

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7129015号
(P7129015)

(45)発行日 令和4年9月1日(2022.9.1)

(24)登録日 令和4年8月24日(2022.8.24)

(51)国際特許分類 F I
G 1 6 B 20/20 (2019.01) G 1 6 B 20/20

請求項の数 10 (全17頁)

(21)出願番号	特願2020-528816(P2020-528816)	(73)特許権者	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(86)(22)出願日	令和1年6月26日(2019.6.26)	(74)代理人	100094112 弁理士 岡部 譲
(86)国際出願番号	PCT/JP2019/025290	(74)代理人	100106183 弁理士 吉澤 弘司
(87)国際公開番号	WO2020/008968	(72)発明者	萩原 久 東京都港区芝五丁目7番1号 日本電気株式会社内
(87)国際公開日	令和2年1月9日(2020.1.9)	(72)発明者	三品 喜典 東京都港区芝五丁目7番1号 日本電気株式会社内
審査請求日	令和2年12月10日(2020.12.10)	(72)発明者	山本 英房 東京都港区芝五丁目7番1号 日本電気株式会社内
(31)優先権主張番号	特願2018-126455(P2018-126455)		
(32)優先日	平成30年7月3日(2018.7.3)		
(33)優先権主張国・地域又は機関	日本国(JP)		

最終頁に続く

(54)【発明の名称】 情報処理システム、変異検出システム、記憶媒体および情報処理方法

(57)【特許請求の範囲】

【請求項1】

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する機能性予測結果取得部と、

前記検査対象遺伝子を含む配列について、CRISPR-Cas9システムを用いた編集に利用可能な、PAM配列およびターゲット配列を含む変異導入部分を特定した結果を取得する変異導入部分特定部と、

該機能性予測結果取得部が取得した結果から人工変異の導入を判定する判定部と、
を備える、ことを特徴とする情報処理システム。

【請求項2】

前記変異導入部分特定部が取得した結果において、前記被検ゲノムの配列中の前記参照ゲノムと異なる配列が前記ターゲット配列中に存在したときに、前記参照ゲノムと異なる配列を有し、前記PAM配列および前記ターゲット配列を含む変異導入部位を前記被検ゲノムの配列から抽出する変異導入部位抽出部をさらに備える、請求項1に記載の情報処理システム。

【請求項3】

前記参照ゲノムが、前記被検ゲノムを有する個体の親株のゲノムである、請求項1または2に記載の情報処理システム。

【請求項4】

前記参照ゲノムが、前記被検ゲノムを有する個体の組織であって、前記被検ゲノムを有

する組織とは異なる組織が有するゲノムである、請求項 1 または 2 に記載の情報処理システム。

【請求項 5】

前記参照ゲノムが、前記被検ゲノムを有する組織と同じ組織から得られたゲノムであって、前記被検ゲノムよりも先に得られたゲノムである、請求項 1 または 2 に記載の情報処理システム。

【請求項 6】

細胞またはウイルスからゲノムを抽出し、精製するゲノム精製部と、
該ゲノム精製部で得られたゲノムの配列を決定するゲノム配列決定部と、
請求項 1 ~ 5 のいずれか 1 項に記載の情報処理システムと、
を備えることを特徴とする、変異検出システム。

10

【請求項 7】

コンピュータに、
被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子の配列について、機能性を予測した結果を取得し、

前記検査対象遺伝子を含む配列について、CRISPR-Cas9 システムを用いた編集に利用可能な、PAM 配列およびターゲット配列を含む変異導入部分を特定した結果を取得し、

該機能性を予測した結果から人工変異の導入を判定する、
ことを実行させる、ことを特徴とする情報処理プログラム。

20

【請求項 8】

さらに、コンピュータに、

前記変異導入部分を特定した結果において、前記被検ゲノムの配列中の前記参照ゲノムと異なる配列が前記ターゲット配列中に存在したときに、前記参照ゲノムと異なる配列を有し、前記 PAM 配列および前記ターゲット配列を含む変異導入部位を前記被検ゲノムの配列から抽出する、

ことを実行させる、請求項 7 に記載の情報処理プログラム。

【請求項 9】

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する機能性予測結果取得ステップと、

前記検査対象遺伝子を含む配列について、CRISPR-Cas9 システムを用いた編集に利用可能な、PAM 配列およびターゲット配列を含む変異導入部分を特定した結果を取得する変異導入部分特定ステップと、

該機能性予測結果取得ステップで取得した結果から人工変異の導入を判定するステップと、

を有する、ことを特徴とする情報処理方法。

30

【請求項 10】

前記変異導入部分特定ステップが取得した結果において、前記被検ゲノムの配列中の前記参照ゲノムと異なる配列が前記ターゲット配列中に存在したときに、前記参照ゲノムと異なる配列を有し、前記 PAM 配列および前記ターゲット配列を含む変異導入部位を前記被検ゲノムの配列から抽出する変異導入部位抽出ステップをさらに有する、請求項 9 に記載の情報処理方法。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理システム、変異検出システム、記憶媒体および情報処理方法に関する。

【背景技術】

【0002】

特許文献 1 には、試料中の大豆事象 MON 8 7 7 0 5 に対応する DNA (DeoxyriboNu

50

cleic Acid) の存在を検出する方法が記載されている。また、特許文献 2 には、ゲノム DNA の任意部位を標的とするガイド RNA (RiboNucleic Acid) 1 およびその発現カセットからなる群より選択される少なくとも 1 種等を細胞または非ヒト生物に導入するステップを含むゲノム編集方法が記載されている。さらに、特許文献 3 には、二本鎖 DNA の標的化された部位を改変する方法が記載されている。

【先行技術文献】

【特許文献】

【0003】

【文献】特表 2012-503989 号公報

特開 2018-011525 号公報

特許第 6206893 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

特許文献 1 に記載の方法では、不特定の人工変異部位を検出することができない。また特許文献 2 および 3 には、人工変異部位を検出する方法について記載されていない。

【0005】

本発明の目的は、上述した課題を鑑み、核酸配列における不特定の人工変異部位の検出を可能にする情報処理システム、変異検出システム、記憶媒体および情報処理方法を提供することにある。

【課題を解決するための手段】

【0006】

本発明の一観点によれば、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する機能性予測結果取得部と、該機能性予測結果取得部で取得した結果から人工変異の導入を判定する判定部と、を備える、ことを特徴とする情報処理システムが提供される。

【0007】

本発明の他の観点によれば、細胞またはウイルスからゲノムを抽出し、精製するゲノム精製部と、該ゲノム精製部で得られたゲノムの配列を決定するゲノム配列決定部と、上記情報処理システムと、を備えることを特徴とする、変異検出システムが提供される。

