(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7141029号
(P7141029)

(45)発行日　令和4年9月22日(2022.9.22)

(24)登録日　令和4年9月13日(2022.9.13)

(51)国際特許分類
Ｇ１６Ｂ　25/00　(2019.01)

ＦＩ
Ｇ１６Ｂ　25/00

請求項の数　43　（全32頁）

(21)出願番号　特願2017-136368(P2017-136368)
(22)出願日　平成29年7月12日(2017.7.12)
(65)公開番号　特開2019-20838(P2019-20838A)
(43)公開日　平成31年2月7日(2019.2.7)
審査請求日　令和2年6月22日(2020.6.22)
特許法第３０条第２項適用　１．発行日　２０１７年２月１日　刊行物　第１４回日本乳癌学会九州地方会　プログラム・抄録集　２．開催日　２０１７年３月４日～２０１７年３月５日　集会名、開催場所　第１４回日本乳癌学会九州地方会　九州大学医学部　百年講堂（住所：福岡県福岡市東区馬出３－１－１）　３．発行日　２０１７年３月５日　刊行物　がんと生殖に関するシンポジウム２０１７抄録集　４．開催日　２０１７年３月５日　集会名、開催場所　がんと生殖に関するシンポジウム２０１７　サンケイプラザ　４Ｆホール（住所：東京都千代田区大手町１－７
最終頁に続く

(73)特許権者　390014960
シスメックス株式会社
兵庫県神戸市中央区脇浜海岸通１丁目５番１号
(73)特許権者　504176911
国立大学法人大阪大学
大阪府吹田市山田丘１番１号
(74)代理人　110000796弁理士法人三枝国際特許事務所
(72)発明者　岸　和希
兵庫県神戸市中央区脇浜海岸通１丁目５番１号　シスメックス株式会社内
(72)発明者　澤　賢一
兵庫県神戸市中央区脇浜海岸通１丁目５番１号　シスメックス株式会社内
最終頁に続く

(54)【発明の名称】　データベースを構築する方法

(57)【特許請求の範囲】

【請求項１】

コンピュータが実施する、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築する方法であって、

前記データベースが、新規マーカーの候補を探索するために使用されるものである、下記工程を含む、方法：

解析対象遺伝子を特定する情報を取得する工程、

解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得する工程、

前記非解析対象遺伝子の遺伝子関連情報をデータベースに出力する工程、及び

非解析対象遺伝子の遺伝子関連情報と、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報とを前記データベースに記憶する工程。

【請求項２】

解析対象遺伝子は、疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測、薬効予測、及び疾患のモニタリングよりなる群から選択される少なくとも一つの解析に使用される、請求項１に記載の方法。

【請求項３】

前記マーカーが、疾患のバイオマーカー又は疾患の治療の標的分子である、請求項１又は２に記載の方法。

【請求項４】
前記疾患のバイオマーカーが、疾患のリスク判定、スクリーニング、鑑別診断、予後予測、及び再発予測よりなる群から選択される少なくとも一種に使用される、請求項３に記載の方法。
【請求項５】
前記生体試料が、血液試料、体液及び組織よりなる群から選択される少なくとも一種である、請求項１から４のいずれか一項に記載の方法。
【請求項６】
前記組織が、新鮮組織、凍結組織、固定組織、及び包埋剤に包埋された組織よりなる群から選択される少なくとも一種である、請求項５に記載の方法。
【請求項７】
前記生体試料が、所定の疾患、所定の疾患型及び所定の疾患の病期よりなる群から選択される少なくとも一種の病巣から採取されたものである、請求項１から６のいずれか一項に記載の方法。
【請求項８】
前記生体試料が、複数であり、前記複数の生体試料は異なる患者の同一疾患の病巣から採取されたものである、請求項５から７のいずれか一項に記載の方法。
【請求項９】
前記遺伝子関連測定データが、ＲＮＡの発現量、ＤＮＡのメチル化量、ＤＮＡの塩基配列情報、ＲＮＡの塩基配列情報、タンパク質の存在量、及びタンパク質の糖鎖修飾情報からなる群から選択される少なくとも一種を含む、請求項１から８のいずれか一項に記載の方法。
【請求項１０】
前記ＤＮＡのメチル化量が、さらにＤＮＡのメチル化部位の位置情報を含み、
前記ＤＮＡの塩基配列情報が、さらにＤＮＡの塩基配列の欠失、置換、融合、コピー数変異、又は挿入の有無と、その位置情報を含み、
前記タンパク質の糖鎖修飾情報が、さらにタンパク質の修飾位置と、糖鎖の種類の情報を含む、
請求項９に記載の方法。
【請求項１１】
前記遺伝子関連測定データが、所定の測定方法により取得される、請求項１から１０のいずれか一項に記載の方法。
【請求項１２】
前記遺伝子関連測定データが、ＲＮＡの発現量、ＤＮＡのメチル化量、ＤＮＡの塩基配列情報、又はＲＮＡの塩基配列情報である場合には、前記所定の測定方法が、塩基配列シーケンス及び／又はマイクロアレイによる測定方法であり、
前記遺伝子関連測定データが、タンパク質の存在量である場合には、前記所定の測定方法が、マイクロアレイ及び／又はＥＬＩＳＡであり、
前記遺伝子関連測定データが、タンパク質の糖鎖修飾である場合には、前記所定の測定方法が、マイクロアレイ及び／又はＥＬＩＳＡである
請求項１１に記載の方法。
【請求項１３】
前記遺伝子関連測定データが、単一の検査機関で取得される、請求項１から１２のいずれか一項に記載の方法。
【請求項１４】
前記遺伝子関連情報が、遺伝子関連測定データの測定日、測定方法、測定試料の量、検査機関、生体試料の保存方法及び生体試料の保存期間よりなる群から選択される少なくとも一種と、各遺伝子関連測定データの遺伝子名、ＧｅｎＢａｎｋのアクセッション番号及び／又は遺伝子を特定するための符号と、前記遺伝子関連測定データを取得した生体試料を特定するための符号とを含む、請求項１から１３に記載の方法。

【請求項１５】
　前記生体試料関連情報が、生体試料を採取した患者の診療関連情報、及び治療関連情報よりなる群から選択される少なくとも一種と、生体試料を特定するための符号と、生体試料を採取した患者を特定するための符号と、生体試料の種類とを含む、請求項１から１４に記載の方法。
【請求項１６】
　前記診療関連情報が、疾患名、疾患型名、疾患の病期、患者の性別、患者の年齢、患者の既往歴、患者の家族歴、再発履歴、転移履歴、問診情報、月経履歴及び遺伝子関連情報以外の検査情報よりなる群から選択される少なくとも一種を含む、請求項１５に記載の方法。
【請求項１７】
　前記治療関連情報が、治療履歴の情報を含む、請求項１５又は１６に記載の方法。
【請求項１８】
　前記マーカーが、疾患のバイオマーカー又は疾患の治療の標的分子であり、前記治療が、治療薬の投与、予防薬の投与、放射線治療及び外科的処置よりなる群から選択される少なくとも一種である、請求項１７に記載の方法。
【請求項１９】
　前記治療が、治療薬の投与又は予防薬の投与である場合には、前記治療履歴に投与した薬剤の名称、用量、投与頻度、投与日及び投与期間を含み、
　前記治療が放射線治療である場合には、前記治療履歴に１回あたりの放射線照射量、頻度、施術日及び総照射放射線量を含み、
　前記治療が外科的処置である場合には、前記治療履歴に主な切除部位、術式、リンパ節等の切除部位周辺組織の郭清の有無、及び施術日を含む、
請求項１８に記載の方法。
【請求項２０】
　前記解析対象遺伝子について前記遺伝子関連測定データを取得する工程を含む、請求項１から１９のいずれか一項に記載の方法。
【請求項２１】
　前記解析対象遺伝子の遺伝子関連情報をデータベースに出力する工程、及び
　前記解析対象遺伝子の遺伝子関連情報をデータベースに記憶する工程を含む、請求項２０に記載の方法。
【請求項２２】
　前記解析対象遺伝子の遺伝子関連情報の報告書を作成する工程を含む、請求項１から２１のいずれか一項に記載の方法。
【請求項２３】
　前記報告書が、
　疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測、薬効予測、及び疾患のモニタリングよりなる群より選択される少なくとも一つの判定結果と、
　各遺伝子名及び／又は各遺伝子を特定するための符号と、
　各遺伝子についての前記遺伝子関連測定データと、
　前記遺伝子関連測定データを取得した生体試料を特定するための符号と、
　遺伝子関連測定データの測定日、測定方法、検査機関、生体試料の保存方法及び生体試料の保存期間よりなる群から選択される少なくとも一種と、
を含む、請求項２２に記載の方法。
【請求項２４】
　前記生体試料が、乳癌組織から採取されたものである、請求項１から２３のいずれか一項に記載の方法。
【請求項２５】
　前記解析対象遺伝子が、Ｃｕｒｅｂｅｓｔ（登録商標）　９５ＧＣ　Ｂｒｅａｓｔ解析対象遺伝子、Ｏｎｃｏｔｙｐｅ（登録商標）　ＤＸ解析対象遺伝子、ＭａｍｍａＰｒｉｎｔ

解析対象遺伝子、ＢｌｕｅＰｒｉｎｔ解析対象遺伝子、ＰＡＭ５０解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ６解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ６ ＋ ＣＯＳＭＩＣ解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ６ ＋ ＵＴＲ解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５ ＋ ＵＴＲｓ対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５ ＋ ＩｎｃＲＮＡ対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５ ＋ Ｒｅｇｕｌａｔｏｒｙ対象遺伝子、ＴｒｕＳｉｇｈｔ Ｃａｎｃｅｒ対象遺伝子、ＴｒｕＳｉｇｈｔ Ｔｕｍｏｒ １５対象遺伝子、及びＴｒｕＳｉｇｈｔ Ｔｕｍｏｒ １７０対象遺伝子よりなる群から選択される少なくとも一種である、請求項１～２４に記載の方法。

【請求項２６】

非解析対象遺伝子が複数である、請求項１から２５に記載の方法。

【請求項２７】

前記解析対象遺伝子は、疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測、薬効予測、及び疾患のモニタリングよりなる群から選択される少なくとも一つの解析に使用され、前記疾患のバイオマーカーが、前記生体試料が採取された患者が患う疾患とは異なる疾患のバイオマーカーである、請求項２から２６に記載の方法。

【請求項２８】

前記解析対象遺伝子は、疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測、薬効予測、及び疾患のモニタリングよりなる群から選択される少なくとも一つの解析に使用され、前記疾患のバイオマーカーが、前記生体試料が採取された患者が患う疾患と同じ疾患のバイオマーカーである、請求項２から２６に記載の方法。

【請求項２９】

コンピュータが実施する、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報に基づき、新規マーカーの候補を探索する方法であって、下記工程を含む方法：

解析対象遺伝子を特定する情報を取得する工程、

解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得する工程、

非解析対象遺伝子の遺伝子関連情報をデータベースに出力する工程、

非解析対象遺伝子の遺伝子関連情報と、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報とを前記データベースに記憶する工程、

前記遺伝子関連情報と、前記生体試料関連情報とを対応させる工程、

前記遺伝子関連情報に含まれる遺伝子関連測定データと、前記生体試料関連情報との関連性の強さを示す数値を遺伝子ごとに取得する工程、及び

前記数値に基づいて、前記生体試料関連情報と関連の強い遺伝子を新規マーカーの候補を決定する工程。

【請求項３０】

前記非解析対象遺伝子の前記遺伝子関連測定データを取得した後、かつ前記数値を取得する前に、各遺伝子関連測定データの値を正規化又は標準化する工程を含む、請求項２９に記載の方法。

【請求項３１】

前記遺伝子関連情報に含まれる遺伝子関連測定データと、前記生体試料関連情報との関連性の強さを示す数値が、ＲＮＡ量（コピー数）、タンパク質量、ＤＮＡメチル化量又はメチル化の割合、ＲＮＡの塩基配列の変化率、ＤＮＡの塩基配列の変化率、タンパク質の糖鎖修飾の割合、有意確率、尤度又はＺスコアである、請求項２９又は３０に記載の方法。

【請求項３２】

前記生体試料関連情報と関連の強い遺伝子を新規マーカーの候補を決定する工程が、請求項２９で取得された数値その絶対数に応じて並べ替える工程と、前記絶対数が高い遺伝

子について前記生体試料関連情報との関連性が強いと決定する工程を含む、請求項３０に記載の方法。

【請求項３３】

前記新規マーカーの候補が複数である、請求項２９又は３０に記載の方法。

【請求項３４】

生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステムであって、

前記データベースが、新規マーカーの候補を探索するために使用されるものであり、

前記システムは、検査機関情報処理装置と、検査機関データベース記憶装置とを備え、

前記検査機関情報処理装置は、

解析対象遺伝子を特定する情報を取得し、

解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、

前記非解析対象遺伝子の遺伝子関連情報を前記検査機関データベース記憶装置に出力し、

前記検査機関データベース記憶装置は、

非解析対象遺伝子の遺伝子関連情報と、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報とを受け付け、記憶する、

