(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105190636 A
(43)申请公布日 2015.12.23

(21)申请号 201480018404.1
(22)申请日 2014.02.21
(30)优先权数据
2013-067770 2013.03.28 JP
(85)PCT国际申请进入国家阶段日
2015.09.25
(86)PCT国际申请的申请数据
PCT/JP2014/054145 2014.02.21
(87)PCT国际申请的公布数据
WO2014/156400 JA 2014.10.02
(71)申请人 三菱宇宙软件株式会社
地址 日本东京都
(72)发明人 谷岛成树 松田规
(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 李辉 龚晓娟
(51)Int.Cl.
G06F 19/22(2006.01)
G06F 17/30(2006.01)
G09C 1/00(2006.01)

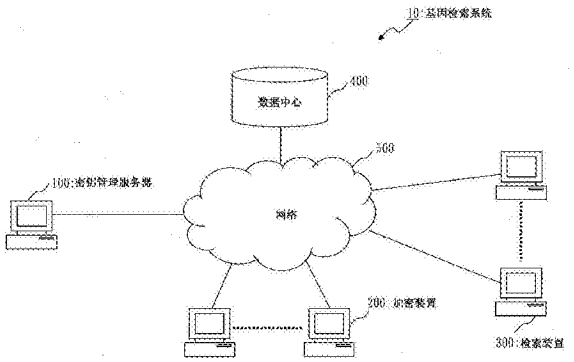
权利要求书5页 说明书20页 附图20页

(54)发明名称

基因信息存储装置、基因信息检索装置、基因信息存储程序、基因信息检索程序、基因信息存储方法、基因信息检索方法以及基因信息检索系统

(57)摘要

本发明的目的在于能够在加密的状态下进行基因信息的检索。加密装置(200)对存储装置存储的基因信息即对象基因进行加密来生成加密基因，并将作为预定的基因信息的基准基因与对象基因进行比较，生成差异信息，生成嵌入了所生成的差异信息并加密而得到的加密标签。数据中心(400)使加密基因和加密标签相关联地存储在存储装置中。检索装置(300)将差异信息作为检索关键字，生成嵌入差异信息并进行加密而得到的检索查询，将所生成的检索查询发送到数据中心(400)。数据中心(400)确定包含检索查询中指定的差异信息的加密标签，提取相关联的加密基因并发送到检索装置(300)。



CN 105190636 A

1. 一种基因信息存储装置，其使存储装置存储基因信息，其特征在于，所述基因信息存储装置具有：

基准基因取得部，其取得作为预定的基因信息的基准基因；

基因输入部，其输入使所述存储装置存储的基因信息即对象基因；

差异生成部，其将所述基准基因取得部取得的基准基因与所述基因输入部输入的对象基因进行比较，生成差异信息；

数据加密部，其对所述对象基因进行加密，生成加密基因；

加密标签生成部，其生成嵌入了所述差异生成部生成的差异信息的加密标签；以及

数据存储部，其使所述存储装置相关联地存储所述数据加密部生成的加密基因和所述加密标签生成部生成的加密标签。

2. 根据权利要求 1 所述的基因信息存储装置，其特征在于，

所述差异信息包含多个类别的信息，

所述基因信息存储装置还具有：

差异信息置换部，其针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将所述差异生成部生成的差异信息中的该类别的值置换为识别该值所属的块的识别信息，

所述加密标签生成部对所述差异信息置换部置换后的差异信息进行加密，生成加密标签。

3. 根据权利要求 2 所述的基因信息存储装置，其特征在于，

所述差异信息置换部以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将所述差异生成部生成的差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

4. 根据权利要求 1 至 3 中任一项所述的基因信息存储装置，其特征在于，

所述加密标签生成部根据当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥解密所述加密数据的加密方式，将能够检索所述加密基因的用户的属性信息和差异信息设定为所述加密属性，对随机数值进行加密，生成所述加密标签。

5. 根据权利要求 4 所述的基因信息存储装置，其特征在于，

根据所述加密方式，将能够解密所述加密基因的用户的属性信息设定为所述加密属性，对所述对象基因进行加密，生成所述加密基因。

6. 一种基因信息检索装置，其对存储在数据管理装置所管理的存储装置内的基因信息进行检索，其特征在于，所述基因信息检索装置具有：

差异信息输入部，其输入要检索的基因信息与作为预定的基因信息的基准基因之间的差异信息；

检索查询生成部，其生成嵌入了所述差异信息输入部输入的差异信息的检索查询；以及

基因信息取得部，其将所述检索查询生成部生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

7. 根据权利要求 6 所述的基因信息检索装置，其特征在于，

所述差异信息包含多个类别的信息，

所述基因信息检索装置还具有：

差异信息置换部，其针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的块的识别信息，

所述检索查询生成部生成嵌入了所述差异信息置换部置换后的差异信息的检索查询。

8. 根据权利要求 7 所述的基因信息检索装置，其特征在于，

所述差异信息置换部以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

9. 根据权利要求 6 至 8 中任一项所述的基因信息检索装置，其特征在于，所述基因信息检索装置还具有：

用户密钥管理部，其管理密钥，该密钥是当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥对所述加密数据进行解密的加密方式中的密钥，是设定了用户的属性信息作为所述密钥属性的密钥，

所述检索查询生成部追加所述差异信息作为所述用户密钥管理部管理的密钥的密钥属性，来生成所述检索查询。

10. 根据权利要求 9 所述的基因信息检索装置，其特征在于，所述基因信息取得部取得如下的加密基因作为包含所述差异信息的基因信息：该加密基因是通过所述加密方式将能够进行解密的用户的属性信息设定为所述加密属性并对所述基因信息进行加密而得到的，

所述基因信息检索装置还具有解密部，该解密部利用所述用户密钥管理部管理的密钥对所述加密基因进行解密。

11. 一种基因信息存储程序，其使存储装置存储基因信息，其特征在于，所述基因信息存储程序使计算机执行：

基准基因取得处理，取得作为预定的基因信息的基准基因；

基因输入处理，输入使所述存储装置存储的基因信息即对象基因；

差异生成处理，将通过所述基准基因取得处理取得的基准基因与通过所述基因输入处理输入的对象基因进行比较，生成差异信息；

数据加密处理，对所述对象基因进行加密，生成加密基因；

加密标签生成处理，生成嵌入了通过所述差异生成处理生成的差异信息的加密标签；以及

数据存储处理，使所述存储装置相关联地存储通过所述数据加密处理生成的加密基因和通过所述加密标签生成处理生成的加密标签。

12. 根据权利要求 11 所述的基因信息存储程序，其特征在于，

所述差异信息包含多个类别的信息，

所述基因信息存储程序还使计算机执行：

差异信息置换处理，针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将通过所述差异生成处理生成的差异信息中的该类别的值置换为识别该值所属的块的识别信息，

在所述加密标签生成处理中，对通过所述差异信息置换处理置换后的差异信息进行加

密，生成加密标签。

13. 根据权利要求 12 所述的基因信息存储程序，其特征在于，

在所述差异信息置换处理中，以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将通过所述差异生成处理生成的差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

14. 根据权利要求 11 至 13 中任一项所述的基因信息存储程序，其特征在于，在所述加密标签生成处理中，根据当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥对所述加密数据进行解密的加密方式，将能够检索所述加密基因的用户的属性信息和差异信息设定为所述加密属性，对随机数值进行加密，生成所述加密标签。

15. 根据权利要求 14 所述的基因信息存储程序，其特征在于，根据所述加密方式，将能够解密所述加密基因的用户的属性信息设定为所述加密属性，对所述对象基因进行加密，生成所述加密基因。

16. 一种基因信息检索程序，其检索存储在数据管理装置管理的存储装置内的基因信息，其特征在于，所述基因信息检索程序使计算机执行 ：

差异信息输入处理，输入要检索的基因信息与作为预定的基因信息的基准基因之间的差异信息 ；

检索查询生成处理，生成嵌入了通过所述差异信息输入处理输入的差异信息的检索查询 ；以及

基因信息取得处理，将通过所述检索查询生成处理生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

17. 根据权利要求 16 所述的基因信息检索程序，其特征在于，

所述差异信息包含多个类别的信息，

所述基因信息检索程序还使计算机执行 ：

差异信息置换处理，针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的块的识别信息，

在所述检索查询生成处理中，生成嵌入了通过所述差异信息置换处理置换后的差异信息的检索查询。

18. 根据权利要求 17 所述的基因信息检索程序，其特征在于，

在所述差异信息置换处理中，以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

19. 根据权利要求 16 至 18 中任一项所述的基因信息检索程序，其特征在于，所述基因信息检索程序还使计算机执行 ：用户密钥管理处理，管理当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥对所述加密数据进行解密的加密方式中的所述密钥，该密钥是设定了用户的属性信息作为所述密钥属性的密钥，

在所述检索查询生成处理中，追加所述差异信息作为通过所述用户密钥管理处理管理的密钥的密钥属性，生成所述检索查询。

20. 根据权利要求 19 所述的基因信息检索程序，其特征在于，

在所述基因信息取得处理中，通过所述加密方式，将能够进行解密的用户的属性信息设定为所述加密属性，取得对所述基因信息进行加密而得到的加密基因作为包含所述差异信息的基因信息，

所述基因信息检索程序还使计算机执行解密处理，在该解密处理中，利用通过所述用户密钥管理处理管理的密钥对所述加密基因进行解密。

21. 一种基因信息存储方法，使存储装置存储基因信息，其特征在于，所述基因信息存储方法具有：

基准基因取得步骤，处理装置取得作为预定的基因信息的基准基因；

基因输入步骤，输入装置输入使所述存储装置存储的基因信息即对象基因；

差异生成步骤，处理装置将通过所述基准基因取得步骤取得的基准基因与通过所述基因输入步骤输入的对象基因进行比较，生成差异信息；

数据加密步骤，处理装置对所述对象基因进行加密，生成加密基因；

加密标签生成步骤，处理装置生成嵌入了通过所述差异生成步骤生成的差异信息的加密标签；以及

数据存储步骤，处理装置使所述存储装置相关联地存储通过所述数据加密步骤生成的加密基因和通过所述加密标签生成步骤生成的加密标签。

22. 一种基因信息检索方法，检索存储在数据管理装置管理的存储装置内的基因信息，其特征在于，所述基因信息检索方法具有：

差异信息输入步骤，输入装置输入要检索的基因信息与作为预定的基因信息的基准基因之间的差异信息；

检索查询生成步骤，处理装置生成嵌入了通过所述差异信息输入步骤输入的差异信息的检索查询；以及

基因信息取得步骤，处理装置将通过所述检索查询生成步骤生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

23. 一种基因信息检索系统，其具有基因信息存储装置和基因信息检索装置，该基因信息存储装置使数据管理装置管理的存储装置存储基因信息，该基因信息检索装置从所述基因信息存储装置存储的基因信息中检索包含检索关键字的基因信息，其特征在于，

所述基因信息存储装置具有：

基准基因取得部，其取得作为预定的基因信息的基准基因；

基因输入部，其输入使所述存储装置存储的基因信息即对象基因；

差异生成部，其将所述基准基因取得部取得的基准基因与所述基因输入部输入的对象基因进行比较，生成差异信息；

数据加密部，其对所述对象基因进行加密，生成加密基因；

加密标签生成部，其生成嵌入了所述差异生成部生成的差异信息的加密标签；以及

数据存储部，其使所述存储装置相关联地存储所述数据加密部生成的加密基因和所述加密标签生成部生成的加密标签，

所述基因信息检索装置具有：

差异信息输入部，其输入要检索的基因信息与所述基准基因之间的差异信息作为所述检索关键字；

检索查询生成部，其生成嵌入了所述差异信息输入部输入的差异信息的检索查询；以及

基因信息取得部，其将所述检索查询生成部生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

基因信息存储装置、基因信息检索装置、基因信息存储程序、基因信息检索程序、基因信息存储方法、基因信息检索方法以及基因信息检索系统

技术领域

[0001] 本发明涉及对作为DNA序列的分析结果而得到的基因组和/或作为基因的分析信息的基因信息进行加密并存储在存储装置内的存储技术。并且，本发明涉及在加密的状态下对利用所述存储技术存储的基因信息进行检索的技术。

背景技术

[0002] 近年来，生命工程学取得了显著的进展，基因组序列的读取能力也取得进展。因此，解读一个人的全部基因组序列的成本降低，从而许多人的基因组序列可以被解读。

[0003] 包含个人的基因组序列和其发现信息等关联信息的基因信息被称为终极的个人信息，对本人以外的信息公开必须安全地进行。必须使从基因组DNA和RNA序列的读取装置输出经分析而成为有意义的信息的基因信息迅速被加密，没有使用者的同意不得被解读。

[0004] 存在如下这样的隐匿检索方式：能够在对蓄积于数据库内的数据和检索关键字进行了加密的状态下，从蓄积于数据库内的数据中检索包含检索关键字的数据（参照非专利文献1）。

[0005] 在该隐匿检索方式中，在将数据蓄积于数据库内时，将被设想为在检索该数据的情况下使用的检索关键字作为标签来提取。然后，对数据和标签分别进行加密，将加密后的标签附加给加密后的数据并蓄积在数据库内。

[0006] 在从蓄积于数据库内的数据中检索包含检索关键字的数据的情况下，输入加密后的检索关键字。然后，检索与加密后的检索关键字对应的加密后的标签。在发现了与检索关键字对应的标签的情况下，将附加有该标签的数据确定为是包含检索关键字的数据。

[0007] 在先技术文献

[0008] 非专利文献

[0009] 非专利文献1：Boneh, Di Crescenzo, Ostrovski, and Persiano “Public key encryption with keyword search” EUROCRYPT 2004, PP506-522

[0010] 非专利文献2：Allison Lewko, Tatsuaki Okamoto, Amit Sahai, Katsuyuki Takashima, Brent Waters, “Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption”, EUROCRYPT2010, Lecture Notes In Computer Science, 2010, Volume 6110/2010.