【0008】

本発明のさらに他の観点によれば、コンピュータに、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子の配列について、機能性を予測した結果を取得し、該機能性を予測した結果から人工変異の導入を判定する、ことを実行させる、ことを特徴とする情報処理プログラムを記憶した記憶媒体が提供される。

【0009】

本発明のまたさらに他の観点によれば、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する機能性予測結果取得ステップと、該機能性予測結果取得ステップで取得した結果から人工変異の導入を判定するステップと、を有する、ことを特徴とする情報処理方法が提供される。

【発明の効果】

【0010】

本発明によれば、核酸配列における不特定の人工変異部位の検出を可能にする情報処理システム、変異検出システム、記憶媒体および情報処理方法を提供することができる。

【図面の簡単な説明】

【0011】

【図 1】第 1 実施形態に係る情報処理システムのハードウェア構成例を示すブロック図である。

【図 2】第 1 実施形態に係る情報処理システムの機能ブロック図である。

【図 3】第 1 実施形態に係る情報処理システムにより行われる処理の概略を示すフローチ

10

20

30

40

50

ャートである。

【図 4】比較解析を説明するための模式図である。

【図 5】検査対象遺伝子を含む領域を一部または全部含む固有配列部分の選抜を説明するための模式図である。

【図 6】変異導入部分を特定するためのアラインメントを説明するための模式図である。

【図 7】参照ゲノムと異なる配列を有し、PAM配列およびターゲット配列を含む変異導入部位の被検ゲノムの配列からの抽出を説明するための模式図である。

【図 8】第 2 実施形態に係る変異検出システムのハードウェア構成例を示すブロック図である。

【図 9】第 2 実施形態に係る変異検出システムの機能ブロック図である。

10

【図 10】第 3 実施形態に係る情報処理システムの機能ブロック図である。

【発明を実施するための形態】

【0012】

以下、図面を参照して、本発明の例示的な実施形態を説明する。図面において同様の要素または対応する要素には同一の符号を付し、その説明を省略または簡略化することがある。

【0013】

[第 1 実施形態]

図 1 は、本実施形態に係る情報処理システム 10 のハードウェア構成例を示すブロック図である。情報処理システム 10 は、例えば、人工変異部位検出装置であり得る。また、情報処理システム 10 は、比較情報処理システムとしてもよい。情報処理システム 10 は、コンピュータの機能を有する。例えば、情報処理システム 10 は、デスクトップ PC (Personal Computer)、ラップトップ PC、タブレット PC、スマートフォン等と一体に構成されていてもよい。情報処理システム 10 は核酸配列における不特定の人工変異部位を検出する機能を備える。情報処理システム 10 は、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果から人工変異の導入を判定することで、人工変異部位を検出できる。

20

【0014】

情報処理システム 10 は、例えば、違法薬物の生産のために編集された植物のゲノムにおける人工変異部位の検出、筋肉増強を目的として人為的に変異が導入された組織中のゲノムにおける人工変異部位の検出、ヒトの組織における個体識別領域の改変を目的とした人工変異部位の検出、生物兵器等の製造を目的として脳などの組織に導入された人工変異部位の検出等に適用され得る。

30

【0015】

情報処理システム 10 は、演算および記憶を行うコンピュータとしての機能を実現するため、CPU (Central Processing Unit) 101、RAM (Random Access Memory) 102、ROM (Read Only Memory) 103 および HDD (Hard Disk Drive) 104 を備える。また、情報処理システム 10 は、通信 I/F (インターフェース) 105、表示装置 106 および入力装置 107 を備える。CPU 101、RAM 102、ROM 103、HDD 104、通信 I/F 105、表示装置 106 および入力装置 107 は、バス 110 を介して相互に接続される。なお、表示装置 106 および入力装置 107 は、これらの装置を駆動するための不図示の駆動装置を介してバス 110 に接続されてもよい。

40

【0016】

図 1 では、情報処理システム 10 を構成する各部が一体の装置として図示されているが、これらの機能の一部は外付け装置により構成されていてもよい。例えば、表示装置 106 および入力装置 107 は、CPU 101 等を含むコンピュータの機能を構成する部分とは別の外付け装置であってもよい。

【0017】

CPU 101 は、ROM 103、HDD 104 等に記憶されたプログラムに従って所定

50

の動作を行うとともに、情報処理システム 10 の各部を制御する機能をも有する。RAM 102 は、揮発性記憶媒体から構成され、CPU 101 の動作に必要な一時的なメモリ領域を提供する。ROM 103 は、不揮発性記憶媒体から構成され、情報処理システム 10 の動作に用いられるプログラム等の必要な情報を記憶する。HDD 104 は、不揮発性記憶媒体から構成され、ゲノム配列等の記憶を行う記憶装置である。

【0018】

通信 I/F 105 は、Wi-Fi (登録商標)、4G 等の規格に基づく通信インターフェースであり、他の装置との通信を行うためのモジュールである。表示装置 106 は、液晶ディスプレイ、OLED (Organic Light Emitting Diode) ディスプレイ等であって、動画、静止画、文字等の表示に用いられる。入力装置 107 は、ボタン、タッチパネル、キーボード、ポインティングデバイス等であって、利用者が情報処理システム 10 を操作するために用いられる。表示装置 106 および入力装置 107 は、タッチパネルとして一体に形成されていてもよい。

10

【0019】

なお、図 1 に示されているハードウェア構成は例示であり、これら以外の装置が追加されていてもよく、一部の装置が設けられていなくてもよい。また、一部の装置が同様の機能を有する別の装置に置換されていてもよい。さらに、一部の機能がネットワークを介して他の装置により提供されてもよく、本実施形態を構成する機能が複数の装置に分散されて実現されるものであってもよい。例えば、HDD 104 は、フラッシュメモリ等の半導体素子を用いた SSD (Solid State Drive) に置換されていてもよく、クラウドストレージに置換されていてもよい。

20

【0020】

図 2 は、本実施形態に係る情報処理システム 10 の機能ブロック図である。情報処理システム 10 は、機能性予測結果取得部 121、変異導入部分特定部 122、変異導入部位抽出部 123、判定部 124、表示部 125 および記憶部 126 を有する。

【0021】

CPU 101 は、ROM 103 等に記憶されたプログラムを RAM 102 にロードして実行することにより、機能性予測結果取得部 121、変異導入部分特定部 122、変異導入部位抽出部 123 および判定部 124 の機能を実現する。これらの各部で行われる処理については後述する。表示部 125 は、機能性予測結果取得部 121、変異導入部分特定部 122、変異導入部位抽出部 123 および判定部 124 で取得あるいは抽出された情報を表示する。CPU 101 は、表示装置 106 を制御することにより表示部 125 の機能を実現する。記憶部 126 は、機能性予測結果取得部 121、変異導入部分特定部 122、変異導入部位抽出部 123 および判定部 124 で取得あるいは抽出されたデータ等を記憶する。CPU 101 は、HDD 104 を制御することにより記憶部 126 の機能を実現する。

