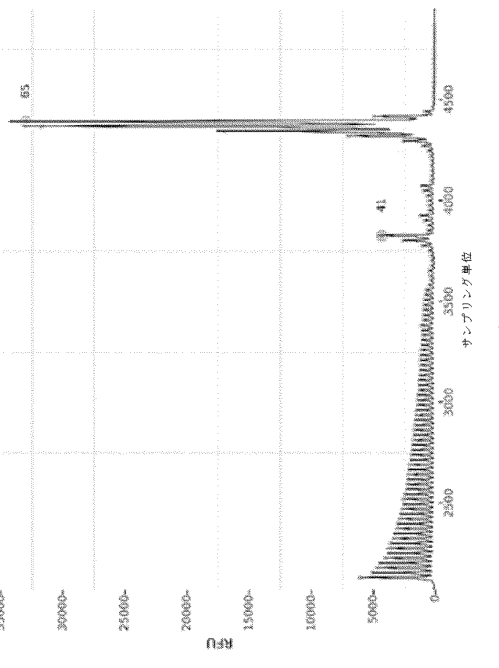


Fig. 22D

10

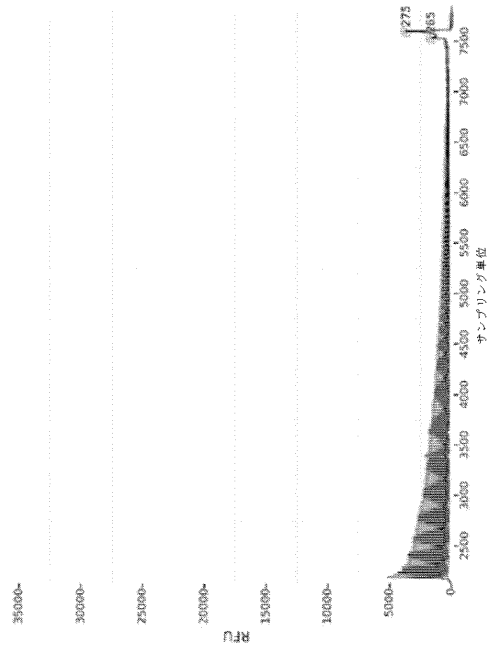
20

30

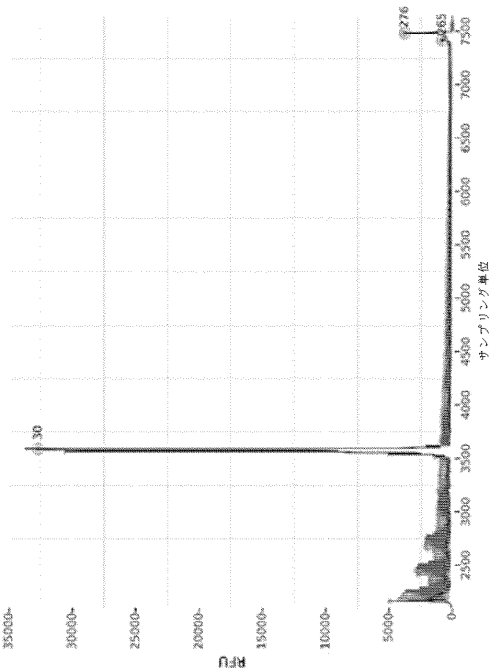
40

50

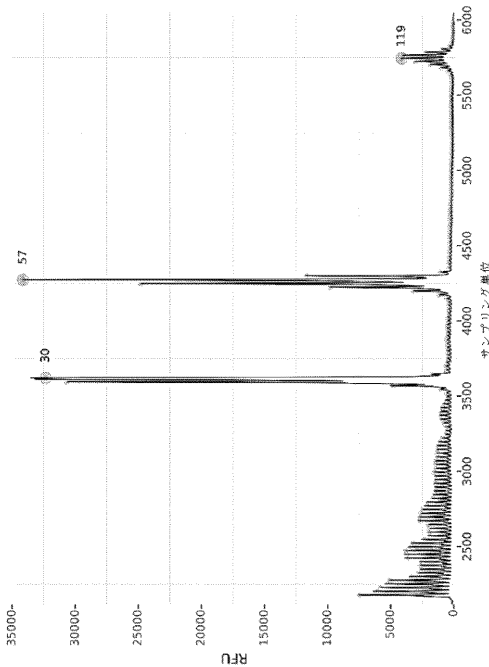
【図 2 3 A】



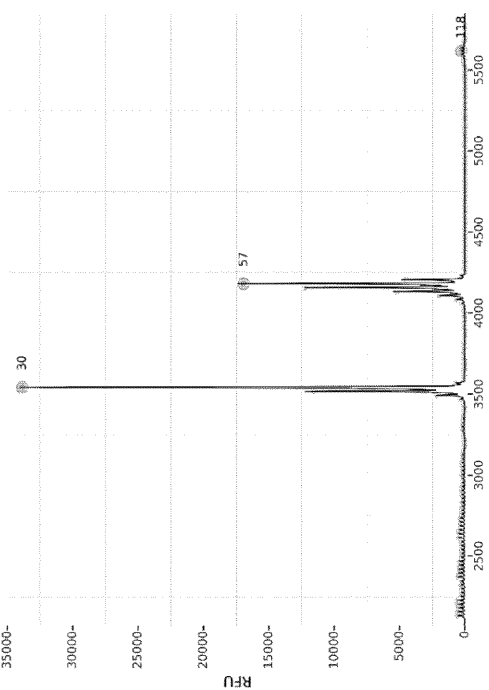
【図 2 3 B】



【図 2 3 C】



【図 2 3 D】



10

20

30

40

50

【図 2 4】

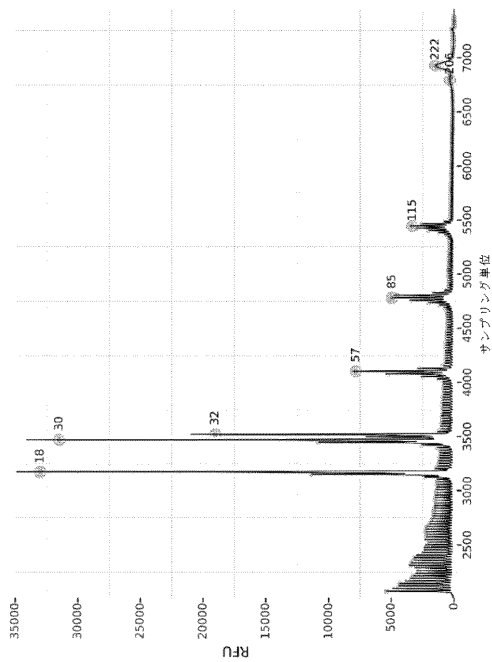


Fig. 24

【図 2 5 A】

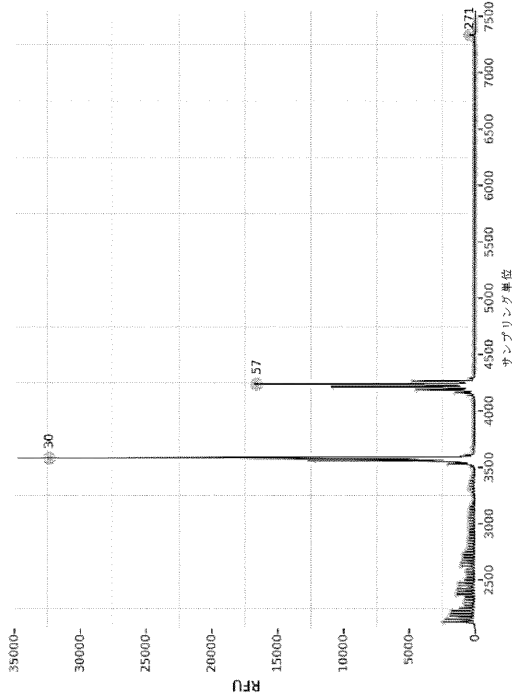


Fig. 25A

【図 2 5 B】

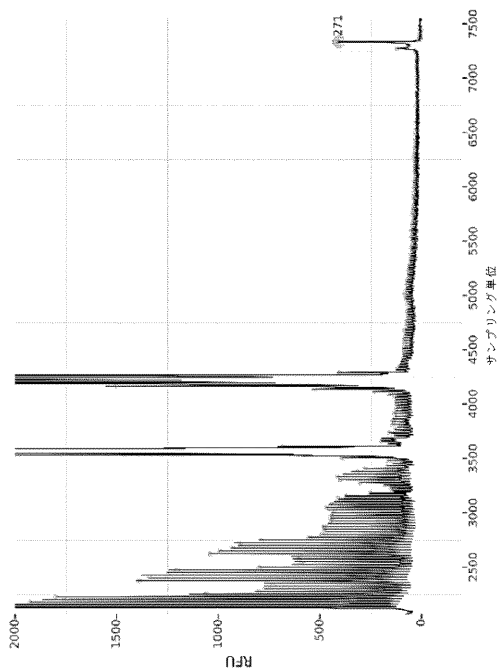


Fig. 25B

【配列表】

0007003035000001.app

10

20

30

40

50

フロントページの続き

- ーダー・パーク、ディア・ホーン・コーブ 1104 番
- (72)発明者 ラガブ・シュロフ
アメリカ合衆国 78735 テキサス州オースティン、デュバル・ストリート 4312 番、アパートメント 104
- (72)発明者 アンドリュー・ハッド
アメリカ合衆国 78735 テキサス州オースティン、リパブリック・オブ・テキサス・ブルバード 5809 番
- (72)発明者 ブレイク・プリンティ
アメリカ合衆国 78759 テキサス州オースティン、ブラフストーン・コーブ 8500 番、スイート・ビー 202
- 審査官 齊藤 貴子
- (56)参考文献 特表 2004-535198 (JP, A)
米国特許出願公開第 2015/0134267 (US, A1)
WILLIAMS, L. C. et al. , Comparative semi-automated analysis of (CAG) repeats in the Huntington disease gene: use of internal standards, Molecular and Cellular Probes, 1999年, Vol.13, No.4, P.283-289
OLEJNICZAK M. et al., Accurate and sensitive analysis of triplet repeat expansions by capillary electrophoresis, Electrophoresis, 2005年, Vol.26, No.11, P.2198-2207
CHEN, L. et al., An information-rich CGG repeat primed PCR that detects the full range of fragile X expanded alleles and minimizes the need for southern blot analysis, The Journal of Molecular Diagnostics, 2010年, Vol.12, No.5, P.589-600
- (58)調査した分野 (Int.Cl., DB名)
C12Q C12N
CAPLUS/MEDLINE/EMBASE/BIOSIS/WPIDS (STN)
JSTPLUS/JMEDPLUS/JST7580 (JDreamII)