发明内容

[0011] 发明所要解决的课题

[0012] 迄今为止，在检索已加密的基因信息的情况下，在执行检索时，暂时对基因进行解密来检索。

[0013] 即使在使用隐匿检索方式的情况下，也将使用平文生成的检索索引赋予给已加密

的基因信息，使用该检索索引来检索。并且，对 1 个基因信息赋予庞大数量的检索索引，检索花费大量的时间。

[0014] 因此，若是蓄积基因信息的计算机的特权用户，则可以在被解密的时间中自由访问数据或者检索索引。因此，不能说基因信息是完全对第三者隐蔽的状态。

[0015] 本发明的目的是能够在使基因信息对第三者隐蔽的状态下进行基因信息的检索。

[0016] 用于解决课题的手段

[0017] 本发明的基因信息存储装置使存储装置存储基因信息，其特征在于，该基因信息存储装置具有：基准基因取得部，其取得作为预定的基因信息的基准基因；基因输入部，其输入使所述存储装置存储的基因信息即对象基因；差异生成部，其将所述基准基因取得部取得的基准基因与所述基因输入部输入的对象基因进行比较，生成差异信息；数据加密部，其对所述对象基因进行加密，生成加密基因；加密标签生成部，其生成嵌入了所述差异生成部生成的差异信息的加密标签；以及数据存储部，其使所述存储装置相关联地存储所述数据加密部生成的加密基因和所述加密标签生成部生成的加密标签。

[0018] 该基因信息存储装置的特征在于，所述差异信息包含多个类别的信息，所述基因信息存储装置还具有：差异信息置换部，其针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将所述差异生成部生成的差异信息中的该类别的值置换为识别该值所属的块的识别信息，所述加密标签生成部对所述差异信息置换部置换后的差异信息进行加密，生成加密标签。

[0019] 该基因信息存储装置的特征在于，所述差异信息置换部以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将所述差异生成部生成的差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

[0020] 该基因信息存储装置的特征在于，所述加密标签生成部根据当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥解密所述加密数据的加密方式，将能够检索所述加密基因的用户的属性信息和差异信息设定为所述加密属性，对随机数值进行加密，生成所述加密标签。

[0021] 该基因信息存储装置的特征在于，根据所述加密方式，将能够解密所述加密基因的用户的属性信息设定为所述加密属性，对所述对象基因进行加密，生成所述加密基因。

[0022] 本发明的基因信息检索装置对存储在数据管理装置所管理的存储装置内的基因信息进行检索，其特征在于，所述基因信息检索装置具有：差异信息输入部，其输入要检索的基因信息与作为预定的基因信息的基准基因之间的差异信息；检索查询生成部，其生成嵌入了所述差异信息输入部输入的差异信息的检索查询；以及基因信息取得部，其将所述检索查询生成部生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

[0023] 该基因信息检索装置的特征在于，所述差异信息包含多个类别的信息，所述基因信息检索装置还具有：差异信息置换部，其针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的块的识别信息，所述检索查询生成部生成嵌入了所述差异信息置换部置换后的差异信息的检索查询。

[0024] 该基因信息检索装置的特征在于，所述差异信息置换部以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将差异信息中的该类别的值

置换为识别该值所属的各块的识别信息。

[0025] 该基因信息检索装置的特征在于，所述基因信息检索装置还具有：用户密钥管理部，其管理当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥对所述加密数据进行解密的加密方式中的所述密钥，即用户的属性信息被设定为所述密钥属性的密钥，所述检索查询生成部追加所述差异信息作为所述用户密钥管理部管理的密钥的密钥属性，来生成所述检索查询。

[0026] 该基因信息检索装置的特征在于，所述基因信息取得部取得如下的加密基因作为包含所述差异信息的基因信息：该加密基因是通过所述加密方式将能够进行解密的用户的属性信息设定为所述加密属性并对所述基因信息进行加密而得到的，所述基因信息检索装置还具有解密部，该解密部利用所述用户密钥管理部管理的密钥对所述加密基因进行解密。

[0027] 本发明的基因信息存储程序使存储装置存储基因信息，其特征在于，所述基因信息存储程序使计算机执行：基准基因取得处理，取得作为预定的基因信息的基准基因；基因输入处理，输入使所述存储装置存储的基因信息即对象基因；差异生成处理，将通过所述基准基因取得处理取得的基准基因与通过所述基因输入处理输入的对象基因进行比较，生成差异信息；数据加密处理，对所述对象基因进行加密，生成加密基因；加密标签生成处理，生成嵌入了通过所述差异生成处理生成的差异信息的加密标签；以及数据存储处理，使所述存储装置相关联地存储通过所述数据加密处理生成的加密基因和通过所述加密标签生成处理生成的加密标签。

[0028] 该基因信息存储程序的特征在于，所述差异信息包含多个类别的信息，所述基因信息存储程序还使计算机执行：差异信息置换处理，针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将通过所述差异生成处理生成的差异信息中的该类别的值置换为识别该值所属的块的识别信息，在所述加密标签生成处理中，对通过所述差异信息置换处理置换后的差异信息进行加密，生成加密标签。

[0029] 该基因信息存储程序的特征在于，在所述差异信息置换处理中，以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将通过所述差异生成处理生成的差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

[0030] 该基因信息存储程序的特征在于，在所述加密标签生成处理中，根据当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥对所述加密数据进行解密的加密方式，将能够检索所述加密基因的用户的属性信息和差异信息设定为所述加密属性，对随机数值进行加密，生成所述加密标签。

[0031] 该基因信息存储程序的特征在于，根据所述加密方式，将能够解密所述加密基因的用户的属性信息设定为所述加密属性，对所述对象基因进行加密，生成所述加密基因。

[0032] 本发明的基因信息检索程序检索存储在数据管理装置管理的存储装置内的基因信息，其特征在于，所述基因信息检索程序使计算机执行：差异信息输入处理，输入要检索的基因信息与作为预定的基因信息的基准基因之间的差异信息；检索查询生成处理，生成嵌入了通过所述差异信息输入处理输入的差异信息的检索查询；以及基因信息取得处理，将通过所述检索查询生成处理生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

[0033] 该基因信息检索程序的特征在于，所述差异信息包含多个类别的信息，所述基因信息检索程序还使计算机执行：差异信息置换处理，针对所述差异信息包含的预定类别，将该类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的块的识别信息，在所述检索查询生成处理中，生成嵌入了通过所述差异信息置换处理置换后的差异信息的检索查询。

[0034] 该基因信息检索程序的特征在于，在所述差异信息置换处理中，以使属于各块的值的一部分也属于其它块的方式，将所述类别可取的值划分为多个块，将差异信息中的该类别的值置换为识别该值所属的各块的识别信息。

[0035] 该基因信息检索程序的特征在于，所述基因信息检索程序还使计算机执行：用户密钥管理处理，管理当加密数据中设定的加密属性与密钥中设定的密钥属性不对应的情况下无法使用所述密钥对所述加密数据进行解密的加密方式中的所述密钥，即用户的属性信息被设定为所述密钥属性的密钥，在所述检索查询生成处理中，追加所述差异信息作为通过所述用户密钥管理处理管理的密钥的密钥属性，生成所述检索查询。

[0036] 该基因信息检索程序的特征在于，在所述基因信息取得处理中，通过所述加密方式，将能够进行解密的用户的属性信息设定为所述加密属性，取得对所述基因信息进行加密而得到的加密基因作为包含所述差异信息的基因信息，所述基因信息检索程序还使计算机执行解密处理，在该解密处理中，利用通过所述用户密钥管理处理管理的密钥对所述加密基因进行解密。

[0037] 本发明的基因信息存储方法，使存储装置存储基因信息，其特征在于，所述基因信息存储方法具有：基准基因取得步骤，处理装置取得作为预定的基因信息的基准基因；基因输入步骤，输入装置输入使所述存储装置存储的基因信息即对象基因；差异生成步骤，处理装置将通过所述基准基因取得步骤取得的基准基因与通过所述基因输入步骤输入的对象基因进行比较，生成差异信息；数据加密步骤，处理装置对所述对象基因进行加密，生成加密基因；加密标签生成步骤，处理装置生成嵌入了通过所述差异生成步骤生成的差异信息的加密标签；以及数据存储步骤，处理装置使所述存储装置相关联地存储通过所述数据加密步骤生成的加密基因和通过所述加密标签生成步骤生成的加密标签。

[0038] 本发明的基因信息检索方法，检索存储在数据管理装置管理的存储装置内的基因信息，其特征在于，所述基因信息检索方法具有：差异信息输入步骤，输入装置输入要检索的基因信息与作为预定的基因信息的基准基因之间的差异信息；检索查询生成步骤，处理装置生成嵌入了通过所述差异信息输入步骤输入的差异信息的检索查询；以及基因信息取得步骤，处理装置将通过所述检索查询生成步骤生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

[0039] 本发明的基因信息检索系统具有基因信息存储装置和基因信息检索装置，该基因信息存储装置使数据管理装置管理的存储装置存储基因信息，该基因信息检索装置从所述基因信息存储装置存储的基因信息中检索包含检索关键字的基因信息，其特征在于，所述基因信息存储装置具有：基准基因取得部，其取得作为预定的基因信息的基准基因；基因输入部，其输入使所述存储装置存储的基因信息即对象基因；差异生成部，其将所述基准基因取得部取得的基准基因与所述基因输入部输入的对象基因进行比较，生成差异信息；数据加密部，其对所述对象基因进行加密，生成加密基因；加密标签生成部，其生成嵌入了所

述差异生成部生成的差异信息的加密标签；以及数据存储部，其使所述存储装置相关联地存储所述数据加密部生成的加密基因和所述加密标签生成部生成的加密标签，所述基因信息检索装置具有：差异信息输入部，其输入要检索的基因信息与所述基准基因之间的差异信息作为所述检索关键字；检索查询生成部，其生成嵌入了所述差异信息输入部输入的差异信息的检索查询；以及基因信息取得部，其将所述检索查询生成部生成的检索查询发送到所述数据管理装置，取得包含所述差异信息的基因信息。

[0040] 发明效果

[0041] 根据本发明，在蓄积于数据库中的基因信息和检索索引即标签、与被用作检索关键字的基因信息中的任意一方都进行了加密的状态下，可以从数据库中提取包含检索关键字的基因信息。因此，基因信息处于完全对第三者隐蔽的状态。

[0042] 特别是，根据本发明，将与基准基因的差异信息作为检索索引和检索关键字。因此，检索索引的数量较少即可，可以高速进行检索。

附图说明

[0043] 图 1 是基因检索系统 10 的结构图。

[0044] 图 2 是密钥管理服务器 100 的结构图。

[0045] 图 3 是加密装置 200 的结构图。

[0046] 图 4 是检索装置 300 的结构图。

[0047] 图 5 是数据中心 400 的结构图。

[0048] 图 6 是使用内积基分级谓词加密 (hierarchical predicate encryption for inner products) 的加密方式的说明图。

[0049] 图 7 是使用内积基分级谓词加密的隐匿检索方式的说明图。

[0050] 图 8 是示出初始设定处理流程的流程图。

[0051] 图 9 是已编码的 SNV 信息的结构图。

[0052] 图 10 是 SNV 信息中的位置信息的分块化的说明图。

[0053] 图 11 是 SNV 信息中的可靠性的分块化的说明图。

[0054] 图 12 是已编码的 SV 信息的结构图。

[0055] 图 13 是 SV 信息中的 CNV 增益的分块化的说明图。

[0056] 图 14 是已编码的 NC 信息的结构图。

[0057] 图 15 是标签 ID 的分层结构的说明图。

[0058] 图 16 是解密者 ID 的分层结构的说明图。

[0059] 图 17 是用户 ID 信息数据库的说明图。

[0060] 图 18 是示出用户密钥发布处理流程的流程图。

[0061] 图 19 是示出患者基因组序列的加密处理流程的流程图。

[0062] 图 20 是带标签加密数据的说明图。

[0063] 图 21 是示出加密数据的存储示例的图。

[0064] 图 22 是示出患者的电子病历的加密处理流程的流程图。

[0065] 图 23 是示出检索处理流程的流程图。

[0066] 图 24 是访问权限管理表的说明图。

[0067] 图25是示出密钥管理服务器100、加密装置200、检索装置300、数据中心400的硬件结构的一例的图。

具体实施方式

[0068] 实施方式1

[0069] 图1是基因检索系统10的结构图。

[0070] 基因检索系统10具有：密钥管理服务器100、多个加密装置200、多个检索装置300以及数据中心400(数据管理装置)。密钥管理服务器100、加密装置200、检索装置300和数据中心400经由网络500连接。