30

【0022】

図 3 は、本実施形態に係る情報処理システム 10 により行われる処理の概略を示すフローチャートである。図 3 のフローチャートに沿って情報処理システム 10 により行われる処理の概略を説明する。なお、以下の説明において、配列という語がゲノムあるいは遺伝子に係る語として用いられるときは、それぞれゲノムあるいは遺伝子の塩基配列を指す場合がある。

40

【0023】

図 3 のステップ S101 において、機能性予測結果取得部 121 は、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する。被検ゲノムは、人為的に変異が導入されたことの有無を検査する対象となるゲノムであり、参照ゲノムは、人為的に変異が導入される前の被検ゲノムと相同的な配列を有するゲノムである。また、検査対象遺伝子は、検出対象とする人工変異の導入により獲得されることが想定される形質に寄与する遺伝子である。

【0024】

50

被検ゲノムを有する個体としては、ゲノムを有する個体であれば特に制限がなく、例えば、ヒト、ヒト以外の動物、植物、酵母、カビ、真性細菌およびウイルス等が挙げられる。

【0025】

参照ゲノムは、被検ゲノムを有する個体の親株のゲノムであることが好ましい。親株としては、被検ゲノムを有する個体の一世代前の個体もしくは被検ゲノムを有する個体のクローンが挙げられる。これら一世代前の個体のゲノムもしくは被検ゲノムを有する個体のクローンのゲノムは、被検ゲノムと同一の配列を有する。つまり、人工変異部位以外の配列が本来同じであるため、人工変異部位の検出の負荷が小さくて済み、また誤検出の可能性も低くすることができる。

【0026】

被検ゲノムを有する個体が複数の組織を有する高等な生物である場合は、同一個体の、被検ゲノムを有する組織とは異なる組織のゲノム配列も本来同一の配列を有する。そのため、先に述べたのと同じ理由で、参照ゲノムは、被検ゲノムを有する個体の組織であって、前記被検ゲノムを有する組織とは異なる組織が有するゲノムであることが好ましい。

【0027】

また、例えば、被検ゲノムを有する組織と同じ組織の一部が、ゲノム編集を受けるよりも前に採取され、保管されていたと推定される場合は、この編集を受ける前の、被検ゲノムを有する組織と同じ組織から参照ゲノムを得ることもできる。この場合、被検ゲノムと参照ゲノムは、互いに同一個体の同一組織に由来するため、本来同一の配列を有する。よって、先に述べたのと同じ理由で、参照ゲノムは、被検ゲノムを有する組織と同じ組織から得られたゲノムであって、被検ゲノムよりも先に得られたゲノムであることが好ましい。

【0028】

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子は、例えば、次のようにして定めることができる。

【0029】

機能性予測結果取得部121は、まず、被検ゲノムの配列と、参照ゲノムの配列とを比較解析することで、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する部分を特定する。なお、比較解析による被検ゲノムの配列中の、参照ゲノムと異なる配列を有する部分の特定は、情報処理システム10とは別の情報処理システムが行っても良い。

【0030】

比較解析に供する被検ゲノムの配列および参照ゲノムの配列としては、全ゲノムの配列でも良いし、変異が導入される可能性のある部位が特定の領域に限られる場合は、その特定の領域のゲノムの配列でも良い。導入された変異をもらさずに全て検出できることから、全ゲノムの配列を取得し、全ゲノムの配列を比較解析に供することが好ましい。ただし、変異の導入部位が特定の領域に限られることの確かさが高いときは、特定の領域に限ってゲノムの配列を取得しても良い。例えば、特定の形質を獲得するのに関わる遺伝子が、特定の候補に限られることが明らかであるような場合は、その候補となる遺伝子に対応する領域のゲノム配列のみを取得しても良い。

【0031】

被検ゲノムの配列および参照ゲノムの配列は、細胞、または個体がウイルスの場合はウイルス本体からゲノムを抽出し、抽出したゲノムの塩基配列を解析することで決定することができる。細胞は、例えば、個体が酵母やカビ等であれば、個体をそのままゲノム抽出に供する細胞として用いることができる。また、例えば、個体がヒト、ヒト以外の動物および植物であれば、組織の一部を採取してゲノム抽出に供する細胞として用いることができる。このとき、例えば個体がヒトやヒト以外の動物であれば、無痛で採取できる口腔内細胞や唾液を、ゲノム抽出に供する組織として用いることができる。

【0032】

細胞、またはウイルス本体からのゲノムの抽出は、ゲノムを有する個体に合わせた適切な処理を施すことにより行うことができる。またゲノムを有する個体に合わせて適切な市販のキット等を用いて行っても良い。例えば、ヒトの口腔内細胞等から抽出する場合であ

10

20

30

40

50

れば、NucleoSpin (登録商標) DNA Forensic (タカラバイオ社製) 等を用いることができる。

【0033】

抽出により得られたゲノムの塩基配列は、例えば、NextSeq シリーズ、HiSeq X シリーズ (イルミナ社製) や PacBio (登録商標) RS II/Sequel (登録商標) システム (パックバイオ社製) 等の市販の DNA シーケンサーを用いることにより決定することができる。

【0034】

参照ゲノムの配列としては、NHGRI (National Human Genome Research Institute)、NCBI (National Center for Biotechnology Information)、DDBJ (DNA Data Bank of Japan) センターおよび東北メディカル・メガバンク機構等の公的機関により公開されているデータベースにある配列を用いてもよい。データベースから配列を取得する場合は、参照ゲノムの配列と相溶性が高い配列を選択して用いる。参照ゲノムの配列と相溶性が高い配列としては、例えば、同じ種に属する個体のゲノム配列が挙げられる。

10

【0035】

比較解析は、例えば、BLASTZ 等の比較解析プログラムにより行うことができる。図4は比較解析を説明するための模式図である。比較解析では、被検ゲノムの配列401と、参照ゲノムの配列402を比較し、参照ゲノム中の部分配列403と対応する部位であり、参照ゲノム中の部分配列403と異なる配列を有する被検ゲノム中の変異部位404を特定する。比較解析により特定した、変異部位404は、具体的には、参照ゲノムと比較したときに1つ以上の塩基が欠失、挿入または置換された部分である。

20

【0036】

変異部位は、人工変異部位、自然変異 (突然変異) 部位および種の多様性に起因する部位を含む。これらのうち、参照ゲノムを、親株のゲノムあるいは同一個体の組織が有するゲノムとすることで、固有配列部分に種の多様性に起因する部位が含まれないようにすることができる。そのため、人工変異部位の検出の負荷が小さくなり、また誤検出の可能性も低くすることができる。

