(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2018年8月2日(02.08.2018)

WIPO | PCT

(10) 国際公開番号
WO 2018/139205 A1

(51) 国際特許分類:
G06F 19/10 (2011.01)　*G06F 19/24* (2011.01)
G06F 17/30 (2006.01)　*C12N 15/09* (2006.01)

(21) 国際出願番号：　PCT/JP2018/000539

(22) 国際出願日：　2018年1月11日(11.01.2018)

(25) 国際出願の言語：　日本語

(26) 国際公開の言語：　日本語

(30) 優先権データ：
特願 2017-010416　2017年1月24日(24.01.2017)　JP

(71) 出願人: 富士通株式会社 (FUJITSU LIMITED) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 Kanagawa (JP).

(72) 発明者: 河場　基行 (KAWABA, Motoyuki); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP).
宇治橋 善史(UJIBASHI, Yoshifumi); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP).

(74) 代理人: 向山　直樹 (MUKOUYAMA, Naoki); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 Kanagawa (JP).

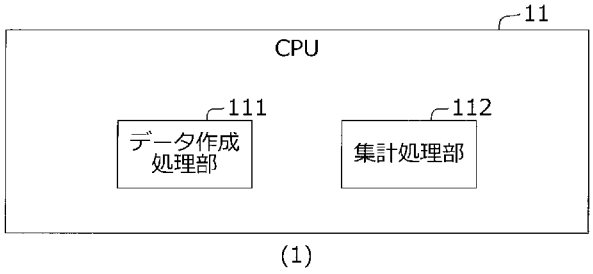
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,

WO 2018/139205 A1

(54) **Title:** INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING SYSTEM, PROGRAM AND INFORMATION PROCESSING METHOD

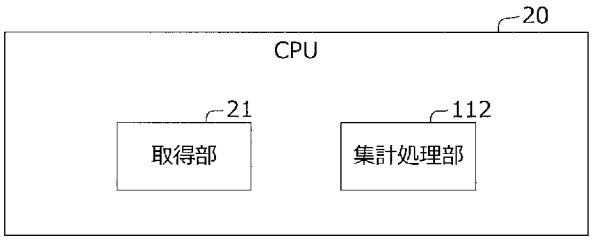
(54) 発明の名称： 情報処理装置、情報処理システム、プログラム及び情報処理方法

[図10]



(1)

21... ACQUISITION UNIT
111... DATA CREATION PROCESSING UNIT
112... AGGREGATION PROCESSING UNIT



(2)

(57) **Abstract:** [Problem] To reduce the amount of data stored in a memory for a plurality of arrays, each of which includes a plurality of mutation patterns. [Solution] This information processing device includes a processing unit 111 and a storage unit. When mutation patterns are identical at the same mutation position among a plurality of arrays, the processing unit 111 excludes the identical mutation patterns from storage targets. The storage unit stores a plurality of arrays in which the exclusion has been processed by the processing unit 111.

[続葉有]

QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

(57) 要約:【課題】それぞれ複数の変異パターンを含む複数の配列において、メモリに記憶されるデータ量を削減する。【解決手段】情報処理装置は、処理部１１１と記憶部とを備える。処理部１１１は、複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行なう。記憶部は、処理部１１１によって除外する処理が施された複数の配列を記憶する。

明　細　書

発明の名称 ：

情報処理装置、情報処理システム、プログラム及び情報処理方法

技術分野

[0001] 本発明は、情報処理装置、情報処理システム、プログラム及び情報処理方法に関する。

背景技術

[0002] 遺伝情報には、個体差を生じさせる部分、すなわち個体によって遺伝情報が相違する部分（「変異」又は「バリアント」と称されてもよい。）が数千万箇所存在しており、このうちの一部の変異についての遺伝情報は特定の疾患の罹患と相関が有る可能性がある。このため、対象とする疾患に罹患している個体群と、罹患していない個体群と、で変異パターンの出現頻度に有意差があるか否かを個々の変異毎に検定することで、前記疾患の罹患と相関が有る変異及びその変異パターンを分析する研究が進められている。

[0003] なお、「遺伝情報」は、「ＤＮＡ（デオキシリボ核酸）の塩基配列」又は「ヒトゲノムの変異情報」と称されてもよい。

先行技術文献

特許文献

[0004] 特許文献1：特開２００４－１６６５６５号公報

特許文献2：特開２００４－２３４１０４号公報

非特許文献

[0005] 非特許文献1：株式会社富士通研究所、“ゲノム情報の解析処理を高速化する技術を開発”、［online］、2016年3月15日、［平成29年1月10日検索］、インターネット〈URL：pr.fujitsu.com/jp/news/2016/03/15.html〉

発明の概要

発明が解決しようとする課題

[0006] ヒトゲノムの変異情報には、約２０００万個の変異が含まれる。例えば、１変異を２ビットの情報で表す場合、１０万人分の変異情報のデータ量は、約５００ＧＢ（ギガバイト）となる。ヒトゲノムの変異情報の検索や解析に使用されるコンピュータの１次記憶装置のデータ容量が変異情報のデータ量に満たない場合には、検索や解析の処理中に２次記憶装置へのアクセスが発生する。

[0007] 上記に例示したように、処理対象の配列データに含まれる変異パターンの数が多く、配列データのデータ量が大きい場合には、配列データの全体を１次記憶装置に格納することができず、２次記憶装置へのアクセスが発生する。これにより、配列データの検索や解析の処理速度が低くなるおそれがある。

[0008] １つの側面では、それぞれ複数の変異パターンを含む複数の配列において、メモリに記憶されるデータ量を削減することを目的とする。

課題を解決するための手段

[0009] このため、この情報処理装置は、複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理装置であって、前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行なう処理部と、前記処理部によって前記除外する処理が施された複数の配列を記憶する記憶部と、を備える。

発明の効果

[0010] １つの側面では、それぞれ複数の変異パターンを含む複数の配列において、メモリに記憶されるデータ量を削減することができる。

図面の簡単な説明

[0011] [図1]特異性がない変異及び特異性がある変異における変異パターンの分布の一例を示すグラフである。

[図2]変異情報の集計処理の概要を示すブロック図である。

[図3]変異情報の一例を示す図である。

[図4]変異配列の抽出処理を説明する図である。

[図5]変異配列の集計処理を説明する図である。

[図6]変異配列の集計処理を説明する図である。

[図7]ゲノム型構造を、変異マスタ情報と共に示す図である。

[図8]変異情報の検索処理を説明する図である。

[図9]実施形態の一例における情報処理システムのハードウェア構成を示すブロック図である。

[図10]実施形態の一例における情報処理装置及び端末の機能構成を示すブロック図である。

[図11]実施形態の一例におけるゲノム型構造を、変異マスタ情報と共に示す図である。

[図12]実施形態の一例におけるグループ統計情報及びグループ分け情報の作成処理を説明する図である。

[図13]実施形態の一例における非圧縮変異情報の圧縮処理を説明する図である。

[図14]実施形態の一例における圧縮済変異情報の集計処理を説明する図である。

[図15]実施形態の一例におけるグループ統計情報をテーブル形式で例示する図である。

[図16]実施形態の一例におけるグループ分け情報の第１の例をテーブル形式で示す図である。

[図17]実施形態の一例におけるグループ分け情報の第２の例をテーブル形式で示す図である。

[図18]実施形態の一例における変異情報の運用例を説明するフローチャートである。

[図19]実施形態の一例における非圧縮変異情報の圧縮処理を説明するフローチャートである。

[図20]実施形態の一例における圧縮サイズ情報の作成処理を説明する図であ

る。

[図21]実施形態の一例における組み合わせ圧縮サイズ情報の作成処理を説明する図である。

[図22]実施形態の一例における圧縮サイズ情報のマージ処理を説明する図である。

[図23]実施形態の一例での非圧縮変異情報の圧縮処理における入力データを例示する図である。

[図24]実施形態の一例での非圧縮変異情報の圧縮処理における出力データを例示する図である。

[図25]実施形態の一例における非圧縮変異情報の圧縮処理の詳細を説明するフローチャートである。

[図26]実施形態の一例におけるゲノム型データの作成処理を説明するフローチャートである。

[図27]実施形態の一例における圧縮済変異情報の集計処理を説明するフローチャートである。

[図28]実施形態の一例での一時集計テーブルの作成処理における入力データを例示する図である。

[図29]実施形態の一例での一時集計テーブルの作成処理における出力データを例示する図である。

[図30]実施形態の一例における一時集計テーブルの作成処理を説明するフローチャートである。

[図31]実施形態の一例での最終集計テーブルの作成処理における入力データを例示する図である。

[図32]実施形態の一例での最終集計テーブルの作成処理における出力データを例示する図である。

[図33]実施形態の一例における最終集計テーブルの作成処理を説明するフローチャートである。

発明を実施するための形態

[0012] 以下、図面を参照して一実施の形態を説明する。ただし、以下に示す実施形態はあくまでも例示に過ぎず、実施形態で明示しない種々の変形例や技術の適用を排除する意図はない。すなわち、本実施形態を、その趣旨を逸脱しない範囲で種々変形して実施することができる。

[0013] また、各図は、図中に示す構成要素のみを備えるという趣旨ではなく、他の機能等を含むことができる。

[0014] 以下、図中において、同一の各符号は同様の部分を示しているので、その説明は省略する。

[0015] 〔Ａ〕関連技術

図１の（１）は、特異性がない変異における変異パターンの分布の一例を示すグラフである。図１の（２）は、特異性がある変異における変異パターンの分布の一例を示すグラフである。

[0016] ヒトのＤＮＡ配列には、アデニン（Ａ），グアニン（Ｇ），シトシン（Ｃ）及びチミン（Ｔ）が含まれる。ＤＮＡ配列における各変異パターンは、Ａ，Ｇ，Ｃ及びＴのうちの２つの組み合わせによって示される。

[0017] 図１の（１）には、Ａ／Ａ，Ａ／Ｃ及びＣ／Ｃの３種類の変異パターンを有する或る変異における、変異パターン毎の人口分布が示されている。また、図１の（２）には、Ｔ／Ｔ，Ｇ／Ｔ及びＧ／Ｇの３種類の変異パターンを有する或る変異における、変異パターン毎の人口分布が示されている。

[0018] 図１の（１）及び（２）において、「罹患者」とは、或る疾患（例えば糖尿病）を有している者である。また、「健常者」とは、或る疾患（例えば糖尿病）を有していない者である。

[0019] 図１の（１）に示されるグラフでは、３つの変異パターンにおいて、健常者と罹患者との分布が相似形である。別言すれば、健常者における変異パターンＡ／Ａ，Ａ／Ｃ及びＣ／Ｃと、罹患者における変異パターンＡ／Ａ，Ａ／Ｃ及びＣ／Ｃとのそれぞれの比は、略一定である。一方、図１の（２）に示されるグラフでは、３つの変異パターンにおいて、健常者と罹患者との分布が相似形でない。別言すれば、健常者における変異パターンＡ／Ａ，Ａ／

Ｃ及びＣ／Ｃと、罹患者における変異パターンＡ／Ａ，Ａ／Ｃ及びＣ／Ｃとのそれぞれの比は、一定でない。

[0020] 図１の（２）に示されるように、或る変異における３つの変異パターンが健常者と罹患者との分布の間で相似形を有しない場合には、当該変異は、当該罹患者が有する疾患に関連する遺伝子であることが想定される。

[0021] 図２は、変異情報３０３の集計処理の概要を示すブロック図である。

[0022] 変異情報３０３は、複数の個体（「ヒト」と称されてもよい。）についてのＤＮＡ配列を示す情報である。変異情報３０３の詳細については、図３を用いて後述される。

[0023] 変異情報の集計処理は、罹患者群の変異情報３０３ａ及び健常者群の変異情報３０３ｂを各々処理対象として行なわれる。このため、図２に示されるように、罹患者群の変異情報３０３ａ及び健常者群の変異情報３０３ｂが、変異情報３０３から各々抽出される（符号Ａ１及びＡ２参照）。そして、罹患者群の変異情報３０３ａ及び健常者群の変異情報３０３ｂからＮ個の変異を有するＤＮＡ配列が各々出力される（符号Ａ３及びＡ４参照）。

[0024] 出力された罹患者群の変異情報３０３ａ及び健常者群の変異情報３０３ｂに基づき、罹患者群と健常者群とで各変異パターンの出現頻度に有意差があるか否かが、カイ自乗検定等の統計手法により個々の変異毎に検定される（符号Ａ５参照）。符号Ａ５で示される検定は、「有意差検定」と称されてもよい。「各変異パターンの出現頻度」は、「変異パターン毎の出現数の分布」と称されてもよい。

[0025] 図３は、変異情報３０３の一例を示す図である。

[0026] 変異情報３０３は、複数のＤＮＡ配列（「変異配列」又は単に「配列」と称されてもよい。）を含む。各ＤＮＡ配列には、複数の変異が含まれる。各変異の内容は、変異パターンによって表わされる。すなわち、変異情報３０３は、各個体におけるＤＮＡ配列に含まれる複数の変異のそれぞれが有する変異パターンを示す。変異情報３０３は、リファレンスゲノム情報との差分情報である。リファレンスゲノム情報はＤＮＡの解析対象の人種と別人種の

ＤＮＡ配列に関する情報である場合があり、例えば日本人を対象に変異情報を収集すると、日本人が共通に有するヒトゲノムの変異情報が抽出される。

[0027] 図３に示される例においては、個体＃０，＃１，＃２，＃３，・・・それぞれにおける変異＃０～＃Ｎ－１の変異パターンが示されている。例えば、個体＃０において、変異＃０の変異パターンはＡ／Ａであり、変異＃１の変異パターンはＡ／Ｃであり、変異＃２の変異パターンはＧ／Ｇである。

[0028] 図４の（１）は、臨床情報３０５をテーブル形式で示す図である。図４の（２）は、変異情報３０３をテーブル形式で示す図である。

[0029] 図４の（１）に示される臨床情報３０５では、各個体（「ヒト」と称されてもよい。）の属性が疾患の有無を示す情報に対応付けられている。

[0030] ゲノム解析においては、臨床情報３０５から、疾患の有無や性別，年齢、人種等の特徴を条件として、個体が抽出されることがある。個体の抽出では、臨床情報３０５と変異情報３０３とを照合（別言すれば、「ＪＯＩＮ」）して、条件に合致した群（「ｃａｓｅ群」と称されてもよい。）と条件に合致しない群（「ｃｏｎｔｒｏｌ群」と称されてもよい。）とが抽出される。

[0031] 臨床情報３０５において、「ＩＤ（識別子）」は、個体を一意に識別するための情報である。「性別」は個体の性別を示す。「年齢」は個体の年齢を示し、「年齢」の単位は“
歳”である。「人種」は、個体の人種を示す。「人種」のカラムにおいて、“ＪＰ”は日本人を示し、“ＵＳ”は米国人を示し、“ＣＮ”は中国人を示す。「糖尿病」は、個体が糖尿病に罹患しているか否かを示す。「糖尿病」のカラムにおいて、“Ｔ”は糖尿病に罹患していることを示し、“Ｆ”は糖尿病に罹患していないことを示す。「癌」は、個体が癌に罹患しているか否かを示す。「癌」のカラムにおいて、“Ｔ”は癌に罹患していることを示し、“Ｆ”は癌に罹患していないことを示す。

[0032] 臨床情報３０５において、例えば、「性別」が男であり、「癌」に罹患している個体が選択される（図４の（１）の下線部参照）。図４の（１）において、「性別」が男であり、「癌」に罹患している個体の「ＩＤ」は、“０

”，“２”及び“４”である。

[0033] 図４の（２）で示される変異情報３０３では、各個体のＩＤが変異パターンに対応付けられている。

[0034] 変異情報３０３において、「ＩＤ」は、個体を一意に識別するための情報であり、臨床情報３０５の「ＩＤ」に対応する。「変異パターン」は、各個体のＤＮＡ配列に含まれる変異のパターンを示す。

[0035] 図４の（２）に示される例において、図４の（１）の上述した説明で選択された「ＩＤ」が“０”，“２”及び“４”の変異パターンがｃａｓｅ群の集計処理のために抽出される。また、図４の（１）の上述した説明で選択されなかった「ＩＤ」が“１”及び“３”の変異パターンがｃｏｎｔｒｏｌ群の集計処理のために抽出される。

[0036] 図５及び図６は、変異配列の集計処理を説明する図である。

[0037] 変異配列の集計処理では、図４の（２）で抽出されたｃａｓｅ群及びｃｏｎｔｒｏｌ群の「ＩＤ」毎に、各変異に含まれる変異パターンをカウントする。

[0038] 図５においては、ｃａｓｅ群の集計処理のために抽出された「ＩＤ」が“０”の個体の変異パターンを入力データ３０４ａとして、各変異の変異パターンを集計テーブル３０４ｂによってカウントされる（符号Ｂ１参照）。

[0039] 集計テーブル３０４ｂにおいては、入力データ３０４ａの変異パターンに対応して、例えば、変異＃０の変異パターン“Ａ／Ａ”，変異＃１の変異パターン“Ａ／Ｃ”及び変異＃２の変異パターン“Ｇ／Ｇ”のカウントが、０から１へカウントアップされている。また、変異＃３～＃Ｎ－１におけるカウントについても、同様に、入力データ３０４ａの変異パターンに対応して、カウントアップされている。

[0040] 次に、図６においては、ｃａｓｅ群の集計処理のために抽出された「ＩＤ」が“２”の個体の変異パターンを入力データ３０４ａとして、各変異の変異パターンを集計テーブル３０４ｂによってカウントされる（符号Ｂ２参照）。

[0041] 集計テーブル３０４ｂにおいては、入力データ３０４ａの変異パターンに対応して、例えば、変異＃０の変異パターン“Ａ／Ａ”及び変異＃１の変異パターン“Ａ／Ｃ”のカウントが、１から２へカウントアップされている。また、例えば、変異＃２の変異パターン“Ｃ／Ｇ”のカウントが、０から１へカウントアップされている。更に、変異＃３～＃Ｎ－１におけるカウントについても、同様に、入力データ３０４ａの変異パターンに対応して、カウントアップされている。

[0042] 図５の処理Ｂ１及び図６の処理Ｂ２で示された処理を、図４の（２）でｃａｓｅ群の集

計処理のために抽出された個体数繰り返すことにより、ｃａｓｅ群の集計処理が完了する。また、ｃｏｎｔｒｏｌ群の集計処理についても、ｃａｓｅ群の集計処理と同様に行なわれる。

[0043] 図７の（１）はゲノム型構造３０１を例示する図であり、図７の（２）は変異マスタ情報３０２をテーブル形式で示す図である。

[0044] ゲノム型構造３０１は、或る変異配列における各変異の変異パターンをそれぞれ２ビットで表わす情報である。

[0045] 変異マスタ情報３０２は、各変異が、ゲノム型構造３０１のどの位置に対応し、どの変異パターンを有するかを管理する情報である。

[0046] ＤＮＡ配列に含まれる各変異の多くは、３つの変異パターン（例えば、図７の（２）の変異＃０はＡ／Ａ，Ａ／Ｃ及びＣ／Ｃ）のいずれかによって表わされる。そこで、各変異に対して、２ビットの格納領域が割り当てられる。これにより、３つの変異パターンを２ビットの格納領域に格納できる。なお、２ビットの格納領域には、最大で４つの変異パターンが格納できる。

[0047] 図７の（２）に示される例では、変異＃０において、パターン＃０はＡ／Ａであり、パターン＃１はＡ／Ｃであり、パターン＃２はＣ／Ｃである。各変異において、パターン＃０は“００”で表わされ、パターン＃１は“０１”で表わされ、パターン＃２は“１０”で表わされる。

[0048] 図７の（２）において下線が付されているように変異＃０～＃５の変異パ

ターンが“Ａ／Ａ，Ａ／Ｃ，Ｃ／Ｇ，Ｃ／Ｃ，Ｃ／Ｔ，Ｔ／Ｔ”である場合には、図７の（１）に示されるようにゲノム型構造３０１は、“０００１０１０００１１０”となる。

[0049] 図８は、変異情報３０３の検索処理（「解析処理」と称されてもよい。）を説明する図である。検索処理は、図９を用いて後述する端末２から情報処理装置１への問い合わせにより実行されてよい。

[0050] 図８に示される例において、１回目の検索条件として、臨床情報３０５において、「性別」が男であり、且つ、「癌」に罹患している個体が検索される（符号Ｃ１の「臨床情報３０５」の下線部参照）。これにより、変異情報３０３において、「ＩＤ」が０，２及び４の変異パターンが問い合わせ結果として抽出される（符号Ｃ１の「変異情報３０３」の下線部参照）。

[0051] 次に、２回目の検索条件として、臨床情報３０５において、「性別」が男であり、且つ、「癌」に罹患しており、且つ、「人種」が日本人である個体が検出される（符号Ｃ２の「臨床情報３０５」の下線部参照）。これにより、変異情報３０３において、「ＩＤ」が０及び２の変異パターンが問い合わせ結果として抽出される（符号Ｃ２の「変異情報３０３」の下線部参照）。

[0052] 以後、問い合わせ結果を見て、検索条件を変更して、再び問い合わせをするという、インタラクティブな処理が繰り返し実行される。

[0053] ヒトゲノムの変異情報には、約２０００万個の変異が含まれる。１変異あたり２ビットの情報を保持するため、１０万人分の変異情報のデータ量は、約５００ＧＢとなる。ヒトゲノムの変異情報の検索や解析に使用されるコンピュータの１次記憶装置のデータ容量が
変異情報のデータ量に満たない場合には、検索や解析の処理中に２次記憶装置へのアクセスが発生する。これにより、ヒトゲノムの変異情報の検索や解析の処理速度が低くなるおそれがある。

[0054] そこで、変異情報３０３を既存のデータ圧縮技術を利用して圧縮し、圧縮されたデータをメモリで展開しながら利用することが想定される。しかしながら、この場合においても、圧縮されたデータをメモリで展開することによ

り、処理速度が遅くなるおそれがある。

[0055] 〔Ｂ〕実施形態の一例

ＤＮＡ配列においては、人種や性別，年齢等でグループ分けを行なうと、グループの全メンバ（「個体」と称されてもよい）で同一の変異パターンを有する変異が多数ある。例えば、日本人のＤＮＡ配列においては、第１染色体における３００万個の変異のうち、８０万個の変異が同一の変異パターンを有する。

[0056] そこで、実施形態の一例においては、複数のＤＮＡ配列間において対応する変異が有する変異パターンが同じ値を有する場合に、当該変異パターンをメモリに記憶させない。これにより、メモリに記憶されるデータ量を削減し、ＤＮＡ配列の解析速度を向上させる。

[0057] 〔Ｂ－１〕ハードウェア構成例

図９は、実施形態の一例における情報処理システム１００のハードウェア構成を示すブロック図である。

[0058] 情報処理システム１００は、情報処理装置１及び端末２を備える。情報処理装置１と端末２とは、ネットワーク３を介して互いに通信可能に接続されてよい。

[0059] 端末２は、ユーザが使用するコンピュータである。ユーザは、実施形態の一例における圧縮処理で圧縮された変異情報に対する解析処理を、この端末２を用いて行なってよい。端末２は、例示的に、ＣＰＵ（Central Processing Unit）２０及びメモリ２２を備える

。なお、端末２は、情報処理装置１と同様に、それぞれ後述する記憶装置１３、媒体読取装置１４、表示制御装置１５、表示装置１６、入力装置１７及び通信制御装置１８を備えてもよい。

[0060] メモリ２２は、記憶部の一例であり、例示的に、ＲＯＭ（Read Only Memory）及びＲＡＭ（Random Access Memory）の少なくとも一方を含む記憶装置である。メモリ２２のＲＯＭには、ＢＩＯＳ（Basic Input/Output System）等のプログラムが書き込まれてよい。

メモリ２２のソフトウェアプログラムは、ＣＰＵ２０に適宜に読み込まれて実行されてよい。また、メモリ２２のＲＡＭは、一次記録メモリあるいはワーキングメモリとして利用されてよい。メモリ２２は、後述するゲノム型構造２０１，変異マスタ情報２０２，元データ変異情報２０３，非圧縮変異情報２０４，臨床情報２０５，圧縮済変異情報２０６，一時集計テーブル２０７，最終集計テーブル２０８を記憶してよい。また、メモリ２２は、後述するグループ統計情報２０９，ＮＵＬＬ変異集計情報２０９ａ，圧縮サイズ情報２０９ｂ，グループ分け情報２１０，組み合わせＮＵＬＬ変異集計情報２１０ａ，組み合わせ圧縮サイズ情報２１０ｂを記憶してよい。更に、メモリ２２は、後述するランキング情報２１１，ＮＵＬＬ変異構造体２１２ａ，２１２ｂ，グループＩＤ対応配列２１３及び組み合わせ２１４を記憶してよい。

[0061] ＣＰＵ２０は、種々の制御や演算を行なう処理装置であり、メモリ２２に格納されたＯＳ（Operating System）やプログラムを実行することにより、種々の機能を実現する。ＣＰＵ２０の機能については、図１０の（２）を用いて後述する。

[0062] 情報処理装置１は、例示的に、ＣＰＵ１１、メモリ１２、記憶装置１３、媒体読取装置１４、表示制御装置１５、表示装置１６、入力装置１７及び通信制御装置１８を備える。ＣＰＵ１１、メモリ１２、記憶装置１３、媒体読取装置１４、表示制御装置１５、入力装置１７及び通信制御装置１８は、バス線１０を介して互いに通信可能に接続されている。

[0063] 記憶装置１３は、例示的に、データを読み書き可能に記憶する装置であり、例えば、ＨＤＤ（Hard Disk Drive）やＳＳＤ（Solid State Drive）、ＳＣＭ（Storage Class Memory）が用いられてよい。記憶装置１３は、後述するゲノム型構造２０１，変異マスタ情報２０２，元データ変異情報２０３，非圧縮変異情報２０４，臨床情報２０５，圧縮済変異情報２０６，一時集計テーブル２０７，最終集計テーブル２０８を記憶してよい。また、記憶装置１３は、後述するグループ統計情報２０９，ＮＵＬＬ変異集計情報２０９ａ

，圧縮サイズ情報２０９ｂ，グループ分け情報２１０，組み合わせＮＵＬＬ変異集計情報２１０ａ，組み合わせ圧縮サイズ情報２１０ｂを記憶してよい。更に、記憶装置１３は、後述するランキング情報２１１，ＮＵＬＬ変異構造体２１２ａ，２１２ｂ，グループＩＤ対応配列２１３及び組み合わせ２１４を記憶してよい。