[0071] 密钥管理服务器100是如下这样的服务器：生成加密用用户密钥和隐匿检索用用户密钥等的用户密钥，发给加密装置200和/或检索装置300。另外，加密用用户密钥是用于对加密数据进行解密的密钥，隐匿检索用用户密钥是用于隐匿检索的密钥。

[0072] 加密装置200是用于对保管在数据中心400的信息进行加密的终端。加密装置200主要是由医院的医师、基因组解读中心的员工或患者等用户利用的终端。

[0073] 检索装置300是用于检索并取得保管在数据中心400中的信息的终端。检索装置300主要由制药公司等研究者、医院的医师等用户利用。

[0074] 数据中心400是保管从患者收集到的基因组信息、记载有患者的病历的电子病历等的服务器。数据中心400根据来自患者、医师、研究者等用户的请求提供检索/阅览基因组信息、电子病历等的服务。

[0075] 网络500是例如互联网那样的公众线路网。

[0076] 图2是密钥管理服务器100的结构图。

[0077] 密钥管理服务器100具有：主密钥生成部110、密钥存储部120、用户密钥生成部130、数据收发部140和用户ID存储部150。

[0078] 主密钥生成部110通过处理装置生成由利用隐匿检索的用户全体共同利用的公开参数，并生成成为生成用户密钥的基础的主密钥。

[0079] 密钥存储部120将主密钥生成部110生成的主密钥、公开参数存储在存储装置内。

[0080] 用户密钥生成部130通过处理装置，使用唯一分配给用户的用户ID，根据主密钥生成用户密钥。

[0081] 数据收发部140将公开参数经由网络500发送到加密装置200、检索装置300和数据中心400。并且，数据收发部140将用户密钥经由网络500发送到检索装置300。并且，数据收发部140根据用户请求，将用户ID对加密装置200、检索装置300、数据中心400的用户进行发送。

[0082] 用户ID存储部150将各用户的用户ID存储在存储装置内。用户ID是用户的姓名、部门、登录ID、邮件地址等属性信息。用户ID存储部150也可以不仅存储当前时刻的属性信息，而且也存储过去的属性信息作为历史记录。

[0083] 图3是加密装置200的结构图。

[0084] 加密装置200具有：基准基因取得部210、对象基因输入部220、公开参数存储部230、差异信息生成部240、差异信息编码部250、数据加密部260、加密标签生成部270以及带标签加密数据生成部280。

[0085] 基准基因取得部 210 取得一般公开的预定的基因组序列作为基准基因组序列（基准基因）。

[0086] 对象基因输入部 220 取得保管在数据中心 400 中的患者基因组序列（对象基因）。并且，对象基因输入部 220 取得患者基因组序列以及表示该患者基因组序列的患者的患者 ID。

[0087] 公开参数存储部 230 接收密钥管理服务器 100 生成的公开参数，将其存储在存储装置内。

[0088] 差异信息生成部 240 通过处理装置，将患者基因组序列与基准基因组序列进行比较，生成多个差异信息。

[0089] 差异信息编码部 250 通过处理装置，将差异信息生成部 240 生成的各差异信息编码成适于在加密的状态下进行检索的形式而生成编码差异信息。关于适于在加密的状态下进行检索的形式，将在后面描述。

[0090] 数据加密部 260 通过处理装置，对对象基因输入部 220 输入的患者基因组序列进行加密，生成加密数据（加密基因）。

[0091] 加密标签生成部 270 通过处理装置，对差异信息编码部 250 生成的编码差异信息进行加密，生成加密标签。

[0092] 带标签加密数据生成部 280 通过处理装置，将数据加密部 260 生成的加密数据、加密标签生成部 270 生成的多个加密标签、以及患者 ID 相结合，而生成带标签加密数据。带标签加密数据生成部 280 委托数据中心 400 保管所生成的带标签加密数据。

[0093] 图 4 是检索装置 300 的结构图。

[0094] 检索装置 300 具有：差异信息输入部 310、用户密钥存储部 320、差异信息编码部 330、检索查询生成部 340、基因信息取得部 350 以及数据解密部 360。

[0095] 差异信息输入部 310 通过输入装置输入作为检索关键字包含与基准基因组序列的差异信息的检索请求。

[0096] 用户密钥存储部 320 将公开参数和密钥管理服务器 100 单独分配给用户的用户密钥存储在存储装置内。

[0097] 差异信息编码部 330 具有与差异信息编码部 250 相同的功能。差异信息编码部 330 通过处理装置，将差异信息输入部 310 输入的检索请求内包含的差异信息编码成适于在加密的状态下进行检索的形式，而生成编码差异信息。

[0098] 检索查询生成部 340 通过处理装置，根据用户密钥存储部 320 存储的用户密钥和公开参数、以及差异信息编码部 330 生成的编码差异信息，生成检索查询（search query）。

[0099] 基因信息取得部 350 将检索查询生成部 340 生成的检索查询经由网络 500 发送到数据中心 400。然后，基因信息取得部 350 从数据中心 400 经由网络 500 接收被加密的包含检索请求内包含的差异信息（或者包含类似的差异信息）的患者基因组序列的加密数据。并且，基因信息取得部 350 接收加密数据和患者 ID。

[0100] 数据解密部 360 通过处理装置，使用用户密钥存储部 320 存储的用户密钥对从数据中心 400 接收到的加密数据进行解密，取得患者基因组序列。

[0101] 图 5 是数据中心 400 的结构图。

[0102] 数据中心 400 具有：保管请求处理部 410、加密数据存储部 420、加密标签存储部

430、检索请求处理部 440、公开参数存储部 450 以及访问权限存储部 460。

[0103] 保管请求处理部 410 从加密装置 200 接收带标签加密数据。保管请求处理部 410 分析接收到的带标签加密数据，分解成加密数据、多个加密标签以及患者 ID。保管请求处理部 410 将共同的管理编号分配给分解后的加密数据和各加密标签，将加密数据与患者 ID 以及管理编号一起发送到加密数据存储部 420，将各加密标签与患者 ID 和管理编号一起发送到加密标签存储部 430。

[0104] 加密数据存储部 420 将从保管请求处理部 410 接收到的加密数据与患者 ID 和管理编号相关联地存储在存储装置内。

[0105] 加密标签存储部 430 将从保管请求处理部 410 接收到的加密标签与患者 ID 和管理编号相关联地存储在存储装置内。

[0106] 检索请求处理部 440 从检索装置 300 接收检索查询。检索请求处理部 440 通过处理装置，对接收到的检索查询和加密数据存储部 430 存储的加密标签进行比较处理。通过该比较处理，判定加密标签内包含的患者基因组序列的差异信息是否符合由检索查询内包含的患者基因组序列的差异信息（检索请求）指定的条件。之后，检索请求处理部 440 从加密数据存储部 420 取得与检索选中的加密标签相关联的加密数据，将其回送到检索装置 300。另外，与加密标签相关联的加密数据是附有与加密标签相同的管理编号的加密数据。

[0107] 公开参数存储部 450 接收密钥管理服务器 100 生成的公开参数，并将其存储在存储装置内。

[0108] 访问权限存储部 460 对患者允许对谁公开患者基因组序列进行管理。

[0109] 对在基因检索系统 10 中利用的加密方式进行说明。

[0110] 在基因检索系统 10 中，使用：在非专利文献 2 等中记载的被称为内积基分级谓词加密(hierarchical predicate encryption for inner products)的加密方式、和在同样使用内积基分级谓词加密的加密的状态能够进行关键字检索的隐匿检索方式。

[0111] 图 6 是使用内积基分级谓词加密的加密方式的说明图。

[0112] 该加密方式由主密钥生成算法、密钥生成算法、转让密钥生成算法、加密算法和解密算法构成。

[0113] 最开始，使用主密钥生成算法生成加密用主密钥和加密用公开参数。加密用主密钥是用于生成解密用户的加密用用户密钥的密钥。加密用公开参数是在加密时使用的公开信息，广泛分发给进行加密的用户。

[0114] 另外，在该处理时，有必要事先决定条件式的结构，将该条件式作为参数来提供。在非专利文献 2 等现有的内积基分级谓词加密的文献中，记载的不是提供条件式的结构，而是提供在将条件式作为矢量来表述的情况下的维数，然而从容易理解的观点出发，这里设为提供条件式。以下相同。

[0115] 接着，使用密钥生成算法，发布根据加密用条件式对解密用户分发的加密用用户密钥。加密用条件式决定解密用户可以对具有怎样属性的文件进行解密，该条件被描述为使用“与”(AND)/“或”(OR) 的逻辑运算的条件式。

[0116] 接着，使用加密算法对数据进行加密。此时，指定附加给加密数据的加密用属性，将该加密用属性嵌入到加密数据内。

[0117] 最后，使用解密算法，对加密数据进行解密。在解密时，指定加密用用户密钥，然而

仅有被赋予了满足嵌入到加密用用户密钥内的条件式的加密用属性的加密数据可以解密。关于不满足条件式的加密数据，全部无法解密。

[0118] 另外，在内积基分级谓词加密中，还存在密钥转让(key delegation)的特征。这是在生成加密用用户密钥时仅设定条件式的一部分，使条件式的一部分处于未指定状态。对该未指定的条件式追加设定条件式的构造是密钥转让。

[0119] 具体而言，使用转让密钥生成算法，指定对加密用用户密钥追加设定的加密用追加条件式，生成加密用转让密钥。该加密用转让密钥与加密用用户密钥一样可以用于加密数据的解密。

[0120] 稍微改变说法，在密钥生成算法中，发布解密用户的属性信息被设定为密钥属性的加密用用户密钥。并且，在加密算法中，生成能够进行解密的用户的属性信息被设定为加密属性的加密数据。于是，在解密算法中，只有在被设定于加密用用户密钥内的密钥属性与被设定于加密数据内的加密属性对应的情况下，才可以使用加密用用户密钥对加密数据进行解密。

[0121] 并且，在转让密钥生成算法中，对加密用用户密钥进一步追加设定属性信息，生成能够仅对能够使用该加密用用户密钥解密的加密数据的一部分进行解密的下位加密用用户密钥。

[0122] 图7是使用内积基分级谓词加密的隐匿检索方式的说明图。

[0123] 该隐匿检索方式由主密钥生成算法、密钥生成算法、转让密钥生成算法、加密标签生成算法和一致判定算法构成。

[0124] 最开始，使用主密钥生成算法生成隐匿检索用主密钥和隐匿检索用公开参数。隐匿检索用主密钥是用于生成检索用户的隐匿检索用用户密钥的密钥。隐匿检索用公开参数是在检索时使用的公开信息，被广泛分发给进行检索的用户。

[0125] 另外，关于条件式的描述，与加密方式相同。

[0126] 接着，使用密钥生成算法，发布根据隐匿检索用条件式对检索用户分发的隐匿检索用用户密钥。隐匿检索用条件式仅决定如下这样的构造：指定由检索用户可以对具有怎样属性的文件进行检索，而且，可以使用怎样的条件式指定检索关键字。该条件被描述为使用“与”(AND)/“或”(OR)的逻辑运算的条件式。检索关键字自身可以之后通过转让密钥生成算法进行设定。

[0127] 接着，使用加密标签生成算法，生成在检索处理中使用的加密标签。首先生成任意的随机数值，为了限制允许检索的用户而决定属性，并决定关键字。将随机数值、属性和关键字作为输入，使用加密标签生成算法来生成加密标签。具体而言，使用加密算法对随机数值进行加密，将加密随机数和随机数值的组合作为加密标签。

[0128] 接着，使用转让密钥生成算法生成检索查询。具体而言，通过针对隐匿检索用用户密钥中的值未被指定的部分嵌入检索关键字，使得之后可以指定关键字，来生成检索查询。该检索查询对应于加密方式的加密转让密钥。

[0129] 最后，使用一致判定算法，进行接收到的检索查询和加密标签的两者中包含的关键字是否相同的一致判定。具体而言，从加密标签中取出加密随机数，使用与加密用转让密钥相当的检索查询进行解密。然后，在解密结果与加密标签内包含的随机数值相同的情况下，即，在使用检索查询正确地进行了解密的情况下，判定为关键字相同。原因是因为，若检

索查询（相当于加密用转让密钥）内包含的关键字与在生成加密随机数时指定的关键字不相同，则无法正确对随机数值进行解密。

[0130] 与上述的加密方式一样，稍许改变说法，在密钥生成算法中，发布检索用户的属性信息被设定为密钥属性的隐匿检索用用户密钥。并且，在加密算法中，生成作为加密属性被设定了能够进行检索的用户的属性信息和关键字的加密数据。在转让密钥生成算法中，对隐匿检索用用户密钥进一步追加设定检索关键字作为密钥属性，生成检索查询。