【0037】

続いて、機能性予測結果取得部121は、変異部位と、変異部位に隣接する参照ゲノムと同一の配列の一部とを含む配列を固有配列部分とし、固有配列部分のうち、検査対象遺伝子を含む領域を一部または全部含む固有配列部分を選抜する。なお、検査対象遺伝子を含む領域を一部または全部含む固有配列部分の選抜は、情報処理システム10とは別の情報処理システムが行っても良い。固有配列部分が含む、参照ゲノムと同一の配列の長さは、任意に定めることができる。選抜された固有配列部分中の、検査対象遺伝子に対応する配列部分が、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子である。

30

【0038】

検査対象遺伝子を含む領域を一部または全部含む固有配列部分の選抜は、具体的には、例えば、次のようにして行うことができる。

【0039】

図5は、検査対象遺伝子を含む領域を一部または全部含む固有配列部分の選抜を説明するための模式図である。まず、機能性予測結果取得部121は、全ての固有配列部分501の配列を母集団502として第1の検査用対照配列503および第2の検査用対照配列504の相溶性検索を行う。

40

【0040】

第1の検査用対照配列503は、検査対象遺伝子の一部の配列または全配列を含む配列である。検査対象遺伝子の配列は、例えば、NHGRI、NCBI、DDBJ センターおよび東北メディカル・メガバンク機構等の公的機関により公開されているデータベースより取得することができる。人工変異部位の検出の感度を高めるため、第1の検査用対照配列503は長いほど好ましく、検査対象遺伝子の全配列を含むことが最も好ましい。

50

【 0 0 4 1 】

また、第2の検査用対照配列504は、検査対象遺伝子の配列に隣接する配列である。第2の検査用対照配列504として用いる検査対象遺伝子の配列に隣接する配列は、検査対象遺伝子の配列の上流の配列でも良いし下流の配列でも良い。また、第2の検査用対照配列504は複数用意してもよい。例えば、図5に示すように、検査対象遺伝子の配列の上流に隣接する配列である第2の検査用対照配列504と、検査対象遺伝子の配列の下流に隣接する配列である第2の検査用対照配列504とを用意して用いても良い。第2の検査用対照配列の長さは任意に定めることができるが、固有配列部分501が含む、参照ゲノムと同一の配列の長さよりも短いことが好ましい。第2の検査用対照配列を、固有配列部分501が含む、参照ゲノムと同一の配列の長さよりも短くすることで、相同性検索による検索漏れを抑制することができる。

10

【 0 0 4 2 】

続いて、機能性予測結果取得部121は、相同性検索により検索された固有配列部分501の配列と、第1の検査用対照配列503との相同性、および/または第2の検査用対照配列504との相同性が、規定値よりも高い固有配列部分501を選抜する。選抜された固有配列部分501は、検査対象遺伝子領域を一部または全部含む部分である。判断基準となる相同性の規定値は、検査対象遺伝子等に応じて任意に定めることができる。

【 0 0 4 3 】

人工変異が検査対象遺伝子に導入されている場合であって、導入された変異が検査対象遺伝子の配列を大きく変更するものでない場合は、その変異が導入された検査対象遺伝子を含む固有配列部分は、第1の検査用対照配列と高い相同性を有し、選抜される。

20

【 0 0 4 4 】

人工変異が検査対象遺伝子に導入されている場合であって、導入された変異が検査対象遺伝子の配列を大きく変更するものである場合は、その変異が導入された検査対象遺伝子を含む固有配列部分は、第1の検査用対照配列との相同性が低くなる。しかし固有配列部分は、参照ゲノムと異なる配列に隣接する参照ゲノムと同一の配列を一部含む。すなわち、固有配列部分は、変異が導入された検査対象遺伝子に隣接する、変異が導入されていない配列を含み、この配列は、第2の検査用対照配列に対応する部分である。そのため、導入された変異が検査対象遺伝子の配列を大きく変更するものである場合、例えば検査対象遺伝子を全て欠失させるような変異であっても、第2の検査用対照配列と高い相同性を有する固有配列部分として選抜することができる。

30

【 0 0 4 5 】

ただし、機能性予測結果取得部121は、固有配列部分中の変異部位が、第1の検査用対照配列に対応する部分に含まれず、第2の検査用対照配列に対応する部分に含まれる場合は、その固有配列部分を選抜しない。そのような固有配列部分は、検出の対象とする人工変異ではないと考えられるためである。

【 0 0 4 6 】

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子の機能性の予測は、対象とする検査対象遺伝子に合わせて予め定めた基準に沿って行うことができる。ここで機能性とは、人工変異の導入により想定される形質の獲得を意味する。

40

【 0 0 4 7 】

すなわち、例えば、検査対象遺伝子が導入された変異により本来有する機能を失うことで、想定された形質が獲得される場合は、検査対象遺伝子が本来有する機能を失う変異であるか否かを判断する基準を予め定める。検査対象遺伝子の特に上流側(5'末端側)における3の倍数でない塩基数の挿入または欠失は、遺伝子発現の翻訳過程でフレームシフトを起こすことから、検査対象遺伝子が本来有する機能を失う変異である可能性が高い。また、検査対象遺伝子の特に上流側(5'末端側)に、塩基置換または挿入により終止コドンが導入される変異も、遺伝子発現の転写過程で未成熟な伝令RNAを生成する原因となることから、検査対象遺伝子が本来有する機能を失う変異である可能性が高い。さらに、検査対象遺伝子の大半または全てを欠失させる変異も、検査対象遺伝子が本来有する機能を

50

失う変異となり得る。

【0048】

また、例えば、被検ゲノム中には本来存在しない検査対象遺伝子の変異として導入され、検査対象遺伝子が有する機能により、想定された形質が獲得される場合は、検査対象遺伝子が導入されたか否かを判断する基準として定めることができる。

【0049】

さらに、例えば、導入された変異により検査対象遺伝子が本来有する機能とは異なる機能を獲得することで、想定された形質が獲得される場合は、検査対象遺伝子が本来有する機能とは異なる機能を獲得するか否かを判断する基準を予め定める。

【0050】

機能性を予測するために用いる基準を定めるにあたっては、例えばPubMed等の論文検索エンジンを利用し、対象とする形質に関するキーワードを元に学術論文を取得し、参照しても良い。また、例えばJpred等のプログラムを利用し、検査対象遺伝子の塩基配列から読み取ったアミノ酸配列を元に翻訳されるペプチド(タンパク質)の構造を予測したり、またPDB(Protein Data Bank)等のデータベースにあるタンパク質の立体構造を参照したりしても良い。