[0064] 媒体読取装置１４は、記録媒体ＲＭが装着可能に構成される。媒体読取装置１４は、記録媒体ＲＭが装着された状態において、記録媒体ＲＭに記録されている情報を読み取り可能に構成される。本例では、記録媒体ＲＭは可搬性を有する。記録媒体ＲＭは、コンピュータ読取可能な記録媒体であって、例えば、フレキシブルディスク，ＣＤ（Compact Disk），ＤＶＤ（Digital Versatile Disk），ブルーレイディスク，磁気ディスク，光ディスク，光磁気ディスク又は、半導体メモリ等である。ＣＤは、ＣＤ－ＲＯＭ（Read Only Memory）やＣＤ－Ｒ（Recordable），ＣＤ－ＲＷ（ReWritable）等であってよい。また、ＤＶＤは、ＤＶＤ－ＲＯＭやＤＶＤ－ＲＡＭ（Random Access Memory），ＤＶＤ－Ｒ，ＤＶＤ＋Ｒ，ＤＶＤ－ＲＷ，ＤＶＤ＋ＲＷ，ＨＤ（High-Definition） ＤＶＤ等であってよ
い。

[0065] 表示制御装置１５は、表示装置１６と通信可能に接続され、表示装置１６の画面表示を制御する。

[0066] 表示装置１６は、液晶ディスプレイやＣＲＴ（Cathode Ray Tube），電子ペーパーディスプレイ等であり、オペレータ等に対する各種情報を表示する。

[0067] 入力装置１７は、例えば、マウス、トラックボール、キーボードであり、この入力装置１７を介して、オペレータが各種の入力操作を行なう。

[0068] 表示装置１６及び入力装置１７は組み合わされてもよく、例えば、タッチパネルでもよい。

[0069] 通信制御装置１８は、情報処理装置１とネットワーク３との間の通信を制御する。通信制御装置１８は、ネットワーク３を介した、情報処理装置１と

端末２等の他のコンピュータとの通信を制御してよい。

[0070] メモリ１２は、記憶部の一例であり、例示的に、ＲＯＭ及びＲＡＭの少なくとも一方を含む記憶装置である。メモリ１２のＲＯＭには、ＢＩＯＳ等のプログラムが書き込まれてよい。メモリ１２のソフトウェアプログラムは、ＣＰＵ１１に適宜に読み込まれて実行されてよい。また、メモリ１２のＲＡＭは、一次記録メモリあるいはワーキングメモリとして利用されてよい。メモリ１２は、後述するゲノム型構造２０１，変異マスタ情報２０２，元データ変異情報２０３，非圧縮変異情報２０４，臨床情報２０５，圧縮済変異情報２０６，一時集計テーブル２０７，最終集計テーブル２０８を記憶してよい。また、メモリ１２は、後述するグループ統計情報２０９，ＮＵＬＬ変異集計情報２０９ａ，圧縮サイズ情報２０９ｂ，グループ分け情報２１０，組み合わせＮＵＬＬ変異集計情報２１０ａ，組み合わせ圧縮サイズ情報２１０ｂを記憶してよい。更に、メモリ１２は、後述するランキング情報２１１，ＮＵＬＬ変異構造体２１２ａ，２１２ｂ，グループＩＤ対応配列２１３及び組み合わせ２１４を記憶してよい。

[0071] ＣＰＵ１１は、種々の制御や演算を行なう処理装置であり、メモリ１２に格納されたＯＳやプログラムを実行することにより、種々の機能を実現する。

[0072] 図１０の（１）は、実施形態の一例における情報処理装置１の機能構成を示すブロック図である。

[0073] ＣＰＵ１１は、図１０の（１）に示すように、データ作成処理部１１１及び集計処理部１１２として機能する。

[0074] なお、これらデータ作成処理部１１１及び集計処理部１１２としての機能を実現するためのプログラムは、例えば前述した記録媒体ＲＭに記録された形態で提供される。そして、コンピュータは媒体読取装置１４を介してその記録媒体ＲＭからプログラムを読み取って内部記憶装置または外部記憶装置に転送し格納して用いる。又、そのプログラムを、例えば磁気ディスク，光ディスク，光磁気ディスク等の記憶装置（記録媒体）に記録しておき、その

記憶装置から通信経路を介してコンピュータに提供してもよい。

[0075] データ作成処理部１１１及び集計処理部１１２としての機能を実現する際には、内部記憶装置（本実施形態ではメモリ１２）に格納されたプログラムがコンピュータのマイクロプロセッサ（本実施形態ではＣＰＵ１１）によって実行される。このとき、記録媒体ＲＭに記録されたプログラムをコンピュータが読み取って実行してもよい。

[0076] なお、情報処理装置１は、ＣＰＵ１１の代わりに、ＭＰＵやＤＳＰ，ＡＳＩＣ，ＰＬＤ，ＦＰＧＡのいずれか１つを備えてもよい。また、情報処理装置１は、ＣＰＵ，ＭＰＵ，ＤＳＰ，ＡＳＩＣ，ＰＬＤ及びＦＰＧＡのうちの２種類以上を組み合わせて備えてもよい。ＭＰＵはMicro Processing Unitの略称であり、ＤＳＰはDigital Signal Processorの略称であり、ＡＳＩＣはApplication Specific Integrated Circuitの略称である。また、ＰＬＤはProgrammable Logic Deviceの略称であり、ＦＰＧＡはField Programmable Gate Arrayの略称である。

[0077] データ作成処理部１１１は、複数のＤＮＡ配列のそれぞれに含まれる複数の変異パターンを、メモリ１２に記憶させる。また、データ作成処理部１１１は、複数の配列間において対応する変異パターンが同じ値である場合に、当該変異パターンをメモリ１２の記憶対象から除外する。

[0078] 別言すれば、データ作成処理部１１１は、処理部の一例であり、それぞれ複数の変異パターンを含む複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、同じ値を有する変異パターンを記憶対象から除外する処理を行なう。また、データ作成処理部１１１は、除外する処理が施された複数の配列をメモリ１２に記憶させる。

[0079] データ作成処理部１１１は、後述するグループ分け情報２１０に基づき、除外する処理の対象となった変異パターンを、除外する処理が施された配列に挿入することにより、除外する処理を施す前の配列を復元してよい。グループ分け情報２１０は、除外する処理の対象となった変異パターンの位置を示す情報と称されてもよい。

[0080] 集計処理部１１２は、データ作成処理部１１１によってメモリ１２に記憶された変異パターンに基づき、ＤＮＡ配列の解析を行なう。なお、ＤＮＡ配列の解析が図９に示された端末２において行なわれる場合においては、集計処理部１１２としての機能は、端末２に備えられてよい。

[0081] なお、データ作成処理部１１１の詳細については、図１１～図１３及び図１５～図１７等を用いて後述する。また、集計処理部１１２の詳細については、図１４等を用いて後述する。

[0082] 図１０の（２）は、実施形態の一例における端末２の機能構成を示すブロック図である。

[0083] ＣＰＵ２０は、図１０の（２）に示すように、取得部２１及び集計処理部１１２として機能する。

[0084] なお、これら取得部２１及び集計処理部１１２としての機能を実現するためのプログラムは、例えば記録媒体に記録された形態で提供される。そして、コンピュータは媒体読取装置（不図示）を介してその記録媒体からプログラムを読み取って内部記憶装置または外部記憶装置に転送し格納して用いる。又、そのプログラムを、例えば磁気ディスク，光ディスク，光磁気ディスク等の記憶装置（記録媒体）に記録しておき、その記憶装置から通信経路を介してコンピュータに提供してもよい。

[0085] 取得部２１及び集計処理部１１２としての機能を実現する際には、内部記憶装置（本実施形態ではメモリ２２）に格納されたプログラムがコンピュータのマイクロプロセッサ（本実施形態ではＣＰＵ２０）によって実行される。このとき、記録媒体に記録されたプログラムをコンピュータが読み取って実行してもよい。

[0086] なお、端末２は、ＣＰＵ２０の代わりに、ＭＰＵやＤＳＰ，ＡＳＩＣ，ＰＬＤ，ＦＰＧＡのいずれか１つを備えてもよい。また、端末２は、ＣＰＵ，ＭＰＵ，ＤＳＰ，ＡＳＩＣ，ＰＬＤ及びＦＰＧＡのうちの２種類以上を組み合わせて備えてもよい。

[0087] 取得部２１は、例えばネットワーク３（図９を参照）を介して、情報処理

装置１から種々のデータを取得し、取得したデータをメモリ２２に記憶させる。種々のデータには、後述するゲノム型構造２０１，変異マスタ情報２０２，元データ変異情報２０３，非圧縮変異情報２０４，臨床情報２０５，圧縮済変異情報２０６，一時集計テーブル２０７，最終集計テーブル２０８が含まれてよい。また、種々のデータには、後述するグループ統計情報２０９，ＮＵＬＬ変異集計情報２０９ａ，圧縮サイズ情報２０９ｂ，グループ分け情報２１０，組み合わせＮＵＬＬ変異集計情報２１０ａ，組み合わせ圧縮サイズ情報２１０ｂが含まれてもよい。更に、種々のデータには、後述するランキング情報２１１，ＮＵＬＬ変異構造体２１２ａ，２１２ｂ，グループＩＤ対応配列２１３及び組み合わせ２１４が含まれてもよい。

[0088] 取得部２１は、情報処理装置１による変異パターンの圧縮に用いられるグループを指定し、指定したグループによって圧縮された変異パターンを情報処理装置１から取得してよい。また、取得部２１は、取得した変異パターンをメモリ２２に記憶させてよい。

[0089] すなわち、取得部２１は、図８を用いて上述したように、性別や人種等のグループに基づいた検索条件を指定して情報処理装置１に対する問い合わせを行なう。そして、取得部２１は、指定した検索条件によって圧縮された変異パターンを情報処理装置１から取得する。

[0090] 取得部２１は、後述するグループ分け情報２１０に基づき、除外する処理の対象となった変異パターンを、除外する処理が施された配列に挿入することにより、除外する処理を施す前の配列を復元してよい。グループ分け情報２１０は、除外する処理の対象となった変異パターンの位置を示す情報と称されてもよい。

[0091] 図１１の（１）はゲノム型構造２０１を示す図であり、図１１の（２）は変異マスタ情報２０２をテーブル形式で示す図である。

[0092] ゲノム型構造２０１は、或る変異配列における各変異の変異パターンをそれぞれ２ビットで表わす情報である。また、ゲノム型構造２０１の先頭領域には、当該変異配列が属するグループを特定するための識別子である「グル

ープＩＤ」が付加されている。

[0093] 変異マスタ情報２０２は、各変異が、ゲノム型構造２０１のどの位置に対応し、どの変異パターンを有するかを管理する情報である。また、変異マスタ情報２０２は、「ゲノム型位置」のカラムを有し、変異パターンが１種類に限られている変異にＮＵＬＬが設定されているとともに、変異パターンが１種類に限られている変異以外の変異がゲノム型構造２０１のどの位置に対応するかの情報を有する。

[0094] ＤＮＡ配列に含まれる各変異の多くは、３つの変異パターン（例えば、図１１の（２）の変異＃０はＡ／Ａ，Ａ／Ｃ及びＣ／Ｃ）のいずれかによって表わされる。そこで、各変異に対して、２ビットの格納領域が割り当てられる。これにより、３つの変異パターンを２ビットの格納領域に格納できる。なお、２ビットの格納領域には、最大で４つの変異パターンを格納できる。

[0095] 図１１の（２）に示される例では、変異＃０において、パターン＃０はＡ／Ａであり、パターン＃１はＡ／Ｃであり、パターン＃２はＣ／Ｃである。各変異において、図４の（２）におけるパターン＃０，＃１及び＃２の変異パターンは、図４の（１）におけるゲノム型構造２０１において、“００”，“０１”及び“１０”にそれぞれ変換されて格納される。

[0096] また、図１１の（２）に示される例で、変異＃３における変異パターンは、パターン＃０のＡ／Ａに限られている。データ作成処理部１１１は、変異パターンが１種類に限られている変異の「ゲノム型位置」をＮＵＬＬに設定する。一方、データ作成処理部１１１は、変異パターンが１種類に限られている変異以外の変異における「ゲノム型位置」に対して、「変異ＩＤ」が小さい変異から順番に、０，１，２，３，４，・・・の値を登録する。

[0097] 図１１の（２）において下線が付されているように変異＃０～＃５の変異パターンが“Ａ／Ａ，Ｃ／Ｔ，Ａ／Ｃ，Ａ／Ａ，Ｃ／Ｃ，Ａ／Ｔ”である場合には、図１１の（１）に示されるようにゲノム型構造３０１は、“０００１０１０００１”となる。図１１の（１）に示されているように、データ作成処理部１１１は、変異パターンが１種類に限られている変異＃３の変異パ

ターンについては、ゲノム型構造２０１に登録しない。一方、データ作成処理部１１１は、変異パターンが１種類に限られている変異以外の変異＃０～＃２，＃４，＃５，・・・の変異パターンについて、ゲノム型構造２０１に登録する。

[0098] 図１２は、実施形態の一例におけるグループ統計情報２０９及びグループ分け情報２１０の作成処理を説明する図である。

[0099] データ作成処理部１１１は、元データ変異情報２０３に基づき、非圧縮変異情報２０４
及び変異マスタ情報２０２を作成する。

[0100] 元データ変異情報２０３は、各個体におけるＤＮＡ配列に含まれる複数の変異のそれぞれが有する変異パターンをＡＧＣＴによって示す情報である。

[0101] 非圧縮変異情報２０４は、各個体におけるＤＮＡ配列に含まれる複数の変異のそれぞれが有する変異パターンをそれぞれ２ビットのデータ示す情報である。元データ変異情報２０３から非圧縮変異情報２０４への変換は、図１１を用いて説明した方法によって行なわれる。

[0102] データ作成処理部１１１は、臨床情報２０５と、作成した非圧縮変異情報２０４及び変異マスタ情報２０２とに基づき、グループ統計情報２０９及びグループ分け情報２１０を作成する。

[0103] 臨床情報２０５は、各個体（「ヒト」と称されてもよい。）の属性と、疾患の有無を示す情報とを対応付ける情報である。

[0104] 臨床情報２０５において、「ＩＤ」は、個体を一意に識別するための情報である。「性別」は個体の性別を示す。「年齢」は個体の年齢を示し、「年齢」の単位は“歳”である。「人種」は、個体の人種を示す。「人種」のカラムにおいて、“ＪＰ”は日本人を示し、“ＵＳ”は米国人を示し、“ＣＮ”は中国人を示す。「糖尿病」は、個体が糖尿病に罹患しているか否かを示す。「糖尿病」のカラムにおいて、“Ｔ”は糖尿病に罹患していることを示し、“Ｆ”は糖尿病に罹患していないことを示す。「癌」は、個体が癌に罹患しているか否かを示す。「癌」のカラムにおいて、“Ｔ”は癌に罹患して

いることを示し、“Ｆ”は癌に罹患していないことを示す。なお、人種は、国籍や出身地等であってもよい。

[0105] グループ統計情報２０９は、臨床情報２０５における「性別」や「人種」等の属性毎にＤＮＡ配列を抽出した場合において、変異パターンが１種類である変異についてはメモリ１２に記憶させないことによって生じる圧縮サイズを示す情報である。グループ統計情報２０９の詳細については、図１５を用いて後述する。なお、本明細書において「圧縮サイズ」とは、データの圧縮処理によってデータ量が削減されるサイズを示す。

[0106] グループ分け情報２１０は、複数の属性の組み合わせについてＤＮＡ配列を抽出した場合において、変異パターンが１種類である変異についてはメモリ１２に記憶させないことによって生じる圧縮サイズを示す情報である。グループ分け情報２１０の詳細については、図１６及び図１７を用いて後述する。

[0107] 図１３は、実施形態の一例における非圧縮変異情報２０４の圧縮処理を説明する図である。

[0108] データ作成処理部１１１は、臨床情報２０５と、作成されたグループ統計情報２０９及びグループ分け情報２１０とに基づき、圧縮済変異情報２０６を作成する。

[0109] 圧縮済変異情報２０６は、各個体におけるＤＮＡ配列に含まれる複数の変異のそれぞれが有する変異パターンをそれぞれ２ビットのデータ示す情報である。圧縮済変異情報２０６の「変異パターン」においては、後述するグループ分け情報２１０において「ＮＵＬＬ変異リスト」に登録されている変異パターンが削除されている。これにより、圧縮済変異情報２０６に登録されている複数の変異パターンのうち、少なくとも一部の変異パターンは、非圧縮変異情報２０４の変異パターンよりも短くなる。

[0110] また、圧縮済変異情報２０６の先頭領域には、当該変異配列が属するグループを特定するための識別子である「グループＩＤ」が付加されている。図１３に示す例において、圧縮済変異情報２０６の「グループＩＤ」には、人

種を示すＪＰ，ＵＳ及びＣＮが関連付けられている。

[0111] 図１４は、実施形態の一例における圧縮済変異情報２０６の集計処理を説明する図である。

[0112] 集計処理部１１２は、臨床情報２０５を照合（「ＪＯＩＮ」と称されてもよい。）することによって、圧縮済変異情報２０６の変異パターンをｃｏｎｔｒｏｌ群の一時集計テーブル２０７ａ及びｃａｓｅ群の一時集計テーブル２０７ｂに登録する。図１４に示される例においては、人種別にグループ分けされた圧縮済変異情報２０６の変異パターンのうち、癌に罹患していない個体の変異パターンが、ｃｏｎｔｒｏｌ群の一時集計テーブル２０７ａに登録される。また、人種別にグループ分けされた圧縮済変異情報２０６の変異パターンのうち、癌に罹患している個体の変異パターンが、ｃａｓｅ群の一時集計テーブル２０７ｂに登録される。

[0113] 図１４に示される例において、ｃｏｎｔｒｏｌ群の一時集計テーブル２０７ａ及びｃａｓｅ群の一時集計テーブル２０７ｂには、それぞれＪＰ集計テーブル，ＣＮ集計テーブル及びＵＳ集計テーブルが含まれる。

[0114] 圧縮済変異情報２０６のＩＤ＝０において、変異パターンにはグループＩＤとしてＪＰが付加されており、臨床情報２０５を照合すると癌に罹患しているため、ＩＤ＝０の変異パターンは、ｃａｓｅ群のＪＰ集計テーブルに登録される。圧縮済変異情報２０６のＩＤ＝１において、変異パターンにはグループＩＤとしてＵＳが付加されており、臨床情報２０５を照合すると癌に罹患していないため、ＩＤ＝１の変異パターンは、ｃｏｎｔｒｏｌ群のＵＳ集計テーブルに登録される。圧縮済変異情報２０６のＩＤ＝２において、変異パターンにはグループＩＤとしてＪＰが付加されており、臨床情報２０５を照合すると癌に罹患しているため、ＩＤ＝２の変異パターンは、ｃａｓｅ群のＪＰ集計テーブルに登録される。圧縮済変異情報２０６のＩＤ＝３において、変異パターンにはグループＩＤとしてＣＮが付加されており、臨床情報２０５を照合すると癌に罹患していないため、ＩＤ＝３の変異パターンは、ｃｏｎｔｒｏｌ群のＣＮ集計テーブルに登録される。圧縮済変異情報２０６のＩＤ

＝４において、変異パターンにはグループＩＤとしてＵＳが付加されており、臨床情報２０５を照合すると癌に罹患しているため、ＩＤ＝４の変異パターンは、ｃａｓｅ群のＵＳ集計テーブルに登録される。

[0115] 集計処理部１１２は、ｃｏｎｔｒｏｌ群のＪＰ集計テーブル，ＣＮ集計テーブル及びＵＳ集計テーブルを組み合わせて、ｃｏｎｔｒｏｌ集計テーブル２０８ａを作成する。また、集計処理部１１２は、ｃａｓｅ群のＪＰ集計テーブル，ＣＮ集計テーブル及びＵＳ集計テーブルを組み合わせて、ｃａｓｅ集計テーブル２０８ｂを作成する。

[0116] なお、一時集計テーブル２０７（別言すれば、「ｃｏｎｔｒｏｌ群の一時集計テーブル２０７ａ」及び「ｃａｓｅ群の一時集計テーブル２０７ｂ」）の詳細については、図３１等を用いて後述する。また、最終集計テーブル２０８（別言すれば、「ｃｏｎｔｒｏｌ集計テーブル２０８ａ」及び「ｃａｓｅ集計テーブル２０８ｂ」）については、図３２等を用いて後述する。

[0117] データ作成処理部１１１は、臨床情報２０５の属性条件の組み合わせの中から、データサイズの圧縮率が高くなる組み合わせを選択し、グルーピングしてよい。また、データ作

成処理部１１１は、グルーピングする組み合わせの数の上限をN_Gに設定し、上限数N_G以下の組み合わせを選択してよい。

[0118] 図１５は、実施形態の一例におけるグループ統計情報２０９をテーブル形式で例示する図である。図１５に例示されるグループ統計情報２０９は、各人種における変異パターンについての圧縮サイズを示す。

[0119] データ作成処理部１１１は、図１５に例示されるグループ統計情報２０９を作成する。データ作成処理部１１１は、「人種」以外の「性別」や「年齢」等の属性についてのグループ統計情報２０９を作成してもよい。

[0120] 「属性値」のカラムには、臨床情報２０５に含まれる複数の属性のうちのいずれかの属性のメンバが登録される。図１５に示される例においては、「属性値」のカラムには、ＪＰ，ＣＮ及びＵＳが登録されている。

[0121] 「ＮＵＬＬ変異数」は、その属性値を有する全個体で同一になる変異（「

ＮＵＬＬ変異」と称されてもよい。）の数を示す。

[0122] 「個体数」は、その属性値を有する個体の数を示す。

[0123] 「圧縮サイズ」は、ＮＵＬＬ変異によって圧縮されるデータサイズを示し、「ＮＵＬＬ変異数」と「個体数」との積によって算出される。各属性値における「圧縮サイズ」が合計されることにより、当該属性でグルーピングした場合における圧縮サイズの合計が算出される。図１５に示される例では、属性「人種」でグルーピングした場合における圧縮サイズの合計が算出される。

[0124] 図１６は、実施形態の一例におけるグループ分け情報２１０の第１の例をテーブル形式で示す図である。

[0125] グループ分け情報２１０は、除外する処理の対象となった変異パターンの位置を示す情報である。データ作成処理部１１１は、作成したグループ統計情報２０９に基づき、図１６に例示されるグループ分け情報２１０を作成する。

[0126] 「組み合わせ」は、複数の属性値の組み合わせを示す。図１６に示される例において、例えば、“ＪＰ　ａｎｄ　男”は、人種が日本人であり、且つ、性別が男である個体を示す。

[0127] 「個体数」は、その属性値を有する個体の数を示す。

[0128] 「ＮＵＬＬ変異リスト」は、ＮＵＬＬ変異の位置（別言すれば、「ゲノム型位置」）と、当該ＮＵＬＬ変異の変異パターンの値とを示す。図１６においては、（ＮＵＬＬ変異の位置，変異パターンの値）の形式で示されている。例えば、（０，２）は、変異＃０がＮＵＬＬ変異であり、変異＃０の変異パターンがパターン＃２であることを示す。

[0129] 「圧縮サイズ」は、ＮＵＬＬ変異によって圧縮されるデータサイズを示し、「ＮＵＬＬ変異リスト」に含まれるＮＵＬＬ変異の数と「個体数」との積によって算出される。

[0130] 図１７は、実施形態の一例におけるグループ分け情報２１０の第２の例をテーブル形式で示す図である。

[0131] データ作成処理部１１１は、圧縮サイズが大きい属性の組み合わせを、組み合わせ数が上限数N_Gを超えるまで、順番にグループ分け情報２１０に登録してよい。そして、データ作成処理部１１１は、組み合わせ数が上限数N_Gを超えた場合に、全ての組み合わせの中で圧縮サイズが下位の複数の組み合わせをマージしてよい。

[0132] 図１７に示される例では、“ＪＰ　ａｎｄ　女”の組み合わせの圧縮サイズが５０００であり、“ＪＰ　ａｎｄ　男”の組み合わせの圧縮サイズが７５００である。そして、“ＪＰ　ａｎｄ　女”及び“ＪＰ　ａｎｄ　男”の組み合わせは、全ての組み合わせの中で、圧縮サイズが下位の２つの組み合わせである。そこで、データ作成処理部１１１は、“ＪＰ　ａｎｄ　女”及び“ＪＰ　ａｎｄ　男”の組み合わせをグループ分け情報２１０から削除する（図１７の取り消し線参照）。また、データ作成処理部１１１は、“ＪＰ　ａｎｄ　女”及び“ＪＰ　ａｎｄ　男”の組み合わせをマージした組み合わせを“（ＪＰ　ａｎｄ　男）ｏｒ（ＪＰ　ａｎｄ　女）”を作成して追加する（図１７の下線部参照）。

[0133] “（ＪＰ　ａｎｄ　男）ｏｒ（ＪＰ　ａｎｄ　女）”の組み合わせにおける「個体数」は、“ＪＰ　ａｎｄ　男”及び“ＪＰ　ａｎｄ　女”の組み合わせの「個体数」の和である５０００となる。また、“（ＪＰ　ａｎｄ　男）ｏｒ（ＪＰ　ａｎｄ　女）”の組み合わせにおける「ＮＵＬＬ変異リスト」には、“ＪＰ　ａｎｄ　男”及び“ＪＰ　ａｎｄ　女”の組み合わせにおける「ＮＵＬＬ変異リスト」に共通に登録されている“（０，２），（５０，０）”となる。更に、“（ＪＰ　ａｎｄ　男）ｏｒ（ＪＰ　ａｎｄ　女）”の組み合わせにおける「圧縮サイズ」は、“（ＪＰ　ａｎｄ　男）ｏｒ（ＪＰ　ａｎｄ　女）”の組み合わせにおける「ＮＵＬＬ変異リスト」に含まれるＮＵＬＬ変異の数と「個体数」との積によって１００００と算出される。

[0134] 〔Ｂ－２〕動作例

上述した実施形態の一例における変異情報の運用例を、図１８に示される

フローチャート（処理Ｄ１～Ｄ５）に従って説明する。

[0135] データ作成処理部１１１は、グルーピング情報を作成する（処理Ｄ１）。具体的には、データ作成処理部１１１は、臨床情報２０５及び元データ変異情報２０３を入力として、グルーピング情報，非圧縮変異情報２０４及び変異マスタ情報２０２を出力する。なお、グルーピング情報については、図２２及び図２３等を用いて後述する。