[0131] 于是，在解密算法中，只有在被设定于检索查询内的密钥属性信息与被设定于加密标签内的属性信息对应的情况下，才可以使用加密标签对随机数值进行解密。也就是说，只有在被设定于检索查询内的用户的属性信息与被设定于加密标签内的用户的属性信息对应、并且被设定于检索查询内的检索关键字与被设定于加密标签内的关键字对应的情况下，才可以使用加密标签对随机数值进行解密。

[0132] 对基因检索系统 10 的处理进行说明。

[0133] 基因检索系统 10 执行初始设定处理、用户密钥发布处理、加密处理和检索处理。另外，加密处理包含有：患者基因组序列的加密处理和患者的电子病历的加密处理。

[0134] 图 8 是示出初始设定处理流程的流程图。

[0135] 初始设定处理是由密钥管理服务器 100 执行的处理，是在利用基因检索系统 10 之前执行一次的处理。

[0136] （S101：编码方法决定处理）

[0137] 主密钥生成部 110 通过处理装置，决定对患者基因组序列的差异信息进行编码的方法。在决定该进行编码的方法时，还对如何进行检索的检索式加以考虑。

[0138] 作为主要的差异信息，有：SNV(Single Nucleotide Variant，单核苷酸变异）、SV(Structural Variants，结构变异）和 NC(Novel Contig，新重叠群）。因此，以这 3 个为例，对该编码方法和检索方法进行说明。

[0139] 图 9 是已编码的 SNV 信息的结构图。

[0140] SNV 信息示出在基因组序列中变化了一个碱基的情况。SNV 信息由患者 ID、染色体编号、位置信息、置换信息 1、置换信息 2 和可靠度构成。

[0141] 患者 ID 是为了与另外管理的电子病历相对应而分配的识别编号。患者 ID 是与病历相对应的编号即可，因而也可以不仅分配在医院赋予给患者的识别编号，而且在病历保管时随机分配编号，并设定该值。

[0142] 染色体编号是检测出 SNV 的染色体的编号。在人的基因组的情况下，设定了例如 1、2、…、22、X、Y、M 等值。

[0143] 位置信息是与检测出 SNV 的位置相关的信息。位置由示出时碱基序列的第几个的数值表示，在人的基因组的情况下为 1 至 30 亿左右的值。一般，位置等数值信息的检索是指定范围来检索。然而，由于在上述的隐匿检索方式中，范围检索实施困难，因而通过进行分块化来实现高速化。例如，图 10 示出在检索时指定的范围一般被抑制为 5000 以下的情况下的编码方法。在位置 7777 检测出 SNV 的情况下，考虑为块 1 和块 2 为相应的位置，对位置信息设定“1”和“2”这两个值。

[0144] 置换信息 1 和置换信息 2 是表示基因组信息的变化的信息。其值是如下这样决定的：例如在由 G 表示的碱基作为 SNV 插入的情况下表记为“G 追加”，在由 A 表示的碱基消失

的情况下表记为“A 删除”等。有两个置换信息是因为，例如在由 G 表示的碱基变化为由 A 表示的碱基的情况下，在置换信息 1 中作为 G 删除来表示，在置换信息 2 中作为 A 追加等来表示。

[0145] 可靠性是表示该 SNV 信息的可靠性的信息，由 0 至 100 的实数来表示。这也与位置信息一样，进行分块化来实现检索的高速化。例如与图 11 一样，以 10％为单位对可靠性进行区分，设定其对应的值。例如，在可靠性是 7.44％的情况下，设定为对应的块的值为 0。

[0146] 在检索的情况下，从上述的患者 ID 到可靠性的值中、指定若干个作为条件想要指定的值，并且未指定的场所指定视为与任何值都一致的“*”（通配符），从而生成差异信息（检索请求）。然后，针对 SNV 信息的各要素，决定在加密标签与检索查询之间全部一致的情况下检索选中的规则。

[0147] 并且，还决定各要素在怎样的条件下一致。

[0148] 例如，在指定了患者 ID 的情况下，将患者 ID 相同的要素视为一致。染色体编号也一样，指定想要检索的染色体编号，将相同的视为一致。在位置信息的情况下，在想要检索例如 7000 ～ 12000 的范围的情况下，确定块 2 作为包含该范围的块，在检索式中指定“2”。并且，在与作为 SNV 信息的位置信息所设定的 2 个值中的任一方相同的情况下，视为一致。在置换信息的情况下，利用与 SNV 信息中指定的值相同的规则，设定想要进行检索的值，在两者相同的情况下，视为一致。可靠性也一样，在指定了相同值的情况下，视为一致。

[0149] 如以上所述，决定 SNV 信息的编码方法、和能够怎样实施检索的方式。

[0150] 图 12 是已编码的 SV 信息的结构图。

[0151] SNV 信息表示 1 个碱基发生变化，而 SV 信息表示连续的多个碱基序列发生变化。SV 信息由以下构成：患者 ID、染色体编号、开始位置信息、结束位置信息、变异类别、CNV 增益、重组染色体编号、重组染色体开始位置信息、重组染色体结束位置信息、分类 ID（插入序列）、版本（插入序列）、染色体编号（插入序列）、开始位置信息（插入序列）和结束位置信息（插入序列）。

[0152] 患者 ID、染色体编号由于与图 9 所示的 SNV 相同，因而省略说明。

[0153] 开始位置信息是表示变化开始的位置的信息。按照与 SNV 信息的位置信息相同的步骤使位置信息分块化，设定表示块的值。

[0154] 结束位置信息是表示变化结束的位置的信息。按照与 SNV 信息的位置信息相同的步骤使位置信息分块化，设定表示块的值。

[0155] 变异类别是表示 SV 的类别的信息。例如，设定了表示基因序列的重复次数的追加 / 削减的“CNV 增益”或“CNV 损失”、表示碱基序列的反转的“Inversion”、表示碱基序列的大规模插入的“Insertion”、表示碱基序列的大规模缺损的“Deletion”、表示与其他染色体的重组的“重组”等的值。

[0156] CNV 增益是在作为变异类别设定有“CNV 增益”和“CNV 损失”的情况下利用的。关于 CNV 增益，在作为变异类别设定了“CNV 增益”的情况下表示重复次数增加几次，在设定了“CNV 损失”的情况下表示重复次数减少几次的信息。由于该值也包含整数值，因而进行分块化并设定值。例如，在如图 13 所示决定了进行分块化的规则且 CNV 增益是 15 的情况下，设定对应的块的编号“3”。

[0157] 重组染色体编号是在作为变异类别指定了“重组”的情况下利用的。重组染色体

编号设定了被重组的其他染色体的编号。能够设定的值与染色体编号相同。

[0158] 重组染色体开始位置信息和重组染色体结束位置信息是在作为变异类别指定了“重组”的情况下利用的。重组染色体开始位置信息和重组染色体结束位置信息由于表示与其他染色体中的哪个碱基序列发生了重组，因而是表示其他染色体中的重组的开始位置和结束位置的信息。能够设定的值与开始位置信息和结束位置信息相同。

[0159] 分类 ID（插入序列）是在作为变异类别指定了“Insertion”的情况下利用的。分类 ID（插入序列）设定有表示所插入的碱基序列是其他哪种生物的碱基序列的编号。

[0160] 版本（插入序列）是用于表示决定上述分类 ID 所使用的基因组数据库的版本的数值。

[0161] 染色体编号（插入序列）、开始位置信息（插入序列）和结束位置信息（插入序列）是表示插入了其他生物的哪个部分的碱基序列的信息，所设定的值利用与染色体编号、开始位置信息、结束位置信息相同的方法进行编码。不过，由于基因组序列的长度根据生物种类而不同，因而优选的是按每个分类 ID 准备不同的块的区分方法。

[0162] 并且，在检索的情况下，从上述的患者 ID 到结束位置信息（插入序列）的值中、指定若干个作为条件想要指定的值，并且未指定的场所指定视为与任何值都一致的“*”（通配符），从而生成差异信息（检索请求），根据该差异信息生成检索查询。然后，针对 SV 信息的各要素，决定在加密标签与检索查询之间全部一致的情况下检索选中的规则。

[0163] 并且，还决定各要素在怎样的条件下一致。

[0164] 患者 ID、染色体编号与 SNV 信息的情况相同。开始位置信息、结束位置信息、重组染色体开始位置信息、重组染色体结束位置信息、开始位置信息（插入序列）和结束位置信息（插入序列）可以与 SNV 信息的位置信息同样处理。变异类别将想要检索何种种类的变异类别指定为检索查询，将设定了相同值的 SV 信息视为一致。CNV 增益在想要检索例如位于 10 至 30 之间的数值的情况下，指定其对应的块的编号“3”。并且，在与 SV 信息内包含的 CNV 增益相同的情况下，视为一致。重组染色体编号、染色体编号（插入序列）可以与染色体编号同样处理。对于分类 ID、版本，在检索查询内包括想要检索的分类 ID 和版本，在与 SV 信息相同的情况下视为一致。

[0165] 如上所述，决定 SV 信息的编码方法和怎样能够实施检索的方式。

[0166] 图 14 是已编码的 NC 信息的结构图。

[0167] NC 信息表示无法映射到基准基因组的碱基。碱基无法映射到基准基因组，在很多情况下认为是由于感染了病毒等而检测出特殊的基因组的情况。因此，NC 信息除了患者 ID 以外，还由以下构成：分类 ID（插入序列）、版本（插入序列）、染色体编号（插入序列）、开始位置信息（插入序列）和结束位置信息（插入序列）。

[0168] 值的设定方法和检索方法由于与 SV 信息相同，因而省略说明。

[0169] （S102：方式决定处理）

[0170] 主密钥生成部 110 通过处理装置，决定要利用的隐匿检索方式和 / 或对数据主体进行加密的加密方式。这里，由于需要可以指定多个检索关键字的隐匿检索方式，因而使用利用上述的内积基分级谓词加密的隐匿检索方式。同样，加密方式也使用利用上述的内积基分级谓词加密的加密方式。

[0171] 并且，主密钥生成部 110 决定隐匿检索方式的利用方法。这里，决定标签 ID 的分层

结构。例如，如图 15 所示，标签 ID 由 3 要素构成，由存储能够检索的用户所属的组的组名的组名栏、存储姓名等的用户名栏、以及存储患者基因组序列的差异信息的差异信息栏构成。在检索时，采用这样的规则：只有在判定为组名、用户名、差异信息全部一致的情况下，才视为检索选中。

[0172] 同样，主密钥生成部 110 决定用于对数据主体进行加密的加密方式的利用方法。这里，决定解密者 ID 的分层结构。例如，如图 16 所示，解密者 ID 由 2 个要素构成，由存储能够解密的用户所属的组的组名的组名栏、和存储姓名等的用户名栏构成。在解密时，采用这样的规则：只有组名、用户名全部一致的才可以解密。

[0173] (S103：用户 ID 存储处理)

[0174] 用户 ID 存储部 150 构建保管用户 ID 的用户 ID 信息数据库。用户 ID 信息数据库存储有：生成用户密钥所需的信息、和在加密装置 200 对数据进行加密时用于确定对方的组名 / 用户名所需的信息。

[0175] 例如，如图 17 所示，在用户 ID 信息数据库内存储有：组名即公司名、用户名即姓名、部门信息、有效期间等。并且，在用户 ID 信息数据库内也可以不仅存储最新的状况，而且存储全部过去的历史记录。

[0176] (S104：主密钥生成处理)

[0177] 主密钥生成部 110 通过处理装置，执行隐匿检索方式的主密钥生成算法，生成隐匿检索用主密钥和隐匿检索用公开参数。同样，主密钥生成部 110 通过处理装置，执行加密方式的主密钥生成算法，生成加密用主密钥和加密用公开参数。

[0178] 以下，将隐匿检索用主密钥和加密用主密钥总称为主密钥，将隐匿检索用公开参数和加密用公开参数总称为公开参数。

[0179] (S105：主密钥存储处理)

[0180] 密钥存储部 120 将主密钥生成部 110 生成的主密钥和公开参数存储在存储装置内。

[0181] (S106：公开参数公开处理)

[0182] 数据收发部 140 将密钥存储部 120 存储的公开参数经由网络 500 对加密装置 200、检索装置 300 和数据中心 400 进行公开。

[0183] 另外，所公开的公开参数在加密装置 200 中，被存储在公开参数存储部 230 内，在检索装置 300 中，被存储在用户密钥存储部 320 内，在数据中心 400 中，被存储在公开参数存储部 450 内。

[0184] 通过以上的步骤，基因检索系统 10 的设置完成。

[0185] 另外，在系统运用中，每当有用户的人事异动、入职、退职时，在 S103 中所生成的用户 ID 信息数据库进行内容维护。

[0186] 图 18 是示出用户密钥发布处理流程的流程图。

[0187] 用户密钥发布处理主要是由密钥管理服务器 100 和检索装置 300 执行的处理，在追加了新用户的情况下，在用户所属的组名改变时等执行。

[0188] (S201：用户 ID 取得处理)

[0189] 用户密钥生成部 130 从用户 ID 存储部 150 保持的用户 ID 信息数据库中取得发布用户密钥的用户的组名和用户名。

[0190] (S202 :用户密钥生成处理)