【0051】

ステップS101において機能性予測結果取得部121は、上記のように一定の基準に沿って機能性を予測した結果を取得する。

【0052】

ステップS102において、変異導入部分特定部122は、検査対象遺伝子を含む配列について、CRISPR-Cas9システムを用いた編集に利用可能な、PAM配列およびターゲット配列を含む変異導入部分を特定した結果を取得する。

【0053】

ここで、検査対象遺伝子を含む配列は、上記における選抜された固有配列部分に対応する。また、PAM配列およびターゲット配列は、それぞれCRISPR-Cas9システムを用いた編集に利用されるプロトスペーサー隣接モチーフ(Protospacer Adjacent Motif)およびPAM配列に隣接する標的配列である。

【0054】

図6は、変異導入部分を特定するためのアラインメントを説明するための模式図である。例えば、変異導入部分特定部122は、変異導入部分の特定を、次のようにして行うことができる。まず、上記の選抜された固有配列部分501に対してPAM配列601をアラインメントする。続いて、PAM配列601の位置を特定し、さらにPAM配列601の上流に隣接する特定の塩基数の配列を、ターゲット配列602として特定する。アラインメントはペアワイズアラインメント等により行うことができる。なお、変異導入部分の特定は、情報処理システム10とは異なる情報処理システムが行っても良い。

【0055】

CRISPR-Cas9システムを用いた編集に利用されるCas9ヌクレアーゼが由来する菌種と、そのCas9ヌクレアーゼの各亜型により認識されるPAM配列の組み合わせとしては、例えば、5'-NGG(*Streptococcus pyogenes*、II型)、5'-CCN(*Sulfolobus solfataricus*、I-A1型)、5'-TCN(*Sulfolobus solfataricus*、I-A2型)、5'-TTC(*Haloquadratum walsbyi*、I-B型)、5'-AWG(*Escherichia coli*、I-E型)、5'-CC(*Escherichia coli*、I-F型)、5'-CC(*Pseudomonas aeruginosa*、I-F型)、5'-NNAAGA(*Streptococcus thermophilus*、II-A型)、5'-NGG(*Streptococcus agalactiae*、II-A型)が挙げられる。

【0056】

ターゲット配列として特定する配列の塩基数は、位置を特定したPAM配列に対応したCas9ヌクレアーゼの各亜型に応じて定める。例えば、CRISPR-Cas9システムを用いた編集に利用されるCas9ヌクレアーゼが*Streptococcus pyogenes*、II

10

20

30

40

50

型由来のものである場合は19または20塩基である。

【0057】

CRISPR-Cas9システムを用いた編集では、PAM配列に隣接するターゲット配列に対応する部分に変異が導入される。そのため、固有配列部分における被検ゲノムの配列と参照ゲノムの配列とで異なっている塩基が、ターゲット配列中に存在する場合は、CRISPR-Cas9システムを利用して人工的に導入された変異であると考えることができる。

【0058】

ステップS103において、変異導入部位抽出部123は、変異導入部分特定部122が取得した結果において、被検ゲノムの配列中の参照ゲノムと異なる配列がターゲット配列中に存在したときに、参照ゲノムと異なる配列を有し、PAM配列およびターゲット配列を含む変異導入部位を被検ゲノムの配列から抽出する。

10

【0059】

変異導入部位抽出部123は、参照ゲノムと異なる配列を有し、PAM配列およびターゲット配列を含む変異導入部位の被検ゲノムの配列からの抽出を、例えば、次のようにして選抜された固有配列部分の情報を取得することで行うことができる。

【0060】

図7は、参照ゲノムと異なる配列を有し、PAM配列およびターゲット配列を含む変異導入部位の被検ゲノムの配列からの抽出を説明するための模式図である。変異導入部位抽出部123は、まず、全ての上記固有配列部分501の配列を母集団502として、被検ゲノムの配列中の参照ゲノムと異なる配列が存在することが特定されたPAM配列601とターゲット配列602の組み合わせについて同一性検索を行う。続いて、変異導入部位抽出部123は、同一性が規定値よりも高い固有配列を選抜する。規定値は任意に定めることができる。なお、PAM配列601とターゲット配列602の組み合わせについての同一性検索および同一性が規定値よりも高い固有配列の選抜は情報処理システム10とは異なる情報処理システムが行ってもよい。

20

【0061】

CRISPR-Cas9システムを用いた編集では、目的とした部位と異なる部位に、非特異的に編集がなされることがあることが知られている。そのため、被検ゲノムの配列に、CRISPR-Cas9システムを用いて人工変異が導入された場合、検査対象遺伝子以外の配列にも同時に変異が導入されている可能性がある。被検ゲノム中の非特異的に変異が導入された部位は、参照ゲノムと異なる配列を有するため、先に述べた比較解析により機能性予測結果取得部121は固有配列部分として特定する。

30

【0062】

また、CRISPR-Cas9システムを用いた編集では、PAM配列およびターゲット配列が含まれるため、非特異的に変異が導入された部位は、上記のように同一性検索を行うことで高い同一性を有する固有配列部分として特定され、選抜することができる。つまり、変異導入部位抽出部123によって抽出された結果が、規定値で定められる一定の値よりも高い同一性を有する固有配列部分を含んでいたときには、CRISPR-Cas9システムを用いた編集が行われたと考えることができる。

40

【0063】

ステップS104において、判定部124は、人工変異の導入を判定する。判定部124は人工変異が導入されたことを判定することで、人工変異部位を検出することができる。判定部124は、例えば、変異導入部位抽出部123によって抽出された結果が、規定値で定められる一定の値よりも高い同一性を有する固有配列部分を1つ以上含むことをもって、人工変異が導入されたと判定することができる。

【0064】

本実施形態では、情報処理システム10が、機能性予測結果取得部121、変異導入部分特定部122および変異導入部位抽出部123を全て含む例を示したが、これに限らない。

50

【 0 0 6 5 】

例えば、情報処理システム 1 0 は、変異導入部位抽出部 1 2 3 を有さず、機能性予測結果取得部 1 2 1 および変異導入部分特定部 1 2 2 のみを有してもよい。このとき判定部 1 2 4 は、例えば、変異導入部分特定部 1 2 2 で取得した結果が、変異がターゲット配列中に存在することを特定したことを含んでいたときに、人工変異が導入されたと判定することができる。

【 0 0 6 6 】

また、例えば、情報処理システム 1 0 は、変異導入部分特定部 1 2 2 および変異導入部位抽出部 1 2 3 を有さず、機能性予測結果取得部 1 2 1 のみを有してもよい。このとき判定部 1 2 4 は、例えば、機能性予測結果取得部 1 2 1 で取得した結果が、変異が導入された検査対象遺伝子が機能性を有すると予測されたことを含んでいたときに、人工変異が導入されたと判定することができる。また、検出対象とする人工変異が導入されるのに用いられる方法は、C R I S P R - C a s 9 システムを用いた編集に限定されない。