[0136] データ作成処理部１１１は、元データ変異情報２０３の圧縮処理を行なう（処理Ｄ２）。具体的には、データ作成処理部１１１は、臨床情報２０５，グルーピング情報，元データ変異情報２０３及び変異マスタ情報２０２を入力として、圧縮済変異情報２０６を出力とする。

[0137] 集計処理部１１２は、非圧縮変異情報２０４の運用処理を行なう（処理Ｄ３）。集計処理部１１２は、エンドユーザによる操作に基づき、変異の検索や，変異の集計，データの挿入及び／又は削除を行なう。

[0138] データ作成処理部１１１は、データの挿入や削除によってデータ分布が変更されるため、グループ分け情報２１０の再作成処理及び圧縮済変異情報２０６の再圧縮処理を行なう（処理Ｄ４）。具体的には、データ作成処理部１１１は、臨床情報２０５，グルーピング情報，圧縮済変異情報２０６及び変異マスタ情報２０２を入力とする。そして、データ作成処理部１１１は、グルーピング情報，圧縮済変異情報２０６及び変異マスタ情報２０２を出力する。

[0139] その後、処理Ｄ３及びＤ４が繰り返し行なわれる（処理Ｄ５）。

[0140] 次に、実施形態の一例における非圧縮変異情報２０４の圧縮処理を、図１９に示されるフローチャート（ステップＳ１～Ｓ５）に従って説明する。

[0141] データ作成処理部１１１は、グループ統計情報２０９を作成する（ステップＳ１）。ステップＳ１の処理の詳細は、図２０を用いて後述する。

[0142] データ作成処理部１１１は、グループ分け情報２１０を作成する（ステップＳ２）。ステップＳ２の処理の詳細は、図２１を用いて後述する。

[0143] データ作成処理部１１１は、作成したグループ分け情報２１０の組み合わ

せをマージする（ステップＳ３）。ステップＳ３の処理の詳細は、図２２を用いて後述する。

[0144] データ作成処理部１１１は、非圧縮変異情報２０４の圧縮処理を行なう（ステップＳ４）。ステップＳ４の処理の詳細は、図２５のフローチャートを用いて後述する。

[0145] データ作成処理部１１１は、ステップＳ１の処理を開始してから所定時間が経過したかを判定する（ステップＳ５）。

[0146] 所定時間が経過していない場合には（ステップＳ５のＮｏルート参照）、ステップＳ５の処理が繰り返し行なわれる。

[0147] 一方、所定時間が経過した場合には（ステップＳ５のＹｅｓルート参照）、処理はステップＳ１へ戻る。

[0148] 図２０は、実施形態の一例における圧縮サイズ情報２０９ｂの作成処理を説明する図である。

[0149] データ作成処理部１１１は、臨床情報２０５と元データ変異情報２０３とに基づき、臨床情報２０５に含まれる各属性についてのＮＵＬＬ変異集計情報２０９ａを作成する。図２０に示される例では、臨床情報２０５に含まれる属性「性別」，「年齢」，「人種」，「糖尿病」及び「癌」についての５つのＮＵＬＬ変異集計情報２０９ａが作成される。なお、図２０に示される例において、属性「年齢」についてのＮＵＬＬ変異集計情報２０９ａ「属性値」は、Ｙｏｕｎｇ（Ｙ），Ｍｉｄｄｌｅ（Ｍ）及びＯｌｄ（Ｏ）を示す。

[0150] データ作成処理部１１１は、各ＮＵＬＬ変異集計情報２０９ａに基づき、圧縮サイズ情報２０９ｂを作成する。圧縮サイズ情報２０９ｂには、属性毎の圧縮サイズの合計値が登録されている。

[0151] なお、図２０に示されるＮＵＬＬ変異集計情報２０９ａ及び圧縮サイズ情報２０９ｂは、図１５に示されたグループ統計情報２０９に対応する。

[0152] 図２１は、実施形態の一例における組み合わせ圧縮サイズ情報２１０ｂの作成処理を説明する図である。

[0153] データ作成処理部１１１は、ランキング情報２１１に基づき、組み合わせ

ＮＵＬＬ変異集計情報２１０ａを作成する。

[0154] ランキング情報２１１は、図２０に示された圧縮サイズ情報２０９ｂに基づき、各属性における圧縮サイズのランキングを示す。「属性値数」は、図２０に示された各属性につ
いてのＮＵＬＬ変異集計情報２０９ａにおいて登録されている属性値の数を示す。

[0155] 図２１に示される例では、属性「性別」と「糖尿病」の組み合わせについて、組み合わせＮＵＬＬ変異集計情報２１０ａが作成される。

[0156] データ作成処理部１１１は、組み合わせＮＵＬＬ変異集計情報２１０ａに基づき、組み合わせ圧縮サイズ情報２１０ｂを作成する。組み合わせ圧縮サイズ情報２１０ｂには、各組み合わせにおける個体数とＮＵＬＬ変異数との積が登録されている。

[0157] なお、図２１に示される組み合わせＮＵＬＬ変異集計情報２１０ａ及び組み合わせ圧縮サイズ情報２１０ｂは、図１６に示されたグループ分け情報２１０に対応する。

[0158] 図２２は、実施形態の一例における組み合わせ圧縮サイズ情報２１０ｂのマージ処理を説明する図である。

[0159] データ作成処理部１１１は、組み合わせ圧縮サイズ情報２１０ｂに含まれる組み合わせの数が上限値N_G以下となるように、組み合わせ圧縮サイズ情報２１０ｂに含まれる組み合わせをマージする。図２２に示される例において、組み合わせ圧縮サイズ情報２１０ｂには、４つの組み合わせが登録されているため、データ作成処理部１１１は、組み合わせの数が上限値N_G（例えば、３）以下となるように、圧縮サイズの小さい複数の組み合わせをマージする。図２２に示される例では、「女　ａｎｄ　Ｆ（糖尿病）」の圧縮サイズが２０であり、「男　ａｎｄ　Ｔ（糖尿病）」の圧縮サイズが６０であり、組み合わせ圧縮サイズ情報２１０ｂに含まれる組み合わせの中での圧縮サイズが小さい。

[0160] そこで、データ作成処理部１１１は、「女　ａｎｄ　Ｆ（糖尿病）」と「

男　ａｎｄ　Ｔ（糖尿病）」とをマージして、マージ後の組み合わせ２１４を得る。マージ後の組み合わせ２１４には、「女　ａｎｄ　Ｔ」，「男　ａｎｄ　Ｆ」及び「（男　ａｎｄ　Ｔ）ｏｒ（女　ａｎｄ　Ｆ）」が含まれている。

[0161] データ作成処理部１１１は、マージ後の組み合わせ２１４に基づき、ＮＵＬＬ変異構造体２１２ａ及び２１２ｂとグループＩＤ対応配列２１３とを作成してよい。ＮＵＬＬ変異構造体２１２ａ及び２１２ｂとグループＩＤ対応配列２１３とは、まとめてグルーピング情報と称されてもよい。このグルーピング情報は、非圧縮変異情報２０４の圧縮処理において使用されてよい。

[0162] ＮＵＬＬ変異構造体２１２ａには、「組み合わせ」，「グループＩＤ」及び「ポインタ」が対応付けて登録されている。「ポインタ」は、対応する「組み合わせ」についての「ＮＵＬＬ変異」と「パターン値」とが登録されたＮＵＬＬ変異構造体２１２ｂを参照する。図２２のＮＵＬＬ変異構造体２１２ａ及び２１２ｂにおいては、「男　ａｎｄ　Ｆ」の組み合わせにはグループＩＤ＝１が付与されており、当該組み合わせのＮＵＬＬ変異は変異＃０，＃５，＃６，＃１０及び＃４３であることが示されている。また、図２２のＮＵＬＬ変異構造体２１２ｂにおいては、グループＩＤ＝１の組み合わせについてのＮＵＬＬ変異＃０，＃５，＃６，＃１０及び＃４３は、パターン＃１，＃０，＃０，＃１及び＃０の変異パターンを有することが示されている。

[0163] グループＩＤ対応配列２１３は、図２０に示された臨床情報２０５における各個体の「ＩＤ」（別言すれば、「個体ＩＤ」）が、ＮＵＬＬ変異構造体２１２ａにおけるどの「グループＩＤ」に対応するかを示す。図２２に示される例において、例えば、個体ＩＤ＝０はグループＩＤ＝２に対応し、個体ＩＤ＝１はグループＩＤ＝２に対応し、個体ＩＤ＝２はグループＩＤ＝０に対応する。

[0164] 図２３は、実施形態の一例での非圧縮変異情報２０４の圧縮処理における入力データを例示する図である。図２４は、実施形態の一例での非圧縮変異

情報２０４の圧縮処理における出力データを例示する図である。

[0165] データ作成処理部１１１は、図２３に示される元データ変異情報２０３と変異マスタ情報２０２とＮＵＬＬ変異構造体２１２ａ及び２１２ｂとグループＩＤ対応配列２１３とに基づき、図２４に示される圧縮済変異情報２０６を作成する。なお、再圧縮処理においては、データ作成処理部１１１は、非圧縮変異情報２０４と変異マスタ情報２０２とＮＵＬＬ変異構造体２１２ａ及び２１２ｂとグループＩＤ対応配列２１３とに基づき、図２４に示される圧縮済変異情報２０６を作成してよい。

[0166] 図２４に示される圧縮済変異情報２０６では、「個体ＩＤ」と「変異パターン」とが対応付けられている。「変異パターン」には、ゲノム型データの前の領域に、グループＩＤ（ｇｒｏｕｐ）が付与されている。

[0167] 次に、実施形態の一例における変異情報の圧縮処理の詳細を、図２５に示されるフローチャート（ステップＳ４１～Ｓ４５）に従って説明する。

[0168] データ作成処理部１１１は、元データ変異情報２０３（再圧縮処理の場合には、「非圧縮変異情報２０４」）から順番にレコードを取り出す（ステップＳ４１）。

[0169] データ作成処理部１１１は、元データ変異情報２０３（再圧縮処理の場合には、「非圧縮変異情報２０４」）における個体ＩＤをグループＩＤに変換する（ステップＳ４２）。

[0170] データ作成処理部１１１は、元データ変異情報２０３（再圧縮処理の場合には、「非圧縮変異情報２０４」）から、グループＩＤに対応するゲノム型データを作成する（ステップＳ４３）。なお、ステップＳ４３の処理の詳細は、図２６のフローチャートを用いて後述する。

[0171] データ作成処理部１１１は、作成したゲノム型データを圧縮済変異情報２０６に挿入する（ステップＳ４４）。

[0172] データ作成処理部１１１は、元データ変異情報２０３（再圧縮処理の場合には、「非圧縮変異情報２０４」）にレコードがまだ存在するかを判定する（ステップＳ４５）。

[0173] レコードがまだ存在する場合には（ステップＳ４５のＹｅｓルート参照）、処理はステップＳ４１へ戻る。

[0174] 一方、レコードがもう存在しない場合には（ステップＳ４５のＮｏルート参照）、処理は終了する。

[0175] 次に、実施形態の一例におけるゲノム型データの作成処理を、図２６に示されるフローチャート（ステップＳ４３１～Ｓ４３６）に従って説明する。

[0176] データ作成処理部１１１は、元データ変異情報２０３（再圧縮処理の場合には、「非圧縮変異情報２０４」）において、変異を１つ選択する（ステップＳ４３１）。

[0177] データ作成処理部１１１は、当該変異はＮＵＬＬ変異であるかを判定する（ステップＳ４３２）。

[0178] 当該変異がＮＵＬＬ変異である場合には（ステップＳ４３２のＹｅｓルート参照）、処理はステップＳ４３１へ戻る。

[0179] 一方、当該変異がＮＵＬＬ変異でない場合には（ステップＳ４３２のＮｏルート参照）、データ作成処理部１１１は、現在実施中の圧縮処理は再圧縮処理であるかを判定する（ステップＳ４３３）。

[0180] 再圧縮処理である場合には（ステップＳ４３３のＹｅｓルート参照）、処理はステップＳ４３５へ進む。

[0181] 一方、再圧縮処理でない場合には（ステップＳ４３３のＮｏルート参照）、データ作成処理部１１１は、変異パターン（別言すれば、「ＡＧＣＴ」）を変異パターン値（別言すれば、「数値」）に変更する（ステップＳ４３４）。

[0182] データ作成処理部１１１は、変更した変異パターン値をゲノム型データに追加する（ステップＳ４３５）。

[0183] データ作成処理部１１１は、元データ変異情報２０３（再圧縮処理の場合には、「非圧縮変異情報２０４」）において次の変異があるかを判定する（ステップＳ４３６）。

[0184] 次の変異がある場合には（ステップＳ４３６のＹｅｓルート参照）、処理

はステップＳ４３１へ戻る。

[0185] 一方、次の変異がない場合には（ステップＳ４３６のＮｏルート参照）、処理は終了する。

[0186] 次に、実施形態の一例における圧縮済変異情報２０６の集計処理を、図２７に示されるフローチャート（ステップＳ６及びＳ７）に従って説明する。

[0187] 集計処理部１１２は、一時集計テーブル２０７の作成処理を行なう（ステップＳ６）。なお、ステップＳ６の処理の詳細は、図３０のフローチャートを用いて後述する。

[0188] 集計処理部１１２は、最終集計テーブル２０８の作成を行ない（ステップＳ７）、処理は終了する。なお、ステップＳ７の処理の詳細は、図３３のフローチャートを用いて後述する。

[0189] 図２８は、実施形態の一例での一時集計テーブル２０７の作成処理における入力データを例示する図である。図２９は、実施形態の一例での一時集計テーブル２０７の作成処理における出力データを例示する図である。

[0190] 集計処理部１１２は、図２８に示される圧縮済変異情報２０６と臨床情報２０５とＮＵＬＬ変異構造体２１２ａ及び２１２ｂと一時集計テーブル２０７とに基づき、図２９に示される一時集計テーブル２０７を作成する。

[0191] 一時集計テーブル２０７は、グループ（例えば、人種の「日本人」，「中国人」及び「米国人」）毎に作成され、各ゲノム型位置においてどの変異パターンがいくつ存在するかを示す。一時集計テーブル２０７においてはＮＵＬＬ変異が省略されているため、グループ毎にゲノム型位置の数が異なっている。図２８における入力に使用される一時集計テーブル２０７には、初期状態として、全ての値が０に設定されている。一方、図２９における出力される一時集計テーブル２０７には、各ゲノム型位置においてパターン＃０～＃２の変異パターンがいくつ存在するかを示す値が登録されている。

[0192] 図２９に示される例において、例えば、グループ＃０の一時集計テーブル２０７では、０番目のゲノム型位置には、パターン＃０の変異パターンが１０つ有り、パターン＃１の変異パターンが３つ有り、パターン＃２の変異パ

ターンが２つ有ることが示されている。

[0193] 次に、実施形態の一例における一時集計テーブル２０７の作成処理を、図３０に示されるフローチャート（ステップＳ６１～Ｓ６７）に従って説明する。

[0194] 集計処理部１１２は、臨床情報２０５及び圧縮済変異情報２０６から、順番に、変異パターン及び群情報を取得する（ステップＳ６１）。なお、群情報は、取得された変異パターンが属するグループがｃａｓｅ群に属するのかｃｏｎｔｒｏｌ群に属するのかを示す。

[0195] 集計処理部１１２は、圧縮済変異情報２０６の変異パターンに付帯されたグループＩＤ（例えば、図２４の「ｇｒｏｕｐ＝０」）を取得する（ステップＳ６２）。

[0196] 集計処理部１１２は、次のゲノム型位置を選択する（ステップＳ６３）。

[0197] 集計処理部１１２は、当該ゲノム型位置のパターン値を取得する（ステップＳ６４）。

[0198] 集計処理部１１２は、処理中の群情報，グループＩＤ，ゲノム型位置及びパターンＩＤに対応する一時集計テーブル２０７の要素をインクリメントする（ステップＳ６５）。

[0199] 集計処理部１１２は、次のゲノム型位置があるかを判定する（ステップＳ６６）。

[0200] 次のゲノム型位置がある場合には（ステップＳ６６のＹｅｓルート参照）、処理はステップＳ６３へ戻る。

[0201] 一方、次のゲノム型位置がない場合には（ステップＳ６６のＮｏルート参照）、集計処理部１１２は、圧縮済変異情報２０６において次のレコードがあるかを判定する（ステップＳ６７）。

[0202] 次のレコードがある場合には（ステップＳ６７のＹｅｓルート参照）、処理はステップＳ６１へ戻る。

[0203] 一方、次のレコードがない場合には（ステップＳ６７のＮｏルート参照）、処理は終了する。

[0204] 図３１は、実施形態の一例での最終集計テーブル２０８の作成処理における入力データを例示する図である。図３２は、実施形態の一例での最終集計テーブル２０８の作成処理における出力データを例示する図である。

[0205] データ作成処理部１１１は、図３１に示される各グループの一時集計テーブル２０７とＮＵＬＬ変異構造体２１２ａ及び２１２ｂとに基づき、図３２に示される最終集計テーブル２０８を作成する。

[0206] 図３２に示される最終集計テーブル２０８は、集計された全てのＤＮＡ配列において、各変異にはどの変異パターンがいくつ存在するかを示す。図３２に示される例において、変異＃０では、パターン＃０の変異パターンが５０個存在し、パターン＃１の変異パターンが１００個存在し、パターン＃２の変異パターンが５０個存在することが示されている。

[0207] 図３２に示される最終集計テーブル２０８における変異毎の集計結果に基づき、各変異についての検定処理が行なわれることで、有意差の度合いを示すｐ値が算出され、ｐ値に基づき変異のランキングが出力される。「検定処理」は、カイ二乗検定やフィッシャー検定等であってよい。

[0208] 医師や医療研究者等のユーザは、疾患等との関連が強いと考えられるランキング上位の変異から、疾患関連遺伝子を特定してよい。

[0209] 疾患関連遺伝子は、ランキング上位の変異のうち、１つの変異である場合や、複数の変異の組み合わせである場合がある。そのため、ランキング上位の複数の変異について、様々な組み合わせの変異で集計処理が行なわれることにより、疾患関連遺伝子が特定されてよい。

[0210] 次に、実施形態の一例における最終集計テーブル２０８の作成処理を、図３３に示されるフローチャート（ステップＳ７１～Ｓ７７）に従って説明する。

[0211] 集計処理部１１２は、一時集計テーブル２０７を１つ選択する（ステップＳ７１）。

[0212] 集計処理部１１２は、最終集計テーブル２０８に登録されるゲノム型位置

を１つ選択する（ステップＳ７２）。

[0213] 集計処理部１１２は、ＮＵＬＬ変異構造体２１２ｂに当該ゲノム型位置が登録されているかを判定する（ステップＳ７３）。

[0214] 当該ゲノム型位置が登録されていない場合には（ステップＳ７３のＮｏルート参照）、集計処理部１１２は、一時集計テーブル２０７に基づき、最終集計テーブル２０８の対応するエントリを加算し（ステップＳ７４）、処理はステップＳ７６へ進む。

[0215] 一方、当該ゲノム型位置が登録されている場合には（ステップＳ７３のＹｅｓルート参照）、集計処理部１１２は、ＮＵＬＬ変異構造体２１２ｂに登録されているパターン値に基づいて、最終集計テーブル２０８の対応するエントリを加算する（ステップＳ７５）。

[0216] 集計処理部１１２は、最終集計テーブル２０８に次のゲノム型位置があるかを判定する（ステップＳ７６）。

[0217] 次のゲノム型位置がある場合には（ステップＳ７６のＹｅｓルート参照）、処理がステップＳ７２へ戻る。

[0218] 一方、次のゲノム型位置がない場合には（ステップＳ７６のＮｏルート参照）、集計処理部１１２は、次のグループについての一時集計テーブル２０７があるかを判定する（ステップＳ７７）。

[0219] 次のグループについての一時集計テーブル２０７がある場合には（ステップＳ７７のＹｅｓルート参照）、処理はステップＳ７１へ戻る。

[0220] 一方、次のグループについての一時集計テーブル２０７がない場合には（ステップＳ７７のＮｏルート参照）、処理は終了する。

[0221] 〔Ｂ－３〕効果

データ作成処理部１１１は、それぞれ複数の変異パターンを含む複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、当該同じ値を有する変異パターンを記憶対象から除外する。また、メモリ１２は、データ作成処理部１１１によって除外する処理が施された複数の配列を記憶する。

[0222] これにより、変異パターンのデータ量を削減することができる。また、変異パターンについての情報を全てメモリ１２に記憶させることができるため、変異パターンの集計処理を高速化できる。

[0223] データ作成処理部１１１は、複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において同じ変異位置における変異パターンが同じ値である場合に、当該変異パターンをメモリ１２の記憶対象から除外する処理を行なう。

[0224] これにより、ＤＮＡ配列について人種や性別，年齢等でグループ分けを行なうと、グループの全メンバで同一の変異パターンを有する変異が多数あるというＤＮＡ配列の特性を利用して、変異パターンのデータ量をより削減することができる。

[0225] データ作成処理部１１１は、２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、当該変異パターンをメモリ１２の記憶対象から除外する処理を行なう。

[0226] これにより、複数のグループ内における同一の変異パターンをメモリ１２の記憶対象から効率的に除外することができる。

[0227] データ作成処理部１１１は、２以上のグループの組み合わせの数が所定数以下となるように、データ削減量が小さい複数の組み合わせをマージする。そして、データ作成処理部１１１は、マージされた複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、当該変異パターンをメモリ１２の記憶対象から除外する処理を行なう。

[0228] これにより、グループの組み合わせの数を制限して、データ圧縮への貢献度が小さい組み合わせについては、まとめてデータ圧縮をすることができるため、データ圧縮を効率的に行なうことができる。

[0229] メモリ１２は、１又は２以上のグループのそれぞれについて、配列における除外する処理の対象となった変異パターンの位置を示す情報を記憶する。また、データ作成処理部１１１は、除外する処理の対象となった変異パター

ンの位置を示す情報に基づき、除外する処理の対象となった変異パターンを、除外する処理が施された配列に挿入することにより、除外する処理を施す前の配列を復元する。

[0230] これにより、圧縮された変異パターンについての情報に基づき、配列に含まれる変異パターンの集計や解析等の処理ができる。

[0231] 〔C〕その他

開示の技術は上述した実施形態に限定されるものではなく、本実施形態の趣旨を逸脱しない範囲で種々変形して実施することができる。本実施形態の各構成及び各処理は、必要に応じて取捨選択することができ、あるいは適宜組み合わせてもよい。

[0232] 〔D〕付記

以上の実施形態及び変形例に関し、さらに以下の付記を開示する。

[0233] （付記１）

複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理装置であって、

前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行なう処理部と、

前記処理部によって前記除外する処理が施された複数の配列を記憶する記憶部と、

を備える、情報処理装置。

[0234] （付記２）

前記処理部は、前記複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記１に記載の情報処理装置。

[0235] （付記３）

前記処理部は、前記２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値

である場合に、前記除外する処理を行なう、

付記２に記載の情報処理装置。

[0236] （付記４）

前記処理部は、前記２以上のグループの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記３に記載の情報処理装置。

[0237] （付記５）

前記記憶部は、前記１又は２以上のグループのそれぞれについて、前記配列における前記除外する処理の対象となった変異パターンの位置を示す情報を記憶し、

前記処理部は、前記情報に基づき、前記除外する処理の対象となった変異パターンを、前記除外する処理が施された前記配列に挿入することにより、前記除外する処理を施す前の配列を復元する、

付記２～４のいずれか１項に記載の情報処理装置。

[0238] （付記６）

前記配列は、デオキシリボ核酸の塩基配列である、

付記１～５のいずれか１項に記載の情報処理装置。

[0239] （付記７）

情報処理装置と端末とを有し、複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理システムであって、

前記情報処理装置は、

前記複数の配列間のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行なう処理部

を備え、

前記端末は、

前記情報処理装置に対して前記同一のグループを指定し、前記除外する処理が施された複数の変異パターンを前記情報処理装置から取得する取得部と、

前記取得部によって取得された前記複数の配列を記憶する記憶部と、

を備える、情報処理システム。

[0240] （付記８）

前記処理部は、前記２以上グループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記７に記載の情報処理システム。

[0241] （付記９）

前記処理部は、前記２以上のグループの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記８に記載の情報処理システム。

[0242] （付記１０）

前記記憶部は、前記１又は２以上のグループのそれぞれについて、前記配列における前記除外する処理の対象となった変異パターンの位置を示す情報を記憶し、

前記取得部は、前記情報に基づき、前記除外する処理の対象となった変異パターンを、前記除外する処理が施された前記配列に挿入することにより、前記除外する処理を施す前の配列を復元する、

付記８～９のいずれか１項に記載の情報処理システム。

[0243] （付記１１）

前記配列は、デオキシリボ核酸の塩基配列である、

付記７～１０のいずれか１項に記載の情報処理システム。

[0244] （付記１２）

複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行するコンピュータに、

前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行ない、

前記除外する処理が施された複数の配列を記憶部に記憶させる、

処理を実行させる、プログラム。

[0245] （付記１３）

前記複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、前記除外する処理を行なう、

処理を前記コンピュータに実行させる、付記１２に記載のプログラム。

[0246] （付記１４）

前記２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

処理を前記コンピュータに実行させる、付記１３に記載のプログラム。

[0247] （付記１５）

前記第２以上のグループの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

処理を前記コンピュータに実行させる、付記１４に記載のプログラム。

[0248] （付記１６）

前記１又は２以上のグループのそれぞれについて、前記配列における前記

除外する処理の対象となった変異パターンの位置を示す情報を前記記憶部に記憶し、

前記情報に基づき、前記除外する処理の対象となった変異パターンを、前記除外する処理が施された前記配列に挿入することにより、前記除外する処理を施す前の配列を復元する、