[0191] 用户密钥生成部 130 通过处理装置,生成在生成检索查询中使用的隐匿检索用用户密钥、和用于对加密数据进行解密的加密用用户密钥。

[0192] 在隐匿检索方式中,有必要在生成隐匿检索用用户密钥时指定标签 ID 分层结构。这里,将在 S201 中取得的组名设定在组名栏,同样将用户名设定在用户名栏,将差异信息以之后进行检索的用户可以进行设定的方式指定为能够转让的要素,从而可以生成隐匿检索用用户密钥。

[0193] 同样,在加密方式中,有必要在生成加密用用户密钥时指定解密者 ID 分层结构。这里,将在 S201 中取得的组名指定给组名栏,同样将用户名指定给用户名栏,从而可以生成加密用用户密钥。

[0194] 将如上所述所生成的隐匿检索用用户密钥和加密用用户密钥总称为用户密钥。

[0195] (S203 :用户密钥发送处理)

[0196] 数据收发部 140 将在 S202 中生成的用户密钥发送到检索装置 300。

[0197] (S204 :用户密钥接收处理)

[0198] 用户密钥存储部 320 接收在 S203 中发送的用户密钥,将其存储在存储装置内。

[0199] 图 19 是示出患者基因组序列的加密处理流程的流程图。

[0200] 患者基因组序列的加密处理主要是由加密装置 200 和数据中心 400 执行的处理,是在对患者基因组序列进行加密并保管在数据中心 400 内时执行的。

[0201] (S301 :差异信息提取处理)

[0202] 基准基因取得部 210 取得例如在互联网上公开的基准基因组序列。并且,对象基因输入部 220 通过输入装置,输入患者基因组序列。

[0203] 差异信息生成部 240 通过处理装置,将患者基因组序列与基准基因组序列进行比较,从而生成 SNV、SV、NC 等差异信息。作为生成该差异信息的方法,公知有 ChIP － seq 法或者 RNA － seq 法或者 MeDIP － seq 法或者变异分析法或者亚硫酸氢盐法 (bisulfite method) 等,使用这些一般的方法。

[0204] (S302 :用户决定处理)

[0205] 数据加密部 260 使操作加密装置 200 的用户输入能够对加密数据进行解密的用户的组名和用户名。同样,加密标签生成部 270 输入能够检索加密数据的用户的组名和用户名。

[0206] 这里输入的组名和用户名无需是一个,在能够解密或检索的用户有多名的情况下也可以输入多个。另外,这里利用的隐匿检索方式和加密方式也能够接受意味着作为组名和用户名谁都可以的通配符。

[0207] (S303 :数据加密处理)

[0208] 数据加密部 260 通过处理装置,使用在 S302 中输入的能够解密的组名和用户名,对在 S301 中输入的患者基因组序列进行加密。

[0209] 具体而言,数据加密部 260 随机生成对话密钥 (session key),利用该对话密钥,使用 AES、Camellia(注册商标) 等公共密钥加密对患者基因组序列进行加密,生成加密数据主体。接着,数据加密部 260 将在 S302 中输入的能够解密的组名和用户名分别指定为解密者 ID 分层结构的组名和用户名,将其作为加密用公开密钥,使用在 S102 中决定的加密方

式对对话密钥进行加密，生成加密对话密钥。然后，数据加密部 260 通过使上述 2 个加密结果（加密数据主体和加密对话密钥）进行组合，生成加密数据。

[0210] 所生成的加密数据的数据结构由图 20 的标号 603 部分表示。另外，在 S302 中接收到多个组名和用户名的情况下，有必要针对组名和用户名的各组生成加密对话密钥。

[0211] （S304：差异信息编码处理）

[0212] 差异信息编码部 250 通过处理装置，依照在 S101 中决定的编码方法对在 S301 中生成的各差异信息进行编码，生成编码差异信息。并且，差异信息编码部 250 使从用户输入患者 ID，将其包含在编码差异信息内。

[0213] （S305：加密标签生成处理）

[0214] 加密标签生成部 270 对编码差异信息进行加密而生成加密标签。

[0215] 具体而言，加密标签生成部 270 通过处理装置，将在 S302 中输入的能够检索的组名和用户名指定为标签 ID 分层结构的组名和用户名，将在 S304 中编码的编码差异信息指定给差异信息栏，使用隐匿检索方式对随机数值进行加密而生成加密标签。并且，加密标签生成部 270 使随机数值在平文的状态下包含在加密标签内。

[0216] 另外，由于上述处理是针对 1 个差异信息的处理，因而针对各编码差异信息来实施该处理。例如，针对 SNV、SV、NC 的各编码差异信息来实施。并且，在 S302 中输入了多个组名和用户名的组的情况下，针对组名和用户名的各组生成加密标签。

[0217] （S306：保管委托处理）

[0218] 带标签加密数据生成部 280 通过处理装置，使在 S303 中生成的加密数据、在 S305 中生成的加密标签和在 S304 中输入的患者 ID 结合而生成带标签加密数据（图 20 的标号 601）。然后，带标签加密数据生成部 280 将所生成的带标签加密数据发送到数据中心 400 并委托保管。

[0219] 此时，为了使带标签加密数据在数据中心 400 中容易保管，带标签加密数据生成部 280 将在 S302 中输入的能够进行解密的组名和用户名与能够进行检索的组名和用户名一起发送。在图 20 所示的带标签加密数据的结构中，将能够进行解密的组名和用户名与能够进行检索的组名和用户名包含在带标签加密数据内。

[0220] （S307：加密数据保管处理）

[0221] 保管请求处理部 410 通过处理装置，对从加密装置 200 接收到的带标签加密数据进行分解，取出加密数据、多个加密标签和患者 ID。然后，保管请求处理部 410 使加密数据存储部 420 保管加密数据和患者 ID。

[0222] 另外，加密数据存储部 420 按带标签加密数据内包含的每个组名和每个用户名分开保管加密数据，再对保管的加密数据赋予管理编号，可以在之后根据管理编号唯一地确定加密数据。在加密数据与多个组名和用户名相关联的情况下，使加密数据与各组名和用户名相关联地进行保管。在与多个组名和用户名相关联的情况下，存储仅一个加密数据，另外仅保管参照信息，从而能够节约存储盘容量。

[0223] 图 21 是示出加密数据的保管例的图。如图 21 所示，保管请求处理部 410 将组名是“A 制药公司”且用户名是“*”（通配符）的加密数据与患者 ID 和管理编号汇总管理，再将组名是“B 医院”且用户名是“*”的加密数据与患者 ID 和管理编号汇总管理。并且，在有对 A 制药公司和 B 医院的双方公开的数据的情况下，例如使患者 ID 和加密数据主体与管理

编号 000001 相关联地进行保管，在管理编号 100002 中除了患者 ID 以外，还将参照管理编号 000001 的指针作为加密数据来保管。

[0224] (S308 :加密标签保管处理)

[0225] 保管请求处理部 410 使加密数据存储部 430 保管在 S307 中取出的多个加密标签以及对应的加密数据的管理编号和患者 ID。加密数据存储部 430 按带标签加密数据内包含的各组名和各用户名分开保管加密标签、管理编号和患者 ID。

[0226] 图 22 是示出患者的电子病历的加密处理流程的流程图。

[0227] 患者的电子病历的加密处理主要是由加密装置 200 和数据中心 400 执行的处理，在对电子病历进行加密并保管在数据中心 400 内时来执行。

[0228] (S401 :用户决定处理)

[0229] 数据加密部 260 使操作加密装置 200 的用户输入能够对电子病历进行解密的用户的组名和用户名。

[0230] 这里输入的组名和用户名无需是一个，在能够解密的用户有多名的情况下也可以输入多个。

[0231] (S402 :数据加密处理)

[0232] 数据加密部 260 使用户输入患者 ID 和电子病历。然后，数据加密部 260 通过处理装置，使用在 S401 中输入的组名和用户名，对电子病历进行加密。具体的加密方法由于与在 S303 中对患者基因组序列进行加密的流程相同，因而省略详情。

[0233] (S403 :保管委托处理)

[0234] 数据加密部 260 将在 S402 中生成的加密数据与表示是谁的电子病历的患者 ID 以及能够进行解密的用户的组名和用户名一起发送到数据中心 400 并委托保管。

[0235] (S404 :加密数据保管处理)

[0236] 保管请求处理部 410 使从加密装置 200 接收到的加密数据与患者 Id 相关联地保管在加密数据存储部 420 内。

[0237] 图 23 是示出检索处理流程的流程图。

[0238] 检索处理是主要由检索装置 300 和数据中心 400 执行的处理，在取得由数据中心 400 保管的加密患者基因组序列时执行。

[0239] (S501 :差异信息输入处理)

[0240] 差异信息输入部 310 使操作检索装置 300 的用户输入包含将与基准基因组序列的差异信息作为检索关键字的检索请求。

[0241] 这里输入的差异信息无需如从患者基因组提取出的差异信息那样指定全部要素，例如仅指定染色体编号、或者仅指定位置信息等即可。

[0242] (S502 :差异信息编码处理)

[0243] 差异信息编码部 330 通过处理装置，对在 S501 中输入的差异信息进行编码，生成编码差异信息。该处理由于与步骤 S304 的处理相同，因而省略详情。不过，需要注意的一点是，将未指定的要素设定为“*”。

[0244] (S503 :检索查询生成处理)

[0245] 检索查询生成部 340 通过处理装置，使用在 S502 中生成的编码差异信息、和存储在用户密钥存储部 320 内的用户密钥生成检索查询。然后，检索查询生成部 340 将生成的

检索查询发送到数据中心 400。

[0246] 此时，还发送用户自身的组名和用户名。另外，为了验证组名和用户名的可靠性，还进行操作检索终端的用户的用户认证。

[0247] (S504 :加密标签提取处理)

[0248] 检索请求处理部 440 通过处理装置，从存储在加密数据存储部 430 内的全部加密标签中取得能够利用在 S503 中与检索查询一起发送的组名和用户名进行检索的全部加密标签。并且，检索请求处理部 440 从访问权限存储部 460 取出相应的组名和用户名的用户能够访问的患者 ID 的列表，将所取得的加密标签缩小到仅包括与该患者 ID 对应的加密标签。

[0249] 另外，访问权限存储部 460 具有图 24 所示的访问权限管理表，将组名和用户名作为访问者信息，确定与可以检索的患者基因组序列对应的患者 ID，输出与该患者 ID 对应的加密标签即可。

[0250] (S505 :一致判定处理)

[0251] 检索请求处理部 440 通过处理装置，对在 S504 中被缩小范围的加密标签进行隐匿检索方式的一致判定处理，判定加密标签内包含的差异信息是否与由在 S503 中发送的检索查询内包含的差异信息指定的条件一致。

[0252] 隐匿检索方式的一致判定处理只能实施 1 个加密标签与 1 个检索查询的比较。因此，对在 S504 中取得的全部加密标签实施一致判定处理。然后，确定与判定处理的结果为被判定为一致的加密标签相关联的管理编号。

[0253] (S506 :加密数据取得处理)

[0254] 检索请求处理部 440 从加密数据存储部 420 取得与在 S505 中确定的管理编号对应的全部加密数据，将其与对应的患者 ID 一起发送到检索装置 300。

[0255] (S507 :加密数据解密处理)

[0256] 数据解密部 360 通过处理装置，使用存储在用户密钥存储部 320 内的加密用用户密钥，对在 S506 中从数据中心 400 接收到的加密数据执行加密方式的解密处理来进行解密。数据解密部 360 对接收到的全部加密数据实施该处理。

[0257] 通过以上的步骤，检索装置 300 可以从用户接收想要检索的差异信息，从数据中心 400 取得与该差异信息一致的加密数据，对其进行解密来阅览患者基因组序列。并且，还可以根据需要，将与加密数据对应的患者 ID 发送到数据中心 400，获得对应的电子病历等。

[0258] 如上所述，在实施方式 1 的基因检索系统 10 中，使用隐匿检索方式对患者基因组进行加密并保管在数据中心 400 内，并且检索请求也使用隐匿检索技术进行加密并对数据中心 400 委托检索。因此，在数据中心 400，尽管无法知道患者基因组的具体内容，也可以提供检索服务。

[0259] 并且，人类基因组信息是由 30 亿碱基构成的非常庞大的数据。因此，如果将人类基因组信息全部作为标签，则可能由于加密而进一步使数据尺寸增加，导致成为存储盘容量和网络容量的负担的原因。

[0260] 然而，在实施方式 1 的基因检索系统 10 中，由于缩小范围至与一般公开的基准基因组序列的差异信息，将其做成标签，因而可以大幅削减存储盘使用量和网络容量。

[0261] 并且，在实施方式 1 的基因检索系统 10 中，通过使 SNV 的位置信息、可靠性等数值

信息分块化，由此使用一致检索实现了通过隐匿检索难以实现的范围检索。因此，也可以应对在基因组检索中利用的范围检索。

[0262] 特别是，在图 10 所示的 SNV 的位置信息中，各块重复。也就是说，块 1 和块 2 从位置 5000 到 10000 重复，块 2 和块 3 从位置 10000 到 15000 重复。由此，在检索时指定的范围是 5000 以下的情况下，使用全部块进行检索，在检索时指定的范围是 10000 以下的情况下，若仅使用奇数块进行检索，则能够更高速地进行处理。