10

【 0 0 6 7 】

判定部 1 2 4 で判定された結果の確度を高める観点から、情報処理システム 1 0 は、変異導入部分特定部 1 2 2 を有することが好ましく、さらに変異導入部位抽出部 1 2 3 を有することがより好ましい。

【 0 0 6 8 】

[第 2 実施形態]

上述した情報処理システム 1 0 は、ゲノム精製部およびゲノム配列決定部とともに変異検出システムを構成することができる。

20

【 0 0 6 9 】

図 8 は、第 2 実施形態に係る変異検出システムのハードウェア構成例を示すブロック図である。変異検出システム 8 0 は、ゲノム精製装置 8 0 1 と、DNAシーケンサー 8 0 2 と、情報処理システム 1 0 を備える。情報処理システム 1 0 の構成は、先ほど述べたのと同様である。なお、図 8 に示されているハードウェア構成は例示であり、これら以外の装置が追加されていてもよく、一部の装置が設けられていなくてもよい。また、一部の装置が同様の機能を有する別の装置に置換されていてもよい。さらに、一部の機能がネットワークを介して他の装置により提供されてもよく、本実施形態を構成する機能が複数の装置に分散されて実現されるものであってもよい。

30

【 0 0 7 0 】

図 9 は、第 2 実施形態に係る変異検出システム 8 0 の機能ブロック図である。ゲノム精製装置 8 0 1 はゲノム精製部 8 9 1 の機能を実現し、DNAシーケンサー 8 0 2 はゲノム配列決定部 8 9 2 の機能を実現する。

【 0 0 7 1 】

ゲノム精製部 8 9 1 は、被検ゲノムを有する細胞あるいは個体からゲノムを精製する。さらに、被検ゲノムを有する個体の親株の細胞あるいは個体、または被検ゲノムを有する個体の組織の細胞からゲノムを精製しても良い。細胞、またはウイルス本体からのゲノムの抽出は、ゲノムを有する個体に合わせた適切な処理を施すことにより行うことができる。

【 0 0 7 2 】

ゲノム配列決定部 8 9 2 では、ゲノム精製部 8 9 1 で精製されたゲノムの塩基配列を決定する。決定する塩基配列はゲノムの全塩基配列でも良いし、ゲノムの特定の領域の塩基配列でも良いが、ゲノムの全塩基配列を決定することが好ましい。ゲノムの塩基配列は、例えば次世代シーケンシングにより決定することができる。

40

【 0 0 7 3 】

情報処理システム 1 0 は、ゲノム配列決定部 8 9 2 で決定したゲノムの塩基配列を用い、人工変異部位を検出する。情報処理システム 1 0 における人工変異部位の検出の詳細は、上で述べたのと同様である。

【 0 0 7 4 】

[第 3 実施形態]

50

図10は、第3実施形態に係る情報処理システム30の機能ブロック図である。情報処理システム30は、機能性予測結果取得部321および判定部324を備える。機能性予測結果取得部321は、被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する。判定部324は、人工変異の導入を判定する。

【0075】

本実施形態によれば、核酸配列における不特定の人工変異部位を検出することができる情報処理システムを提供することができる。

【0076】

なお、上述の実施形態は、いずれも本発明を実施するにあたっての具体化の例を示したものに過ぎず、これらによって本発明の技術的範囲が限定的に解釈されてはならないものである。すなわち、本発明はその技術思想、又はその主要な特徴から逸脱することなく、様々な形で実施することができる。

10

【0077】

上述の実施形態の一部または全部は、以下の付記のようにも記載されうるが、以下には限られない。

【0078】

(付記1)

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能性を予測した結果を取得する機能性予測結果取得部と、

20

該機能性予測結果取得部が取得した結果から人工変異の導入を判定する判定部と、
を備える、ことを特徴とする情報処理システム。

【0079】

(付記2)

前記検査対象遺伝子を含む配列について、CRISPR-Cas9システムを用いた編集に利用可能な、PAM配列およびターゲット配列を含む変異導入部分を特定した結果を取得する変異導入部分特定部をさらに備える、付記1に記載の情報処理システム。

【0080】

(付記3)

前記変異導入部分特定部が取得した結果において、前記被検ゲノムの配列中の前記参照ゲノムと異なる配列が前記ターゲット配列中に存在したときに、前記参照ゲノムと異なる配列を有し、前記PAM配列および前記ターゲット配列を含む変異導入部位を前記被検ゲノムの配列から抽出する変異導入部位抽出部をさらに備える、付記2に記載の情報処理システム。

30

【0081】

(付記4)

前記参照ゲノムが、前記被検ゲノムを有する個体の親株のゲノムである、付記1～3のいずれか1項に記載の情報処理システム。

【0082】

(付記5)

前記参照ゲノムが、前記被検ゲノムを有する個体の組織であって、前記被検ゲノムを有する組織とは異なる組織が有するゲノムである、付記1～3のいずれか1項に記載の情報処理システム。

40

【0083】

(付記6)

前記参照ゲノムが、前記被検ゲノムを有する組織と同じ組織から得られたゲノムであって、前記被検ゲノムよりも先に得られたゲノムである、付記1～3のいずれか1項に記載の情報処理システム。

【0084】

(付記7)

50

細胞またはウイルスからゲノムを抽出し、精製するゲノム精製部と、
 該ゲノム精製部で得られたゲノムの配列を決定するゲノム配列決定部と、
 付記 1 ~ 6 のいずれか 1 項に記載の情報処理システムと、
 を備えることを特徴とする、変異検出システム。

【 0 0 8 5 】

(付記 8)

コンピュータに、

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子の配列につい
 て、機能性を予測した結果を取得し、

該機能性を予測した結果から人工変異の導入を判定する、

ことを実行させる、ことを特徴とする情報処理プログラムが記憶された記憶媒体。

10

【 0 0 8 6 】

(付記 9)

さらに、コンピュータに、

前記検査対象遺伝子を含む配列について、C R I S P R - C a s 9 システムを用いた編
 集に利用可能な、P A M 配列およびターゲット配列を含む変異導入部分を特定した結果を
 取得する、

ことを実行させる、情報処理プログラムが記憶された付記 8 に記載の記憶媒体。

【 0 0 8 7 】

(付記 1 0)

さらに、コンピュータに、

前記変異導入部分を特定した結果において、前記被検ゲノムの配列中の前記参照ゲノム
 と異なる配列が前記ターゲット配列中に存在したときに、前記参照ゲノムと異なる配列を
 有し、前記 P A M 配列および前記ターゲット配列を含む変異導入部位を前記被検ゲノムの
 配列から抽出する、