処理を前記コンピュータに実行させる、付記１２～１５のいずれか１項に記載のプログラム。

[0249] （付記１７）

複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理方法であって、

前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行ない、

前記除外する処理が施された複数の配列を記憶部に記憶させる、

情報処理方法。

[0250] （付記１８）

前記複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記１７に記載の情報処理方法。

[0251] （付記１９）

前記２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記１８に記載の情報処理方法。

[0252] （付記２０）

前記２以上のグループとの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせ

に含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

付記１９に記載の情報処理方法。

符号の説明

[0253] １　：情報処理装置

２　：端末

３　：ネットワーク

１０　：バス線

１１　：ＣＰＵ

１２　：メモリ

１３　：記憶装置

１４　：媒体読取装置

１５　：表示制御装置

１６　：表示装置

１７　：入力装置

１８　：通信制御装置

１００　：情報処理システム

１１１　：データ作成処理部

１１２　：集計処理部

２０　：ＣＰＵ

２１　：取得部

２２　：メモリ

２０１　：ゲノム型構造

２０２　：変異マスタ情報

２０３　：元データ変異情報

２０４　：非圧縮変異情報

２０５　：臨床情報

２０６　：圧縮済変異情報

２０７　：一時集計テーブル

２０７ａ　：一時集計テーブル

２０７ｂ　：一時集計テーブル

２０８　：最終集計テーブル

２０８ａ　：ｃｏｎｔｒｏｌ集計テーブル

２０８ｂ　：ｃａｓｅ集計テーブル

２０９　：グループ統計情報

２０９ａ　：ＮＵＬＬ変異集計情報

２０９ｂ　：圧縮サイズ情報

２１０　：グループ分け情報

２１０ａ　：組み合わせＮＵＬＬ変異集計情報

２１０ｂ　：組み合わせ圧縮サイズ情報

２１１　：ランキング情報

２１２ａ　：ＮＵＬＬ変異構造体

２１２ｂ　：ＮＵＬＬ変異構造体

２１３　：グループＩＤ対応配列

２１４　：組み合わせ

３０１　：ゲノム型構造

３０２　：変異マスタ情報

３０３　：変異情報

３０３ａ　：罹患者の変異情報

３０３ｂ　：健常者の変異情報

３０４ａ　：入力データ

３０４ｂ　：集計テーブル

３０５　：臨床情報

ＲＭ　：記録媒体

請求の範囲

[請求項1] 複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理装置であって、

前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行なう処理部と、

前記処理部によって前記除外する処理が施された複数の配列を記憶する記憶部と、

を備える、情報処理装置。

[請求項2] 前記処理部は、前記複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項１に記載の情報処理装置。

[請求項3] 前記処理部は、前記２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項２に記載の情報処理装置。

[請求項4] 前記処理部は、前記２以上のグループの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項３に記載の情報処理装置。

[請求項5] 前記記憶部は、前記１又は２以上のグループのそれぞれについて、前記配列における前記除外する処理の対象となった変異パターンの位置を示す情報を記憶し、

前記処理部は、前記情報に基づき、前記除外する処理の対象となった変異パターンを、前記除外する処理が施された前記配列に挿入することにより、前記除外する処理を施す前の配列を復元する、

請求項２～４のいずれか１項に記載の情報処理装置。

[請求項6] 前記配列は、デオキシリボ核酸の塩基配列である、

請求項１～５のいずれか１項に記載の情報処理装置。

[請求項7] 情報処理装置と端末とを有し、複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理システムであって、

前記情報処理装置は、

前記複数の配列間のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行なう処理部

を備え、

前記端末は、

前記情報処理装置に対して前記同一のグループを指定し、前記除外する処理が施された複数の変異パターンを前記情報処理装置から取得する取得部と、

前記取得部によって取得された前記複数の配列を記憶する記憶部と、

を備える、情報処理システム。

[請求項8] 前記処理部は、前記２以上グループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項７に記載の情報処理システム。

[請求項9] 前記処理部は、前記２以上のグループの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前

記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項８に記載の情報処理システム。

[請求項10] 前記記憶部は、前記１又は２以上のグループのそれぞれについて、前記配列における前記除外する処理の対象となった変異パターンの位置を示す情報を記憶し、

前記取得部は、前記情報に基づき、前記除外する処理の対象となった変異パターンを、前記除外する処理が施された前記配列に挿入することにより、前記除外する処理を施す前の配列を復元する、

請求項８～９のいずれか１項に記載の情報処理システム。

[請求項11] 前記配列は、デオキシリボ核酸の塩基配列である、

請求項７～１０のいずれか１項に記載の情報処理システム。

[請求項12] 複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行するコンピュータに、

前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行ない、

前記除外する処理が施された複数の配列を記憶部に記憶させる、

処理を実行させる、プログラム。

[請求項13] 前記複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、前記除外する処理を行なう、

処理を前記コンピュータに実行させる、請求項１２に記載のプログラム。

[請求項14] 前記２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値で

ある場合に、前記除外する処理を行なう、

処理を前記コンピュータに実行させる、請求項１３に記載のプログラム。

[請求項15] 前記第２以上のグループの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

処理を前記コンピュータに実行させる、請求項１４に記載のプログラム。

[請求項16] 前記１又は２以上のグループのそれぞれについて、前記配列における前記除外する処理の対象となった変異パターンの位置を示す情報を前記記憶部に記憶し、

前記情報に基づき、前記除外する処理の対象となった変異パターンを、前記除外する処理が施された前記配列に挿入することにより、前記除外する処理を施す前の配列を復元する、

処理を前記コンピュータに実行させる、請求項１２～１５のいずれか１項に記載のプログラム。

[請求項17] 複数の配列それぞれに含まれる複数の変異パターンに応じて前記複数の配列に関する処理を実行する情報処理方法であって、

前記複数の配列間において、同じ変異位置における変異パターンが同じ場合に、同じ変異パターンを記憶対象から除外する処理を行ない、

前記除外する処理が施された複数の配列を記憶部に記憶させる、

情報処理方法。

[請求項18] 前記複数の配列のうち、１又は２以上のグループのうちの同一のグループに含まれる複数の配列間において、同じ変異位置における変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項１７に記載の情報処理方法。

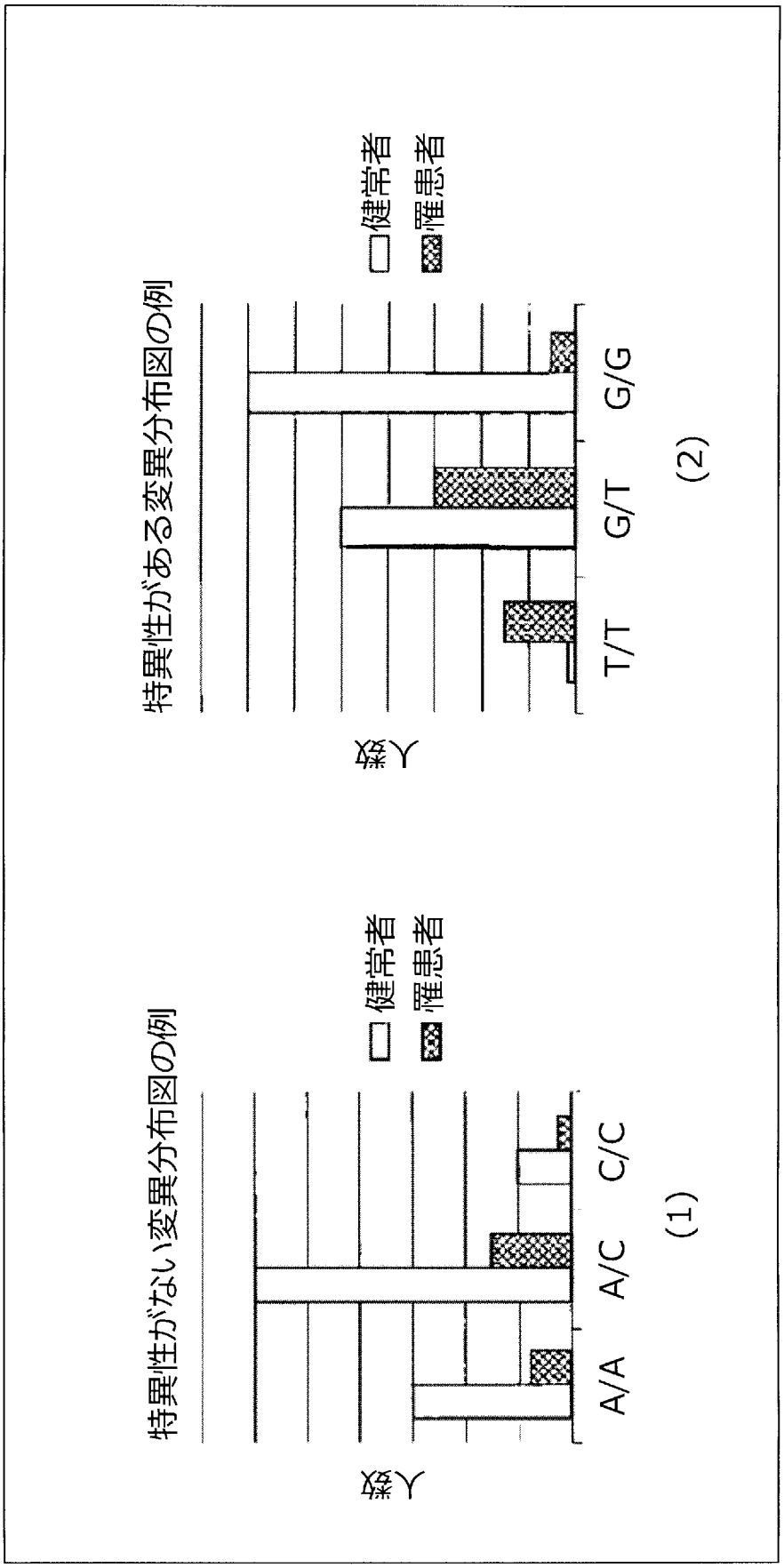
[請求項19] 前記２以上のグループのうちの第１のグループ且つ第２のグループに含まれる複数の配列間において、対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項１８に記載の情報処理方法。

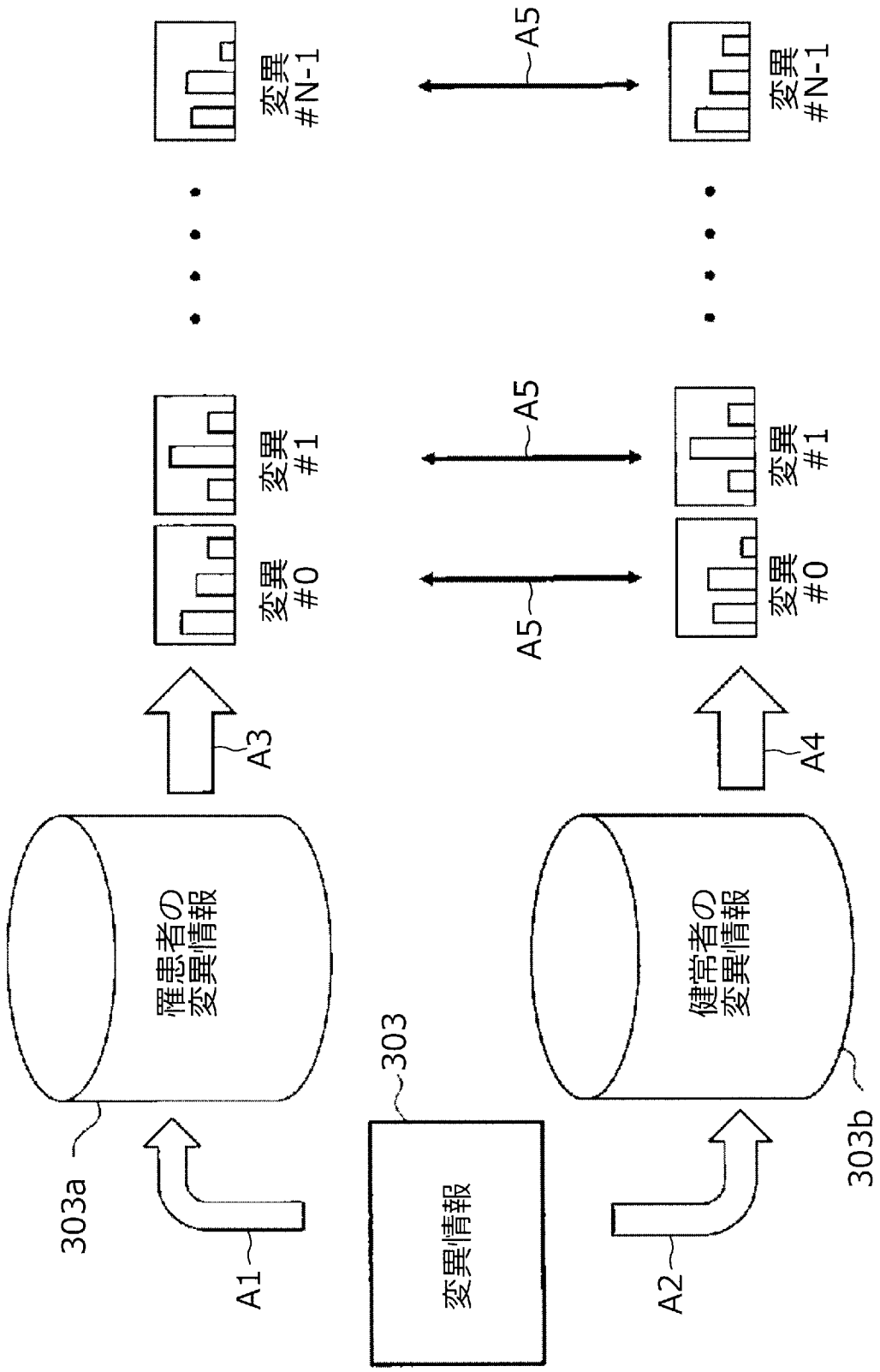
[請求項20] 前記２以上のグループとの組み合わせの数が所定数以下となるように、前記組み合わせのうち、前記除外する処理による前記記憶部に記憶されるデータ量の削減量が小さい複数の組み合わせについては、当該複数の組み合わせに含まれる複数の配列間において対応する変異パターンが同じ値である場合に、前記除外する処理を行なう、

請求項１９に記載の情報処理方法。

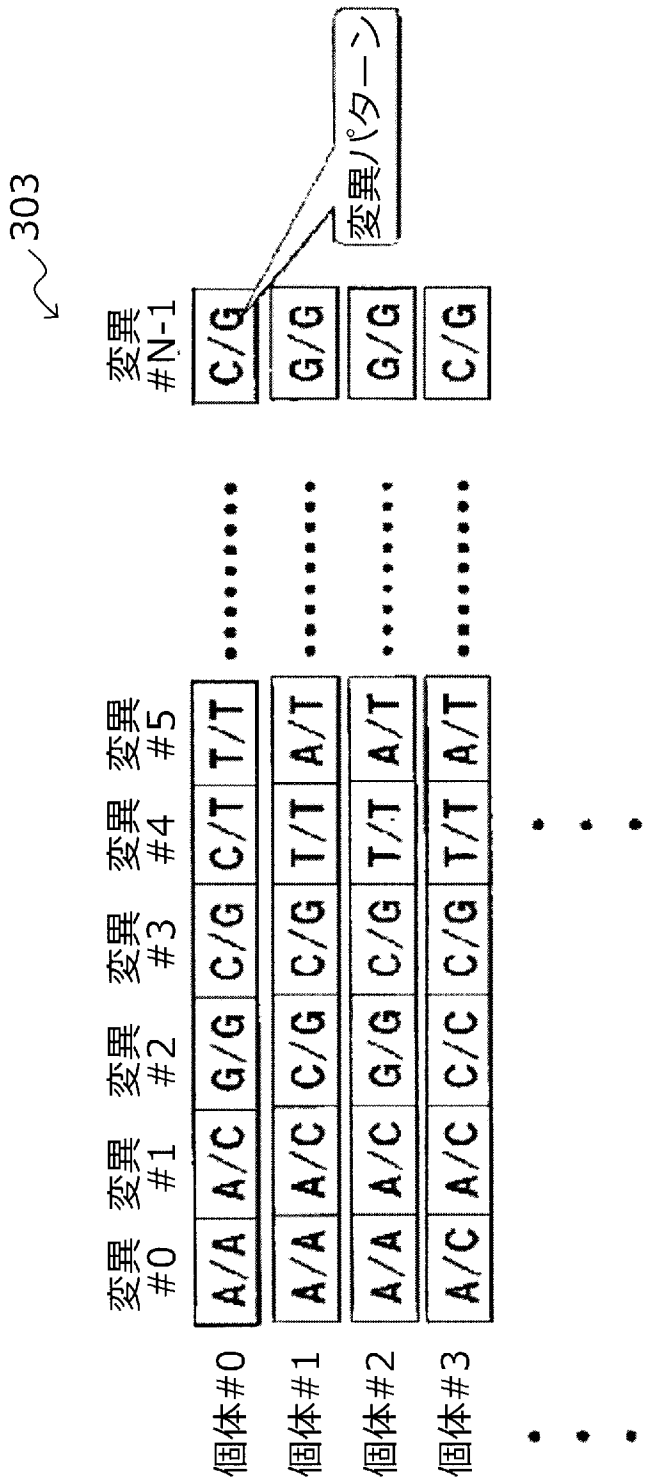
[図1]


特異性がない変異分布図の例
人数
□健常者
▨罹患者
A/A
A/C
C/C
(1)
特異性がある変異分布図の例
人数
□健常者
▨罹患者
T/T
G/T
G/G
(2)

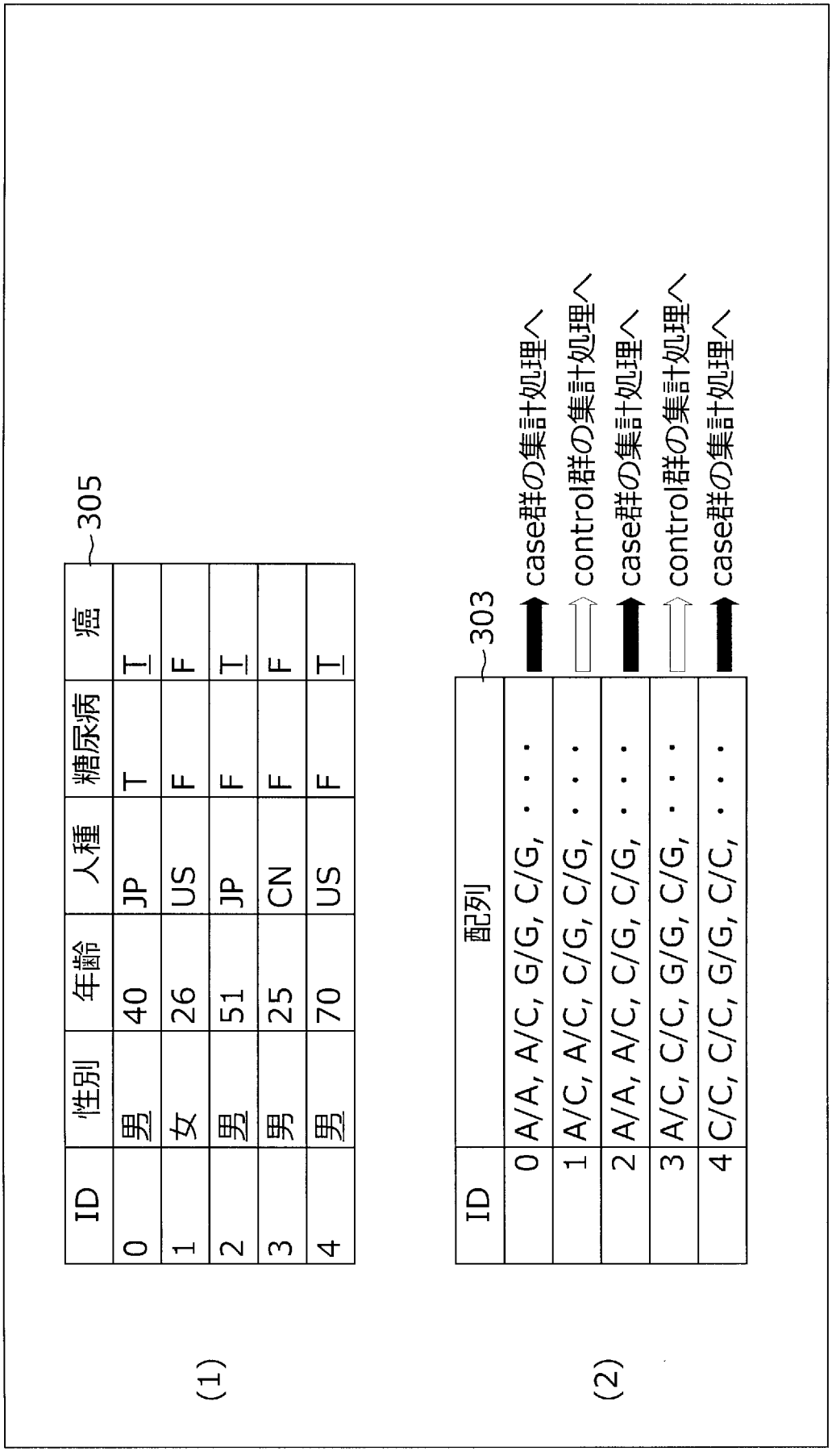
[図2]


罹患者の変異情報
303a
A1
変異情報
303
A2
健常者の変異情報
303b
A3
変異 #0
変異 #1
変異 #N-1
A4
A5
A5
A5
変異 #0
変異 #1
変異 #N-1

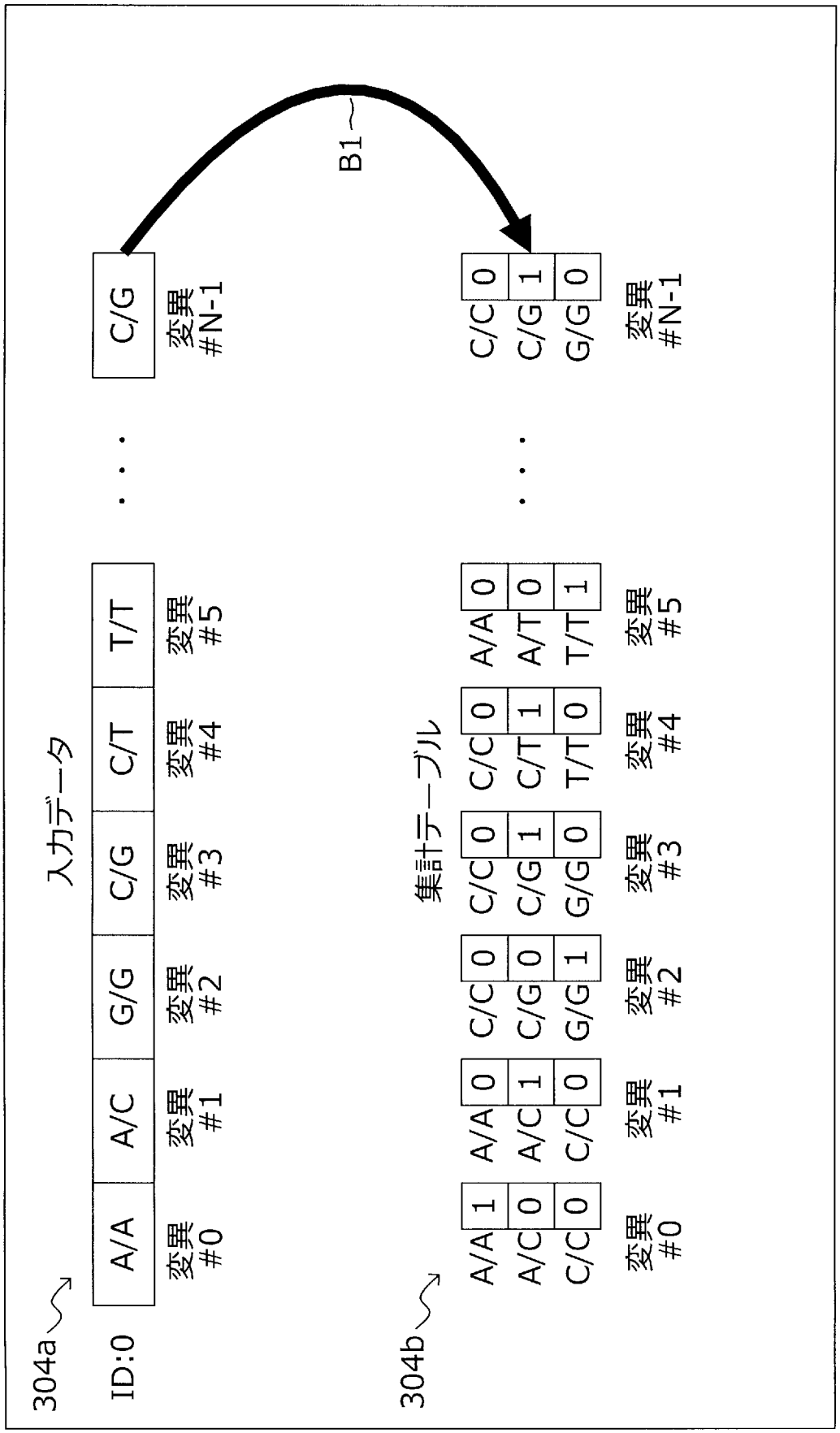
[図3]


303
変異パターン
変異 #0
変異 #1
変異 #2
変異 #3
変異 #4
変異 #5
変異 #N-1
個体#0
個体#1
個体#2
個体#3
A/A A/C G/G C/G C/T T/T C/G
A/A A/C C/G C/G T/T A/T G/G
A/A A/C G/G C/G T/T A/T G/G
A/C A/C C/C C/G T/T A/T C/G