システム。

【請求項３５】

前記生体試料関連情報は、生体試料を採取した医療機関において生成される、請求項３４に記載のシステム。

【請求項３６】

生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステムであって、

前記データベースが、新規マーカーの候補を探索するために使用されるものであり、

前記システムは、医療機関情報処理装置と、検査機関情報処理装置と、医療機関データベース記憶装置とを備え、

前記検査機関情報処理装置は、

解析対象遺伝子を特定する情報を取得し、

解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、

前記非解析対象遺伝子の遺伝子関連情報を前記医療機関データベース記憶装置に出力し、

前記医療機関情報処理装置は、

前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を前記医療機関データベース記憶装置に出力し、

前記医療機関データベース記憶装置は、

前記非解析対象遺伝子の遺伝子関連情報と、前記生体試料関連情報とを受け付け、記憶する、

システム。

【請求項３７】

前記検査機関情報処理装置は、

解析対象遺伝子について前記遺伝子関連測定データを取得し、

前記解析対象遺伝子の遺伝子関連情報を前記医療機関データベース記憶装置に出力し、

前記医療機関データベース記憶装置は、

前記解析対象遺伝子の遺伝子関連情報を受け付け、記憶する、

請求項３６に記載のシステム。

【請求項３８】

生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステムであって、

前記データベースが、新規マーカーの候補を探索するために使用されるものであり、

前記システムは、医療機関情報処理装置と、検査機関情報処理装置と、データベース記憶装置とを備え、
前記検査機関情報処理装置は、
解析対象遺伝子を特定する情報を取得し、
解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、
前記非解析対象遺伝子の遺伝子関連情報を前記データベース記憶装置に出力し、
前記医療機関情報処理装置は、
前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を前記データベース記憶装置に出力し、
前記データベース記憶装置は、
前記非解析対象遺伝子の遺伝子関連情報と、前記生体試料関連情報とを受け付け、記憶する、
システム。
【請求項３９】
前記検査機関情報処理装置は、
解析対象遺伝子について前記遺伝子関連測定データを取得し、
前記解析対象遺伝子の遺伝子関連情報を出力する、
請求項３８に記載のシステム。
【請求項４０】
コンピュータが実施する、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築する方法であって、
前記データベースに記憶されたデータが、新規マーカーを探索するための人工知能の訓練データ又は検証データとして使用される、下記工程を含む、方法：
解析対象遺伝子を特定する情報を取得する工程、
前記解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得する工程、
前記非解析対象遺伝子の遺伝子関連情報をデータベースに記憶する工程、及び
前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を前記データベースに記憶する工程。
【請求項４１】
解析対象遺伝子について前記遺伝子関連測定データを取得する工程、及び
前記解析対象遺伝子の遺伝子関連情報をデータベースに記憶する工程をさらに含む、請求項４０に記載の方法。
【請求項４２】
コンピュータが実施する、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築する方法であって、
前記データベースが、新規マーカーの候補を探索するために使用されるものである、下記工程を含む、方法：
解析対象遺伝子以外の非解析対象遺伝子を含む複数の遺伝子について取得された前記遺伝子関連情報を、検査機関情報処理装置及び／又は医療機関情報処理装置から取得する工程、
前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を、検査機関情報処理装置及び／又は医療機関情報処理装置から取得する工程、及び
前記遺伝子関連情報と、前記生体試料関連情報とを前記データベースに記憶する工程。
【請求項４３】
生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステムであって、
前記データベースが、新規マーカーの候補を探索するために使用されるものであり、

前記システムは、データベース記憶装置を備え、
前記データベース記憶装置は、
解析対象遺伝子以外の非解析対象遺伝子を含む複数の遺伝子について取得された前記遺伝子関連情報を、検査機関情報処理装置及び／又は医療機関情報処理装置から取得し、
前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を、検査機関情報処理装置及び／又は医療機関情報処理装置から取得し、
前記遺伝子関連情報と、前記生体試料関連情報とを記憶する、システム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、データベースを構築する方法、及びデータベースを構築するシステムに関する。

【背景技術】

【０００２】

近年、乳癌を中心に、遺伝子発現レベル等の患者の分子レベルに基づいて治療方針を決定することが試みられている。例えば、特許文献１には、９５個の遺伝子発現に基づいて、リンパ節転移陰性かつエストロゲン受容体陽性の乳癌の予後を予測する方法が記載されている。

【０００３】

このような予後予測が可能となった背景には、全遺伝子にわたって遺伝子の発現を網羅的に解析するための、次世代シーケンシングやマイクロアレイ等による検出技術及び解析技術が急速に発展したことがある。

【先行技術文献】

【特許文献】

【０００４】

【文献】特開第２０１１－２２３９５７号公報

【発明の概要】

【発明が解決しようとする課題】

【０００５】

次世代シーケンシング解析やマイクロアレイ解析により、現代では、膨大な数の遺伝子の発現量やＤＮＡの塩基配列変異を解析することが可能である。また、ＮＣＢＩ Ｇｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ等、パブリックドメインで使用できるデータベースも構築されている。しかし、その一方で、各データベースに蓄積されているデータは、必ずしも一定の条件でサンプルが採取され解析が行われたものではなく解析誤差等を含むため、純粋にサンプルの遺伝子発現等の状態を反映しているデータベースであるということは難しい。また、サンプルを採取した個体の状態も臨床的な背景も均質ではない。

【０００６】

さらに、疾患の予後予測や、薬剤の治療効果の予測に使用される遺伝子の数は限られているのに対して、次世代シーケンシング解析やマイクロアレイ解析では、測定の必要のない遺伝子やタンパク質までも大量に解析されるという問題も含んでいる。

【０００７】

本発明は、次世代シーケンシング解析やマイクロアレイ解析におけるこのような問題に鑑み、次世代シーケンシング解析やマイクロアレイ解析で取得される、測定対象の遺伝子及び測定対象以外の遺伝子の発現、又は前記遺伝子産物の機能を反映するデータを有効活用することを課題とする。

【課題を解決するための手段】

【０００８】

本発明の課題を解決するための第１の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築する方法であって、前記データベースが、新規マーカーの候補を探索するために

使用されるものである、下記工程を含む、方法：解析対象遺伝子を特定する情報を取得する工程、解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得する工程、前記非解析対象遺伝子の遺伝子関連情報をデータベースに出力する工程、及び非解析対象遺伝子の遺伝子関連情報と、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報とを前記データベースに記憶する工程である。

【０００９】

本発明の課題を解決するための第２の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報に基づき、新規マーカーの候補を探索する方法であって、下記工程を含む方法：解析対象遺伝子を特定する情報を取得する工程、解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得する工程、非解析対象遺伝子の遺伝子関連情報をデータベースに出力する工程、非解析対象遺伝子の遺伝子関連情報と、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報とを前記データベースに記憶する工程、前記遺伝子関連情報と、前記生体試料関連情報とを対応させる工程、前記遺伝子関連情報に含まれる遺伝子関連測定データと、前記生体試料関連情報との関連性の強さを示す数値を遺伝子ごとに取得する工程、及び前記数値に基づいて、前記生体試料関連情報と関連の強い遺伝子を新規マーカーの候補を決定する工程である。

【００１０】

本発明の課題を解決するための第３－１の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステム５００であって、前記データベースが、新規マーカーの候補を探索するために使用されるものであり、前記システムは、検査機関情報処理装置２０と、検査機関データベース記憶装置１００とを備え、前記検査機関情報処理装置２０は、解析対象遺伝子を特定する情報を取得し、解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、前記非解析対象遺伝子の遺伝子関連情報を前記検査機関データベース記憶装置に出力し、前記検査機関データベース記憶装置１００は、非解析対象遺伝子の遺伝子関連情報と、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報とを受け付け、記憶する、システムである。

【００１１】

本発明の課題を解決するための第３－２の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステム６００であって、前記データベースが、新規マーカーの候補を探索するために使用されるものであり、前記システムは、医療機関情報処理装置５０と、検査機関情報処理装置２０と、医療機関データベース記憶装置１０１とを備え、前記検査機関情報処理装置２０は、解析対象遺伝子を特定する情報を取得し、解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、前記非解析対象遺伝子の遺伝子関連情報を前記医療機関データベース記憶装置１０１に出力し、前記医療機関情報処理装置５０は、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を前記医療機関データベース記憶装置１０１に出力し、前記医療機関データベース記憶装置は、前記非解析対象遺伝子の遺伝子関連情報と、前記生体試料関連情報とを受け付け、記憶する、システムである。

【００１２】

本発明の課題を解決するための第３－３の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステム７００であって、前記データベースが、新規マーカーの候補を探索するために使用されるものであり、前記システムは、医療機関情報処理装置５０と、検査機関情報処理装置２０と、データベース記憶装置１０２とを備え、前記検査機関情報処理装置２０は、解析対象遺伝子を特定する情報を取得し、解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、前記非解析対象遺伝子の遺伝子関連情報を前記データベース記憶装置に出力し、前記医療機関情報処理装置５０は、前記

遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を前記データベース記憶装置に出力し、前記データベース記憶装置１０２は、前記非解析対象遺伝子の遺伝子関連情報と、前記生体試料関連情報とを受け付け、記憶する、システムである。

【００１３】

第１、第２、第３－１、第３－２、第３－３の実施形態によれば、次世代シーケンシング解析やマイクロアレイ解析で取得される、測定対象の遺伝子及び測定対象以外の遺伝子の発現、又は前記遺伝子産物の機能を反映するデータを有効活用することができる。

【００１４】

本発明の課題を解決するための第４の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築する方法であって、前記データベースに記憶されたデータが、新規マーカーを探索するための人工知能の訓練データ又は検証データとして使用される、下記工程を含む、方法：測定対象遺伝子を特定する情報を取得する工程、測定対象遺伝子について前記遺伝子関連測定データを取得する工程、前記測定対象遺伝子の遺伝子関連情報をデータベースに記憶する工程、及び前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を前記データベースに記憶する工程である。

本発明によれば、大量の人工知能の訓練データ又は検証データを提供することができる。

【００１５】

本発明の課題を解決するための第５の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築する方法であって、前記データベースが、新規マーカーの候補を探索するために使用されるものである、下記工程を含む、方法：解析対象遺伝子以外の非解析対象遺伝子を含む複数の遺伝子について取得された前記遺伝子関連情報を、検査機関情報処理装置及び／又は医療機関情報処理装置から取得する工程、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を、検査機関情報処理装置及び／又は医療機関情報処理装置から取得する工程、及び前記遺伝子関連情報と、前記生体試料関連情報とを前記データベースに記憶する工程である。

【００１６】

本発明の課題を解決するための第６の実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報のデータベースを構築するシステム５００，６００，７００であって、前記データベースが、新規マーカーの候補を探索するために使用されるものであり、前記システムは、データベース記憶装置１００，１０１，１０２を備え、前記データベース記憶装置は、解析対象遺伝子以外の非解析対象遺伝子を含む複数の遺伝子について取得された前記遺伝子関連情報を、検査機関情報処理装置２０及び／又は医療機関情報処理装置５０から取得し、前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報を、検査機関情報処理装置２０及び／又は医療機関情報処理装置５０から取得し、前記遺伝子関連情報と、前記生体試料関連情報とを記憶する、システムである。

第５、第６の実施形態によれば、次世代シーケンシング解析やマイクロアレイ解析で取得される、測定対象の遺伝子及び測定対象以外の遺伝子の発現、又は前記遺伝子産物の機能を反映するデータを有効活用することができる。

【発明の効果】

【００１７】

本発明によれば、次世代シーケンシング解析やマイクロアレイ解析で取得される、測定対象の遺伝子及び測定対象以外の遺伝子の発現、又は前記遺伝子産物の機能を反映するデータを有効活用することができる。

【図面の簡単な説明】

【００１８】

【図１】図１は、本発明の第１の実施形態の概要を示す図である。

【図２】図２は、生体試料の採取から測定用試料の前処理までの流れを示す図である。
【図３】図３は、測定用試料の前処理産物を用いてデータベースを構築するまでを示すフローチャートである。
【図４】図４は、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔの解析対象遺伝子の一部を示す図である。
【図５】図５は、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔの図４に示された解析対象遺伝子以外の解析対象遺伝子を示す図である。
【図６】図６は、遺伝子関連情報の例を示す図である。
【図７】図７は、生体試料関連情報の例を示す図である。
【図８】図８は、報告書の例を示す図である。
【図９】図９は、測定用試料の前処理産物を用いて訓練データ又は検証データのデータベースを構築するまでを示すフローチャートである。
【図１０】図１０は、第３－１の実施形態のデータベース構築システムの概要を示す図である。
【図１１】図１１は、第３－２の実施形態のデータベース構築システムの概要を示す図である。
【図１２】図１２は、第３－３の実施形態のデータベース構築システムの概要を示す図である。
【図１３】図１３は、検査機関情報処理装置のブロック図である。
【図１４】図１４は、医療機関情報処理装置のブロック図である。
【図１５】図１５は、第１から第３のデータベース記憶装置のブロック図である。
【図１６】図１６は、新規マーカーの候補の探索方法を示すフローチャートである。
【図１７】図１７は、新規マーカー候補探索装置のブロック図である。

【発明を実施するための形態】
【００１９】
　以下、本発明の各実施形態を、添付の図面を参照して詳細に説明する。なお、本発明におけるデータベースを構築する方法、データベースを構築するためのシステム、及びデータベース記憶装置は、以下に説明する具体的な実施形態に限定されるものではない。また、以下の説明において同一の構成には、同一の符号を付す。したがって、同一符号が付された各構成についての説明は、同一符号間で共有され得る。さらに、各実施形態において共通して使用される用語については、各実施形態における用語の説明は、他の実施形態にも援用される。
【００２０】
［１．データベースの構築方法］
　初めに、図１を用いて本発明の一実施形態の概要を説明する。本実施形態は、生体試料における遺伝子の発現、又は遺伝子産物の機能を指標として疾患の診断や疾患の予後の予測、投薬の要否を判定する検査において、前記検査の目的を達成するために測定される解析対象遺伝子以外の、非解析対象遺伝子の遺伝子関連情報１を記憶したデータベースを構築する。例えば、生体試料として乳癌組織を用いて、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔ（シスメックス株式会社）による検査を行う際、一般的には、検査項目に含まれる解析対象遺伝子（９５ＧＣ）についてＲＮＡの発現量等の遺伝子関連測定データを取得する。本発明においては、９５ＧＣのＲＮＡの発現量を測定するのと同様の方法により、９５ＧＣ以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、前記非解析対象遺伝子の遺伝子関連測定データ含む遺伝子関連情報をデータベース化する。これらのデータベースは、疾患のバイオマーカーや疾患の治療標的分子等の新規マーカーを探索するために、例えば前記新規マーカーの再解析（リプロファイリング）に使用することができる。
【００２１】
　また、これらのデータベースは、人工知能を用いて前記新規マーカーの探索等を行う際に、人工知能に機械学習を行わせるための訓練データ、検証データを提供するために使用

することが可能である。さらに、前記データベースは、統計学的な手法を用いて、新規マーカーの探索を行う際の検証データを提供するために使用することが可能である。

【００２２】

［１－１．リプロファイリング用データベースの構築］

本発明の第１の実施形態は、新規マーカーの候補を探索するリプロファイリングに使用されるデータベースの構築方法に関する。具体的には、前記データベースは、生体試料における、遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報を不揮発性に記憶したものである。