[0263] 并且，在实施方式 1 的基因检索系统 10 中，即使在指定了染色体编号和位置信息等多个检索条件的情况下，也不单独进行条件的一致判定，而使用内积基分级谓词加密一并判定全部检索条件是否成立。因此，服务器不知晓部分性地检索命中的情况，安全性高。

[0264] 并且，在实施方式 1 的基因检索系统 10 中，与加密数据一起还包含患者 ID。因此，可以从作为检索结果获得的患者 ID 导出关联的电子病历等信息。因此，在发生相应的碱基变异的情况下，也能够研究其与怎样的病相关联。

[0265] 并且，在实施方式 1 的基因检索系统 10 中，在对患者基因组序列的差异信息进行加密时，标签 ID 分层结构和解密者 ID 分层结构包含组名和用户名。因此，能够限制可以进行检索和 / 或解密的研究者和医生。例如，若将“A 制药公司”指定为组名，将通配符“*”指定为用户名，则可以将该患者的信息限定于 A 制药公司的职员。并且，通过将无论是组名还是用户名都作为通配符“*”进行加密，则只要是在该系统中登记的医生或研究者，都可以利用该患者的信息。

[0266] 并且，在实施方式 1 的基因检索系统 10 中，与基于加密的访问控制不同，数据中心 400 保管访问权限管理表，也能够进行基于该信息的访问控制。因此，可以根据来自患者的期望，细致管理“仅能阅览基因组序列”或者“能够阅览病历”等。也就是说，由于根据该信息判定是否允许从患者基因组序列的检索请求，因而极其细致的访问控制成为可能。

[0267] 并且，在实施方式 1 的基因检索系统 10 中，针对检索者的用户密钥包含组名和 / 或用户名。因此，通过确认根据该用户密钥生成的检索查询内包含的组名和 / 或用户名，也能够进行认证。

[0268] 另外，在上述说明中，示出了以公司为单位进行访问限制的情况的例子。然而，该访问控制的单位仅是一例。例如，也可以输入医生、护士等国家资格的条件，也可以建立是否是国家项目的参加者的标志。这些 ID 分层结构仅是一例，因而也可以追加各种要素，或者删除各种要素。

[0269] 并且，在上述说明中，设想了如下这样的情况 ：国家对唯一的密钥管理服务器 100 进行管理，作为医生的代理进行基因组分析的序列操作者利用加密装置 200。然而，可以根据利用系统，灵活变更系统结构。例如，从序列操作者接收到检测结果的医生也可以操作加密装置 200 对患者基因组序列进行加密。

[0270] 并且，在上述说明中，在检索装置 300 内保管用户密钥，还实施检索查询生成和加密数据的解密。然而，为了进一步提高安全性，也可以使用 IC 卡等器件而不是检索装置 300 来实施用户密钥的管理。在该情况下，由 IC 卡安全地管理用户密钥，因而实现了安全性的提高。

[0271] 并且，在上述说明中，检索终端利用的隐匿检索用用户密钥作为在 SNV 信息、SV 信息、NC 信息等的检索中可以共同使用的密钥来生成。然而，也可以是，由于各差异信息的长

度不同，因而不使用共同的隐匿检索用用户密钥，而根据用途单独生成 SNV 信息隐匿检索用用户密钥、SV 信息隐匿检索用用户密钥、NC 信息隐匿检索用用户密钥等隐匿检索用用户密钥。在该情况下，由于隐匿检索用用户密钥的长度为最适于各信息的长度的长度，因而运算时间高速化。

[0272] 并且，在上述说明中，在表示组名和用户名时使用了“A 制药公司”、“田中”等文字串进行表述。这是以实施例容易理解作为优先考虑的，实际上，也可以不仅利用文字串，还可以利用编号等 ID。染色体编号等其他要素也一样。

[0273] 并且，在上述说明中，示出作为隐匿检索方式或加密方式使用内积基分级谓词加密的情况的例子。然而，若是具有相同功能的加密，则无需限于内积基分级谓词加密。并且，隐匿检索方式和加密方式也可以利用不同的方式。

[0274] 并且，在上述说明中，在用户 ID 存储部 150 中，还可以管理用户过去的属性信息。这只有在管理上需要的情况下才实施即可，也可以仅管理当前时刻的属性信息。

[0275] 并且，在上述说明中，示出假定全员共同使用 1 个基准基因组序列来实施的情况。然而，还能够使用不同的基准基因组来实施。在该情况下，还能够在进行加密时针对全部的基准基因组序列生成差异信息，将与任意一个基准基因组序列的差异信息作为检索查询，还能够在进行加密时生成与一个基准基因组序列的差异信息，在进行检索的情况下将与所有的基准基因组序列的差异信息作为检索查询。

[0276] 并且，在上述说明中，关于需要位置信息、可靠性、CNV 增益等范围检索的信息，可以通过进行分块化而通过关键字的完全一致进行判定。然而，由于设想了想要检索的范围根据用途而不同的情况，因而不一定需要通过完全一致实现检索。例如，也可以在加密时指定属于块 10 的同时，在检索时如指定“块 10 或块 11”那样将多个块指定为检索范围。

[0277] 并且，在上述说明中，设定了密钥管理服务器 100 只有 1 台的情况。然而，用作隐匿检索方式或加密方式的内积基分级谓词加密还能够对密钥管理服务器 100 进行分层化并分散为多个来运用。因此，这里，还能够将密钥管理服务器 100 分层化为多个来运用。

[0278] 并且，在上述说明中，示出将 SNV 信息、SV 信息、NC 信息作为差异信息进行加密的例子。然而，还能够根据相同的构造对所述以外的差异信息进行加密。

[0279] 图 25 是示出密钥管理服务器 100、加密装置 200、检索装置 300 和数据中心 400 的硬件结构的一例的图。

[0280] 如图 25 所示，密钥管理服务器 100、加密装置 200、检索装置 300 和数据中心 400 具有执行程序的 CPU 911(Central Processing Unit(中央处理单元)，也称为中央处理装置、处理装置、运算装置、微处理器、微计算机、处理器)。CPU 911 经由总线 912 与 ROM 913、RAM 914、LCD 901(Liquid Crystal Display，液晶显示器)、键盘 902(K/B)、通信板 915、磁盘装置 920 连接，并控制这些硬件器件。也可以是光盘装置、存储卡读写装置等存储装置，而取代磁盘装置 920(固定磁盘装置)。磁盘装置 920 经由预定的固定磁盘接口连接。

[0281] ROM 913、磁盘装置 920 是非易失性存储器的一例。RAM 914 是易失性存储器的一例。ROM 913、RAM 914 和磁盘装置 920 是存储装置(存储器)的一例。并且，键盘 902、通信板 915 是输入装置的一例。并且，通信板 915 是通信装置的一例。而且，LCD 901 是显示装置的一例。

[0282] 在磁盘装置 920 或 ROM 913 等中存储有：操作系统 921(OS)、窗口系统 922、程序组

923 和文件组 924。程序组 923 的程序由 CPU 911、操作系统 921、窗口系统 922 执行。

[0283] 在程序组 923 内存储有在上述的说明中执行作为以下等所说明的功能的软件、程序和其他程序，即，“主密钥生成部 110”、“用户密钥生成部 130”、“数据收发部 140”、“基准基因取得部 210”、“对象基因输入部 220”、“差异信息生成部 240”、“差异信息编码部 250”、“数据加密部 260”、“加密标签生成部 270”、“带标签加密数据生成部 280”、“差异信息输入部 310”、“差异信息编码部 330”、“检索查询生成部 340”、“基因信息取得部 350”、“数据解密部 360”、“保管请求处理部 410”、“检索请求处理部 440”。程序由 CPU 911 读出并执行。

[0284] 在文件组 924 内，作为“文件”、“数据库”的各项目存储有在上述说明中存储在“密钥存储部 120”、“用户 ID 存储部 150”、“公开参数存储部 230”、“用户密钥存储部 320”、“加密数据存储部 420”、“加密标签存储部 430”、“公开参数存储部 450”、“访问权限存储部 460”等中的信息、数据、信号值、变量值、参数。“文件”、“数据库”被存储在存储盘、存储器等存储介质内。存储在存储盘、存储器等存储介质内的信息、数据、信号值、变量值、参数经由读写电路由 CPU 911 读出到主存储器、高速缓冲存储器内，用于提取、检索、参照、比较、运算、计算、处理、输出、打印、显示等 CPU 911 的动作。在提取、检索、参照、比较、运算、计算、处理、输出、打印、显示的 CPU 911 的动作期间，信息、数据、信号值、变量值、参数被暂时存储在主存储器、高速缓冲存储器、缓冲存储器内。

[0285] 并且，上述说明中的流程图的箭头部分主要表示数据和信号的输入输出，数据和信号值被记录在 RAM 914 的存储器、其他的光盘等存储介质、IC 芯片内。并且，数据和信号通过总线 912、信号线、线缆、其他传送介质、电波被在线传送。

[0286] 并且，在上述说明中作为“～部”进行说明的也可以是“～电路”、“～装置”、“～设备”、“～单元”、“功能”，并且也可以是“～步骤”、“～进程”、“～处理”。并且，作为“～装置”进行说明的也可以是“～电路”、“～设备”、“～单元”、“功能”，并且也可以是“～步骤”、“～进程”、“～处理”。而且，作为“～处理”进行说明的也可以是“～步骤”。即，作为“～部”进行说明的也可以使用存储在 ROM 913 内的固件来实现。或者，也可以通过仅软件、或者仅元件、器件、基板、布线等的硬件、或者软件和硬件的组合、以及它们与固件的组合来实施。固件和软件作为程序被存储在 ROM 913 等记录介质内。程序由 CPU 911 读出，由 CPU 911 执行。即，程序使计算机等作为上文描述的“～部”发挥功能。或者，使计算机等执行在上文描述的“～部”的进程和方法。

[0287] 标号说明

[0288] 10 :基因检索系统 ;100 :密钥管理服务器 ;110 :主密钥生成部 ;120 :密钥存储部 ;130 :用户密钥生成部 ;140 :数据收发部 ;150 :用户 ID 存储部 ;200 :加密装置 ;210 :基准基因取得部 ;220 :对象基因输入部 ;230 :公开参数存储部 ;240 :差异信息生成部 ;250 :差异信息编码部 ;260 :数据加密部 ;270 :加密标签生成部 ;280 :带标签加密数据生成部 ;300 :检索装置 ;310 :差异信息输入部 ;320 :用户密钥存储部 ;330 :差异信息编码部 ;340 :检索查询生成部 ;350 :基因信息取得部 ;360 :数据解密部 ;400 :数据中心 ;410 :保管请求处理部 ;420 :加密数据存储部 ;430 :加密标签存储部 ;440 :检索请求处理部 ;450 :公开参数存储部 ;460 :访问权限存储部 ;500 :网络。