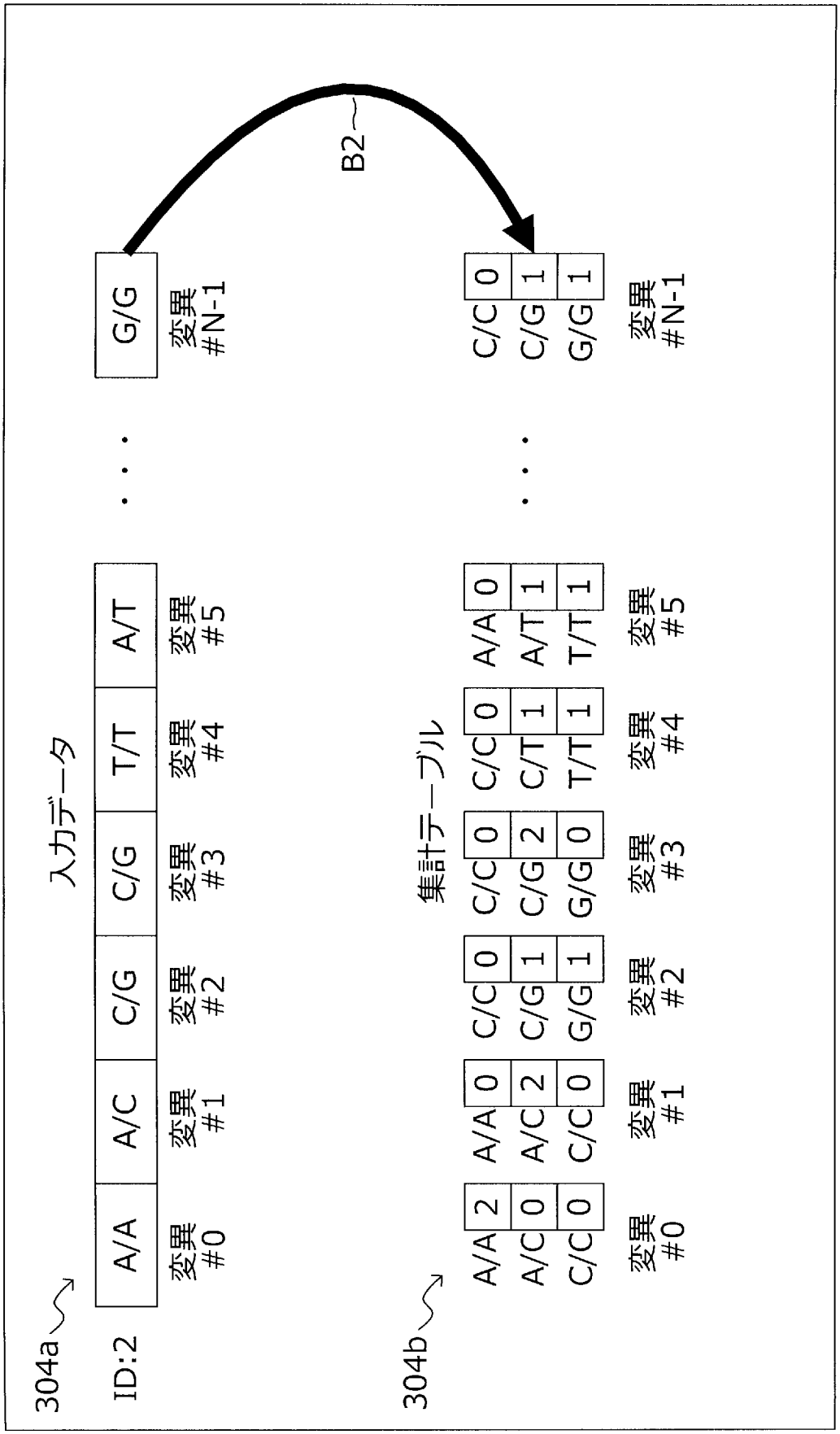
[図4]


(1)
305
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
(2)
303
ID 配列
0 A/A, A/C, G/G, C/G, ・・・
1 A/C, A/C, C/G, C/G, ・・・
2 A/A, A/C, C/G, C/G, ・・・
3 A/C, C/C, G/G, C/G, ・・・
4 C/C, C/C, G/G, C/C, ・・・
case群の集計処理へ
control群の集計処理へ
case群の集計処理へ
control群の集計処理へ
case群の集計処理へ

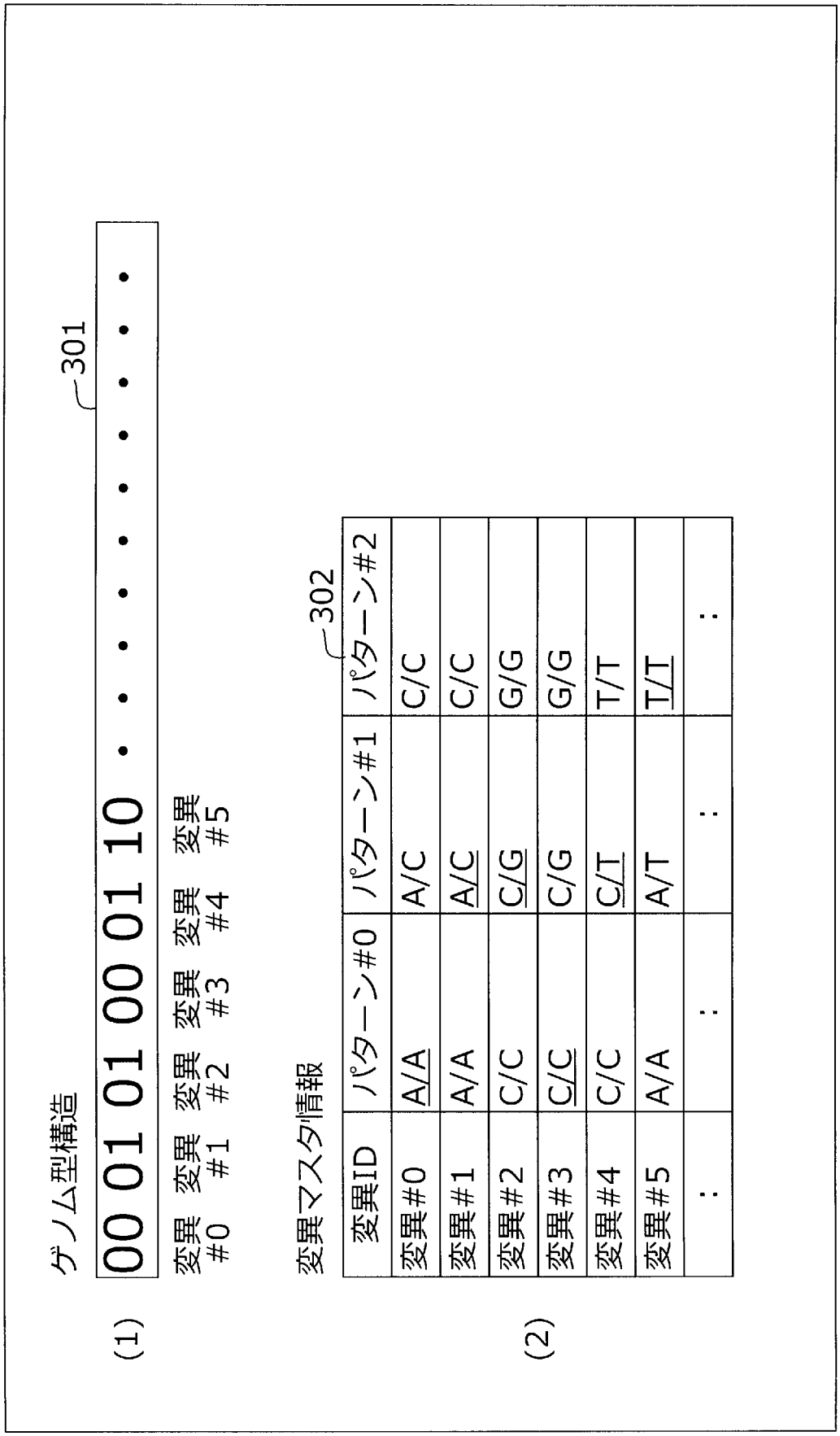
[図5]


入力データ
304a
ID:0
A/A
A/C
G/G
C/G
C/T
T/T
変異 #0
変異 #1
変異 #2
変異 #3
変異 #4
変異 #5
C/G
変異 #N-1
B1
集計テーブル
304b
A/A 1
A/C 0
C/C 0
A/A 0
A/C 1
C/C 0
C/C 0
C/G 0
G/G 1
C/C 0
C/G 1
G/G 0
C/C 0
C/T 1
T/T 0
A/A 0
A/T 0
T/T 1
C/C 0
C/G 1
G/G 0
変異 #N-1

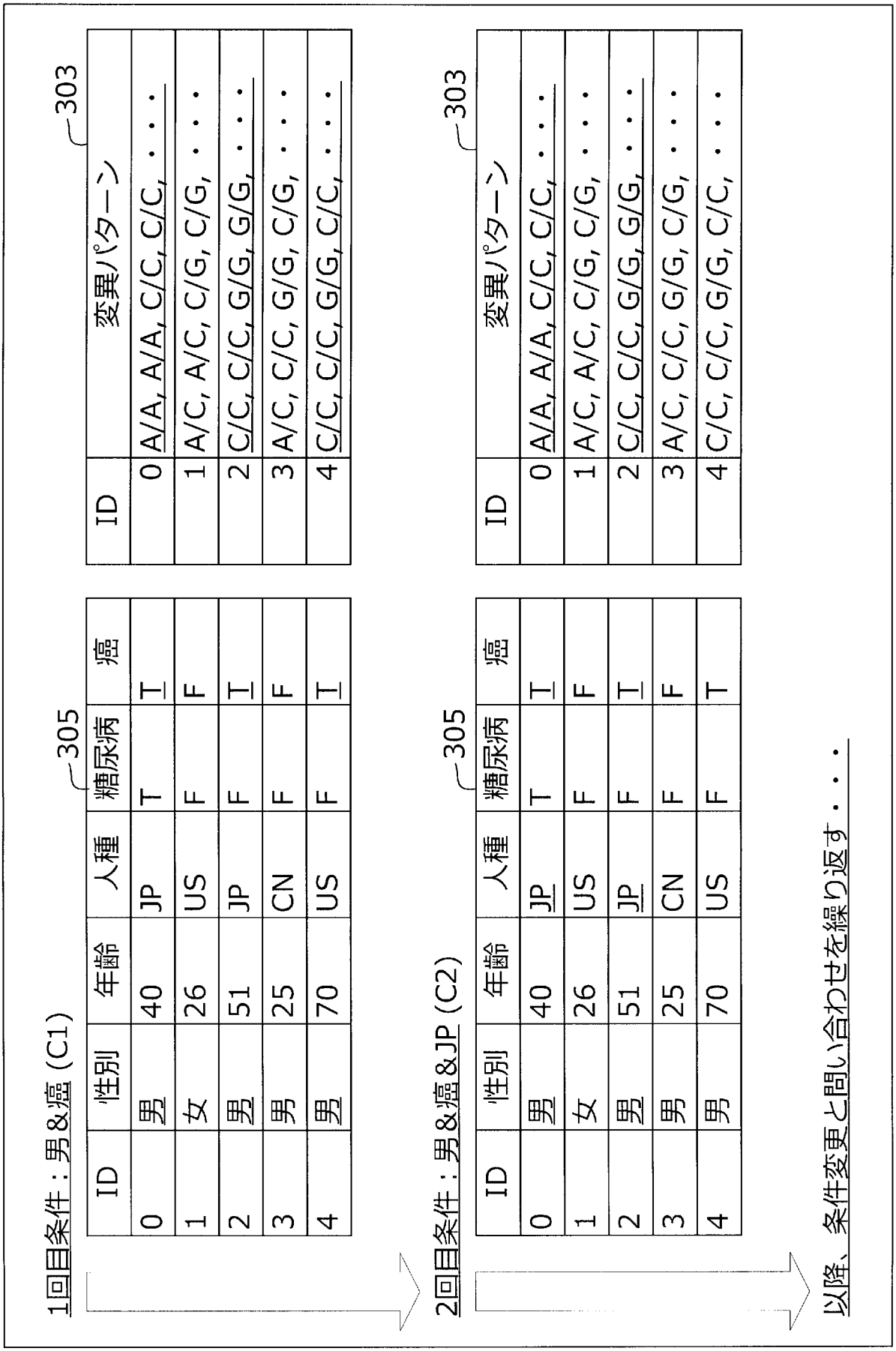
[図6]


入力データ
304a
ID:2
A/A
A/C
C/G
C/G
T/T
A/T
変異 #0
変異 #1
変異 #2
変異 #3
変異 #4
変異 #5
G/G
変異 #N-1
B2
集計テーブル
304b
A/A 2
A/C 0
C/C 0
A/A 0
A/C 2
C/C 0
C/C 0
C/G 1
G/G 1
C/C 0
C/G 2
G/G 0
C/C 0
C/T 1
T/T 1
A/A 0
A/T 1
T/T 1
C/C 0
C/G 1
G/G 1
変異 #0
変異 #1
変異 #2
変異 #3
変異 #4
変異 #5
変異 #N-1

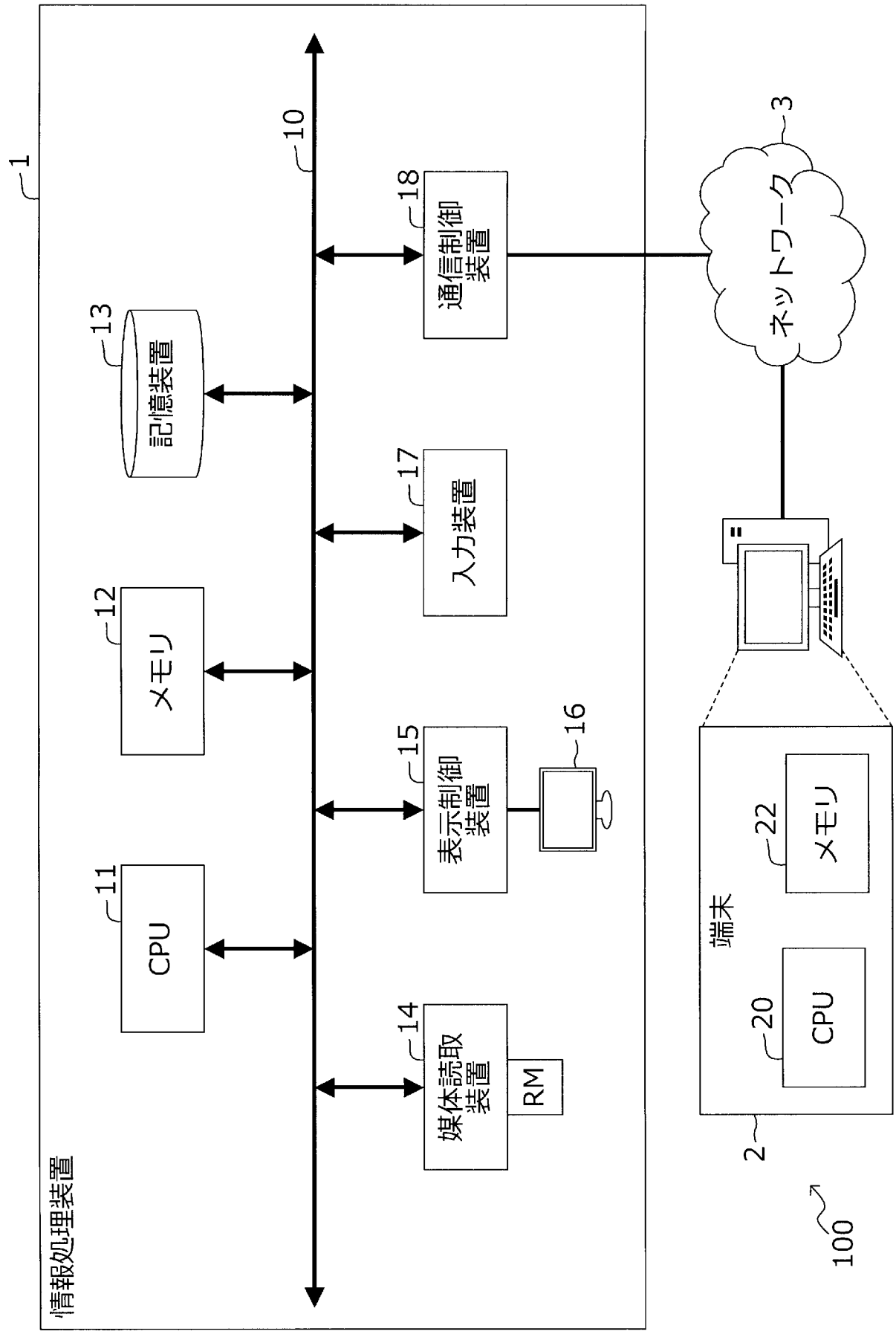
[図7]


(1)
ゲノム型構造
301
00 01 01 00 01 10 ・・・・・・・・・・
変異 #0
変異 #1
変異 #2
変異 #3
変異 #4
変異 #5
(2)
変異マスタ情報
302
変異ID
パターン#0
パターン#1
パターン#2
変異#0 A/A A/C C/C
変異#1 A/A A/C C/C
変異#2 C/C C/G G/G
変異#3 C/C C/G G/G
変異#4 C/C C/T T/T
変異#5 A/A A/T T/T

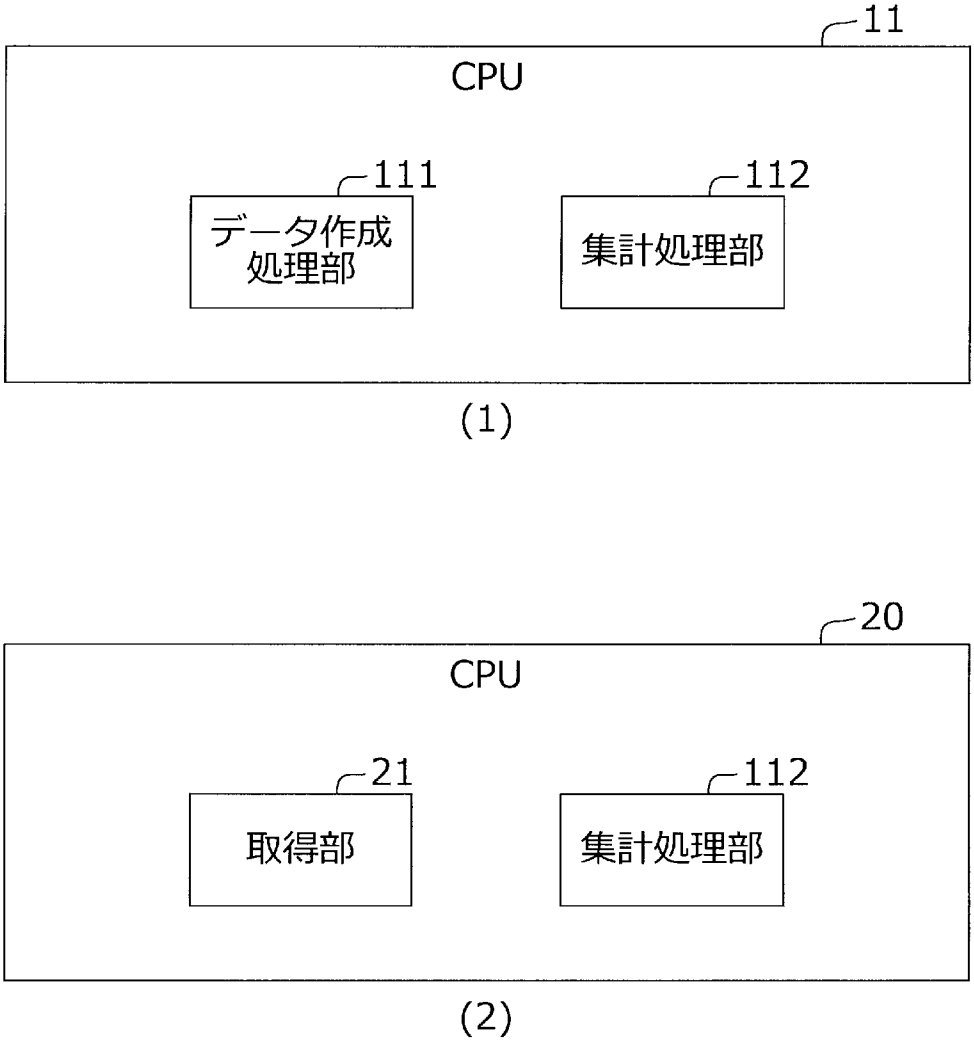
[図8]


1回目条件：男&癌 (C1)
305
303
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
ID 変異パターン
0 A/A, A/A, C/C, C/C, ・・・
1 A/C, A/C, C/G, C/G, ・・・
2 C/C, C/C, G/G, G/G, ・・・
3 A/C, C/C, G/G, C/G, ・・・
4 C/C, C/C, G/G, C/C, ・・・
2回目条件：男&癌&JP (C2)
305
303
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
ID 変異パターン
0 A/A, A/A, C/C, C/C, ・・・
1 A/C, A/C, C/G, C/G, ・・・
2 C/C, C/C, G/G, G/G, ・・・
3 A/C, C/C, G/G, C/G, ・・・
4 C/C, C/C, G/G, C/C, ・・・
以降、条件変更と問い合わせを繰り返す・・・

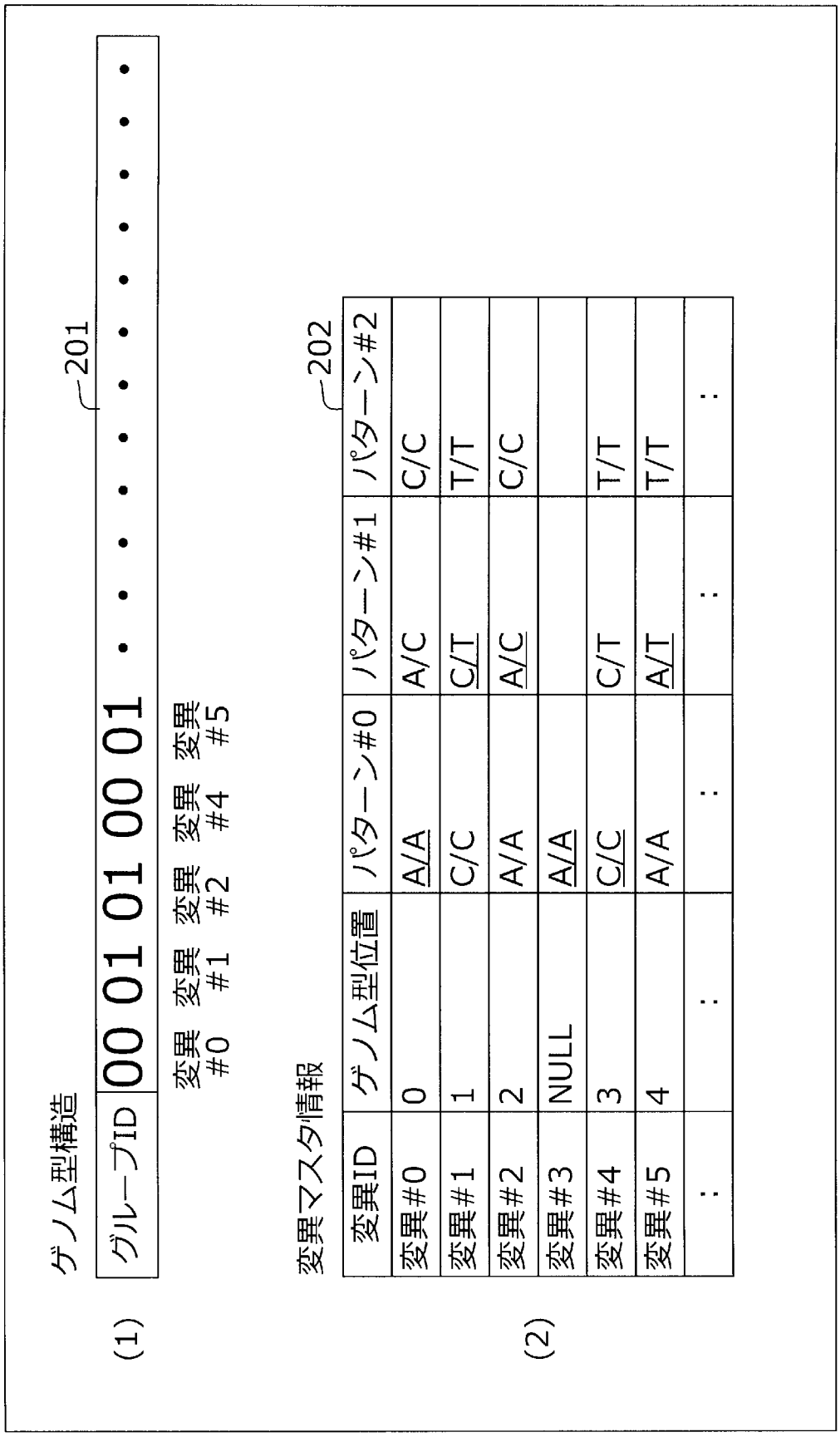
[図9]


1
情報処理装置
10
11
CPU
12
メモリ
13
記憶装置
14
媒体読取装置
RM
15
表示制御装置
16
17
入力装置
18
通信制御装置
3
ネットワーク
2
端末
20
CPU
22
メモリ
100

[図10]

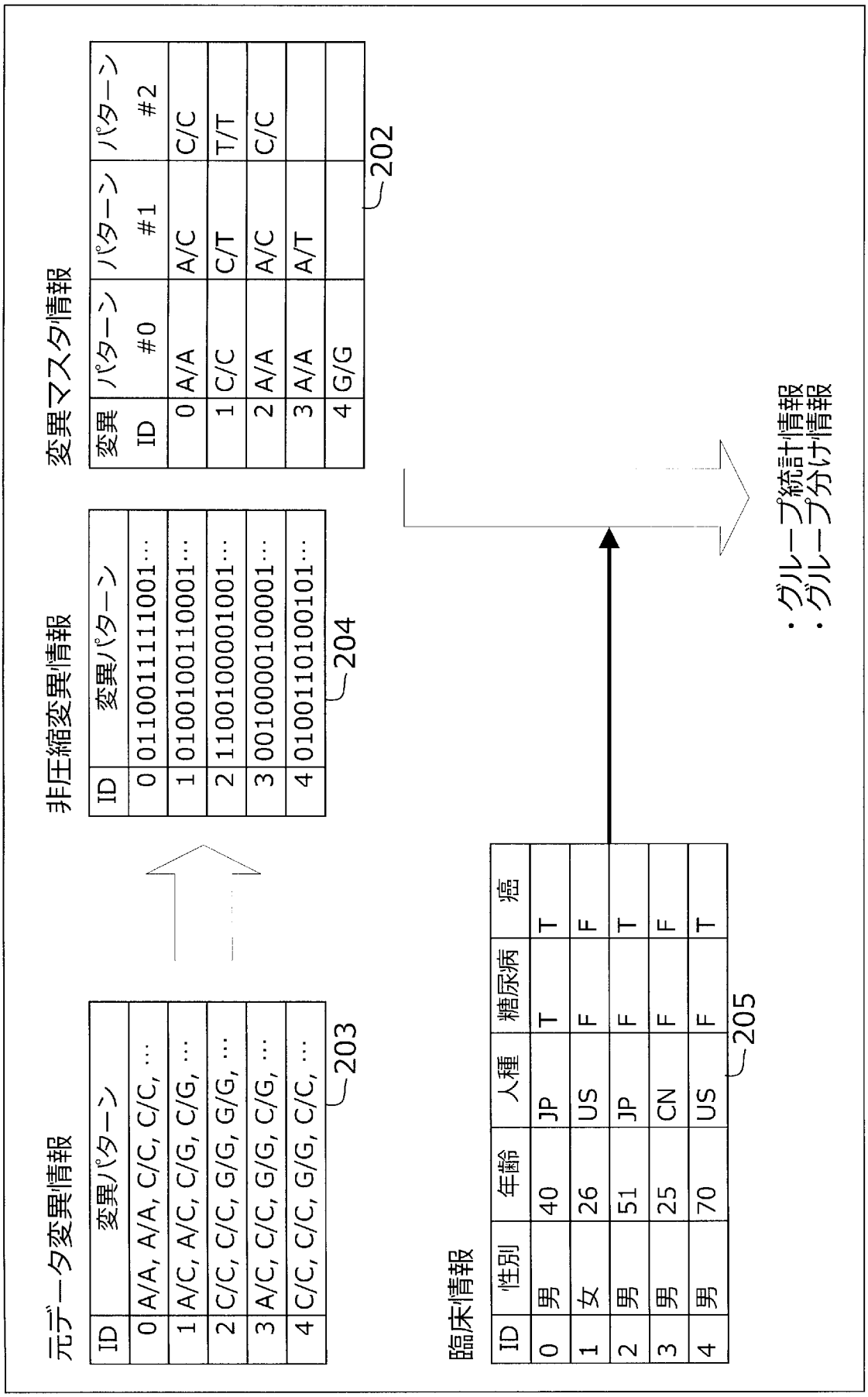


[図11]