ことを実行させる、情報処理プログラムが記憶された付記 9 に記載の記憶媒体。

20

【 0 0 8 8 】

(付記 1 1)

被検ゲノムの配列中の、参照ゲノムと異なる配列を有する検査対象遺伝子について機能
 性を予測した結果を取得する機能性予測結果取得ステップと、

該機能性予測結果取得ステップで取得した結果から人工変異の導入を判定するステッ
 プと、

を有する、ことを特徴とする情報処理方法。

30

【 0 0 8 9 】

(付記 1 2)

前記検査対象遺伝子を含む配列について、C R I S P R - C a s 9 システムを用いた編
 集に利用可能な、P A M 配列およびターゲット配列を含む変異導入部分を特定した結果を
 取得する変異導入部分特定ステップをさらに有する、付記 1 1 に記載の情報処理方法。

【 0 0 9 0 】

(付記 1 3)

前記変異導入部分特定ステップが取得した結果において、前記被検ゲノムの配列中の前
 記参照ゲノムと異なる配列が前記ターゲット配列中に存在したときに、前記参照ゲノムと
 異なる配列を有し、前記 P A M 配列および前記ターゲット配列を含む変異導入部位を前記
 被検ゲノムの配列から抽出する変異導入部位抽出ステップをさらに有する、付記 1 2 に記
 載の情報処理方法。

40

【 0 0 9 1 】

この出願は、2018年7月3日に出願された日本出願特願2018-126455を
 基礎とする優先権を主張し、その開示の全てをここに取り込む。

【符号の説明】

【 0 0 9 2 】

50

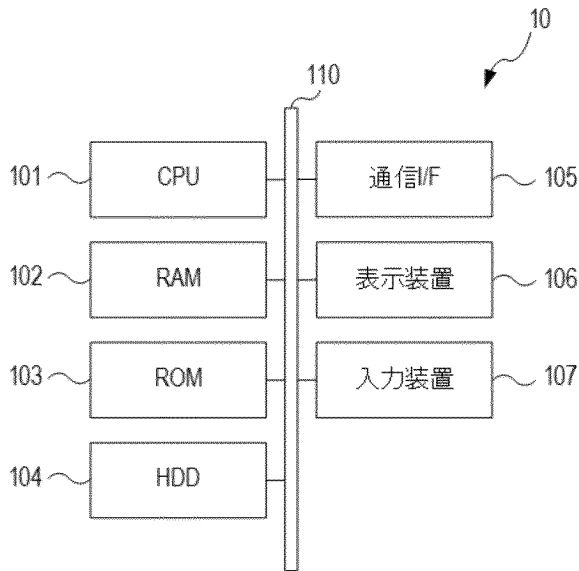
- 10、30 情報処理システム
- 80 変異検出システム
- 101 CPU
- 102 RAM
- 103 ROM
- 104 HDD
- 105 通信I/F
- 106 表示装置
- 107 入力装置
- 110 バス
- 121、321 機能性予測結果取得部
- 122 変異導入部分特定部
- 123 変異導入部位抽出部
- 124、324 判定部
- 125 表示部
- 126 記憶部
- 801 ゲノム精製装置
- 802 DNAシーケンサー
- 891 ゲノム精製部
- 892 ゲノム配列決定部