【００２３】

前記新規マーカーは、例えば、疾患のバイオマーカー又は疾患の治療の標的分子である。前記疾患のバイオマーカーは、疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測等に使用することができる。また、前記疾患の治療の標的分子は、前記標的分子の機能を制御することにより、疾患を予防、治療、又は疾患の進行を遅延させることができる分子である。さらに、前記標的分子は、治療効果を予測するために用いられてもよい。

【００２４】

（１）生体試料の採取から測定用試料の前処理

次に、図２を用いてデータベース構築に使用される生体試料の採取から遺伝子関連情報を取得するまでの工程を説明する。

【００２５】

本実施形態において、生体試料としては、生体から採取されたものである限り制限されない。例えば、前記生体試料は、血液試料（全血、血漿、血清等）、尿、体液（汗、皮膚からの分泌液、涙液、唾液、髄液、腹水及び胸水）及び組織（新鮮組織、凍結組織、固定組織、及びパラフィン等の包埋剤に包埋された組織）であり得る。

【００２６】

また、生体試料は、所定の疾患、所定の疾患型及び所定の疾患の病期よりなる群から選択される少なくとも一種の病巣から採取されたものであることが好ましい。前記疾患は制限されないが、疾患として、好ましくは腫瘍（良性上皮性腫瘍、良性非上皮性腫瘍、悪性上皮性腫瘍、悪性非上皮性腫瘍）であり、より好ましくは悪性上皮性腫瘍、又は悪性非上皮性腫瘍であり、さらに好ましくは悪性上皮性腫瘍であり、さらにより好ましくは乳癌である。最も好ましくはリンパ節転移陰性かつエストロゲン受容体（ＥＲ）陽性乳癌である。

【００２７】

前記生体試料は、好ましくは複数であり、前記複数の生体試料は異なる患者の病巣から採取されたものである。より好ましくは前記複数の生体試料は、異なる患者の同一疾患の病巣から採取されたものであり、さらに好ましくは異なる患者の同一病期の病巣から採取されたものである。

【００２８】

また、生体試料は、前記病巣部位に対する陰性対照となりうる、正常と思われる組織を採取しても良い。この場合、前記正常と思われる組織は、前記病巣部位が属する組織の正常部位であることが好ましい。前記病巣部位が属する組織の正常部位は、複数の患者又は前記病巣を有していない者から採取されてもよい。

【００２９】

生体試料は、患者が属する医療機関等において、手術時又は生検時に採取することができる。採取された生体試料は、チューブ等の容器に収容される。前記容器には、サーモフィッシャー・サイエンティフィック（ＴｈｅｒｍｏＦｉｓｈｅｒＳｃｉｅｎｔｉｆｉｃ）社製、商品名：ＲＮＡｌａｔｅｒ（登録商標）等の保存液又はホルムアルデヒド等の固定液が入っていてもよい。容器に収容された生体試料は、冷蔵、冷凍してもよい。前記保存液又は固定液は公知のものを使用することができるが、保存又は輸送中の生体試料内の分子の分解や構造変化を防ぎ、生体試料をある程度一定の状態に保つ観点から、市販のキット又は市販の試薬を使用することが好ましい。例えば、生体試料の採取及び生体試料の収

容容器としては、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔ（シスメックス株式会社）に付属の容器を使用することができる。

容器に収容された生体試料は、医療機関、又は検査を受託する検査機関で遺伝子関連測定データを取得するために、前処理される。

【００３０】

ここで、遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データとしては、例えば、各遺伝子についてのＲＮＡ（ｍＲＮＡ及び／又はｍｉｃｒｏＲＮＡ）の発現量、ＲＮＡの塩基配列情報、ＤＮＡ（ゲノムＤＮＡ及び／又はミトコンドリアＤＮＡ）のメチル化量、ＤＮＡ（ゲノムＤＮＡ及び／又はミトコンドリアＤＮＡ）の塩基配列情報、又は遺伝子産物であるタンパク質（単量体タンパク質、複合体タンパク質、単量体ペプチド及び複合体ペプチドを含む）の存在量、タンパク質（単量体タンパク質、複合体タンパク質、単量体ペプチド及び複合体ペプチドを含む）の糖鎖修飾情報等を挙げることができる。例えば、遺伝子関連測定データがＤＮＡのメチル化量である場合には、前記遺伝子関連測定データには、各遺伝子におけるＤＮＡのメチル化量の他、少なくともそのＤＮＡのメチル化部位の位置情報が含まれる。また、遺伝子関連測定データがＤＮＡの塩基配列情報である場合には、遺伝子関連測定データには、塩基配列情報の他、少なくとも各遺伝子のＤＮＡの塩基配列の欠失、置換、融合、コピー数変異、又は挿入の有無と、その位置の情報が含まれる。前記ＤＮＡの配列情報には、１塩基多型、２塩基多型、３塩基多型等の遺伝子多型の情報も含まれる。さらに、遺伝子関連測定データがタンパク質の糖鎖修飾情報である場合には、前記遺伝子関連測定データには、各タンパク質の修飾の有無の他、各タンパク質の修飾位置と、前記タンパク質を修飾している糖鎖の種類の情報が含まれる。

【００３１】

したがって、遺伝子関連測定データを取得するための生体試料の前処理は、上記遺伝子関連測定データを取得するために、ＲＮＡ、ＤＮＡ又はタンパク質等の測定用試料を抽出できる限り制限されない。

【００３２】

例えば、遺伝子関連測定データを取得するためにＲＮＡを使用する場合には、公知の方法によって生体試料からＲＮＡを取得することができる。生体試料からのＲＮＡ抽出には、キアゲン（Ｑｉａｇｅｎ）社製、商品名：Ｑｉａｇｅｎ　ＲＮｅａｓｙ　ｋｉｔ（登録商標）等の市販のキットを使用することもできる。また、遺伝子関連測定データを取得するためにＤＮＡを取得する場合にも、公知の方法によって生体試料からＤＮＡを取得することができる。生体試料からのＤＮＡ抽出には、キアゲン（Ｑｉａｇｅｎ）社製、商品名：ＱＩＡａｍｐ　ＤＮＡ　Ｍｉｎｉ　Ｋｉｔ（登録商標）等の市販のキットを使用することもできる。さらに遺伝子関連測定データを取得するためにタンパク質を使用する場合にも、公知の方法によって生体試料からタンパク質を抽出することができる。生体試料からのタンパク質の抽出は、ＧＥヘルスケア・ジャパン株式会社、商品名：Ｍａｍｍａｌｉａｎ　Ｐｒｏｔｅｉｎ　Ｅｘｔｒａｃｔｉｏｎ　Ｂｕｆｆｅｒ等の市販試薬を使用することもできる。また、生体試料がパラフィン包埋されたものである場合には、キアゲン（Ｑｉａｇｅｎ）社製、商品名：ＱＩＡａｍｐ　ＤＮＡ　ＦＦＰＥ　Ｔｉｓｓｕｅ　Ｋｉｔ（登録商標）等を使用して生体試料からＤＮＡを抽出することができる。

【００３３】

生体試料の前処理は、その工程でのＲＮＡやＤＮＡの分解やタンパク質の構造変化等を防ぎ、測定用試料の均質化を図る点から、市販のキット又は市販の試薬を使用することが好ましい。

【００３４】

次に、遺伝子関連測定データを取得する前に、前記測定用試料は必要に応じて、前処置されてもよい。前記前処理には、遺伝子関連測定データを取得する際の検出に必要な蛍光標識やビオチン標識等を測定用試料のＲＮＡ、ＤＮＡ又はタンパク質又は以下で述べる測定用試料の前処理産物に施すことを含む。例えば、測定用試料がＲＮＡである場合には、測定用試料の前処理には、前記測定用試料のＲＮＡを鋳型として、ｃＤＮＡ又はｃＲＮＡ

を合成することが含まれてもよい。さらに、前記ｃＤＮＡ又はｃＲＮＡをＰＣＲによって増幅することが含まれてもよい。また、測定用試料がＤＮＡである場合には、測定用試料の前処理には、必要に応じて前記測定用試料のＤＮＡをＰＣＲによって増幅することが含まれてもよい。さらに、測定用試料の前処理には、測定用試料のＤＮＡ又は測定用試料のＤＮＡを鋳型として増幅されたＰＣＲ産物を制限酵素で切断することが含まれてもよい。測定用試料がタンパク質である場合には、必要に応じてドデシル硫酸ナトリウム、ＮＰ－４０、Ｔｒｉｔｏｎ　Ｘ－１００、Ｔｗｅｅｎ－２０等の界面活性剤及び／又はβ－メルカプトエタノール、ジチオスレイトール等の還元剤で変性することが含まれてもよい。前記前処理方法は、公知である。

【００３５】

測定用試料のＲＮＡ、ＤＮＡ又はタンパク質又は以下で述べる測定用試料の前処理産物に蛍光やビオチンを標識する方法も、公知である。例えば、サーモフィッシャー・サイエンティフィック社製、商品名：３’ＩＶＴ　ＰＬＵＳ　Ｒｅａｇｅｎｔ　Ｋｉｔを使用することができる。

【００３６】

上記の方法により測定用試料を前処理した前処理産物は、遺伝子関連測定データを取得するための測定に供される。

【００３７】

上述した生体試料の採取、生体試料からの測定用試料の抽出及び測定用試料の前処理は、均質化されたデータベースを構築する目的から、それぞれの工程における品質を管理するため、市販のキット、又は市販の試薬等を統一して使用することが望ましい。

【００３８】

次に、図３を用いて遺伝子関連測定データを取得するための各工程を説明する。遺伝子関連測定データの取得は、後述する第３の実施形態に係る検査機関情報処理装置２０によって行ってもよい。

【００３９】

（２）遺伝子関連測定データの取得

初めに医療機関が記入するする検査依頼書から、検査者、又は後述する検査機関情報処理装置２０の処理部２１が解析対象遺伝子を特定するための情報を取得する（ステップＳ１）。例えば、解析対象遺伝子は、疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測、薬効予測、及び疾患のモニタリングからなる群より選択される少なくとも一つの解析に使用される１又は複数の遺伝子を挙げることができる。さらに、前記解析対象遺伝子は、予め検査機関及び／又は医療機関等において、どの遺伝子について解析を行うか、例えば疾患ごと、疾患の病期ごとに応じて定められていることが好ましい。例えば、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔを例にして説明すると、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔには、専用の検査依頼書が貼付されている。必要事項が記入された検査依頼書は、医療機関から検査機関に郵送又はオンライン等で送付される。検査機関の検査者は、前記検査依頼書を受領することにより、検査項目がＣｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔを把握し、必要に応じて、処理部２１がＣｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔの検査を開始するための情報の入力を受け付ける。Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔは、図４及び図５に記載される９５個の遺伝子を解析対象遺伝子とするように規定されている。したがって、検査者、あるいは処理部２１は、Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔの解析対象遺伝子が図４及び図５に記載される９５遺伝子であると特定することができる。

【００４０】

ここで図４及び図５に記載の「プローブセット．ＩＤ」は、サーモフィッシャー・サイエンティフィック社製のマイクロアレイ〔商品名：ＧｅｎｅＣｈｉｐ（登録商標）　Ｓｙｓｔｅｍ〕において、基材上に固定されたプローブの１１～２０個をまとめたプローブセットそれぞれにつけられているＩＤ番号を示す。前記プローブセット．ＩＤで示された核

酸（プローブセット）の塩基配列は、ウェブページｈｔｔｐｓ：／／ｗｗｗ．ａｆｆｙｍｅｔｒｉｘ．ｃｏｍ／ａｎａｌｙｓｉｓ／ｎｅｔａｆｆｘ／ｉｎｄｅｘ．ａｆｆｘにより容易に入手することができる（２００９年６月３０日更新のデータベース）。「ＵｎｉＧｅｎｅ．ＩＤ」は、ＮＣＢＩが公開しているデータベースであるＵｎｉＧｅｎｅのＩＤ番号を示す。ＧｅｎＢａｎｋアクセッション番号は、前記サーモフィッシャー・サイエンティフィック社製のマイクロアレイ（商品名：ＧｅｎｅＣｈｉｐ（登録商標）　Ｓｙｓｔｅｍ）において、基材上に固定されたプローブそれぞれの配列の設計に用いられた公開データベースＧｅｎＢａｎｋのアクセッション番号を示す。前記ＧｅｎＢａｎｋアクセッション番号は、２００９年６月３０日時点での番号を示す。

【００４１】

次に、ステップＳ２では、検査者、あるいは処理部２１が、遺伝子関連測定データを所定の測定方法により取得する。遺伝子関連測定データの取得方法は制限されない。遺伝子関連測定データが、ＲＮＡの発現量、ＲＮＡの塩基配列情報、ＤＮＡのメチル化量、又はＤＮＡの塩基配列情報である場合には、塩基配列シーケンス及び／又はマイクロアレイにより測定することができる。より具体的には、ＲＮＡの発現量を測定するためには、次世代シーケンサーを使用したＲＮＡ－ｓｅｑ解析（Ｉｌｌｕｍｉｎａ，　Ｉｎｃ．）、ＲＮＡ発現解析が可能なマイクロアレイであるサーモフィッシャー・サイエンティフィック社製、商品名：Ｈｕｍａｎ　Ｇｅｎｏｍｅ　Ｕ１３３　Ｐｌｕｓ　２．０　Ａｒｒａｙ等を使用することができる。またＤＮＡのメチル化量を測定するためには、マイクロアレイを利用するＩｎｆｉｎｉｕｍ　ＭｅｔｈｙｌａｔｉｏｎＥＰＩＣ　Ｋｉｔ（Ｉｌｌｕｍｉｎａ，Ｉｎｃ．）等を使用することができる。また、ＤＮＡの塩基配列情報を測定（あるいは検出）するためには、サーモフィッシャー・サイエンティフィック社製、商品名：Ｇｅｎｏｍｅ－Ｗｉｄｅ　Ｈｕｍａｎ　ＳＮＰ　Ａｒｒａｙ　６．０又はＧｅｎｅＣｈｉｐ（登録商標）　Ｈｕｍａｎ　Ｇｅｎｏｍｅ　Ｕ１３３　Ｐｌｕｓ　２．０　Ａｒｒａｙ等を用いたマイクロアレイ測定、次世代シーケンサーによるエクソンシーケンスや全ゲノムシーケンス等を使用することができる。