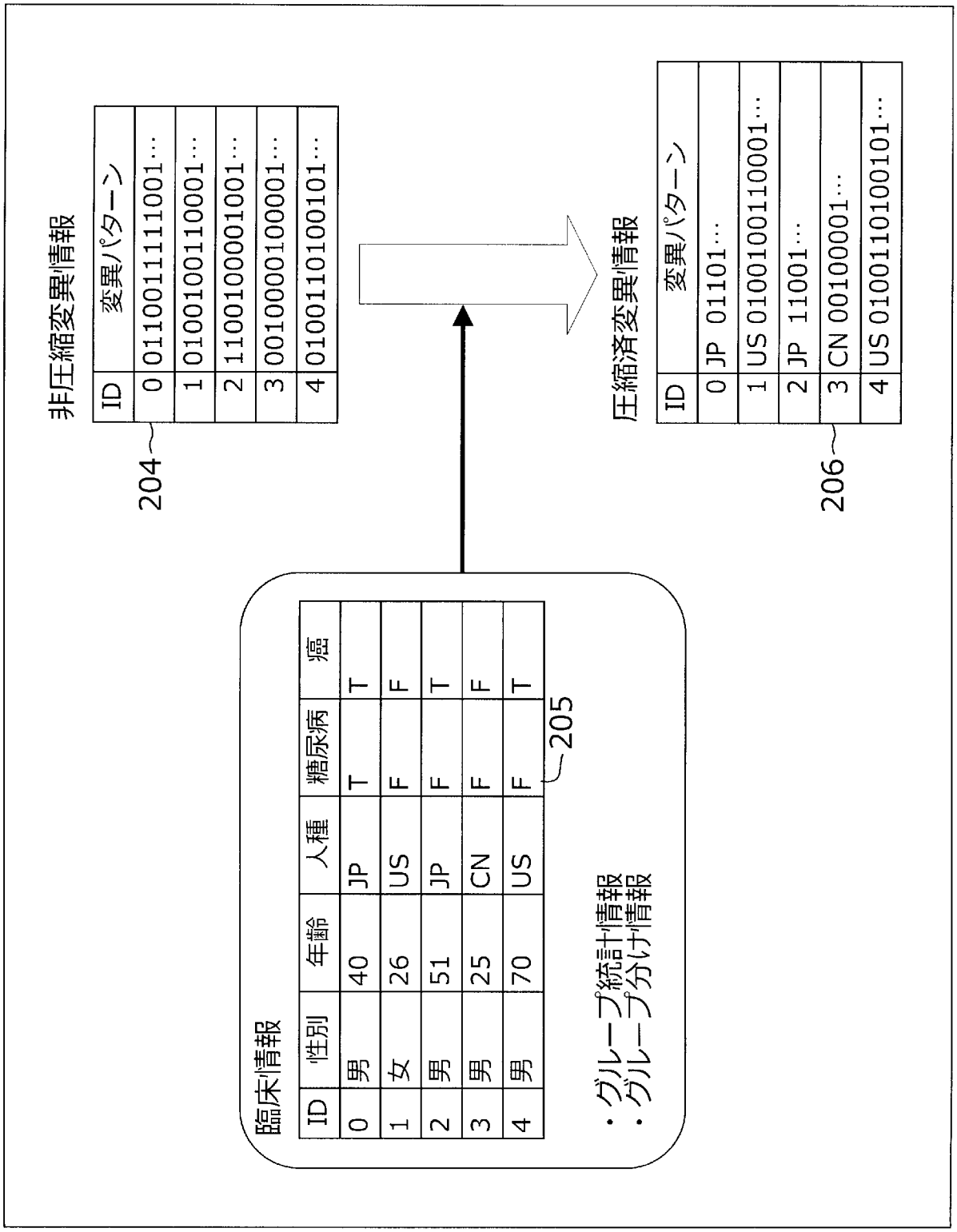


変異ID	ゲノム型位置	パターン#0	パターン#1	パターン#2
変異#0	0	A/A	A/C	C/C
変異#1	1	C/C	C/T	T/T
変異#2	2	A/A	A/C	C/C
変異#3	NULL	A/A		
変異#4	3	C/C	C/T	T/T
変異#5	4	A/A	A/T	T/T
:	:	:	:	:

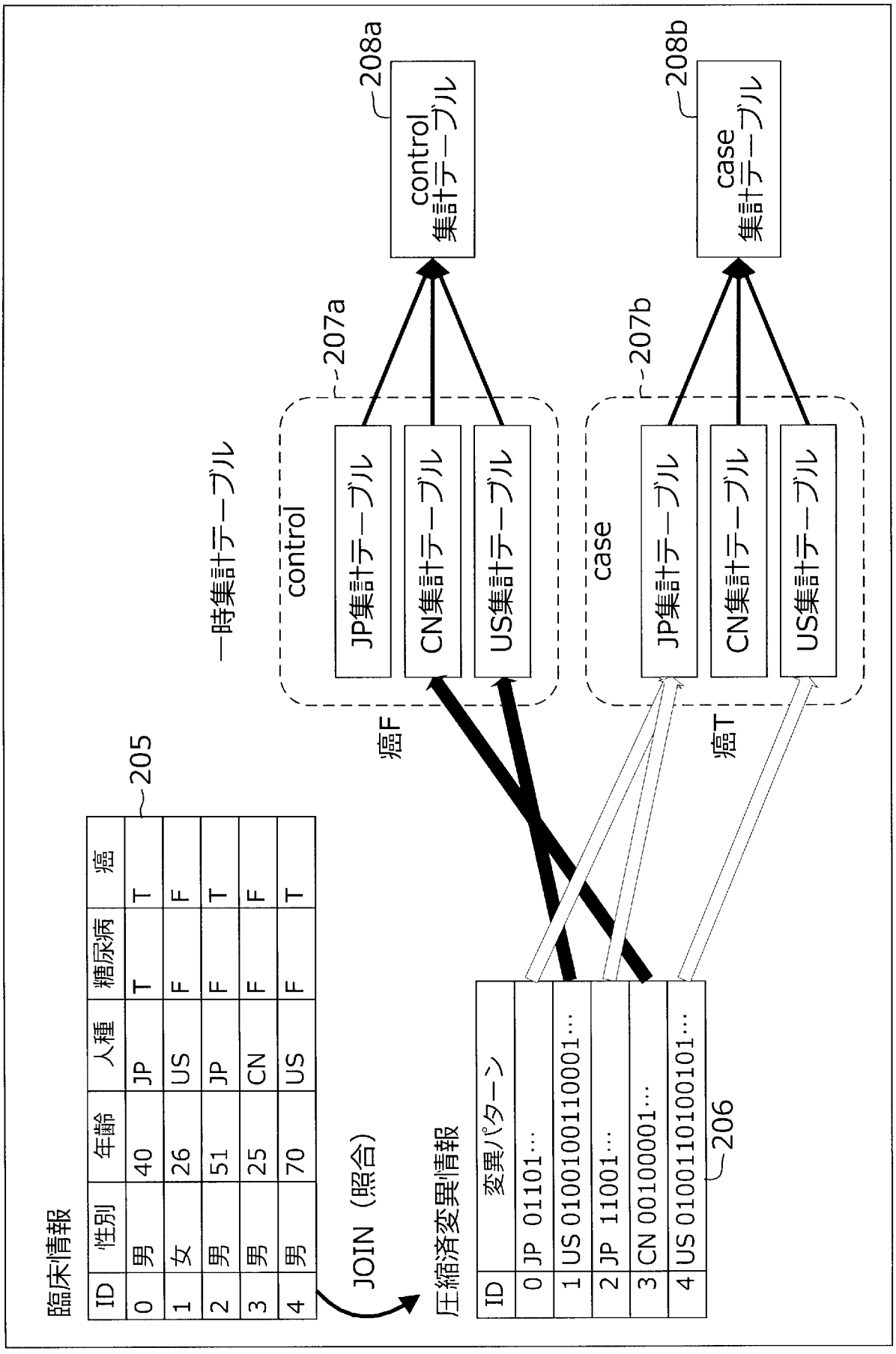
[図12]


元データ変異情報
ID
変異パターン
0 A/A, A/A, C/C, C/C, …
1 A/C, A/C, C/G, C/G, …
2 C/C, C/C, G/G, G/G, …
3 A/C, C/C, G/G, C/G, …
4 C/C, C/C, G/G, C/C, …
203
非圧縮変異情報
ID
変異パターン
0 0110011111110011…
1 0100100110001…
2 1100100001001…
3 0010000100001…
4 0100110100101…
204
変異マスタ情報
変異ID
パターン#0
パターン#1
パターン#2
0 A/A A/C C/C
1 C/C C/T T/T
2 A/A A/C C/C
3 A/A A/T
4 G/G
202
臨床情報
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
205
・グループ統計情報
・グループ分け情報

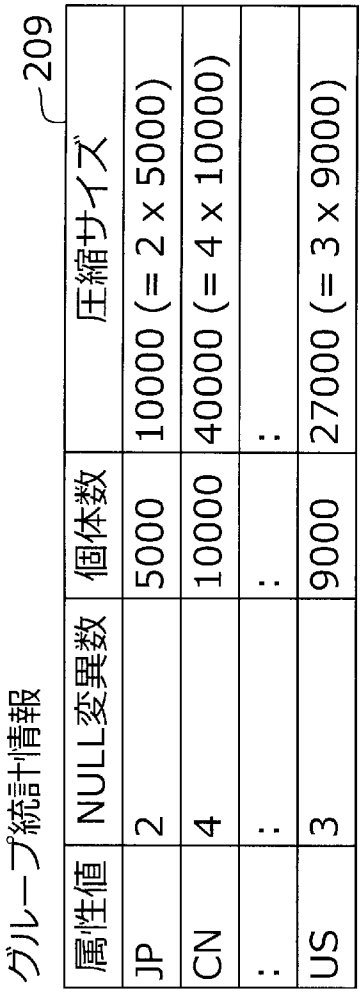
[図13]


非圧縮変異情報
ID
変異パターン
0 0110011111100 1…
1 0100100110001…
2 1100100001001…
3 0010000100001…
4 0100110100101…
204
圧縮済変異情報
ID
変異パターン
0 JP 01101…
1 US 0100100110001…
2 JP 11001…
3 CN 00100001…
4 US 0100110100101…
206
臨床情報
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
205
・グループ統計情報
・グループ分け情報

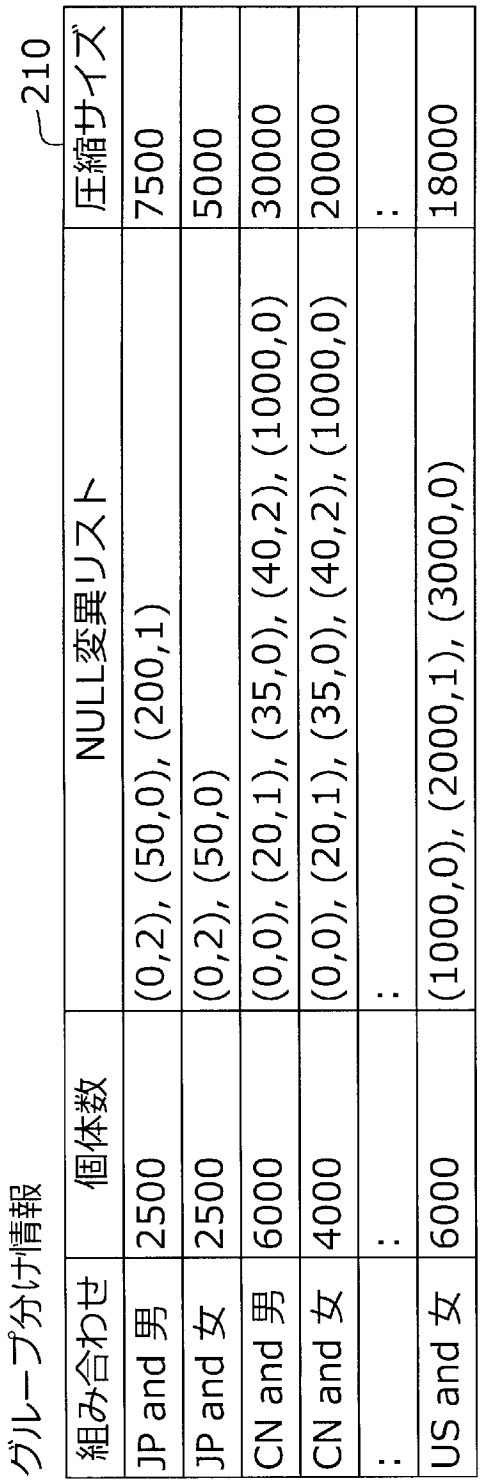
[図14]


臨床情報
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
205
JOIN（照合）
圧縮済変異情報
ID 変異パターン
0 JP 01101…
1 US 0100100110001…
2 JP 11001…
3 CN 00100001…
4 US 0100110100101…
206
一時集計テーブル
control
JP集計テーブル
CN集計テーブル
US集計テーブル
207a
癌F
case
JP集計テーブル
CN集計テーブル
US集計テーブル
207b
癌T
control
集計テーブル
208a
case
集計テーブル
208b

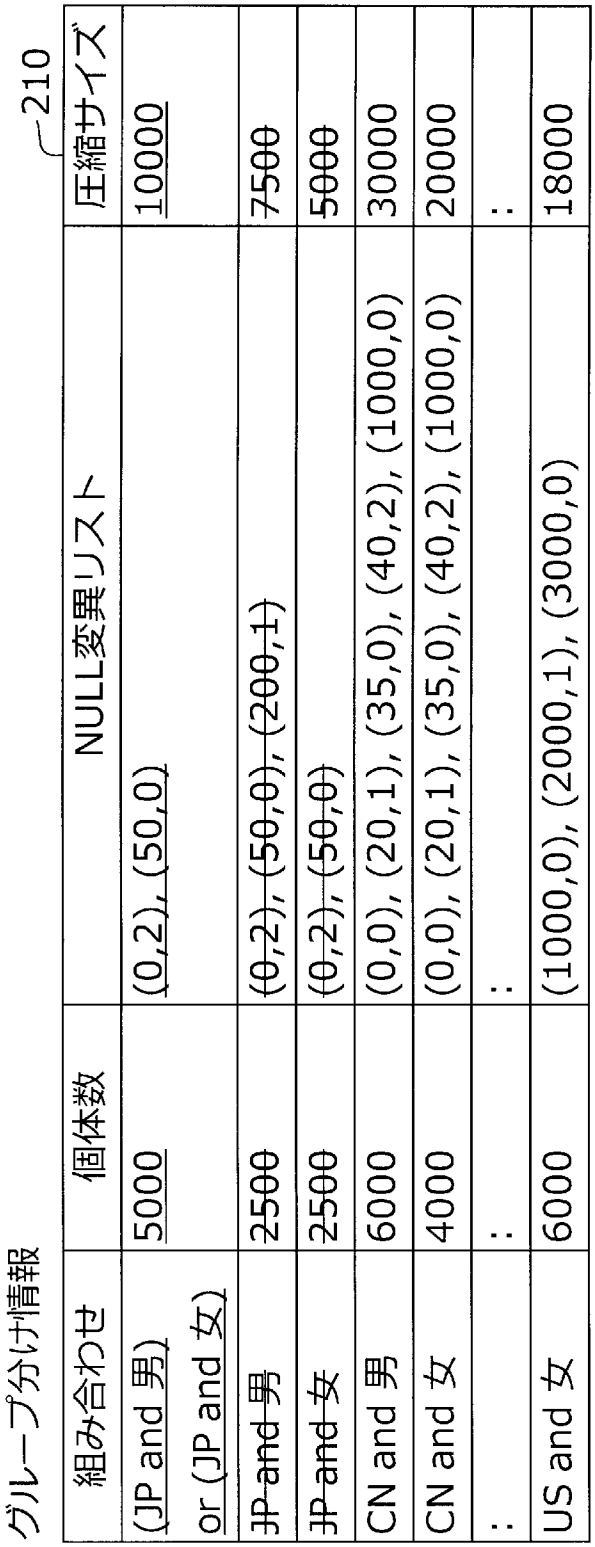
[図15]


209
グループ統計情報
属性値
NULL変異数
個体数
圧縮サイズ
JP
2
5000
10000 (= 2 x 5000)
CN
4
10000
40000 (= 4 x 10000)
:
:
:
:
US
3
9000
27000 (= 3 x 9000)

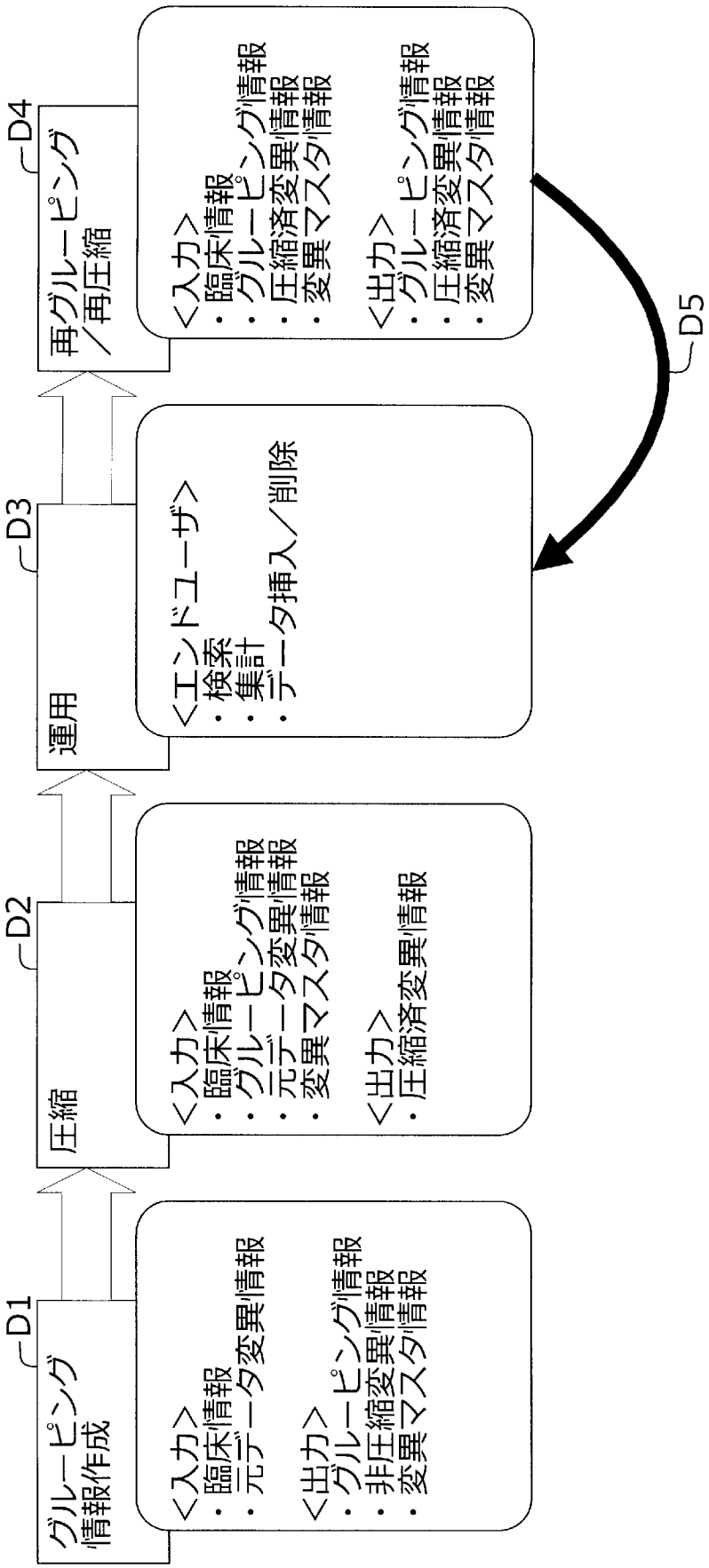
[図16]


グループ分け情報
210
組み合わせ
個体数
NULL変異リスト
圧縮サイズ
JP and 男
2500
(0,2), (50,0), (200,1)
7500
JP and 女
2500
(0,2), (50,0)
5000
CN and 男
6000
(0,0), (20,1), (35,0), (40,2), (1000,0)
30000
CN and 女
4000
(0,0), (20,1), (35,0), (40,2), (1000,0)
20000
:
:
:
:
US and 女
6000
(1000,0), (2000,1), (3000,0)
18000

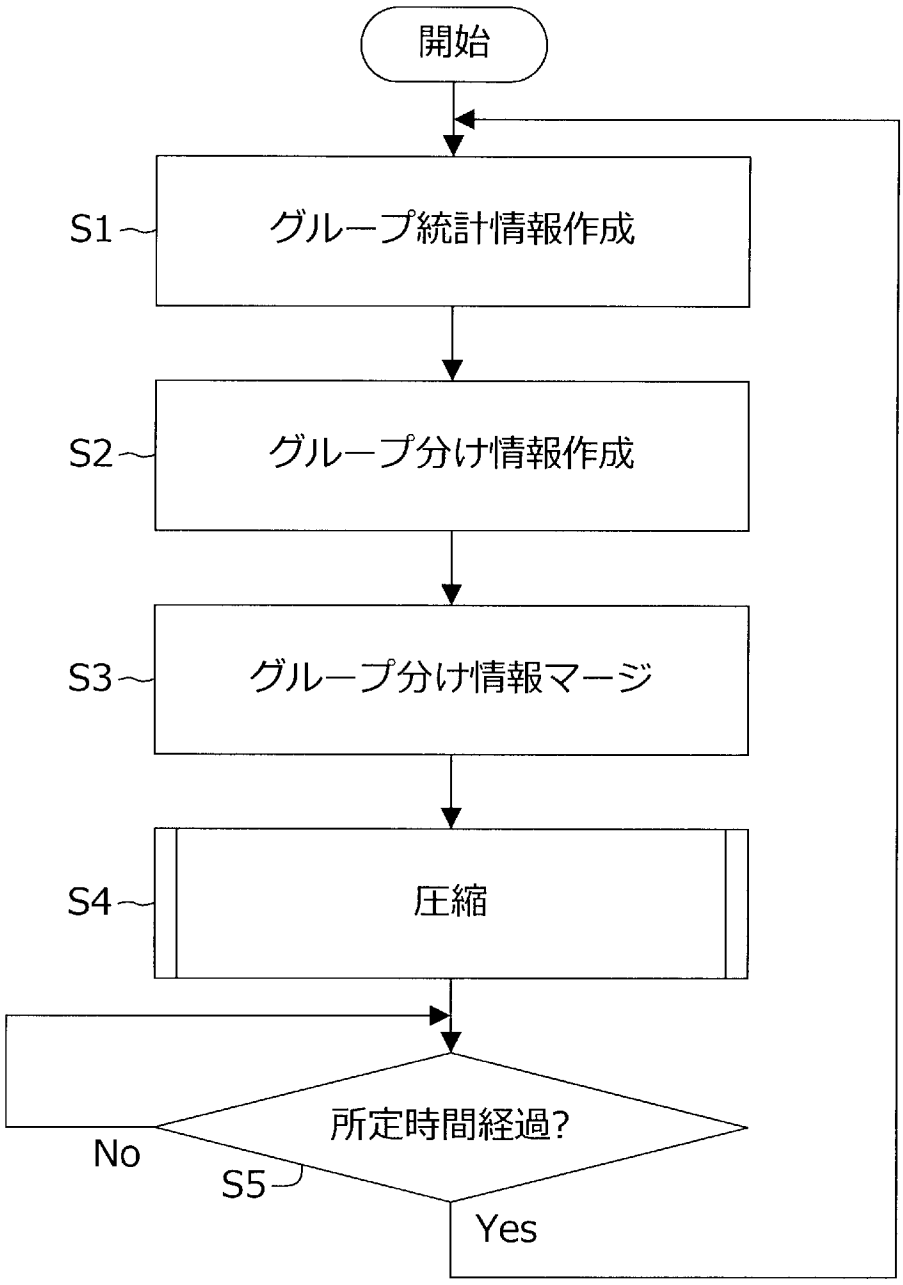
[図17]


グループ分け情報
210
組み合わせ
個体数
NULL変異リスト
圧縮サイズ
(JP and 男) or (JP and 女)
5000
(0,2), (50,0)
10000
JP and 男
2500
(0,2), (50,0), (200,1)
7500
JP and 女
2500
(0,2), (50,0)
5000
CN and 男
6000
(0,0), (20,1), (35,0), (40,2), (1000,0)
30000
CN and 女
4000
(0,0), (20,1), (35,0), (40,2), (1000,0)
20000
:
:
:
:
US and 女
6000
(1000,0), (2000,1), (3000,0)
18000