10

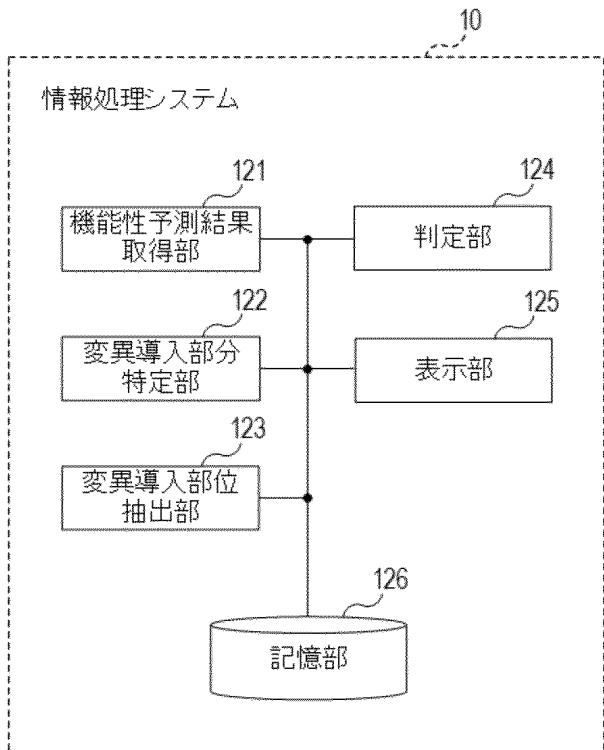
20

【図面】

【図1】



【図2】

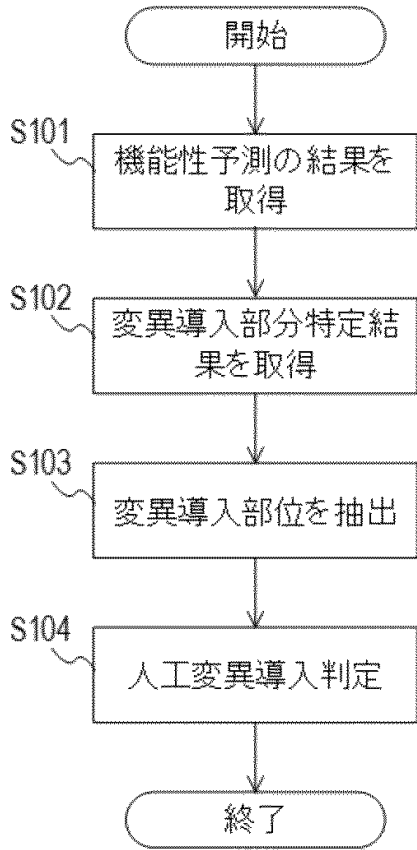


30

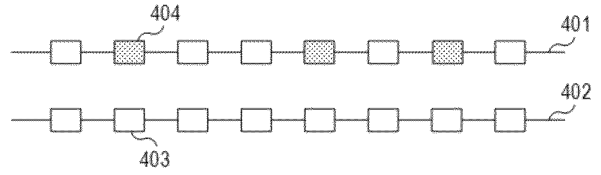
40

50

【図3】



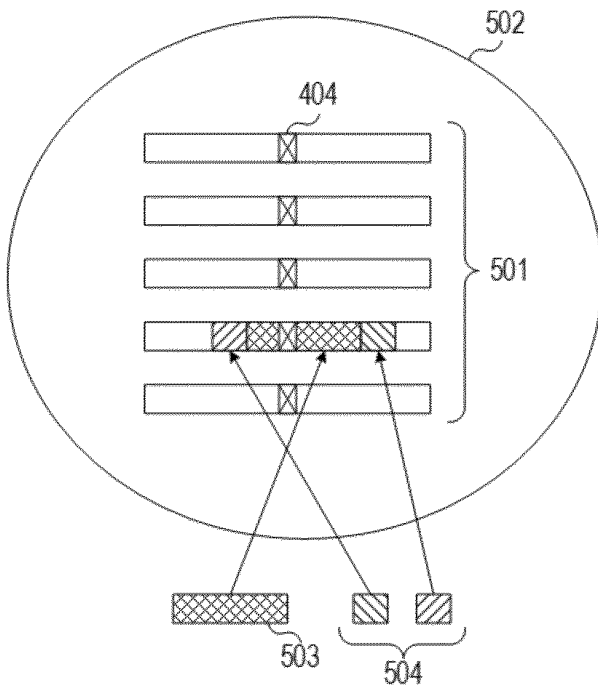
【図4】



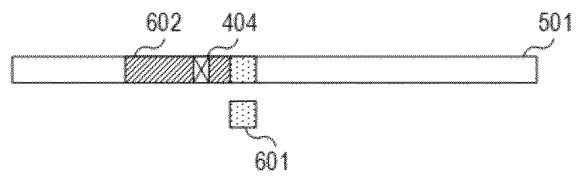
10

20

【図5】



【図6】

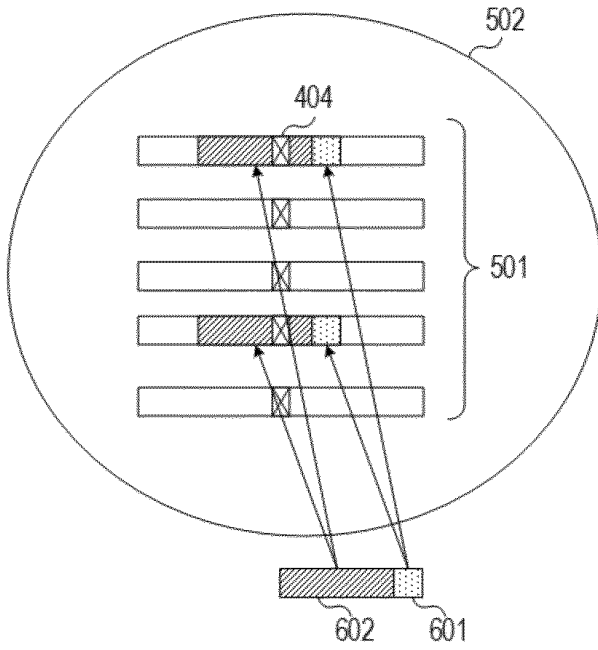


30

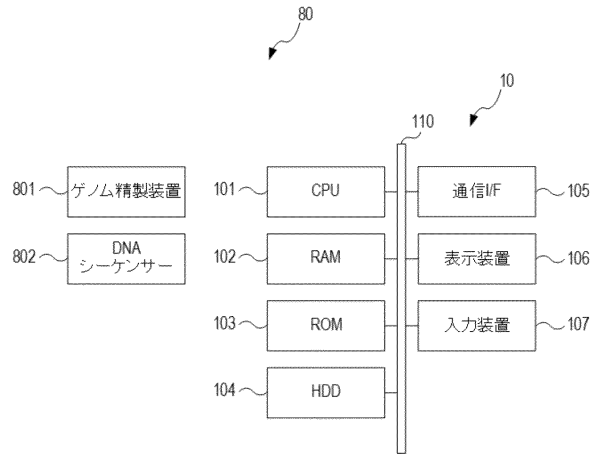
40

50

【図7】



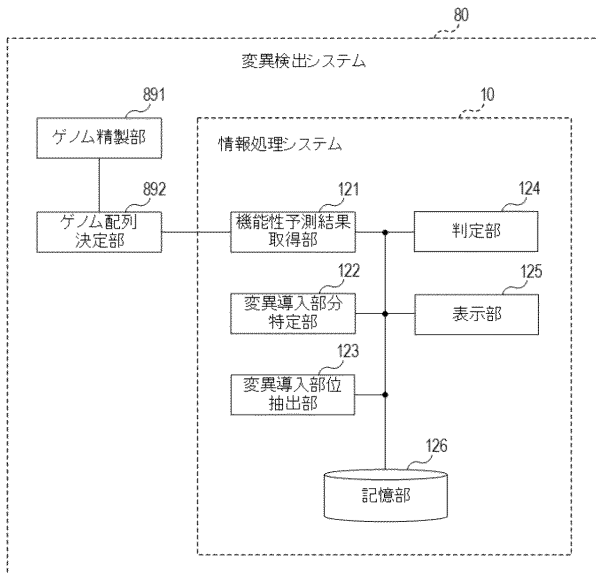
【図8】



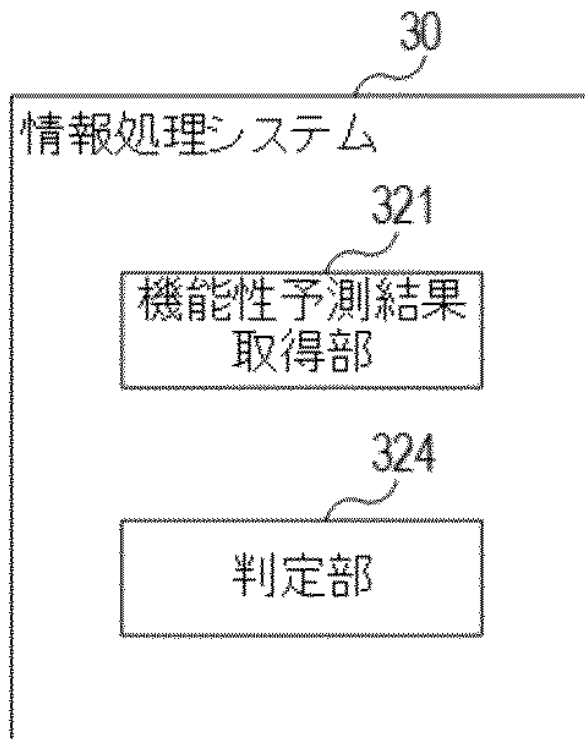
10

20

【図9】



【図10】



30

40

50

フロントページの続き

株式会社内

(72)発明者 菅 侑子

東京都港区芝五丁目7番1号 日本電気株式会社内

審査官 山内 裕史

(56)参考文献 特表2015-501974(JP,A)

特表2016-531569(JP,A)

特表2016-536021(JP,A)

米国特許出願公開第2015/0356239(US,A1)

(58)調査した分野 (Int.Cl., DB名)

G16B 5/00 - 99/00