【００４２】

また、遺伝子関連測定データが、タンパク質の存在量である場合には、マイクロアレイ及び／又はＥＬＩＳＡ（ＥＩＡを含む）により測定することができる。より具体的には、ＲａｙＢｉｏｔｅｃｈ社製の抗体アレイ（Ｃ－シリーズ、Ｇ－シリーズ、Ｌ－シリーズ、Ｑｕａｎｔｉｂｏｄｙ）及びＰｒｏｔｅｉｎ　Ａｒｒａｙシリーズ等を用いて測定することができる。

【００４３】

さらに、遺伝子関連測定データが、タンパク質の糖鎖修飾である場合には、マイクロアレイ及び／又はＥＬＩＳＡ（ＥＩＡを含む）により測定することができる。より具体的には、ＲａｙＢｉｏｔｅｃｈ社製のレクチンアレイ等を用いて測定することができる。

【００４４】

ステップＳ２では、測定用試料又はこれを前処理して得られた産物が核酸である場合には、上記測定行う前に、これらの核酸を熱変性することを含んでもよい。

【００４５】

上記測定方法は、取得される遺伝子関連測定データの均質性を保つ観点から、遺伝子関連測定データの再現性が担保される測定方法を選択することが好ましい。例えばマイクロアレイやその他の測定試薬は、一定のものを使用することが好ましい。このように、測定方法の均質化を図ることにより、上記測定試料及び／又は測定試料の前処理産物の均質化とあわせて、遺伝子関連測定データの品質を一定に保つことができる。また、遺伝子関連測定データの品質さらに一定に保つために、遺伝子関連測定データを取得する検査機関は、単一の機関（一定の検査精度を保ったブランチラボも含む）であるか、一定の検査精度を保った１又は複数の機関であることが好ましい。前記検査機関は、医療機関内に設置されていてもよい。

【００４６】

上記測定方法による遺伝子関連測定データの取得は、上記各測定方法において蛍光等のシグナルを測定するために適した後述する測定装置１０が、上記測定においてシグナルを取得し、上記処理部２１が当該シグナルの強度を算出することにより行われる。また前記シグナルの強度はＲＮＡ量（コピー数）、タンパク質量、ＤＮＡメチル化量又はメチル化の割合、ＲＮＡの塩基配列の変化率、ＤＮＡの塩基配列の変化率、タンパク質の糖鎖修飾の割合等に換算されて、遺伝子関連測定データとして取得されてもよい。

【００４７】

上記測定方法により取得された遺伝子関連測定データは、図４又は図５に示すように、少なくとも遺伝子名（あるいはＧｅｎＢａｎｋのアクセッション番号）又は遺伝子を特定するための符号（例えば、ＧｅｎｅＣｈｉｐ（登録商標）　Ｓｙｓｔｅｍのプローブセット．ＩＤ）と紐付けられている。したがって、遺伝子名又は遺伝子を特定するための符号から、検査者又は処理部２１は、どの遺伝子関連測定データが非解析対象遺伝子のものであるかを特定することができ（ステップＳ３）、検査者、又は処理部２１が、非解析対象遺伝子の遺伝子関連測定データを取得することができる（ステップＳ４）。

【００４８】

上記遺伝子関連測定データの取得は、解析対象遺伝子以外の非解析対象遺伝子についてのみ行ってもよいが、例えば、マイクロアレイ上に搭載されている全ての解析対象や、全ＲＮＡ、全ＤＮＡ又は全タンパク質に対して測定を行い、例えば遺伝子関連測定データに非解析対象遺伝子の遺伝子関連測定データのみを抽出してもよい。

【００４９】

取得された遺伝子関連測定データは、図３のステップＳ５において、図６に示すように遺伝子名（あるいはＧｅｎＢａｎｋのアクセッション番号）又は遺伝子を特定するための符号に加え、遺伝子関連測定データの測定日、測定方法、測定試料の量、検査機関、生体試料の保存方法及び生体試料の保存期間よりなる群から選択される少なくとも一種、及び生体試料を特定するための符号（例えばＩＤ）等の他の遺伝子関連情報と紐付けられ、検査者、又処理部２１によって後述する第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２に出力される（ステップＳ６）。

【００５０】

上記遺伝子関連測定データは、複数の非解析対象遺伝子及び／又は複数の解析対象遺伝子について取得されることが好ましい。前記複数の非解析対象遺伝子は、例えば解析対象遺伝子としては選択されなかったものの、所定の疾患、所定の疾患型又は所定の疾患の病期との関連が示唆された遺伝子を選択してもよい。非解析対象遺伝子は、解析対象遺伝子以外であって、かつ上記各測定方法において解析可能な遺伝子としてもよい。

【００５１】

さらに、上記方法により、検査者、あるいは処理部２１は、解析対象遺伝子の遺伝子関連測定データをさらに取得してもよい（ステップＳ９）。また、解析対象遺伝子の遺伝子関連測定データは、非解析対象遺伝子の遺伝子関連測定データと同様に、他の遺伝子関連情報と紐付けられて（ステップＳ１０）、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２に出力されてもよい（ステップＳ１０）。

【００５２】

上記遺伝子関連データは、正規化又は標準化されて第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２に記憶されてもよい。正規化の方法としては、例えば測定方法がマイクロアレイの場合には、総インテンシティ正規化、Ｌｏｗｅｓｓ正規化等の大域的正規化及び／又は局所的正規化を挙げることができる。より具体的には、ＲＭＡアルゴリズム、ＭＡＳ５アルゴリズム、ＰＬＩＥＲアルゴリズム等によって正規化することができる。前記ＲＭＡアルゴリズムを使用した解析ソフトウェアとしては、商品名：Ａｆｆｙｍｅｔｒｉｘ　Ｅｘｐｒｅｓｓｉｏｎ　Ｃｏｎｓｏｌｅソフトウェア（サーモフィッシャー・サイエンティフィック社）等を挙げ

ることができる。また、測定方法が次世代シーケンサーを使用する方法である場合には、Ｒｅａｄｓ Ｐｅｒ Ｍｉｌｌｉｏｎ ｍａｐｐｅｄ ｒｅａｄｓ （ＲＰＭ）、Ｒｅａｄ ｐｅｒ ｋｉｌｏｂａｓｅ ｏｆ ｅｘｏｎ ｍｏｄｅｌ ｐｅｒ ｍｉｌｌｉｏｎ ｍａｐｐｅｄ ｒｅａｄｓ （ＲＰＫＭ）、Ｔｒｉｍｍｅｄ ｍｅａｎ ｏｆ Ｍ ｖａｌｕｅｓ （ＴＭＭ）法等を挙げることができる。

【００５３】

上記遺伝子関連データの標準化は、生体試料の内部標準であるハウスキーピング遺伝子（ＧＡＰＤＨ：ｇｌｙｃｅｒａｌｄｅｈｙｄｅ－３－ｐｈｏｓｐｈａｔｅ ｄｅｈｙｄｒｏｇｅｎａｓｅ、β－アクチン、β２－マイクログロブリン、ＨＰＲＴ １：ｈｙｐｏｘａｎｔｈｉｎｅ ｐｈｏｓｐｈｏｒｉｂｏｓｙｌｔｒａｎｓｆｅｒａｓｅ １等）又はその遺伝子産物の発現量に基づいて遺伝子関連測定データの値を相対化する方法、マイクロアレイ実験の遺伝子発現情報データベースＮＣＢＩ Ｇｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ （ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／）に登録されているＤａｔａＳｅｔ Ｒｅｃｏｒｄ ＧＤＳ３８３４ （Ｍｕｌｔｉｐｌｅ ｎｏｒｍａｌ ｔｉｓｓｕｅｓ）等のデータを基準値として、Ｚスコア、有意確率（ｐ値）、又は尤度等を求める統計学的処理により行うことができる。また、前記基準値となるデータも、均質化された方法で取得されたものであることが好ましい。

【００５４】

ここで、複数の解析対象遺伝子の組み合わせとしては、例えば、Ｃｕｒｅｂｅｓｔ（登録商標） ９５ＧＣ Ｂｒｅａｓｔ解析対象遺伝子、Ｏｎｃｏｔｙｐｅ（登録商標） ＤＸ解析対象遺伝子、ＭａｍｍａＰｒｉｎｔ解析対象遺伝子、ＢｌｕｅＰｒｉｎｔ解析対象遺伝子、ＰＡＭ５０解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ６解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ６ ＋ ＣＯＳＭＩＣ解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ６ ＋ ＵＴＲ解析対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５ ＋ ＵＴＲｓ対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５ ＋ ＩｎｃＲＮＡ対象遺伝子、ＳｕｒｅＳｅｌｅｃｔ Ｈｕｍａｎ Ａｌｌ Ｅｘｏｎ Ｖ５ ＋ Ｒｅｇｕｌａｔｏｒｙ対象遺伝子、ＴｒｕＳｉｇｈｔ Ｃａｎｃｅｒ対象遺伝子、ＴｒｕＳｉｇｈｔ Ｔｕｍｏｒ １５対象遺伝子、及びＴｒｕＳｉｇｈｔ Ｔｕｍｏｒ １７０対象遺伝子よりなる群から選択される少なくとも一種を挙げることができる。

【００５５】

上記解析対象遺伝子は、２０遺伝子から１００遺伝子程度であることが一般的である。しかし、実際にマイクロアレイ等で測定される遺伝子は、３８，５００遺伝子程度であり、遺伝子産物のバリアント等も含めると５０，０００以上の遺伝子産物について解析が行われている。したがって、上記解析対象遺伝子を測定する際に、取得した非解析対象遺伝子の遺伝子関連情報や、これに対応する生体試料関連情報は非常に膨大なもとなる。したがって、これらの情報を集めたデータベースは、非常に膨大な情報を有し有用である。

【００５６】

また、上記遺伝子関連測定データを取得するにあたり、どのような疾患や病期の患者から生体試料を採取するか、どのような測定方法で遺伝子関連測定データを取得するか、生体試料についてどのような部位を採取するか、どのくらいの試料を採取するか、生体試料をどのように採取するか、測定まで採取された生体試料をどのように保存するか等の検査基準を予め定めておき、この基準に適合する生体試料について遺伝子関連測定データを取得してもよい。前記検査基準としては、前記診療関連情報、前記治療関連情報、生体試料の種類、測定方法、測定される前記生体試料の量、生体試料の採取方法、生体試料の保管方法よりなる群から選択される少なくとも一つに対して設定されている基準を挙げることができる。当該基準は、検査機関及び／又は医療機関が定めてもよい。

【００５７】

（３）データベースの構築

上記遺伝子関連情報を記憶する第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２の処理部１０１は、図３のステップ６で出力された遺伝子関連情報を取得し（ステップＳ７）、取得した前記遺伝子関連情報と、ステップ１２で医療機関から提供され取得した生体試料関連情報５とを不揮発性に記憶する（ステップＳ８）。前記生体試料関連情報５には、図７に示すように、少なくとも生体試料を特定するための符号が含まれる。また生体試料を特定するための符号（例えばＩＤ）には、前記生体試料を採取した患者を特定するための符号（例えば患者ＩＤ）と、生体試料の種類が紐付けられる。さらに、生体試料関連情報５には、前記患者の診療関連情報、及び治療関連情報よりなる群から選択される少なくとも一種が含まれる。前記診療関連情報には、疾患名、疾患型名、疾患の病期、患者の性別、患者の年齢、患者の既往歴、患者の家族歴、再発履歴、転移履歴、問診情報、月経履歴及び遺伝子関連情報以外の検査情報よりなる群から選択される少なくとも一種が含まれる。また、前記治療関連情報には、例えば、図７に示すように、治療薬の投与、予防薬の投与、放射線治療及び外科的処置よりなる群から選択される少なくとも一種の治療履歴が含まれる。より具体的には、前記治療が、治療薬の投与又は予防薬の投与である場合には、前記治療履歴には、投与した薬剤の名称、用量、投与頻度、投与日、投与期間等が含まれる。また、前記治療が放射線治療である場合には、前記治療履歴には、１回あたりの放射線照射量、頻度、施術日、総照射放射線量等が含まれる。前記治療が外科的処置である場合には、前記治療履歴には、主な切除部位、術式、リンパ節等の切除部位周辺組織の郭清の有無、施術日等が含まれる。

【００５８】

前記遺伝子関連情報と前記生体試料関連情報５は、生体試料を特定するための符号をキーとして対応させることが可能である。このため、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２において、前記遺伝子関連情報と前記生体試料関連情報５とは、一つのファイルに結合される必要はないが、一つのファイルに結合されてもよい。また、別の態様として、前記遺伝子関連情報と前記生体試料関連情報５とは、ネットワークを介して例えばデータベースのユーザの端末から呼び出し可能に接続された２つのデータベース記憶装置にそれぞれが個別に記憶されていてもよい。

【００５９】

さらに、本実施形態において構築されたデータベースは、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等の記憶媒体に記憶されていてもよい。前記記憶媒体へのデータベースの記憶形式は、前記提示装置が前記データベースを読み取り可能である限り制限されない。前記記憶媒体への記憶は、不揮発性であることが好ましい。この場合、前記データベースの構築方法は、前記データベースを記憶した記憶媒体の製造方法と読み替えることができる。

【００６０】

（４）その他の態様

上記データベースの構築方法においては、上記１－１．（２）で取得された解析対象遺伝子の遺伝子関連情報２、又は解析対象遺伝子の遺伝子関連情報２と非解析対象遺伝子の遺伝子関連情報１を医療機関に報告するための報告書３，４を作成する工程を含んでいてもよい。前記報告書３，４には、例えば図８に示すように、各遺伝子の名称（あるいはＧｅｎＢａｎｋのアクセッション番号）及び／又は各遺伝子を特定するための符号と、各遺伝子についての前記遺伝子関連測定データと、前記遺伝子関連測定データを取得した生体試料を特定するための符号と、遺伝子関連測定データの測定日、測定方法、検査機関の名称、生体試料の保存方法及び生体試料の保存期間よりなる群から選択される少なくとも一種とが含まれる。さらに、報告書３，４は、例えば疾患のリスク判定、スクリーニング、鑑別診断、予後予測、再発予測、薬効予測、及び疾患のモニタリングよりなる群より選択される少なくとも一つの判定結果を含んでいてもよい。Ｃｕｒｅｂｅｓｔ（登録商標）９５ＧＣ　Ｂｒｅａｓｔでは、乳癌の術前化学療法に対する感受性、リンパ節転移陰性か