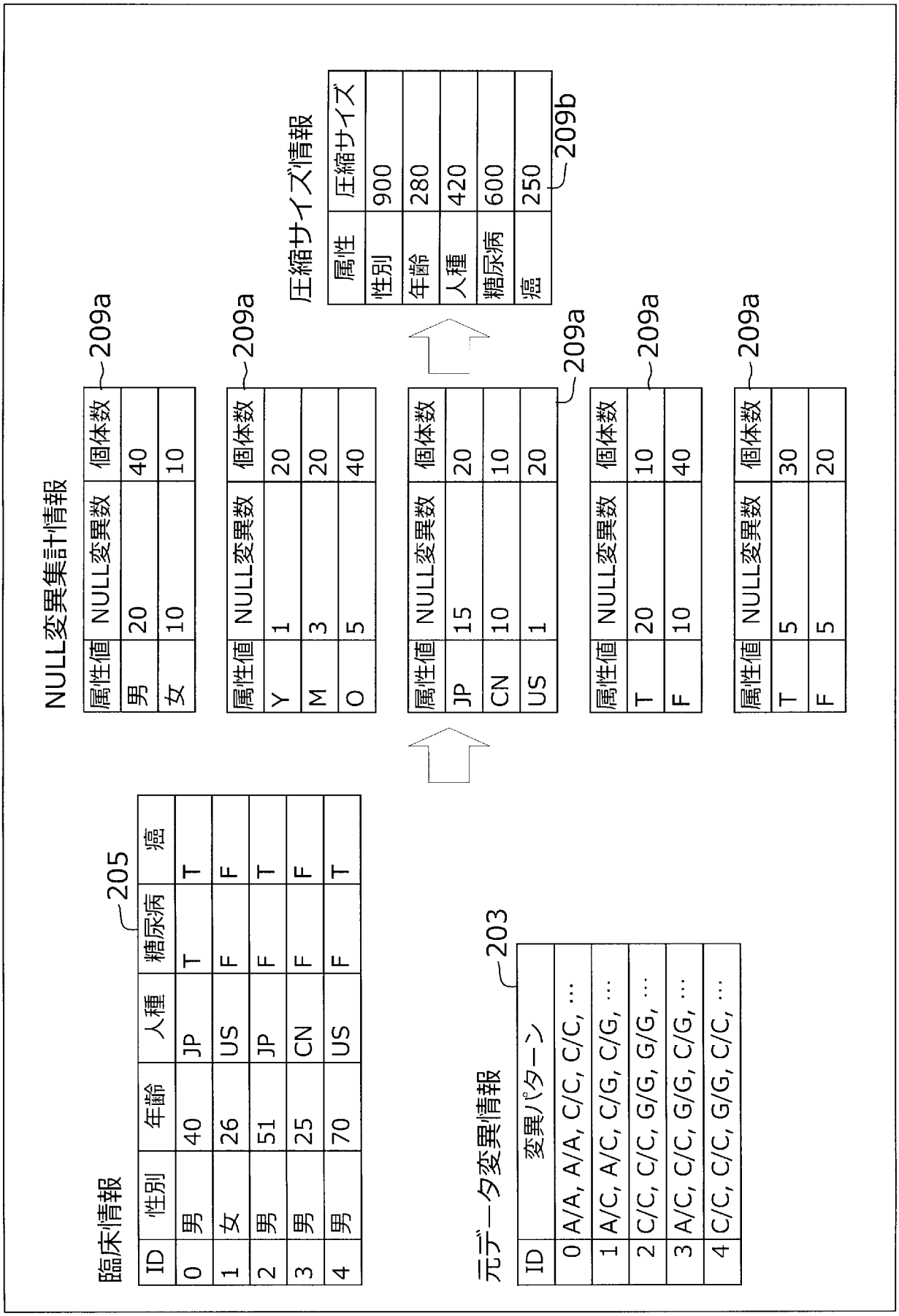
[図18]


D1
グルーピング
情報作成
<入力>
・臨床情報
・元データ変異情報
<出力>
・グルーピング情報
・非圧縮変異情報
・変異マスタ情報
D2
圧縮
<入力>
・臨床情報
・グルーピング情報
・元データ変異情報
・変異マスタ情報
<出力>
・圧縮済変異情報
D3
運用
<エンドユーザ>
・検索
・集計
・データ挿入／削除
D4
再グルーピング
／再圧縮
<入力>
・臨床情報
・グルーピング情報
・圧縮済変異情報
・変異マスタ情報
<出力>
・グルーピング情報
・圧縮済変異情報
・変異マスタ情報
D5

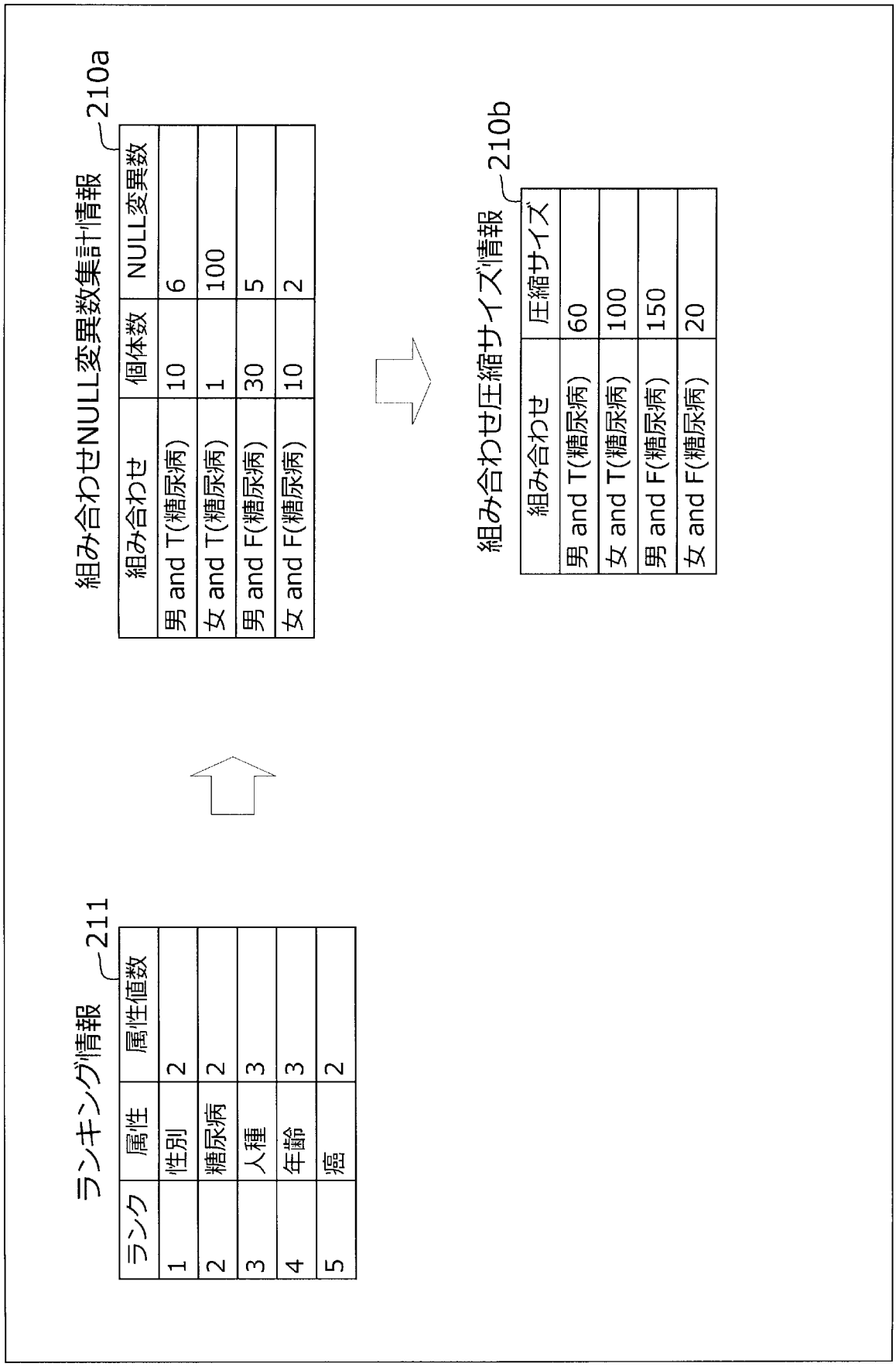
[図19]


開始
S1
グループ統計情報作成
S2
グループ分け情報作成
S3
グループ分け情報マージ
S4
圧縮
所定時間経過?
No
S5
Yes

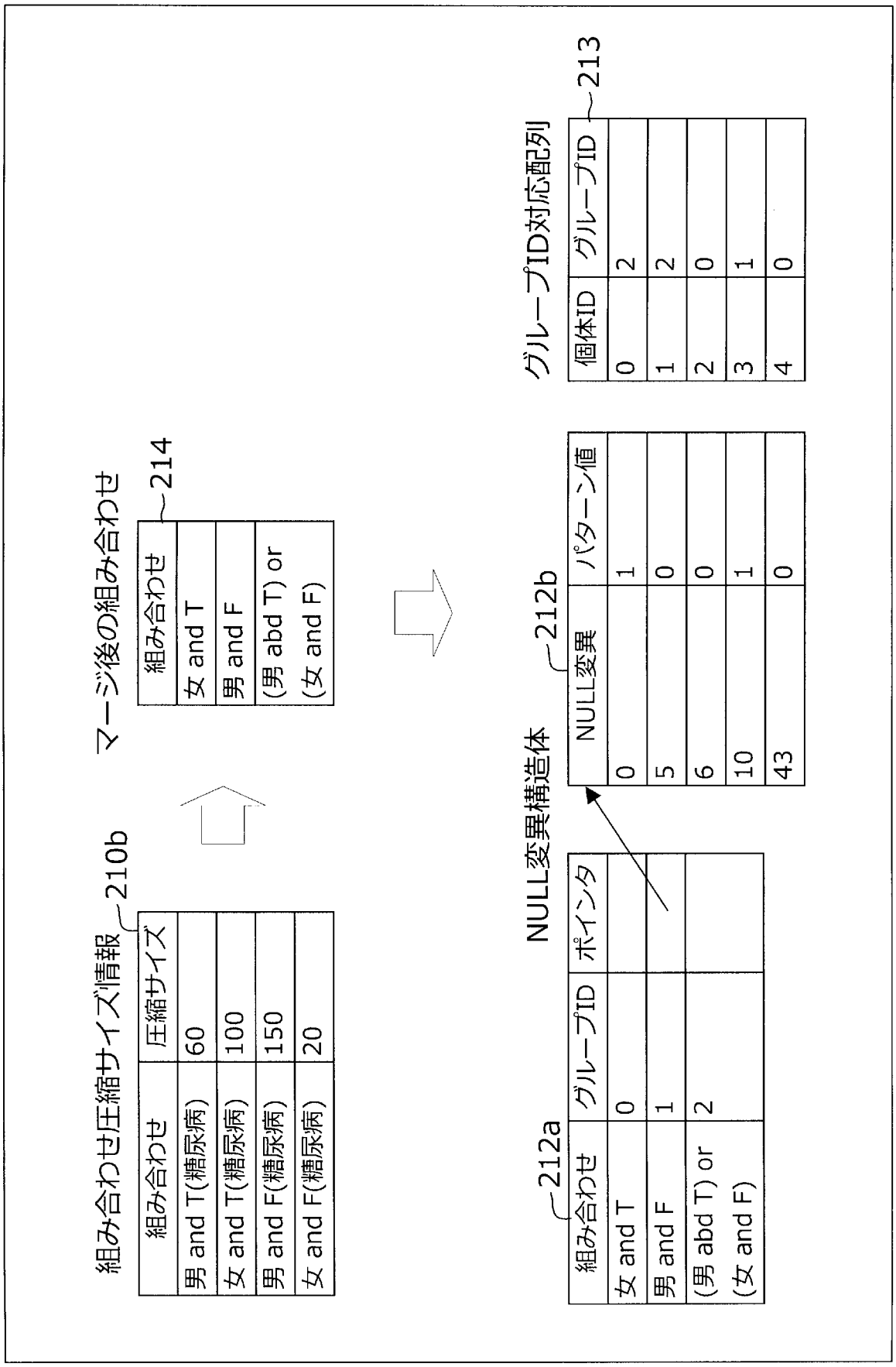
[図20]


NULL変異集計情報
209a
属性値 NULL変異数 個体数
男 20 40
女 10 10
209a
属性値 NULL変異数 個体数
Y 1 20
M 3 20
O 5 40
209a
属性値 NULL変異数 個体数
JP 15 20
CN 10 10
US 1 20
209a
属性値 NULL変異数 個体数
T 20 10
F 10 40
209a
属性値 NULL変異数 個体数
T 5 30
F 5 20
圧縮サイズ情報
属性 圧縮サイズ
性別 900
年齢 280
人種 420
糖尿病 600
癌 250
209b
臨床情報
205
ID 性別 年齢 人種 糖尿病 癌
0 男 40 JP T T
1 女 26 US F F
2 男 51 JP F T
3 男 25 CN F F
4 男 70 US F T
元データ変異情報
203
ID 変異パターン
0 A/A, A/A, C/C, C/C, …
1 A/C, A/C, C/G, C/G, …
2 C/C, C/C, G/G, G/G, …
3 A/C, C/C, G/G, C/G, …
4 C/C, C/G, G/G, C/C, …

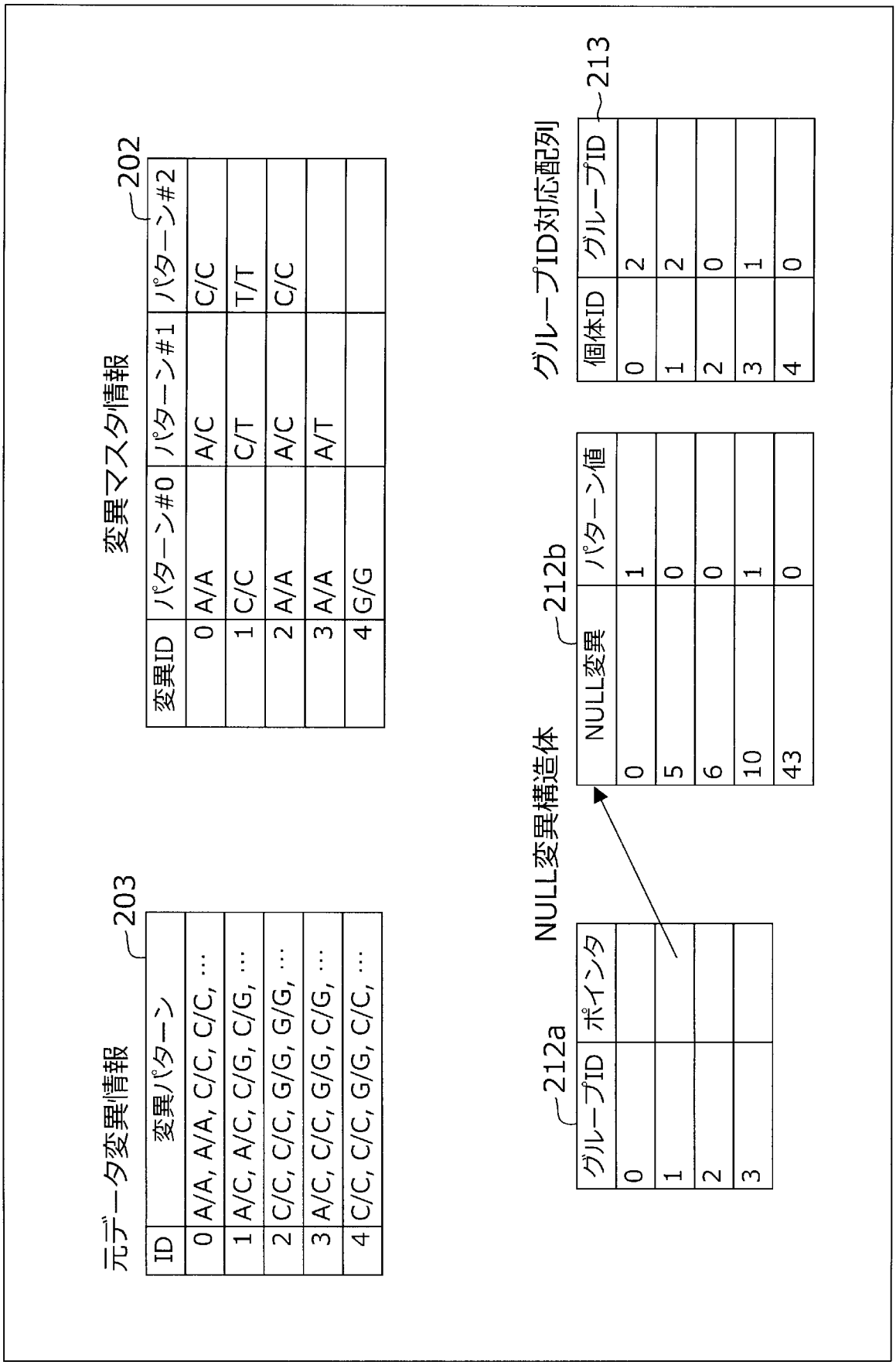
[図21]


ランキング情報 211
ランク | 属性 | 属性値数
1 | 性別 | 2
2 | 糖尿病 | 2
3 | 人種 | 3
4 | 年齢 | 3
5 | 癌 | 2
組み合わせNULL変異数集計情報 210a
組み合わせ | 個体数 | NULL変異数
男 and T(糖尿病) | 10 | 6
女 and T(糖尿病) | 1 | 100
男 and F(糖尿病) | 30 | 5
女 and F(糖尿病) | 10 | 2
組み合わせ圧縮サイズ情報 210b
組み合わせ | 圧縮サイズ
男 and T(糖尿病) | 60
女 and T(糖尿病) | 100
男 and F(糖尿病) | 150
女 and F(糖尿病) | 20

[図22]


組み合わせ圧縮サイズ情報
210b
組み合わせ
圧縮サイズ
男 and T(糖尿病)
60
女 and T(糖尿病)
100
男 and F(糖尿病)
150
女 and F(糖尿病)
20
マージ後の組み合わせ
214
組み合わせ
女 and T
男 and F
(男 abd T) or (女 and F)
NULL変異構造体
212a
組み合わせ
グループID
ポインタ
女 and T
0
男 and F
1
(男 abd T) or (女 and F)
2
212b
NULL変異
パターン値
0
1
5
0
6
0
10
1
43
0
グループID対応配列
213
個体ID
グループID
0
2
1
2
2
0
3
1
4
0

[図23]


変異マスタ情報
202
変異ID
パターン#0
パターン#1
パターン#2
0 A/A A/C C/C
1 C/C C/T T/T
2 A/A A/C C/C
3 A/A A/T
4 G/G
元データ変異情報
203
ID
変異パターン
0 A/A, A/A, C/C, C/C, …
1 A/C, A/C, C/G, C/G, …
2 C/C, C/C, G/G, G/G, …
3 A/C, C/C, G/G, C/G, …
4 C/C, C/C, G/G, C/C, …
グループID対応配列
213
個体ID
グループID
0 2
1 2
2 0
3 1
4 0
NULL変異構造体
212b
NULL変異
パターン値
0 1
5 0
6 0
10 1
43 0
212a
グループID
ポインタ
0
1
2
3

[図24]

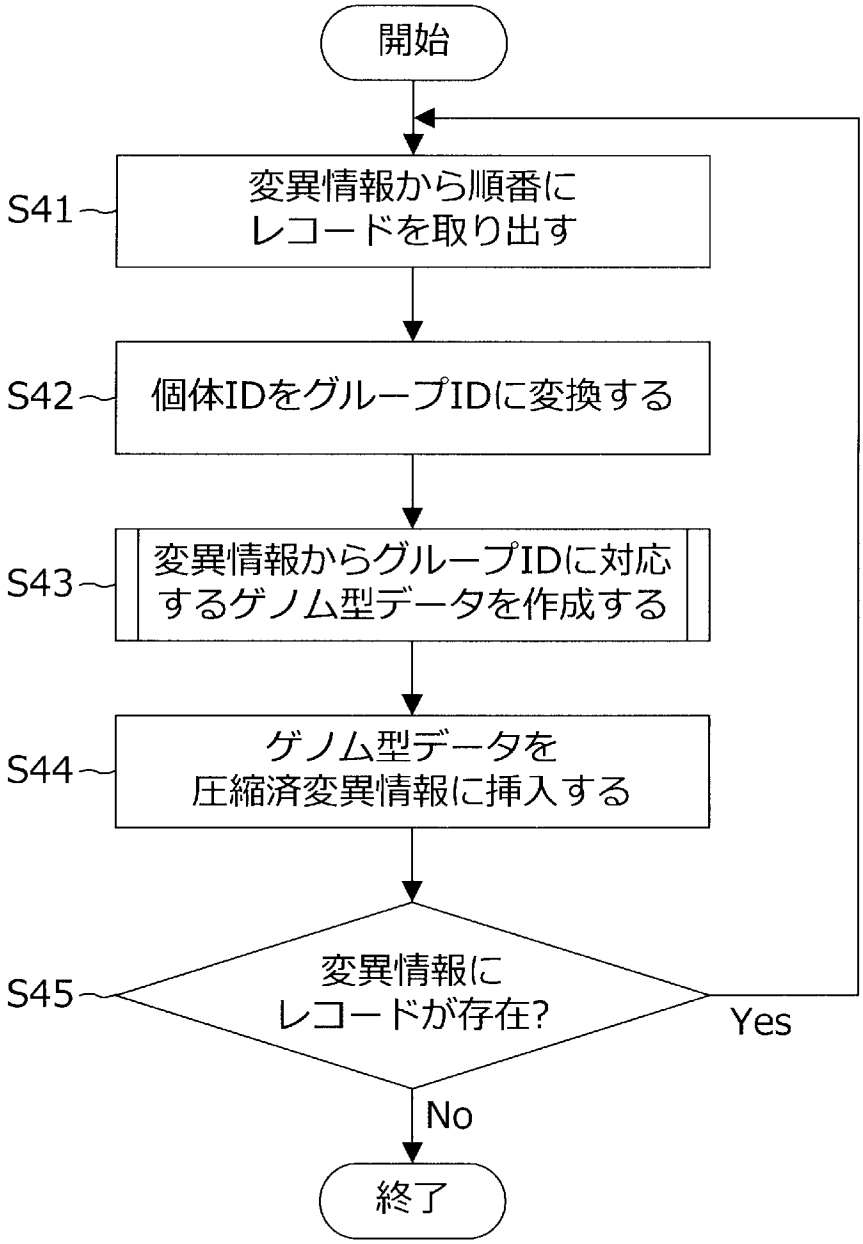
圧縮済変異情報

個体ID	変異パターン
0	group=0; 0,1,0,2,1,…
1	group=1; 1,2,1,0,0,0,…
2	group=0; 0,0,0,0,0,…
3	group=1; 1,0,1,0,0,2,…
4	group=3; 1,0,1,0,…

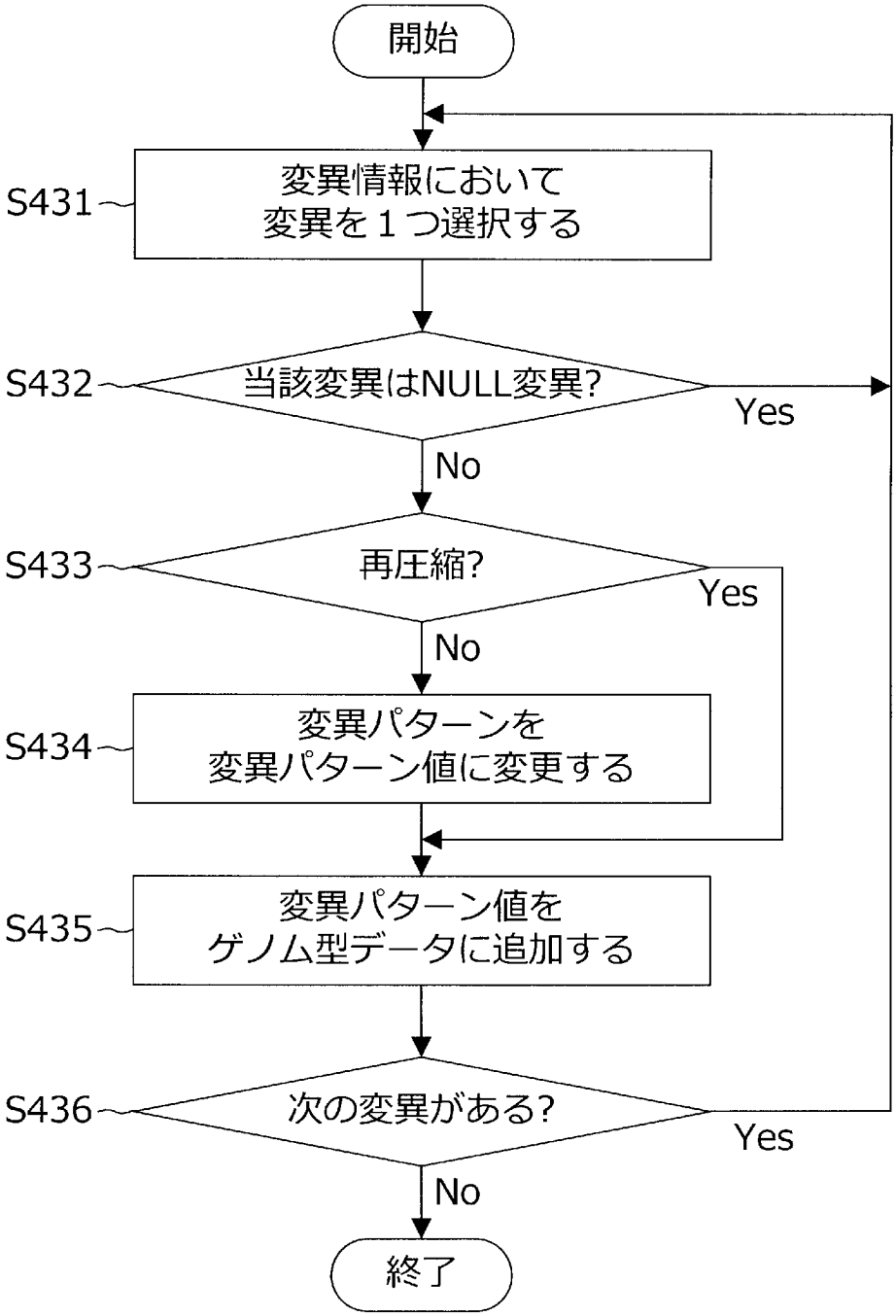
206

グループID ゲノム型データ

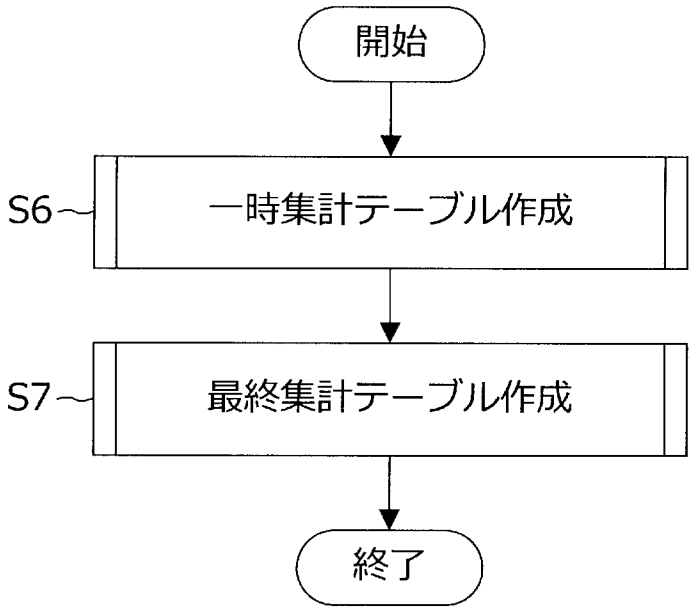
[図25]