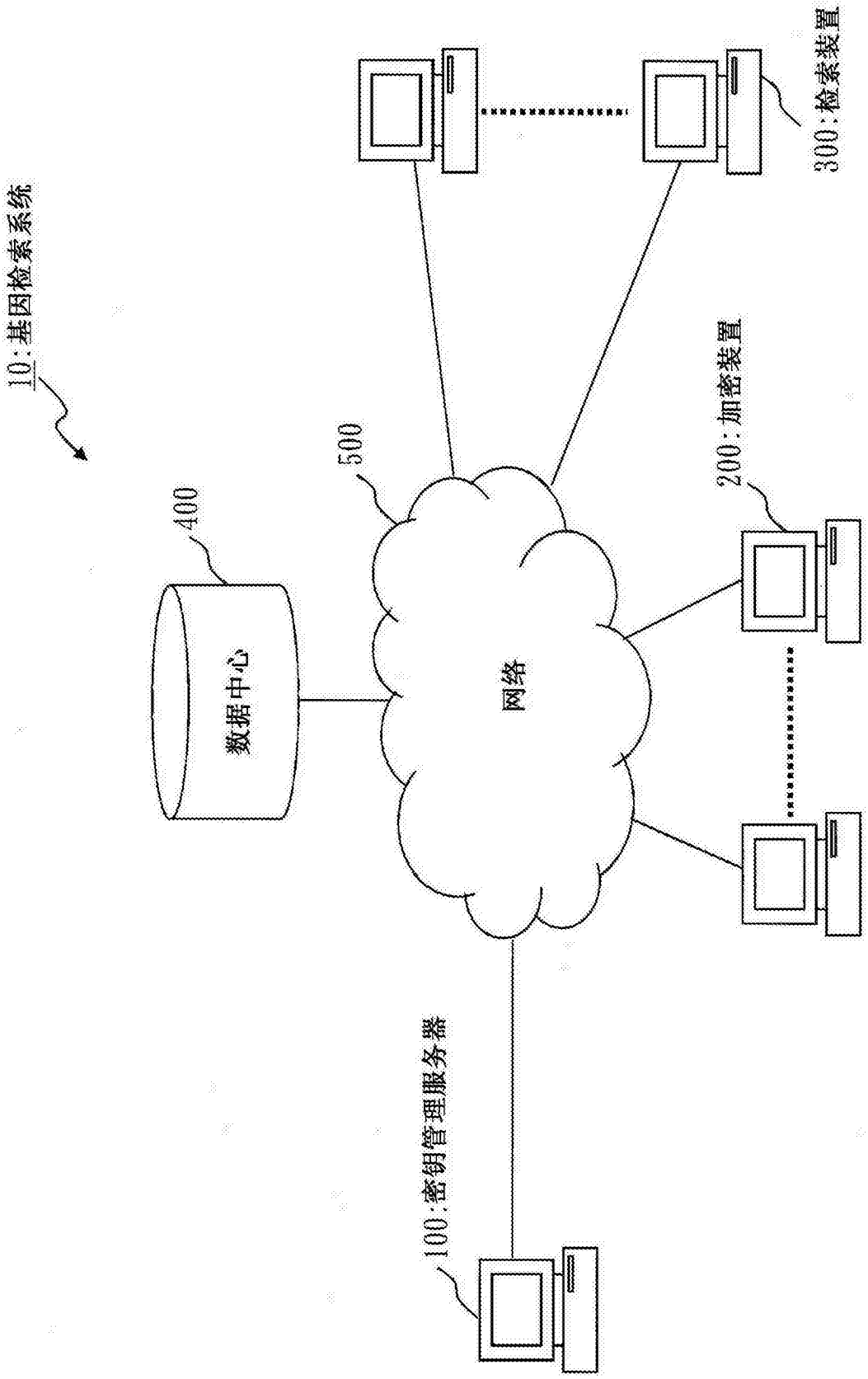


图 1

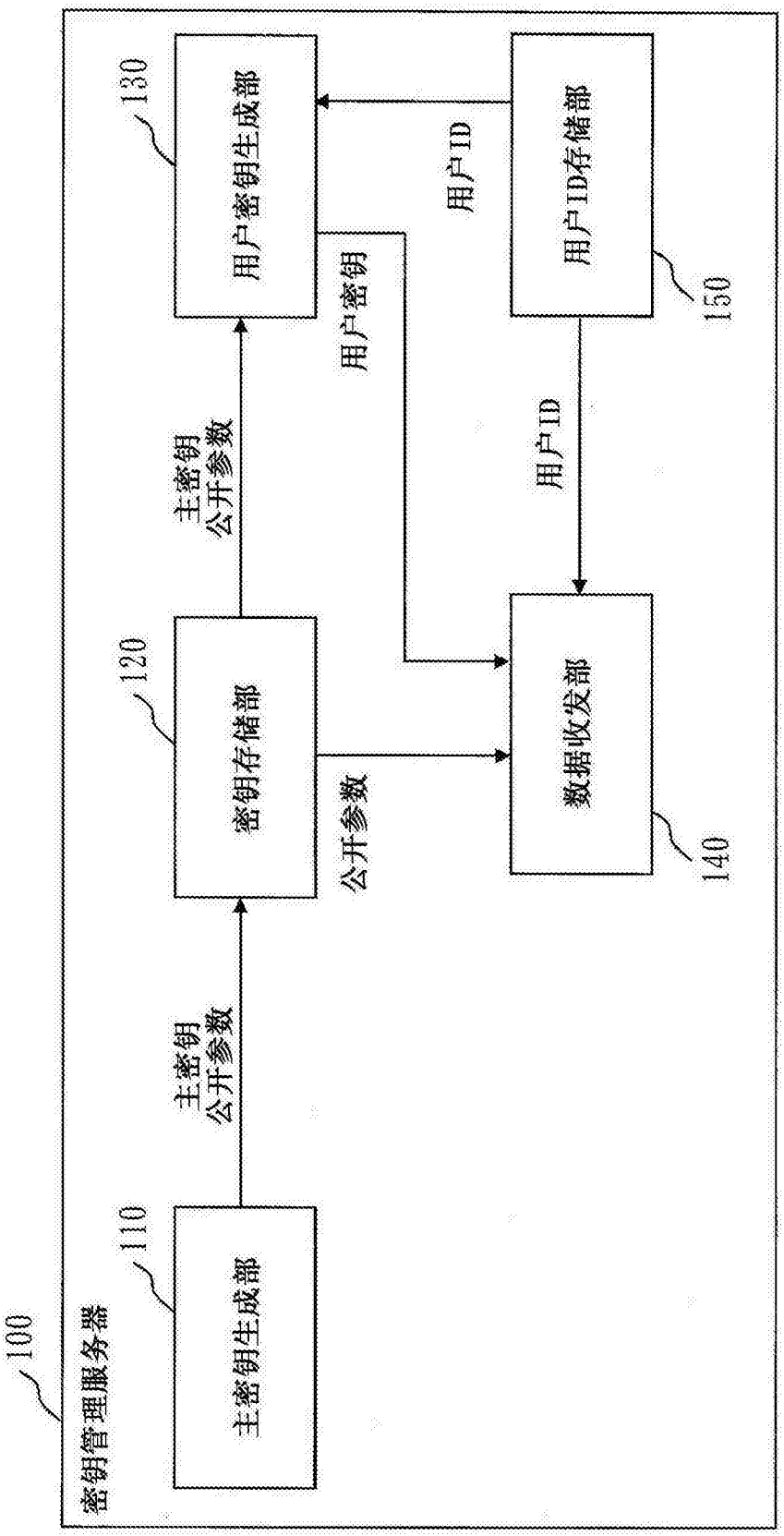

密钥管理服务器
100
主密钥生成部
110
密钥存储部
120
用户密钥生成部
130
数据收发部
140
用户ID存储部
150
主密钥
公开参数
主密钥
公开参数
公开参数
用户密钥
用户ID
用户ID

图 2

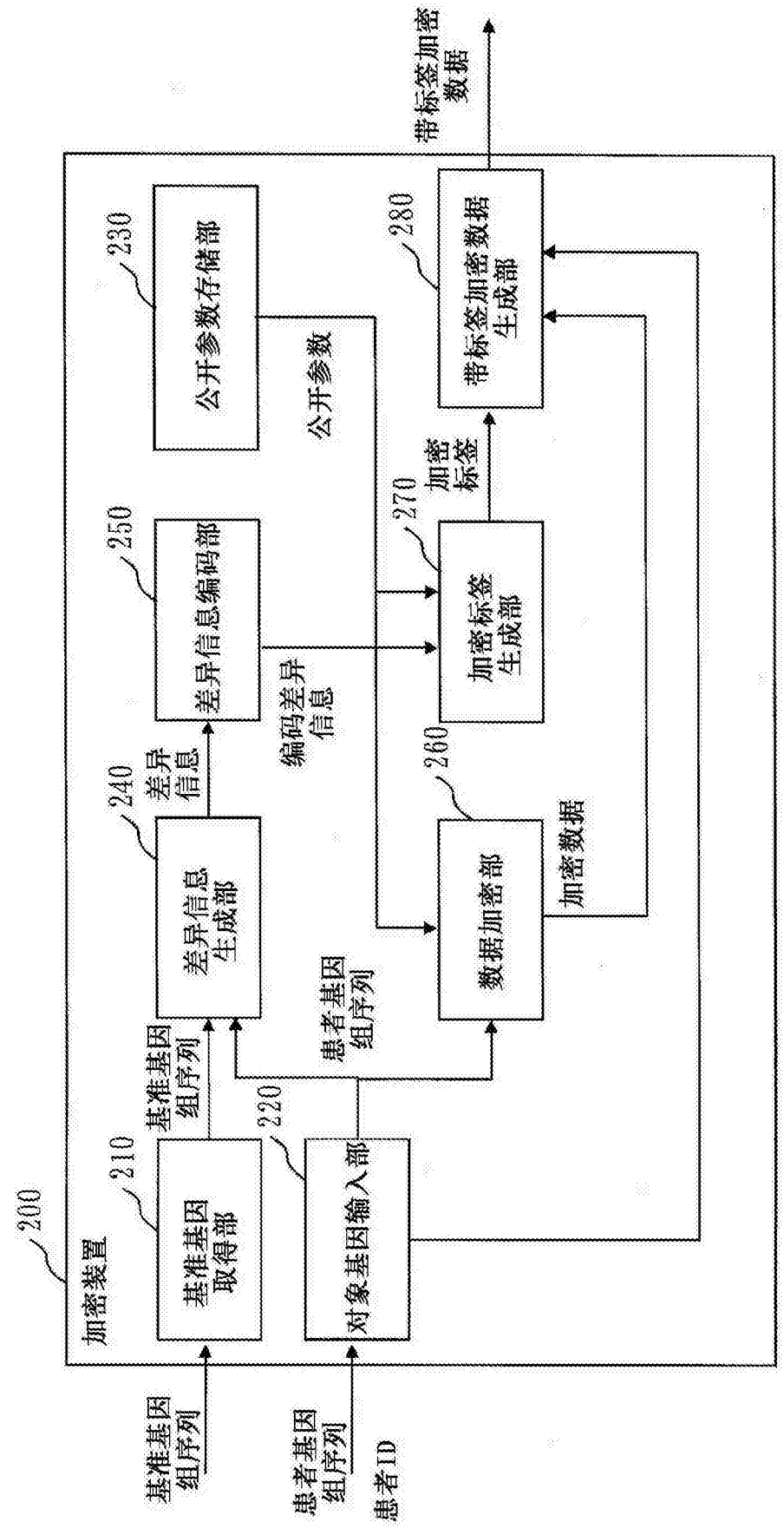

200
加密装置
210
基准基因取得部
220
对象基因输入部
230
公开参数存储部
240
差异信息生成部
250
差异信息编码部
260
数据加密部
270
加密标签生成部
280
带标签加密数据生成部
基准基因组序列
患者基因组序列
患者ID
基准基因组序列
患者基因组序列
差异信息
编码差异信息
公开参数
加密标签
加密数据
带标签加密数据

图 3

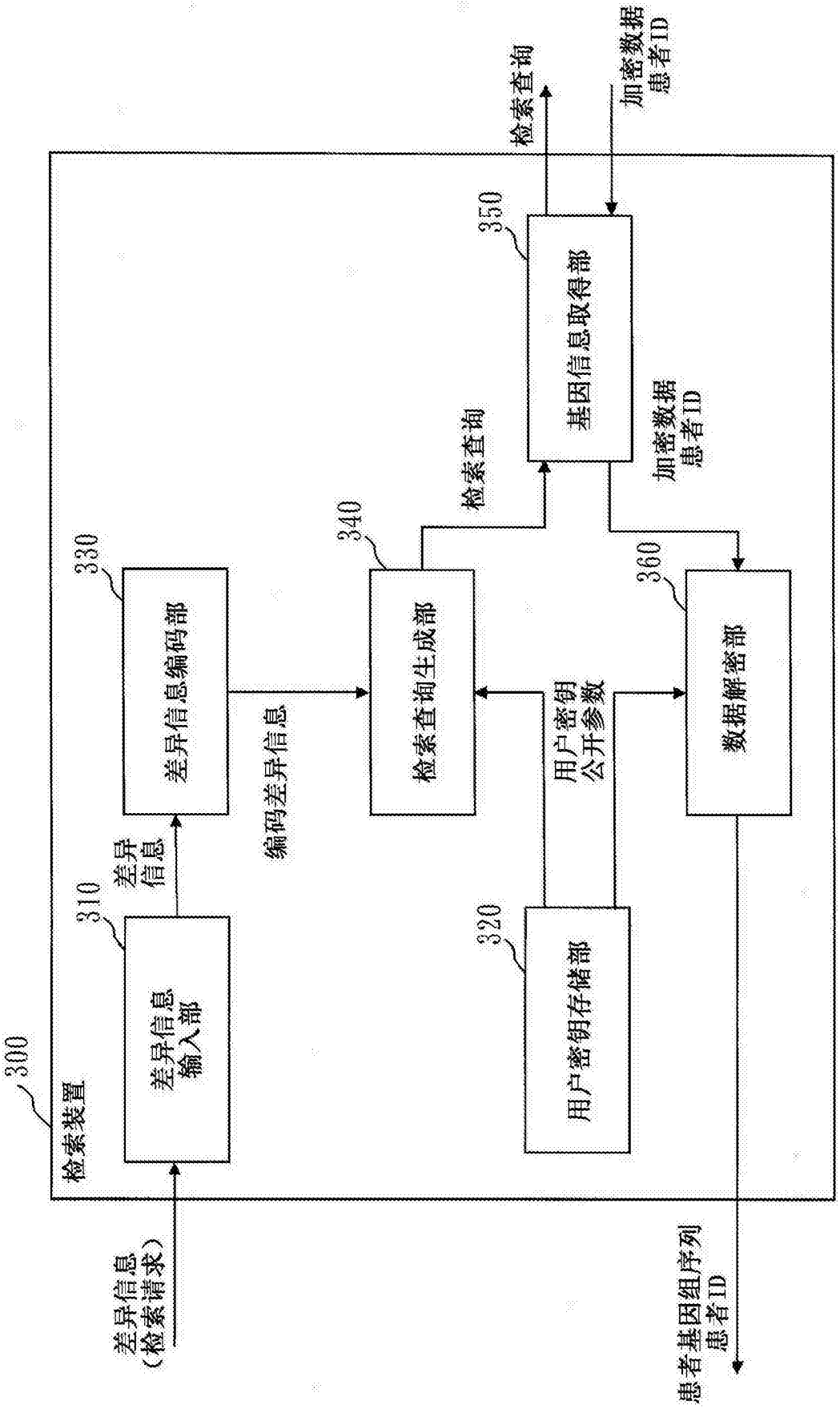

300
检索装置
差异信息输入部
310
差异信息
差异信息编码部
330
编码差异信息
检索查询生成部
340
检索查询
基因信息取得部
350
检索查询
加密数据患者ID
用户密钥存储部
320
用户密钥
公开参数
数据解密部
360
加密数据患者ID
差异信息（检索请求）
患者基因组序列患者ID