つエストロゲン受容体（ＥＲ）陽性乳癌患者について乳癌の再発伴う予後を予測することができる。さらには、前記予後予測から、手術後にホルモン療法を適用するのみでよいか、化学療法を併用すべきかの予測を行うこともできる。例えば、Ｃｕｒｅｂｅｓｔ（登録商標）　９５ＧＣ　Ｂｒｅａｓｔでは、報告書３には、リンパ節転移陰性かつエストロゲン受容体（ＥＲ）陽性乳癌患者について、乳癌再発の予後予測結果がＨ（再発Ｈｉｇｈ－ｒｉｓｋ群）又はＬ（再発Ｌｏｗ－ｒｉｓｋ群）として表示される。また、報告書３，４には、生体試料に検査に必要な量の癌細胞が含まれていたかを示すための癌細胞の含有率（有無）を示す値を表示してもよい。

【００６１】

本実施形態において、検査機関情報処理装置２０の処理部２１が行う各ステップ（ステップＳ１からステップＳ６、又はステップＳ１からステップＳ６、ステップＳ９及びステップＳ１０）は、コンピュータプログラムによって実行される。第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２の処理部１０１が行う各ステップ（ステップＳ７、ステップＳ１２及びステップＳ８）もまた、コンピュータプログラムによって実行される。前記コンピュータプログラムは、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等の記憶媒体に記憶されていてもよい。前記記憶媒体へのプログラムの記憶形式は、前記提示装置が前記プログラムを読み取り可能である限り制限されない。前記記憶媒体への記憶は、不揮発性であることが好ましい。

【００６２】

また、本実施形態の一例において、リプロファイリングにより探索される疾患のバイオマーカーは、前記生体試料が採取された患者が患う疾患とは異なる疾患のバイオマーカーであっても、前記生体試料が採取された患者が患う疾患と同じ疾患のバイオマーカーであってもよい。

【００６３】

本実施形態によれば、生体試料の採取からデータベース構築までの工程を均質化するよう、測定試料、遺伝子関連測定データの品質を管理した条件で行うことも可能である。このように品質が管理された条件で取得される遺伝子関連測定データは、生体試料の保存状態による測定試料の品質不良等を考慮する必要がないため、生体試料を採取した患者の病変組織の状態を反映する。したがって、第１の実施形態に従って構築されるデータベースは、患者の病変組織の状態を反映するという点において、他のデータベースよりも信頼性が高い。

【００６４】

［１－２．訓練データ、検証データ用のデータベースの構築］

本発明の第２の態様は、人工知能を用いて前記新規マーカーの探索等を行う際に、人工知能に判別式、決定木、最近傍法、サポートベクターマシン、ニューラルネットワーク、深層学習等の機械学習を行わせるための訓練データ（教師データ、学習データともいう）、構築した学習モデルが有効か否かを判定するための検証データ（テストデータ）を提供するためのデータベースの構築方法に関する。また、本実施形態において構築されるデータベースは、回帰分析、重回帰分析、分散分析、主成分分析などの統計学的手法によって求められる数理モデルの検証（バリデーション）に使用することができる。

【００６５】

本発明のデータベースの構築方法においては、第１の実施形態で述べたように、生体試料の採取からデータベース構築までの工程を均質化するよう、測定試料、遺伝子関連測定データの品質を管理した条件で行うことも可能である。このため、上記第１の実施形態に記載の生体試料の採取、生体試料の前処理、前記前処理によって取得された測定試料の前処理方法、及び遺伝子関連測定データの取得方法にしたがって取得された解析対象遺伝子及び非解析対象遺伝子の遺伝子関連測定データは、患者の病変組織の状態を反映するという点において、他のデータベースよりも信頼性が高い。このため、訓練データ、又は構築した学習モデルが有効か否かを判定するための検証データとして、信頼性の高いデータを

提供することができる。

【００６６】

具体的には、第２の実施形態は、図９に示すように、検査者、又は検査機関情報処理部２０の処理部２１が解析対象遺伝子を特定する情報を取得するステップＳ２１と、検査者、又は処理部２１が、解析対象遺伝子について前記遺伝子関連測定データを取得するステップＳ２２と、前記解析対象遺伝子の遺伝子関連情報２を第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２に出力するステップＳ２３を含む。また、第２の実施形態は、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２の処理部１０１が、ステップ２３で出力された遺伝子関連情報を取得し（ステップＳ２４）、取得した前記遺伝子関連情報と、ステップＳ２６で医療機関から提供され取得した生体試料関連情報５とを不揮発性に記憶するステップＳ２５を含む。

【００６７】

さらに、本実施形態において構築されたデータベースは、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等の記憶媒体に記憶されていてもよい。前記記憶媒体へのデータベースの記憶形式は、前記提示装置が前記データベースを読み取り可能である限り制限されない。前記記憶媒体への記憶は、不揮発性であることが好ましい。この場合、前記データベースの構築方法は、前記データベースを記憶した記憶媒体の製造方法と読み替えることができる。

【００６８】

また、第２の実施形態ではさらに、検査者、又は処理部２１が、ステップ２２において、非解析対象遺伝子について前記遺伝子関連測定データを取得し、ステップ２３において、前記非解析対象遺伝子の遺伝子関連情報１を第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２に出力し、ステップ２４において、前記非解析対象遺伝子の遺伝子関連情報１を第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２に記憶してもよい。また、第２の実施形態では、ステップＳ２２からステップＳ２５において非解析対象遺伝子の遺伝子関連情報１のみからデータベースを構築してもよい。

【００６９】

本実施形態において、検査機関情報処理装置２０の処理部２１が行う各ステップ（ステップＳ２１からステップＳ２３、又はステップＳ１からステップＳ２３、ステップＳ２６及びステップＳ２７）は、コンピュータプログラムによって実行される第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１、又は第３のデータベース記憶装置１０２の処理部１０１が行う各ステップ（ステップＳ２４、ステップＳ２６及びステップＳ２５）もまた、コンピュータプログラムによって実行される。前記コンピュータプログラムは、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等の記憶媒体に記憶されていてもよい。前記記憶媒体へのプログラムの記憶形式は、前記提示装置が前記プログラムを読み取り可能である限り制限されない。前記記憶媒体への記憶は、不揮発性であることが好ましい。

【００７０】

上記方法により構築されたデータベースは、人工知能に学習させるために、又は人工知能が構築したモデルを検証するために使用することができる。データベースに記憶された解析対象遺伝子の遺伝子関連情報２及び非解析対象遺伝子の遺伝子関連情報１は、目的に応じて一方又は両方を人工知能に学習させるために使用してもよい。例えば１疾患について、データベースに記憶されている解析対象遺伝子の遺伝子関連情報２とこれらに対応する生体資料関連情報５を２群に分け、一方を訓練データとして使用し、もう一方を検証データとして使用してもよい。また、１疾患について、データベースに記憶されている解析対象遺伝子の遺伝子関連情報２を全て訓練データとして使用し、Ｌｅａｖｅ－Ｏｎｅ－Ｏｕｔ Ｃｒｏｓｓ－Ｖａｌｉｄａｔｉｏｎを行う場合にも、Ｌｅａｖｅ－Ｏｎｅ－Ｏｕｔ Ｃｒｏｓｓ－Ｖａｌｉｄａｔｉｏｎに使用される解析対象遺伝子の遺伝子関連情報２とこ

れらに対応する生体資料関連情報５は、検証データとして扱うことができる。本段落において、解析対象遺伝子の遺伝子関連情報２は、非解析対象遺伝子の遺伝子関連情報１と置き換えることができる。

【００７１】

［２．データベースを構築するシステム］

本発明の第３の実施形態は、上記第１の実施形態及び第２の実施形態で説明したデータベースを構築するためのシステムに関する。

【００７２】

第３の実施形態の実施形態には、検査機関においてデータベースを構築する第３－１の実施形態、医療機関においてデータベースを構築する第３－２の実施形態、及び検査機関及び医療機関が協働でデータベースを構築する第３－３の実施形態を含む。

以下、図１０から図１２に示すシステムの概略図と、図１３から図１５を用いて各実施形態について説明する。

【００７３】

［２－１．各ハードウェアの構成］

図１３に記載の検査機関情報処理装置２０、図１４に記載の医療機関情報処理装置５０、図１５に記載の第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１及び第３のデータベース記憶装置１０２は、ハードウェア構成の一例である。ハードウェアは、パーソナルコンピュータ、タブレット型端末でありうる。また、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１及び第３のデータベース記憶装置１０２を構成するハードウェアは、いわゆるサーバとしての役割を有するものであってもよく、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）又はＭＰＵ（Ｍｉｃｒｏ－ｐｒｏｃｅｓｓｉｎｇ　ｕｎｉｔ）であり、例えば、Ｌｉｎｕｘ（登録商標）、ＵＮＩＸ（登録商標）、マイクロソフト　ウインドウズ　サーバ（登録商標）等のサーバ　オペレーティングシステム（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ：ＯＳ）を使って前記記憶装置１００、１０１、１０２を制御する。

【００７４】

検査機関情報処理装置２０は、処理部（ＣＰＵ）２１、主記憶部２２、ＲＯＭ（ｒｅａｄ　ｏｎｌｙ　ｍｅｍｏｒｙ）２３；補助記憶部２４；通信Ｉ／Ｆ（ｉｎｔｅｒｆａｃｅ）２５；入力Ｉ／Ｆ２６；出力Ｉ／Ｆ２７；メディアＩ／Ｆ２８；バス２９を備える。また、検査機関情報処理装置２０は、入力部３０、及び表示部３１を備える。また、検査機関情報処理装置２０は、記憶媒体３２を備えていてもよい。

【００７５】

医療機関情報処理装置５０は、処理部（ＣＰＵ）５１、主記憶部５２、ＲＯＭ５３；補助記憶部５４；通信Ｉ／Ｆ５５；入力Ｉ／Ｆ５６；出力Ｉ／Ｆ５７；メディアＩ／Ｆ５８；バス５９を備える。また、医療機関情報処理装置５０は、入力部６０、及び表示部６１を備える。また、医療機関情報処理装置５０は、記憶媒体６２を備えていてもよい。

【００７６】

第１のデータベース記憶装置（検査機関データベース記憶装置）１００、第２のデータベース記憶装置（医療機関データベース記憶装置）１０１及び第３のデータベース記憶装置１０２は、処理部（ＣＰＵ）２０１、主記憶部２０２、ＲＯＭ２０３；補助記憶部２０４；通信Ｉ／Ｆ２０５；入力Ｉ／Ｆ２０６；出力Ｉ／Ｆ２０７；メディアＩ／Ｆ２０８；バス２０９を備える。また、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１及び第３のデータベース記憶装置１０２は、入力部２１０、及び表示部２１１を備える。また、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１及び第３のデータベース記憶装置１０２は、記憶媒体２１２を備えていてもよい。

【００７７】

ＣＰＵ２１、５１、２０１は、ＲＯＭ２３、５３、２０３、及び補助記憶部２４、５４、２０４に記憶されたプログラムに基づいて、各部を制御する。ＣＰＵ２１、５１、２０１はＭＰＵ２１、５１、２０１としてもよい。

【００７８】
ＲＯＭ２３、５３、２０３は、マスクＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭなどによって構成され、検査機関情報処理装置１０、医療機関情報処理装置５０、第１のデータベース記憶装置１００、第２のデータベース記憶装置１０１及び第３のデータベース記憶装置１０２の起動時に、ＣＰＵ２１、５１、２０１によって実行されるブートプログラムや前記装置のハードウェアの動作に関連するプログラムや設定を記憶する。
【００７９】
主記憶部２２、５２、２０２は、ＳＲＡＭ又はＤＲＡＭなどのＲＡＭから構成され、入力部３０、６０、２１０から受け付けた情報を揮発性に記憶する。補助記憶部２４、５４、２０４はアプリケーションソフトや、前記各装置２０、５０、１００、１０１，１０２の動作中に入力又は生成される情報を不揮発性に記憶する（不揮発性の記憶は、「記録」ともいう）。補助記憶部２４、５４、２０４は、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等によって構成される。
【００８０】
通信Ｉ／Ｆ２５、５５、２０５は、外部機器からの情報を受信し、また各装置２０、５０、１００、１０１，１０２が保存又は生成する情報を外部に送信する。通信Ｉ／Ｆ２５、５５、２０５は、ＵＳＢ、ＩＥＥＥ１３９４、ＲＳ－２３２Ｃなどのシリアルインタフェース、ＳＣＳＩ、ＩＤＥ、ＩＥＥＥ１２８４などのパラレルインタフェース、及びＤ／Ａ変換器、Ａ／Ｄ変換器などからなるアナログインタフェース、ネットワークインタフェースコントローラ（Ｎｅｔｗｏｒｋ ｉｎｔｅｒｆａｃｅ ｃｏｎｔｒｏｌｌｅｒ：ＮＩＣ）等から構成される。
【００８１】
入力Ｉ／Ｆ２６、５６、２０６は、入力部３０、６０、２１０からの文字入力、クリック、音声入力等を受け付ける。例えば入力Ｉ／Ｆ２６、５６、２０６は、ＵＳＢ、ＩＥＥＥ１３９４、ＲＳ－２３２Ｃなどのシリアルインタフェース、ＳＣＳＩ、ＩＤＥ、ＩＥＥＥ１２８４などのパラレルインタフェース、及びＤ／Ａ変換器、Ａ／Ｄ変換器などからなるアナログインタフェースなどから構成される。受け付けた入力内容は、主記憶部２２、５２、２０２又は補助記憶部２４、５４、２０４に記憶される。
【００８２】
出力Ｉ／Ｆ２７、５７、２０７は、例えば、入力Ｉ／Ｆ２６、５６、２０６と同様のインタフェースから構成され、ＣＰＵ２１、５１、２０１が生成した情報を表示部３１、５１、２１１に出力する。出力Ｉ／Ｆ２７、５７、２０７は、ＣＰＵ２１、５１、２０１が生成し、補助記憶部２４、５４、２０４に記憶した情報を、表示部３１、５１、２１１に出力する。ここで表示部３１、５１、２１１は、ディスプレイ又はプロジェクタであってもよいが、プリンターであってもよい。
【００８３】
メディアＩ／Ｆ２８、５８、２０８は、記憶媒体３２、６２、２１２に記憶された例えばアプリケーションソフト等を読み出す。読み出されたアプリケーションソフト等は、主記憶部２２、５２、２０２又は補助記憶部２４、５４、２０４に記憶される。また、メディアＩ／Ｆ２８、５８、２０８は、ＣＰＵ２１、５１、２０１が生成した情報を記憶媒体３２、６２、２１２に書き込む。メディアＩ／Ｆ２８、５８、２０８は、ＣＰＵ２１、５１、２０１が生成し、補助記憶部２４、５４、２０４に記憶した情報を、記憶媒体３２、６２、２１２に書き込む。記憶媒体３２、６２、２１２は、フレキシブルディスク、ＣＤ－ＲＯＭ、又はＤＶＤ－ＲＯＭ等で構成される。記憶媒体３２、６２、２１２は、フレキシブルディスクドライブ、ＣＤ－ＲＯＭドライブ、又はＤＶＤ－ＲＯＭドライブ等によってメディアＩ／Ｆ２８、５８、２０８と接続される。
ＣＰＵ２１、５１、２０１による各ハードウェア構成の制御は、バス２９、５９、２０９によって各ハードウェア構成に伝達される。
【００８４】
［２－２．検査機関においてデータベースを構築するシステム］