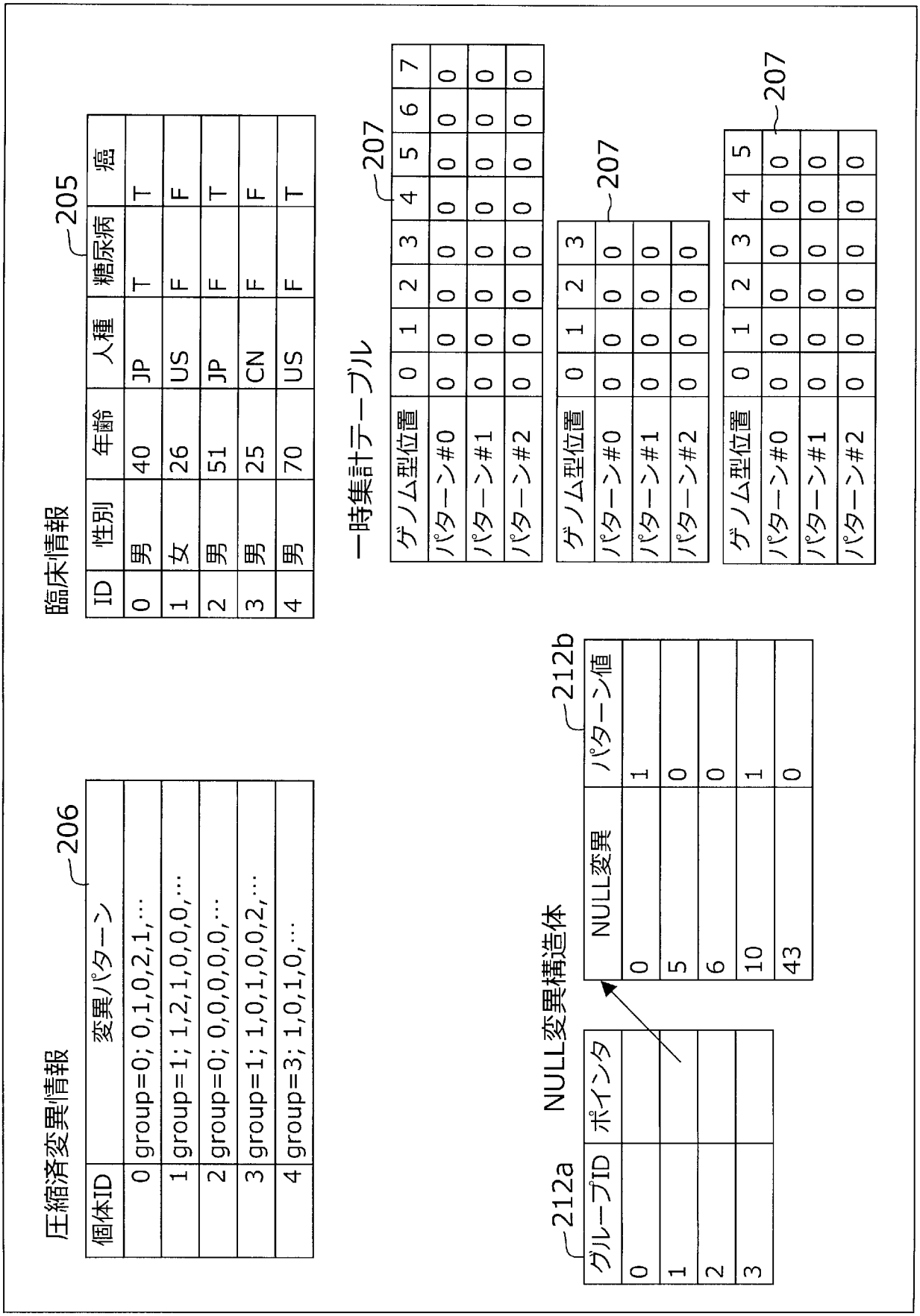
[図26]


開始
S431
変異情報において
変異を１つ選択する
S432
当該変異はNULL変異?
Yes
No
S433
再圧縮?
Yes
No
S434
変異パターンを
変異パターン値に変更する
S435
変異パターン値を
ゲノム型データに追加する
S436
次の変異がある?
Yes
No
終了

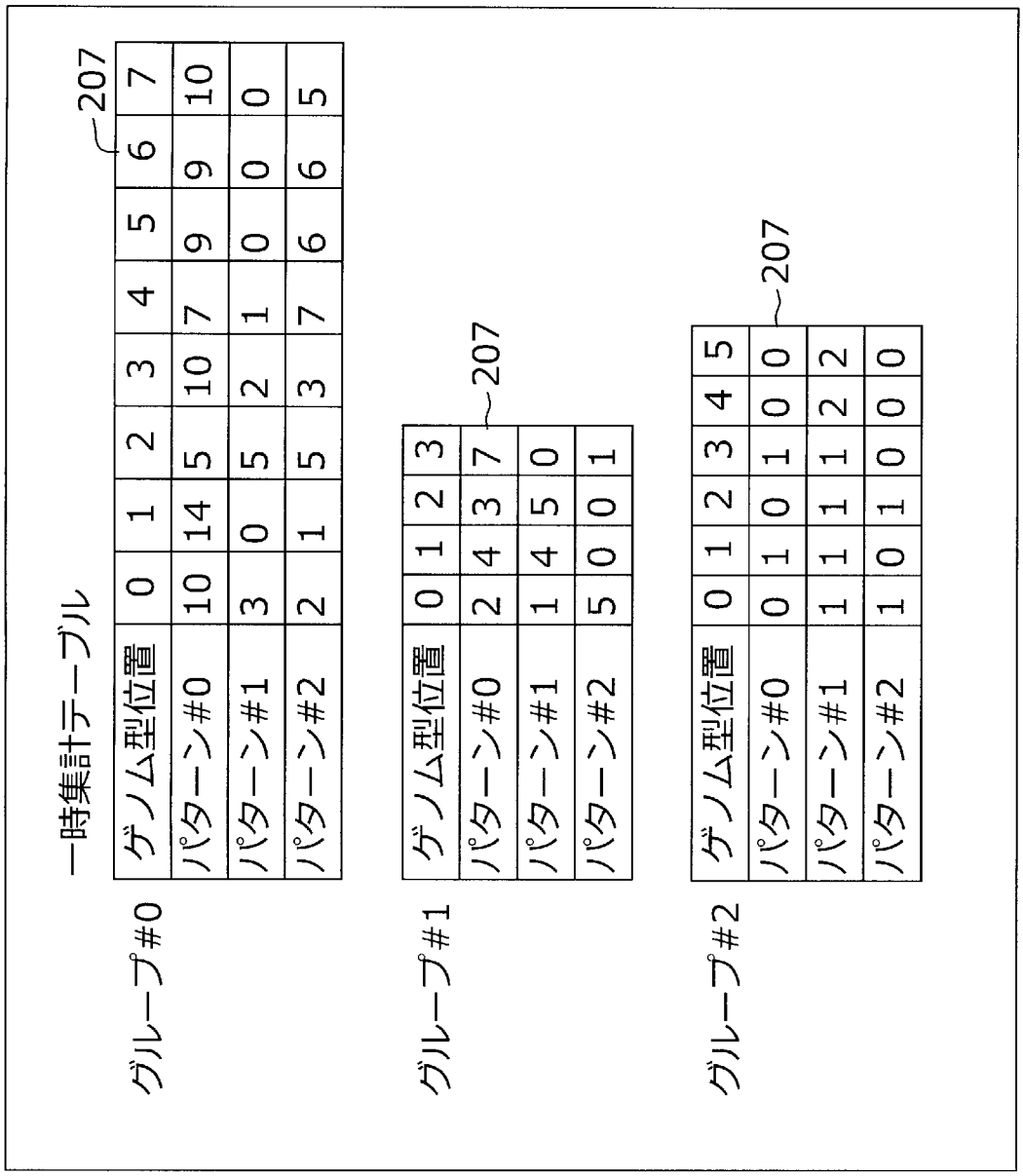
[図27]



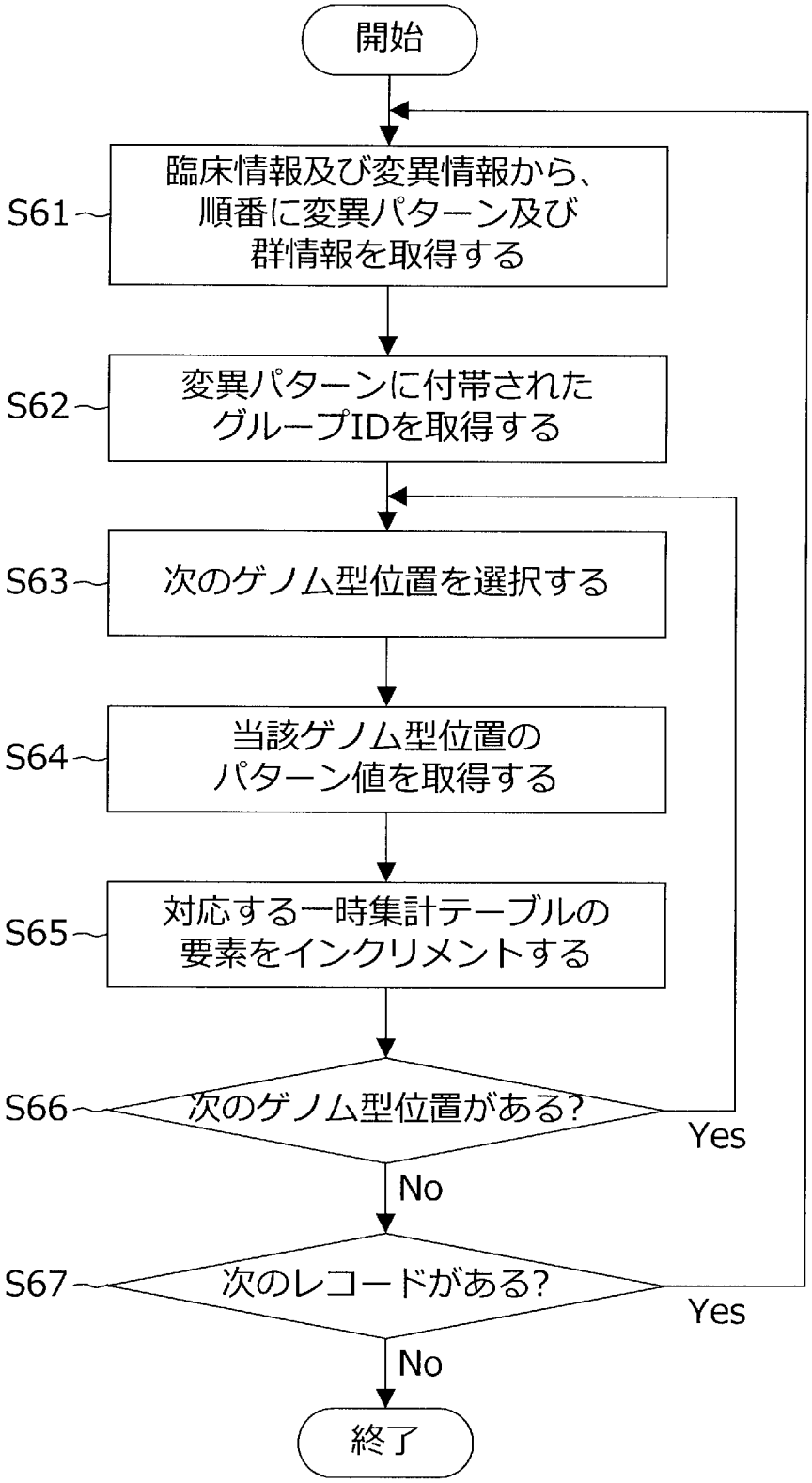
[図28]


圧縮済変異情報
206
個体ID
変異パターン
0
group=0; 0,1,0,2,1,…
1
group=1; 1,2,1,0,0,0,…
2
group=0; 0,0,0,0,0,0,…
3
group=1; 1,0,1,0,0,2,…
4
group=3; 1,0,1,0,…
臨床情報
205
ID
性別
年齢
人種
糖尿病
癌
0
男
40
JP
T
T
1
女
26
US
F
F
2
男
51
JP
F
T
3
男
25
CN
F
F
4
男
70
US
F
T
一時集計テーブル
207
ゲノム型位置
0
1
2
3
4
5
6
7
パターン#0
パターン#1
パターン#2
NULL変異構造体
212a
グループID
ポインタ
212b
NULL変異
パターン値
0
5
6
10
43
1
0
0
1
0

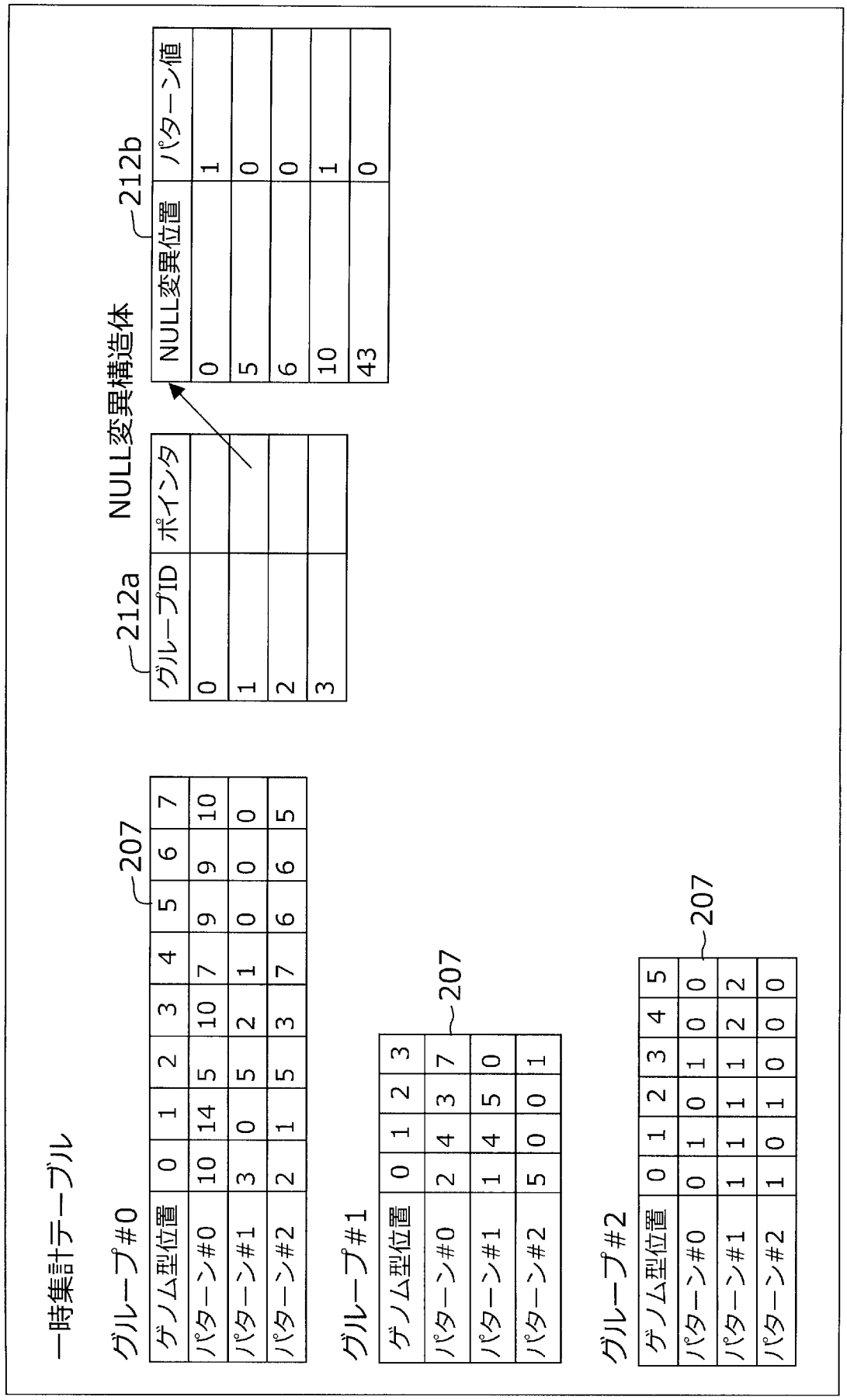
[図29]


一時集計テーブル
207
グループ#0
ゲノム型位置 0 1 2 3 4 5 6 7
パターン#0 10 14 5 10 7 9 9 10
パターン#1 3 0 5 2 1 0 0 0
パターン#2 2 1 5 3 7 6 6 5
グループ#1
207
ゲノム型位置 0 1 2 3
パターン#0 2 4 3 7
パターン#1 1 4 5 0
パターン#2 5 0 0 1
グループ#2
207
ゲノム型位置 0 1 2 3 4 5
パターン#0 0 1 0 1 0 0
パターン#1 1 1 1 1 2 2
パターン#2 1 0 1 0 0 0

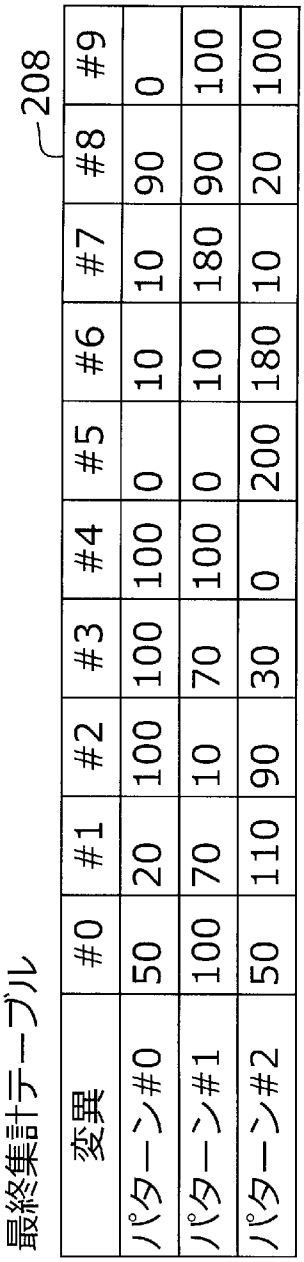
[図30]


開始
S61
臨床情報及び変異情報から、順番に変異パターン及び群情報を取得する
S62
変異パターンに付帯されたグループIDを取得する
S63
次のゲノム型位置を選択する
S64
当該ゲノム型位置のパターン値を取得する
S65
対応する一時集計テーブルの要素をインクリメントする
S66
次のゲノム型位置がある?
Yes
No
S67
次のレコードがある?
Yes
No
終了

[図31]

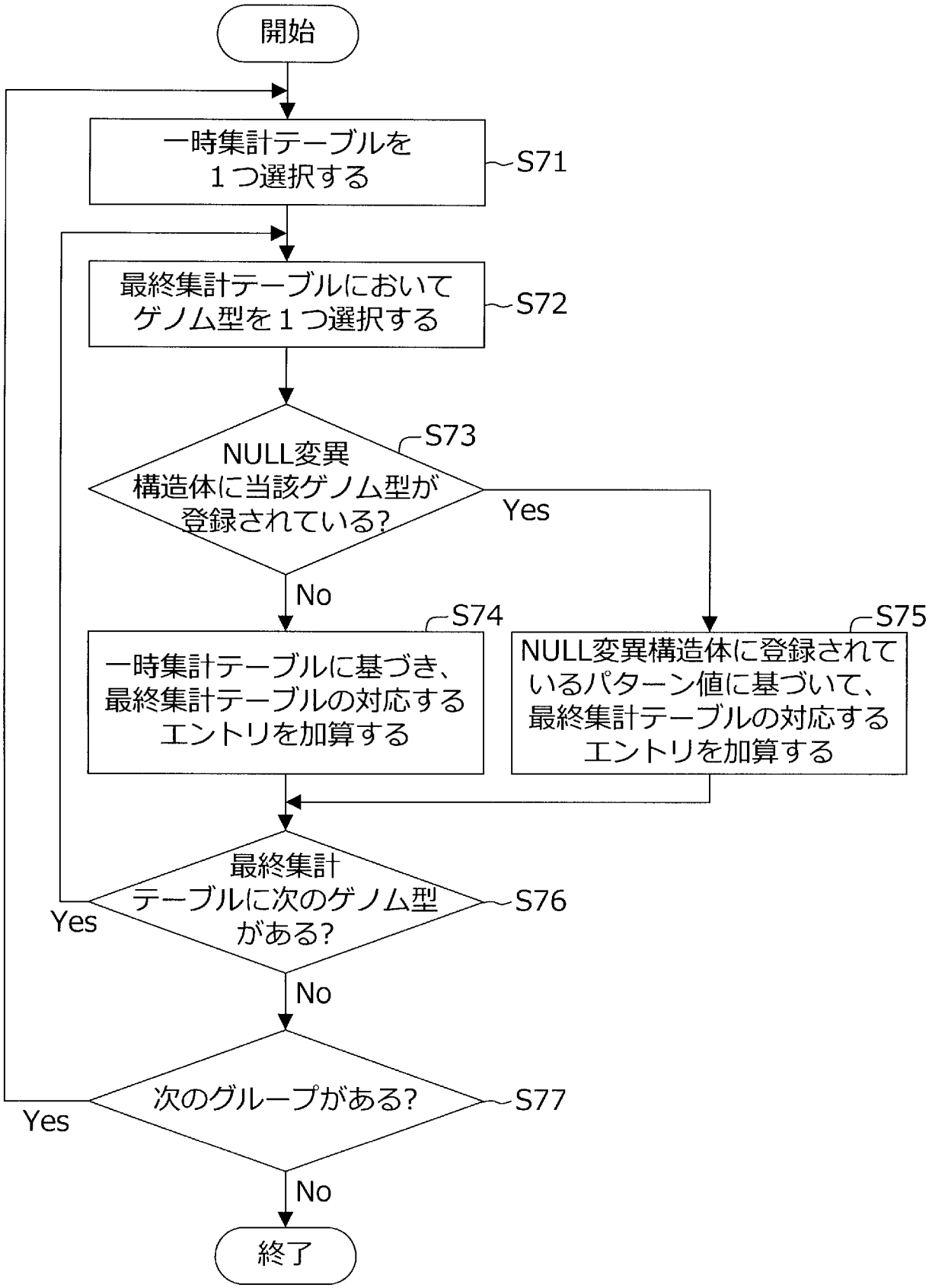

一時集計テーブル
グループ#0
207
ゲノム型位置 0 1 2 3 4 5 6 7
パターン#0 10 14 5 10 7 9 9 10
パターン#1 3 0 5 2 1 0 0 0
パターン#2 2 1 5 3 7 6 6 5
NULL変異構造体
212a
グループID ポインタ
0
1
2
3
212b
NULL変異位置 パターン値
0 1
5 0
6 0
10 1
43 0
グループ#1
207
ゲノム型位置 0 1 2 3
パターン#0 2 4 3 7
パターン#1 1 4 5 0
パターン#2 5 0 0 1
グループ#2
207
ゲノム型位置 0 1 2 3 4 5
パターン#0 0 1 0 1 0 0
パターン#1 1 1 1 1 2 2
パターン#2 1 0 1 0 0 0

[図32]



変異	#0	#1	#2	#3	#4	#5	#6	#7	#8	#9
パターン#0	50	20	100	100	100	0	10	10	90	0
パターン#1	100	70	10	70	100	0	10	180	90	100
パターン#2	50	110	90	30	0	200	180	10	20	100

[図33]


開始
一時集計テーブルを
１つ選択する
S71
最終集計テーブルにおいて
ゲノム型を１つ選択する
S72
NULL変異
構造体に当該ゲノム型が
登録されている?
S73
Yes
No
一時集計テーブルに基づき、
最終集計テーブルの対応する
エントリを加算する
S74
NULL変異構造体に登録されて
いるパターン値に基づいて、
最終集計テーブルの対応する
エントリを加算する
S75
最終集計
テーブルに次のゲノム型
がある?
S76
Yes
No
次のグループがある?
S77
Yes
No
終了

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/000539

A. CLASSIFICATION OF SUBJECT MATTER
Int. Cl. G06F19/10(2011.01)i, G06F17/30(2006.01)i, G06F19/24(2011.01)i, C12N15/09(2006.01)n

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int. Cl. G06F19/10, G06F17/30, G06F19/24, C12N15/09

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2018
Registered utility model specifications of Japan 1996-2018
Published registered utility model applications of Japan 1994-2018

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2007-193708 A (FUJITSU LTD.) 02 August 2007, abstract, claim 1, paragraphs [0035], [0053], fig. 4 (Family: none)	1, 12
Y		2-11, 13-20
Y	WO 2015/050174 A1 (TOHOKU UNIVERSITY) 09 April 2015, claim 34 & US 2015/0193588 A1, claim 34 & EP 3054412 A1 & CN 104704526 A	2-11, 13-20

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20.02.2018	06.03.2018
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１８／０００５３９

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G06F19/10(2011.01)i, G06F17/30(2006.01)i, G06F19/24(2011.01)i, C12N15/09(2006.01)n

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　G06F19/10, G06F17/30, G06F19/24, C12N15/09

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１８年
日本国実用新案登録公報	１９９６－２０１８年
日本国登録実用新案公報	１９９４－２０１８年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

Ｃ．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	JP 2007-193708 A（富士通株式会社）2007.08.02, [要約]、[請求項１]、段落[００３５]、[００５３]、図４	１，１２
Y	(ファミリーなし)	２～１１， １３～２０
Y	WO 2015/050174 A1（国立大学法人東北大学）2015.04.09, [請求項３４] & US 2015/0193588 A1,[請求項３４] & EP 3054412 A1 & CN 104704526 A	２～１１， １３～２０

☐　Ｃ欄の続きにも文献が列挙されている。

☐　パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
２０．０２．２０１８	０６．０３．２０１８

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
東京都千代田区霞が関三丁目４番３号

特許庁審査官（権限のある職員）　５Ｌ　３７８６
山本　雅士
電話番号　０３－３５８１－１１０１　内線　３５６２