第３－１の実施形態に係るシステム５００は、図１０に示すように、検査機関情報処理装置２０と、第１のデータベース記憶装置１００とを備える。また、本実施形態に係るシステム５００は、医療機関情報処理装置５０を備えていてもよい。検査機関情報処理装置２０は、測定装置１０と直接、又はネットワークを介して接続され、測定システム３００を構築してもよい。前記システムにおいて、少なくとも検査機関情報処理装置２０と第１のデータベース記憶装置１００とは、ネットワークを介して接続されていてもよい。また、検査機関情報処理装置２０と医療機関情報処理装置５０とは、ネットワークを介して接続されていてもよい。

【００８５】

前記検査機関情報処理装置２０の処理部２１は、例えば入力部３０からの入力により、あるいは通信Ｉ／Ｆ２５又はメディアＩ／Ｆ２８を介して解析対象遺伝子を特定する情報を取得し、主記憶部２２、ＲＯＭ２３又は補助記憶部２４に記憶する。また、処理部２１は、測定装置１０から遺伝子関連測定データを取得する。次に処理部２１は、解析対象遺伝子及び／又は解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、各遺伝子について遺伝子関連情報を生成する。続いて、処理部２１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を通信Ｉ／Ｆ２５を介して、前記第１のデータベース記憶装置１００に出力する。

【００８６】

前記第１のデータベース記憶装置１００の処理部２０１は、解析対象遺伝子及び／又は非解析対象遺伝子の遺伝子関連情報１を、通信Ｉ／Ｆ２０５を介して取得する。また、第１のデータベース記憶装置１００の処理部２０１は、入力部２１０からの入力により、あるいは通信Ｉ／Ｆ２０５又はメディアＩ／Ｆ２０８を介して前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報５を取得する。第１のデータベース記憶装置１００の処理部２０１は取得した前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１と前記生体試料関連情報５を補助記憶部２０４に記憶する。

【００８７】

ここで、検査機関情報処理装置２０の処理部２１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を前記第１のデータベース記憶装置１００に出力するために、記憶媒体３２に記憶してもよい。前記第１のデータベース記憶装置１００の処理部２０１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１をメディアＩ／Ｆ２０８を介して取得してもよい。また、検査機関情報処理装置２０の処理部２１は、前記生体試料関連情報５を取得して、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１とともに、前記第１のデータベース記憶装置１００に出力してもよい。

前記［１－１．リプロファイリング用データベースの構築］の各工程の説明は、ここに援用される。

【００８８】

［２－３．医療機関においてデータベースを構築するシステム］

第３－２の実施形態に係るシステム６００は、図１１に示すように、検査機関情報処理装置２０と、医療機関情報処理装置５０と、第２のデータベース記憶装置１０１とを備える。前記システム６００において、検査機関情報処理装置２０と、医療機関情報処理装置５０及び／又は第２のデータベース記憶装置１０１とは、ネットワークを介して接続されていてもよい。

【００８９】

前記検査機関情報処理装置２０の処理部２１は、例えば入力部３０からの入力により、あるいは通信Ｉ／Ｆ２５又はメディアＩ／Ｆ２８を介して解析対象遺伝子を特定する情報を取得し、主記憶部２２、ＲＯＭ２３又は補助記憶部２４に記憶する。また、処理部２１は、測定装置１０から遺伝子関連測定データを取得する。次に処理部２１は、解析対象遺伝子及び／又は解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定デー

タを取得し、各遺伝子について遺伝子関連情報を生成する。続いて、処理部２１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を通信Ｉ／Ｆ２５を介して、前記第２のデータベース記憶装置１０１に出力する。

【００９０】

前記医療機関情報処理部５０の処理部５１は、医療機関において医師等により入力部６０から入力された前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報５を受け付け、前記生体試料関連情報５を通信Ｉ／Ｆ５５を介して、前記第２のデータベース記憶装置１０１に出力する。

【００９１】

前記第２のデータベース記憶装置１０１の処理部２０１は、解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を、通信Ｉ／Ｆ２０５を介して取得する。また、第２のデータベース記憶装置１０１の処理部２０１は、通信Ｉ／Ｆ２０５又を介して前記生体試料関連情報５を取得する。第２のデータベース記憶装置１０１の処理部２０１は取得した前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１と前記生体試料関連情報５を補助記憶部２０４に記憶する。

【００９２】

ここで、検査機関情報処理装置２０の処理部２１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を前記第２のデータベース記憶装置１０１に出力するために、記憶媒体３２に記憶してもよい。医療機関情報処理装置５０の処理部５１は、前記生体試料関連情報５を前記第２のデータベース記憶装置１０１に出力するために、記憶媒体５２に記憶してもよい。前記第２のデータベース記憶装置１０１の処理部２０１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１及び前記生体試料関連情報５をメディアＩ／Ｆ２０８を介して取得してもよい。

前記［１－１．リプロファイリング用データベースの構築］の各工程の説明は、ここに援用される。

【００９３】

［２－４．検査機関及び医療機関が協働でデータベースを構築するシステム］

第３－３の実施形態に係るシステム７００は、図１２に示すように、検査機関情報処理装置２０と、医療機関情報処理装置５０と、第３のデータベース記憶装置１０２とを備える。前記システム７００において、前記検査機関情報処理装置２０と第３のデータベース記憶装置１０２、及び／又は、医療機関情報処理装置５０と第３のデータベース記憶装置１０２とは、ネットワークを介して接続されていてもよい。

【００９４】

前記検査機関情報処理装置２０の処理部２１は、例えば入力部３０からの入力により、あるいは通信Ｉ／Ｆ２５又はメディアＩ／Ｆ２８を介して解析対象遺伝子を特定する情報を取得し、主記憶部２２、ＲＯＭ２３又は補助記憶部２４に記憶する。また、処理部２１は、測定装置１０から遺伝子関連測定データを取得する。次に処理部２１は、解析対象遺伝子及び／又は解析対象遺伝子以外の非解析対象遺伝子について前記遺伝子関連測定データを取得し、各遺伝子について遺伝子関連情報を生成する。続いて、処理部２１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を通信Ｉ／Ｆ２５を介して、前記第３のデータベース記憶装置１０２に出力する。

【００９５】

前記医療機関情報処理部５０の処理部５１は、医療機関において医師等により入力部６０から入力された前記遺伝子関連測定データを取得した生体試料に関連する情報である生体試料関連情報５を受け付け、前記生体試料関連情報５を通信Ｉ／Ｆ５５を介して、前記第３のデータベース記憶装置１０２に出力する。

【００９６】

前記第３のデータベース記憶装置１０２の処理部２０１は、解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を、通信Ｉ／Ｆ２０５を介して

取得する。また、第３のデータベース記憶装置１０２の処理部２０１は、通信Ｉ／Ｆ２０５又を介して前記生体試料関連情報５を取得する。第３のデータベース記憶装置１０２の処理部２０１は取得した前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１と前記生体試料関連情報５を補助記憶部２０４に記憶する。

【００９７】

ここで、検査機関情報処理装置２０の処理部２１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１を前記第３のデータベース記憶装置１０２に出力するために、記憶媒体３２に記憶してもよい。医療機関情報処理装置５０の処理部５１は、前記生体試料関連情報５を前記第３のデータベース記憶装置１０２に出力するために、記憶媒体５２に記憶してもよい。前記第３のデータベース記憶装置１０２の処理部２０１は、前記解析対象遺伝子の遺伝子関連情報２及び／又は非解析対象遺伝子の遺伝子関連情報１及び前記生体試料関連情報５をメディアＩ／Ｆ２０８を介して取得してもよい。

前記［１－１．リプロファイリング用データベースの構築］の各工程の説明は、ここに援用される。

【００９８】

上記第３－１の実施形態、第３－２の実施形態、及び第３－３の実施形態において、検査機関情報処理装置２０の処理部２１は、解析対象遺伝子及び／又は非解析対象遺伝子についての報告書３、４を生成してもよい。

【００９９】

［３．新規マーカーの候補を探索する方法］

本発明の第４の実施形態は、第１の実施形態により構築されたデータベースを使用して、生体試料における遺伝子の発現、又は遺伝子産物の機能を反映する遺伝子関連測定データを含む遺伝子関連情報をリプロファイリングし、新規マーカーの候補を探索する方法に関する。したがって、本実施形態において第１の実施形態と共通する用語や説明は、第１の実施形態の記載を援用する。また、第４の実施形態は、後述する第５の実施形態に係る新規マーカー探索装置８０によって実施してもよい。

【０１００】

本実施形態では、図１６に示すように、検査者、又は新規マーカー探索装置８０の処理部８１は、第１の実施形態において、非解析対象遺伝子の遺伝子関連情報１と、前記生体試料関連情報５とを記憶したデータベースから非解析対象遺伝子の遺伝子関連情報１と、前記生体試料関連情報５を取得し、例えば、双方の情報に含まれる生体試料を特定するための情報をキーとして、各非解析対象遺伝子の遺伝子関連情報１と、前記生体試料関連情報５とを対応させる（ステップＳ３１）。次に、検査者、又は新規マーカー探索装置８０の処理部８１は、前記遺伝子関連情報に含まれる遺伝子関連測定データと前記生体試料関連情報５との関連性の強さを示す数値を各遺伝子について取得する（ステップＳ３２）。例えば、前記数値は、ＲＮＡ量（コピー数）、タンパク質量、ＤＮＡメチル化量又はメチル化の割合、ＲＮＡの塩基配列の変化率、ＤＮＡの塩基配列の変化率、タンパク質の糖鎖修飾の割合等であり得る。前記数値は、ＲＮＡ量（コピー数）、タンパク質量、ＤＮＡメチル化量又はメチル化の割合、ＲＮＡの塩基配列の変化率、ＤＮＡの塩基配列の変化率、タンパク質の糖鎖修飾の割合等の値を統計学的に処理して、標準化したデータを前記数値としてもよい。具体的には、前記標準化は、有意確率（ｐ値）、尤度、又はＺスコア等である。前記統計学的処理は、公知の方法に従って行うことができる。例えば有意確率（ｐ値）は、ステューデントｔ検定、ウェルチのｔ検定、ウィルコクソンの符号順位検定及びこれらの改良方法から選択される有意差検定等で求めることができる。尤度は、最尤推定法、尤度検定等で求めることができる。ｚスコアを求める場合には、統計解析用ソフトウェア「Ｒ」で用いられる追加パッケージ集「ＢｉｏＣｏｎｄｕｃｔｏｒ」ｖｅｒ．２．４に含まれるパッケージ「ＧｅｎｅＭｅｔａ　ｖ１．１６．０」（ｈｔｔｐ：／／ｗｗｗ．ｂｉｏｃｏｎｄｕｃｔｏｒ．ｏｒｇ／ｐａｃｋａｇｅｓ／２．４／ｂｉｏｃ／ｈｔｍｌ／ＧｅｎｅＭｅｔａ．ｈｔｍｌ）を用い、ジュン・キョン・チェ（Ｊｕｎｇ　Ｋｙｏｏｎ　Ｃ

ｈｏｉ）らの文献〔「複数のマイクロアレイ研究の統合及び研究間バリデーションのモデリング（Ｃｏｍｂｉｎｉｎｇ ｍｕｌｔｉｐｌｅ ｍｉｃｒｏａｒｒａｙ ｓｔｕｄｉｅｓ ａｎｄ ｍｏｄｅｌｉｎｇ ｉｎｔｅｒｓｔｕｄｙ ｖａｒｉａｔｉｏｎ）」バイオインフォマティックス（Ｂｉｏｉｎｆｏｒｍａｔｉｃｓ）、第１９巻、補遺１、２００３年、ｐ．ｉ８４－ｉ９０〕にしたがって、求めることができる。

【０１０１】

また、前記統計学的処理において、健常組織の基準データが必要な場合には、例えば、ＤａｔａＳｅｔ Ｒｅｃｏｒｄ ＧＤＳ３８３４ （Ｍｕｌｔｉｐｌｅ ｎｏｒｍａｌ ｔｉｓｓｕｅｓ）等のデータを使用することができる。また、統計学的な解析において疾患の基準となるデータが必要な場合には、ＮＣＢＩ Ｇｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ （ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／）に登録されているデータを使用することができる。また、好ましくは、均質化したデータを得るために、上記第１の実施形態における遺伝子関連測定データの取得方法にしたがって、健常組織、又は疾患の病巣を有する組織の基準データを取得してもよい。

【０１０２】

続いて、検査者、又は新規マーカー探索装置８０の処理部８１は、前記数値に基づいて、各生体試料関連情報と関連の強い遺伝子を新規マーカーの候補を決定する。具体的には、検査者、又は新規マーカー探索装置８０の処理部８１は、上記数値について、例えば、その絶対値をとり、その絶対値に基づいて当該絶対値に対応する遺伝子関連測定データを並べ変え（ステップＳ３３）、いずれの遺伝子の絶対値が高いかを決定する（ステップＳ３４）。そして、検査者、又は新規マーカー探索装置８０の処理部８１は、絶対値が高い遺伝子を新規マーカーの候補として決定し（ステップＳ３５）、絶対値が低い遺伝子を新規マーカーの候補ではないと決定することができる（ステップＳ３６）。前記新規マーカーは複数であってもよい。

【０１０３】

各生体試料関連情報と、複数の遺伝子との関連性を求める場合には、前記数値に対して、総計学的処理等を施して関連性を求めることができる。例えば、ステップＳ３３において前記数値の絶対値に基づいて並べられた遺伝子について最上位から所定の順位までの複数の遺伝子について、ＦＡＬＳＥ ＤＩＳＣＯＶＥＲＹ ＲＡＴＥやファミリーワイズエラー率、Ｂｏｎｆｅｒｒｏｎｉ法、Ｈｏｌｍ法等の多重比較；Ｐｅｒｍｕｔａｔｉｏｎ ｔｅｓｔ、Ｂｏｏｔｓｔｒａｐ法、Ｃｒｏｓｓ Ｖａｌｉｄａｔｉｏｎ等のリサンプリング法等によって生体資料関連情報を関連性のある（有意差が認められる）遺伝子を推定する方法を挙げることができる。

【０１０４】

また、各遺伝子を生体内での機能（例えばアポトーシス関連遺伝子等）ごとに分類し、前記生体内の機能と各診療関連情報又は各治療関連情報等との関連性を求めてもよい。このような関連は、Ｇｅｎｅ Ｓｅｔ Ｅｎｒｉｃｈｍｅｎｔ Ａｎａｌｙｓｉｓ等により求めることができる。あるいは、超幾何分布等により、生体試料関連情報との関連性が強い遺伝子群を選出した後で、各遺伝子を生体内機能に基づいて分類された遺伝子群との重なり度合を指標に各遺伝子と生体試料関連情報との関連性を求めることができる。

【０１０５】

さらに、例えば家族歴の有無等の診療関連情報、又は疾患の予後が良好であるか否か等の治療関連情報と遺伝子関連測定データの関連の強さに基づいて、新規マーカーの候補を探索してもよい。このような探索は、取得された前記遺伝子関連測定データと生体試料関連情報との関連性を示す数値を使って、回帰分析、分散分析、主成分分析等の統計学的処理により；又は階層的クラスタリング、ｋ－ｍｅａｎｓ、ｍｅａｎ－ｓｈｉｆｔ等のクラスタ解析により、数理モデルを求め、得られた数理モデルを、前記数値の一部を使って検証（バリデーション）し、その検証データから生体試料関連情報と関連性の強い複数の遺伝子を決定することができる。

【０１０６】

本実施形態において、新規マーカー探索装置８０の処理部８１が行う各ステップ（ステップＳ３１からステップＳ３６）を行う場合には、コンピュータプログラムによって実行される。前記コンピュータプログラムは、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等の記憶媒体に記憶されていてもよい。前記記憶媒体へのプログラムの記憶形式は、前記提示装置が前記プログラムを読み取り可能である限り制限されない。前記記憶媒体への記憶は、不揮発性であることが好ましい。

【０１０７】

［４．新規マーカーの候補を探索装置］

図１７に記載の新規マーカー探索装置８０は、ハードウェア構成の一例である。ハードウェアは、パーソナルコンピュータ、タブレット型端末でありうる。

【０１０８】

新規マーカー探索装置８０は、処理部（ＣＰＵ）８１、主記憶部８２、ＲＯＭ８３；補助記憶部８４；通信Ｉ／Ｆ８５；入力Ｉ／Ｆ８６；出力Ｉ／Ｆ８７；メディアＩ／Ｆ８８；バス８９を備える。また、新規マーカー探索装置８０は、入力部９０、及び表示部９１を備える。また、新規マーカー探索装置８０は、記憶媒体９２を備えていてもよい。各構成の説明は、［２－１．ハードウェアの構成］の記載をここに援用する。

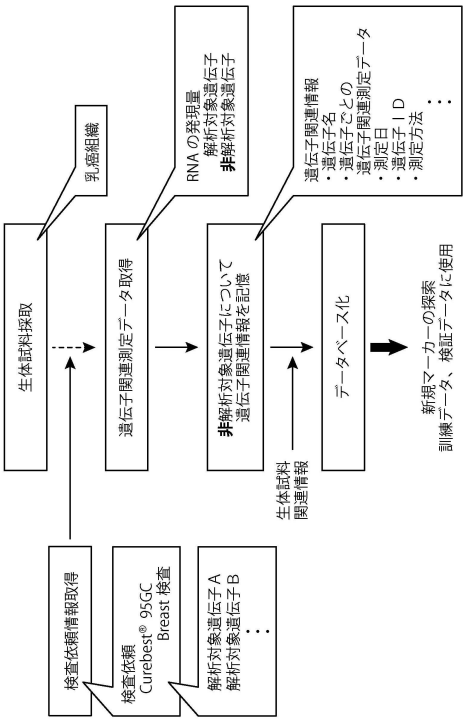
【符号の説明】

【０１０９】

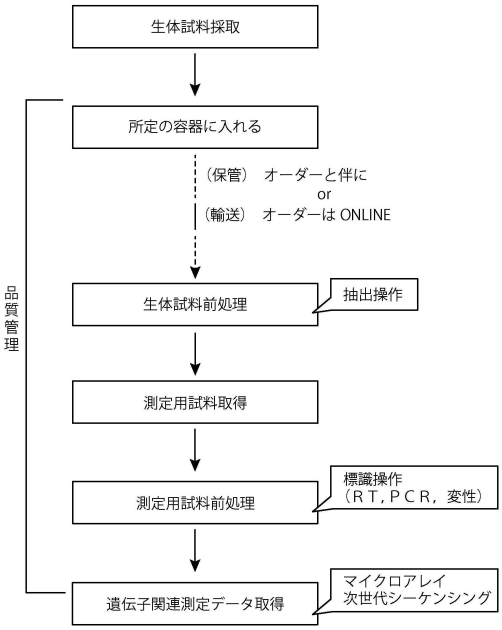
２０　検査機関情報処理装置
５０　医療機関情報処理装置
１００　第１のデータベース記憶装置
１０１　第２のデータベース記憶装置
１０２　第３のデータベース記憶装置
５００，６００，７００　システム

【図面】

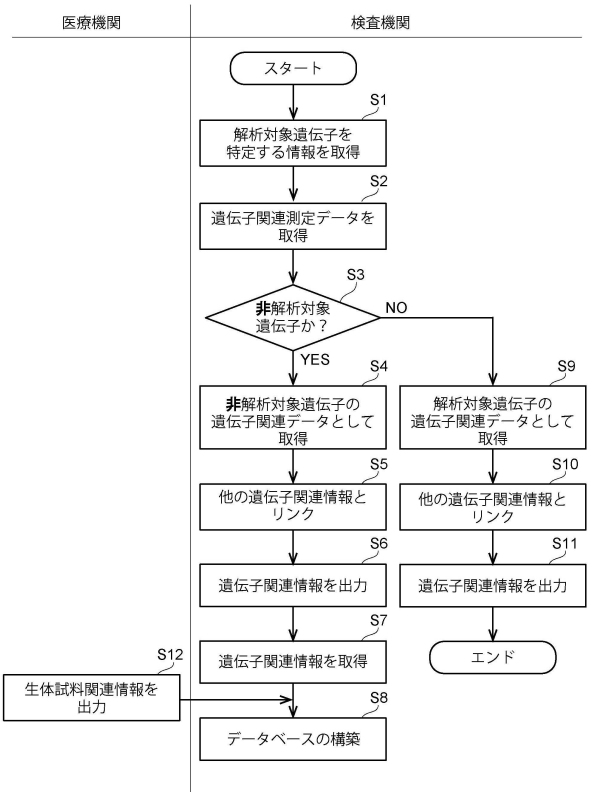
【図１】



【図２】



【図３】



【図４】

遺伝子番号	プローブセット.ID	遺伝子記号	UniGene.ID	GenBank アクセッション番号	重み係数
1	219306_at	KIF15	Hs.646856	NM_020242	0.5960
2	218585_s_at	DTL	Hs.656473	NM_016448	1.4097
3	221677_s_at	DONSON	Hs.436341	AF232674	0.4683
4	201088_at	KPNA2	Hs.594238	NM_002266	1.0480
5	209034_at	PNRC1	Hs.75969	AF279899	−1.4059
6	202610_s_at	MED14	Hs.407604	AF135802	−0.0393
7	218906_x_at	KLC2	Hs.280792	NM_022822	0.0880
8	212723_at	JMJD6	Hs.514505	AK021780	0.3141
9	222231_s_at	LRRC59	Hs.370927	AK025328	0.6264
10	208838_at	CAND1	Hs.546407	AB020636	0.2207
11	218039_at	NUSAP1	Hs.615092	NM_016359	1.5846
12	209472_at	CCBL2	Hs.481898	BC000819	−1.7235
13	212898_at	KIAA0406	Hs.655481	AB007866	0.1172
14	202620_s_at	PLOD2	Hs.477866	NM_000935	1.3605
15	201059_at	CTTN	Hs.596164	NM_005231	0.3591
16	201841_s_at	HSPB1	Hs.520973	NM_001540	1.2420
17	203755_at	BUB1B	Hs.631699	NM_001211	0.9909
18	211750_x_at	TUBA1C	Hs.719091	BC005946	0.0145
19	38158_at	ESPL1	Hs.153479	D79987	0.5325
20	204709_s_at	KIF23	Hs.270845	NM_004856	0.0798
21	201589_at	SMC1A	Hs.211602	D80000	0.3106
22	218460_at	HEATR2	Hs.535896	NM_017802	0.0198
23	207430_s_at	MSMB	Hs.255462	NM_002443	1.9177
24	212139_at	GCN1L1	Hs.298716	D86973	−0.0501
25	211596_s_at	LRIG1	Hs.518055	AB050468	−2.0999
26	212160_at	XPOT	Hs.85951	AI984005	0.3461
27	219238_at	PIGV	Hs.259605	NM_017837	−1.2689
28	203432_at	TMPO	Hs.11355	AW272611	0.4665
29	201377_at	UBAP2L	Hs.490551	NM_014847	0.1269
30	218875_s_at	FBXO5	Hs.520506	NM_012177	0.1012
31	221922_at	GPSM2	Hs.584901	AW195581	0.4423
32	218727_at	SLC38A7	Hs.10499	NM_018231	−0.0411
33	207469_s_at	PIR	Hs.495728	NM_003662	0.8827
34	218483_s_at	C11orf60	Hs.533738	NM_020153	−1.3198
35	204641_at	NEK2	Hs.153704	NM_002497	1.5825
36	219502_at	NEIL3	Hs.405467	NM_018248	−0.1883
37	209054_s_at	WHSC1	Hs.113876	AF083389	0.0465
38	220318_at	EPN3	Hs.670090	NM_017957	0.3073
39	210297_s_at	MSMB	Hs.255462	U22178	1.6681
40	209186_at	ATP2A2	Hs.506759	M23114	0.2014
41	219787_s_at	ECT2	Hs.518299	NM_018098	0.8181
42	45633_at	GINS3	Hs.47125	AI421812	−0.2363
43	200848_at	AHCYL1	Hs.705418	AA479488	−1.5895
44	200822_x_at	TPI1	Hs.524219	NM_000365	0.0814
45	211072_x_at	TUBA1B	Hs.719075	BC006481	0.0380
46	200811_at	CIRBP	Hs.634522	NM_001280	−1.4620
47	202864_s_at	SP100	Hs.369056	NM_003113	−1.3947
48	202154_x_at	TUBB3	Hs.511743	NM_006086	0.1241
49	213152_s_at	SFRS2B	Hs.476680	AI343248	−1.2495
50	209368_at	EPHX2	Hs.212088	AF233336	−1.8835

【図５】

遺伝子番号	プローブセットID	遺伝子記号	UniGene.ID	GenBank アクセッション番号	重み係数
51	211058_x_at	TUBA1B	Hs.719075	BC006379	0.0646
52	209251_x_at	TUBA1C	Hs.719091	BC004949	0.0453
53	213646_x_at	TUBA1B	Hs.719075	BE300252	0.0396
54	204540_at	EEF1A2	Hs.433839	NM_001958	1.8487
55	202026_at	SDHD	Hs.719164	NM_003002	−1.3587
56	201090_x_at	TUBA1B	Hs.719075	NM_006082	0.0733
57	213119_at	SLC36A1	Hs.269004	AW058600	0.0680
58	217840_at	DDX41	Hs.484288	NM_016222	0.0313
59	206559_x_at	EEF1A1	---	NM_001403	−0.9727
60	202066_at	PPFIA1	Hs.530749	AA195259	0.7385
61	203108_at	GPRC5A	Hs.631733	NM_003979	1.0799
62	218697_at	NCKIPSD	Hs.655006	NM_016453	−0.0693
63	222039_at	KIF18B	Hs.135094	AA292789	0.6820
64	202069_s_at	IDH3A	Hs.591110	AI826060	0.2302
65	203362_s_at	MAD2L1	Hs.591697	NM_002358	0.8095
66	202666_s_at	ACTL6A	Hs.435326	NM_004301	0.2162
67	204892_x_at	EEF1A1	Hs.520703	NM_001402	−0.9566
68	205682_x_at	APOM	Hs.534468	NM_019101	−1.0558
69	209714_s_at	CDKN3	Hs.84113	AF213033	0.9594
70	218381_s_at	U2AF2	Hs.528007	NM_007279	−0.0076
71	201947_s_at	CCT2	Hs.189772	NM_006431	0.2632
72	212722_s_at	JMJD6	Hs.514505	AK021780	0.0968
73	204825_at	MELK	Hs.184339	NM_014791	1.1379
74	203184_at	FBN2	Hs.519294	NM_001999	0.7174
75	201266_at	TXNRD1	Hs.708065	NM_003330	0.2610
76	202969_at	DYRK2	Hs.173135	AI216690	0.2560
77	204817_at	ESPL1	Hs.153479	NM_012291	0.4866
78	209523_at	TAF2	Hs.122752	AK001618	0.3803
79	218491_s_at	THYN1	Hs.13645	NM_014174	−1.3652
80	217363_x_at	---	---	AL031313	−0.9838
81	218009_s_at	PRC1	Hs.567385	NM_003981	1.6691
82	204026_s_at	ZWINT	Hs.591363	NM_007057	0.9942
83	218355_at	KIF4A	Hs.648326	NM_012310	1.1017
84	202153_s_at	NUP62	Hs.574492	NM_016553	−0.0983
85	213011_s_at	TPI1	Hs.524219	BF116254	0.1005
86	217966_s_at	FAM129A	Hs.518662	NM_022083	−2.4459
87	214782_at	CTTN	Hs.596164	AU155105	0.2306
88	217967_s_at	FAM129A	Hs.518662	AF288391	−2.7067
89	204649_at	TROAP	Hs.524399	NM_005480	0.1495
90	35671_at	GTF3C1	Hs.371718	U02619	0.0169
91	213502_x_at	LOC91316	Hs.148656	AA398569	−2.1336
92	221285_at	ST8SIA2	Hs.302341	NM_006011	−0.9209
93	221519_at	FBXW4	Hs.500822	AF281859	−1.1897
94	202551_s_at	CRIM1	Hs.699247	BG546884	−2.0141
95	217138_x_at	IGL@	Hs.449585	AJ249377	−1.0505

【図６】

1

遺伝子関連測定データ	遺伝子名（GenBank アクセッション番号）	遺伝子を特定するための符号	その他の情報
０．９	○○○	1	・測定日 ・測定方法 ・生体試料を特定するための符号 ・生体試料の保存方法 ・生体試料の保存期間 ・測定試料の料 ・検査機関名
０．１	△△△	2	
０．５	□□□	3	
・・・	・・・	・・・	

【図７】

5

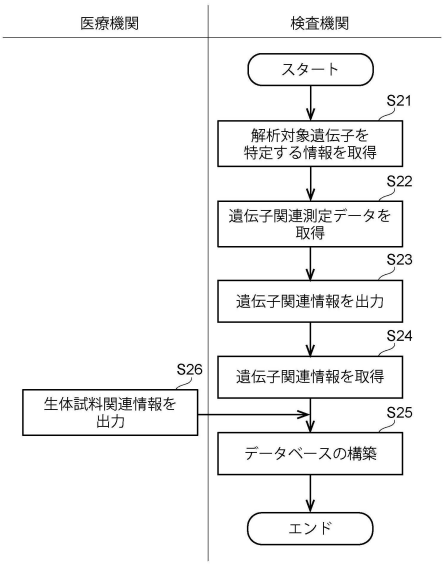
生体試料を特定するための符号	患者ID	生体試料の種類	診療関連情報	治療関連情報		
				治療	治療薬の投与 予防薬の投与	薬の名称 用量 投与頻度 投与日 投与期間
					放射線治療	照射線量／回 頻度 施術日 総放射線量
					外科的処置	切除部位 術式 周辺組織郭清 施術日
1	○○	乳癌組織	疾患名 疾患型名 疾患の病期 性別 年齢 既往歴 家族歴 再発履歴 問診情報 他の検査情報			
2	△△	・・・	・・・	・・・	・・・	・・・

【図８】

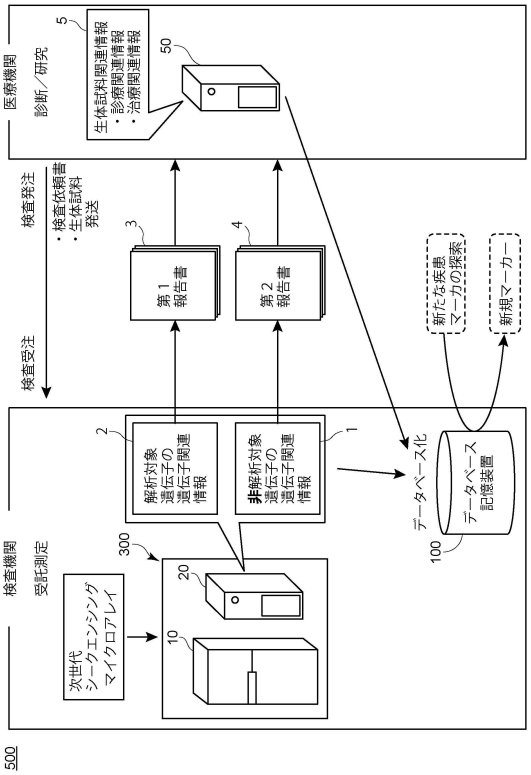
3，4

生体試料の符号：1				測定日：○年○月○日		検査機関の名称
測定方法 ○○マイクロアレイ				生体試料の保存方法		生体試料の保存期間
遺伝子の名称	遺伝子の符号	測定データ	遺伝子の名称	遺伝子の符号	測定データ	
○○○	209034_at	10.523	□□□	221519_at	8.173	95GC 判定 L
△△△	209472_at	9.222	××× ・・・・・	202551_at ・・・・	8.266 ・・・・・	癌細胞の有無 ５０％
・・・・・	・・・・・	・・・・・				

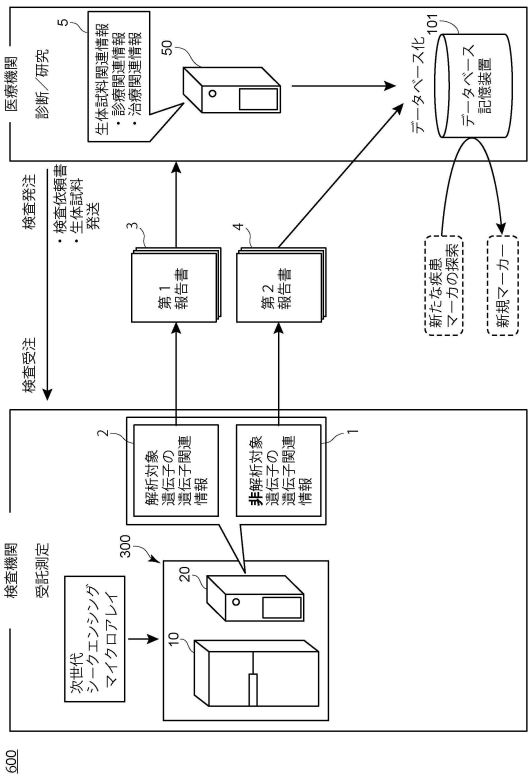
【図９】



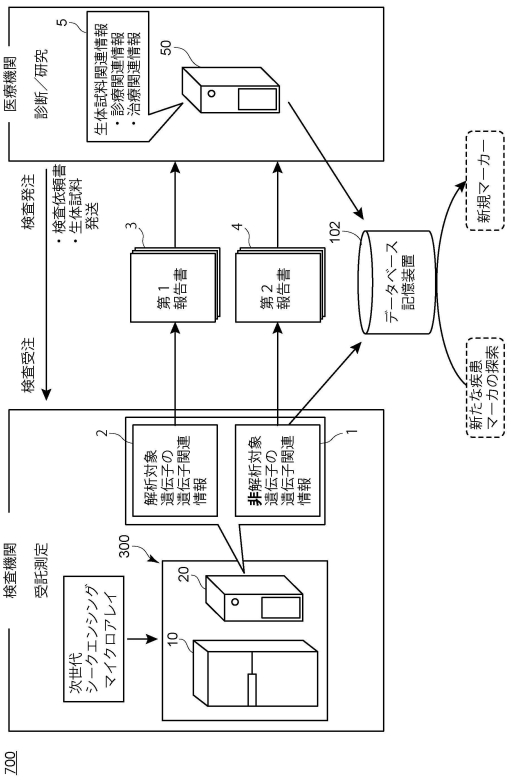
【図１０】



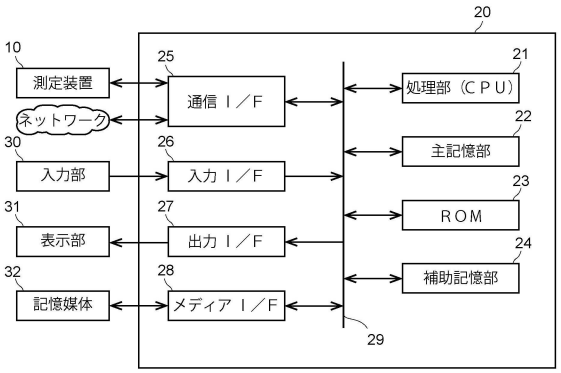
【図１１】



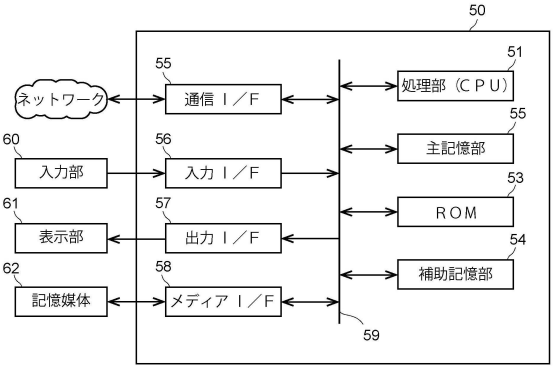
【図１２】



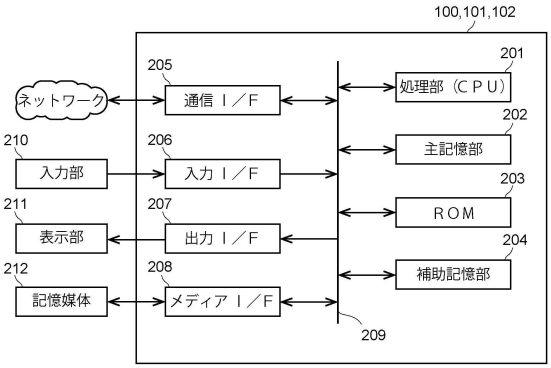
【図１３】



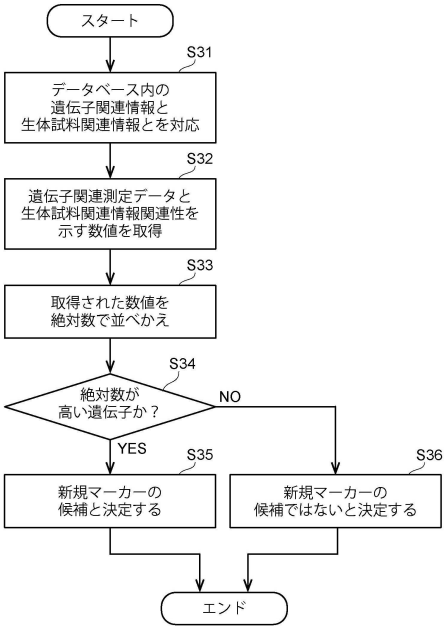
【図１４】



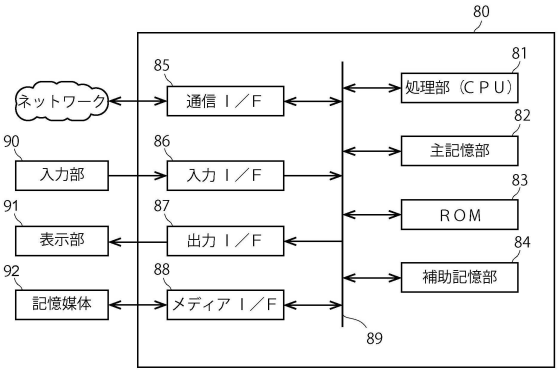
【図１５】



【図１６】



【図１７】


80
81
処理部（ＣＰＵ）
82
主記憶部
83
ＲＯＭ
84
補助記憶部
85
通信Ｉ／Ｆ
86
入力Ｉ／Ｆ
87
出力Ｉ／Ｆ
88
メディアＩ／Ｆ
89
ネットワーク
90
入力部
91
表示部
92
記憶媒体

フロントページの続き

－２）　５．発行日　２０１７年６月２２日　刊行物　第２４回アジア太平洋癌学会　ＡＰＣＣ　２０１７　Ａｂｓｔｒａｃｔ　Ｂｏｏｋ　６．開催日　２０１７年６月２２日～２０１７年６月２４日　集会名、開催場所　第２４回アジア太平洋癌学会（ＡＰＣＣ２０１７）ソウルＣＯＥＸ展示コンベンションセンター（住所：大韓民国　ソウル特別市　江南区　三成洞１５９永東大路５１３）　７．発行日　２０１７年４月１日　刊行物　日本臨床７５巻　増刊号３（通巻１１１９号）　８．ウェブサイトの掲載日　２０１７年６月２９日　ウェブサイトのアドレス　ｈｔｔｐ：／／ｗｗｗ２．ｃｏｎｖｅｎｔｉｏｎ．ｃｏ．ｊｐ／２５ｊｂｃｓ／　ｈｔｔｐ：／／ｗｗｗ２．ｃｏｎｖｅｎｔｉｏｎ．ｃｏ．ｊｐ／２５ｊｂｃｓ／ｓｙｏｒｏｋｕ．ｈｔｍｌ　ｈｔｔｐ：／／ｗｗｗ２．ｃｏｎｖｅｎｔｉｏｎ．ｃｏ．ｊｐ／２５ｊｂｃｓ／ｄｌ／Ｐｒｏｇｒａｍ_Ｓｙｏｒｏｋｕｓｙｕ．ｐｄｆ

(72)発明者　野口　眞三郎
大阪府吹田市山田丘１番１号　国立大学法人大阪大学内
(72)発明者　直居　靖人
大阪府吹田市山田丘１番１号　国立大学法人大阪大学内

審査官　田中　秀樹
(56)参考文献　特開２０１０－２１８１５０（ＪＰ，Ａ）
特開２０１１－２２３９５７（ＪＰ，Ａ）
特開２００４－１８７５６２（ＪＰ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１６Ｂ　５／００－９９／００