图 4

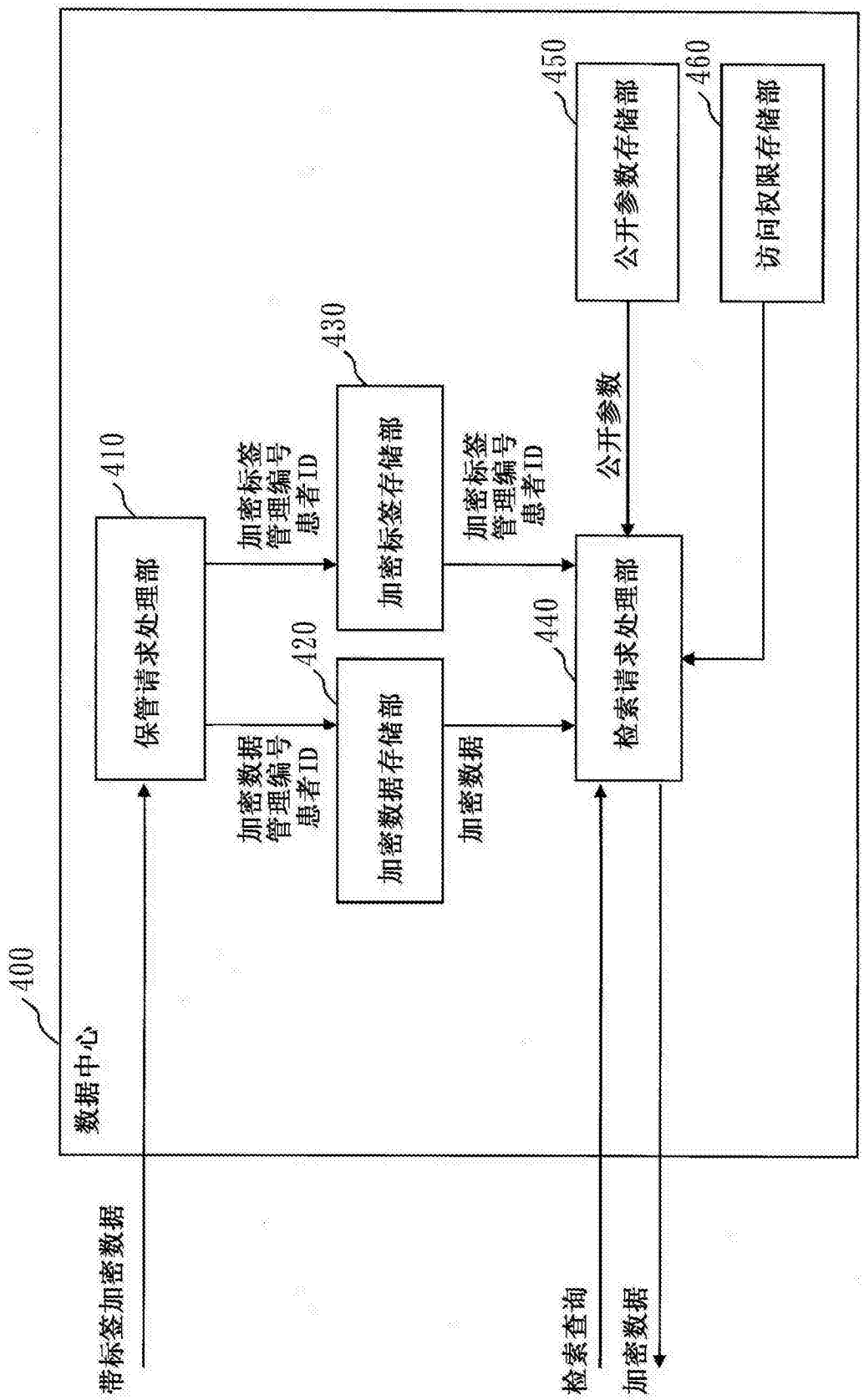

400
数据中心
带标签加密数据
410
保管请求处理部
加密数据
管理编号
患者ID
加密标签
管理编号
患者ID
420
加密数据存储部
430
加密标签存储部
440
检索请求处理部
检索查询
加密数据
公开参数
450
公开参数存储部
460
访问权限存储部

图 5

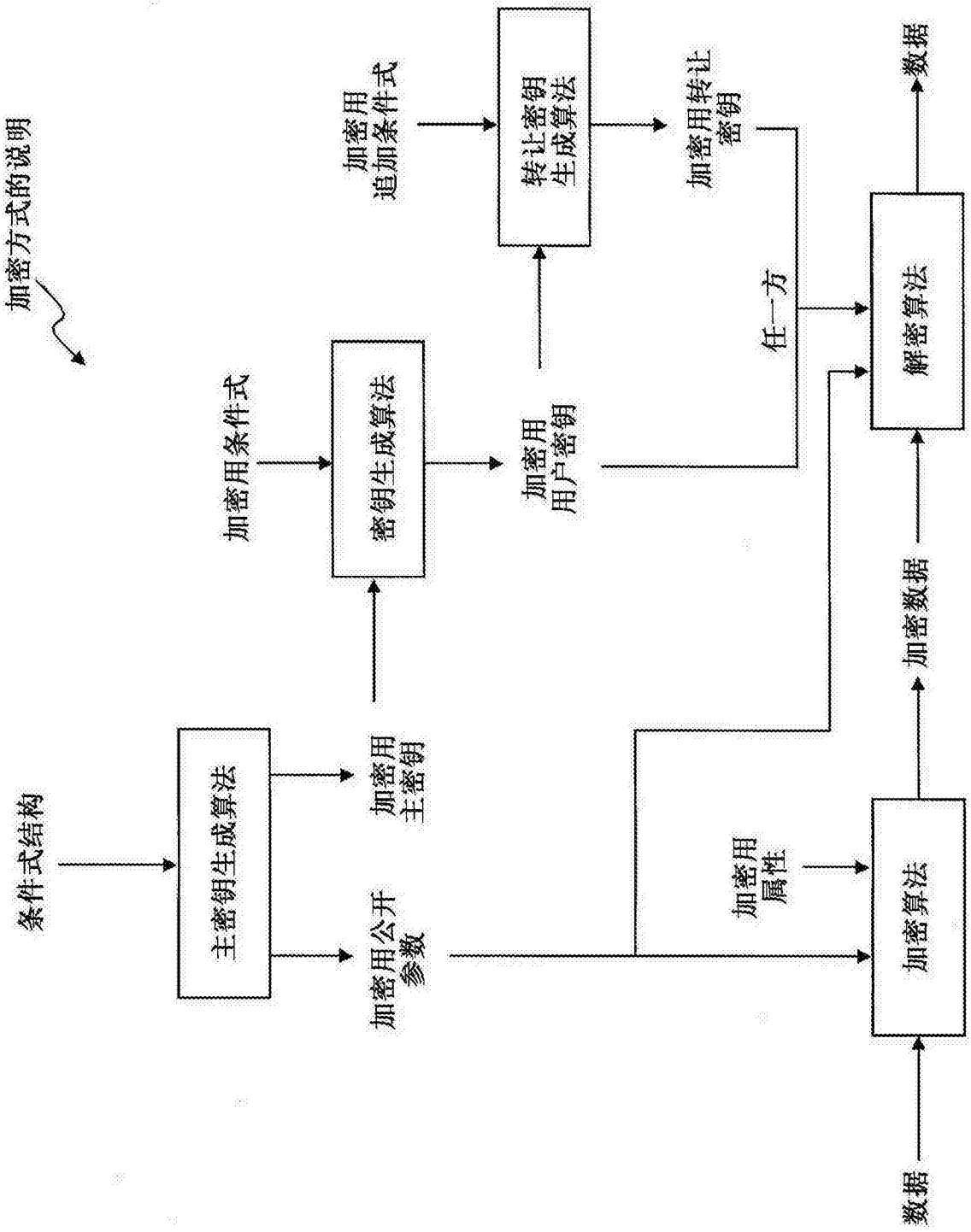

加密方式的说明
条件式结构
主密钥生成算法
加密用主密钥
加密用公开参数
加密用条件式
密钥生成算法
加密用用户密钥
加密用追加条件式
转让密钥生成算法
加密用转让密钥
任一方
加密用属性
加密算法
加密数据
解密算法
数据
数据

图 6

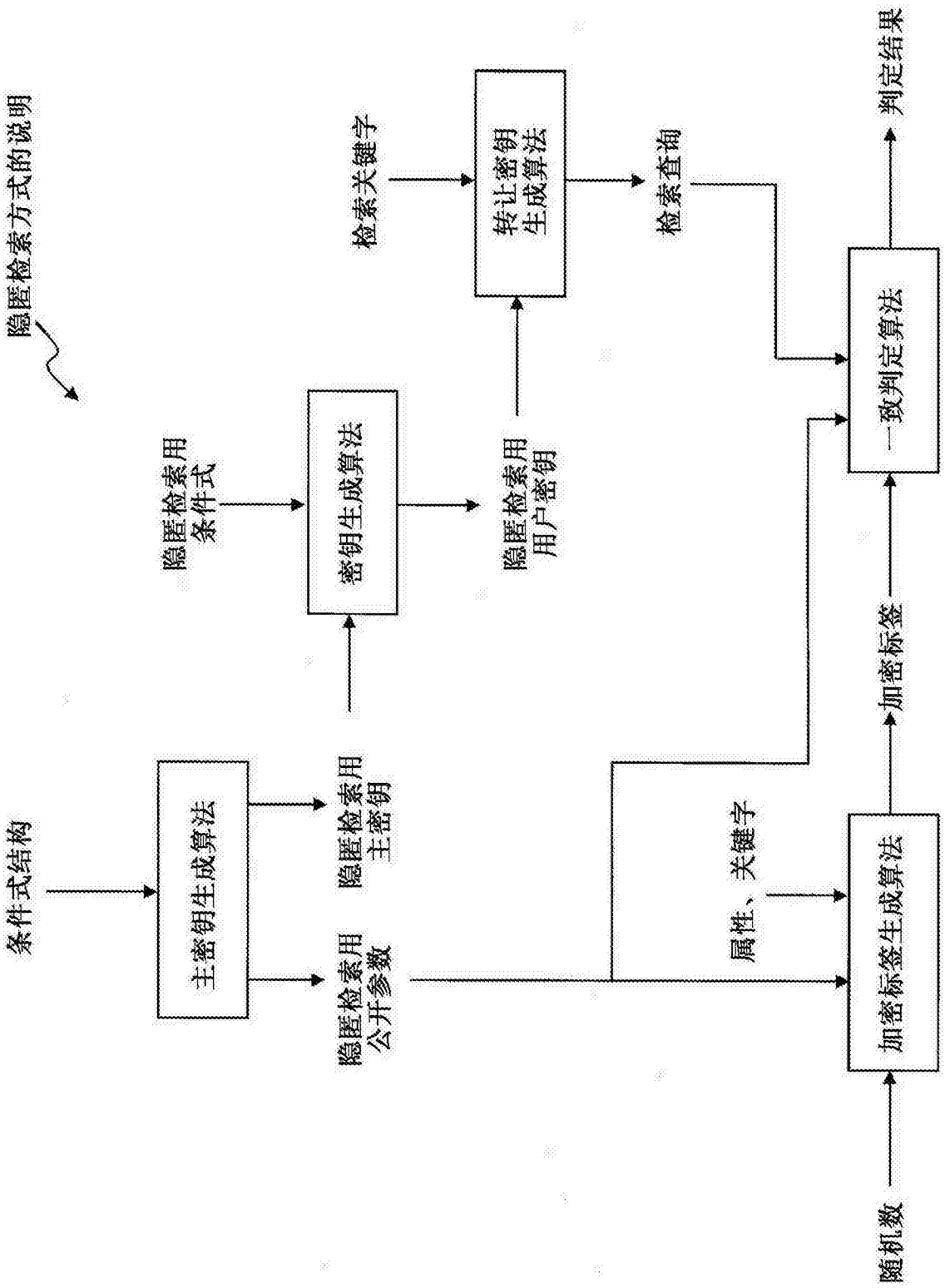

隐匿检索方式的说明
条件式结构
主密钥生成算法
隐匿检索用主密钥
隐匿检索用公开参数
隐匿检索用条件式
密钥生成算法
隐匿检索用户密钥
检索关键字
转让密钥生成算法
检索查询
属性、关键字
加密标签生成算法
加密标签
一致判定算法
判定结果
随机数

图 7

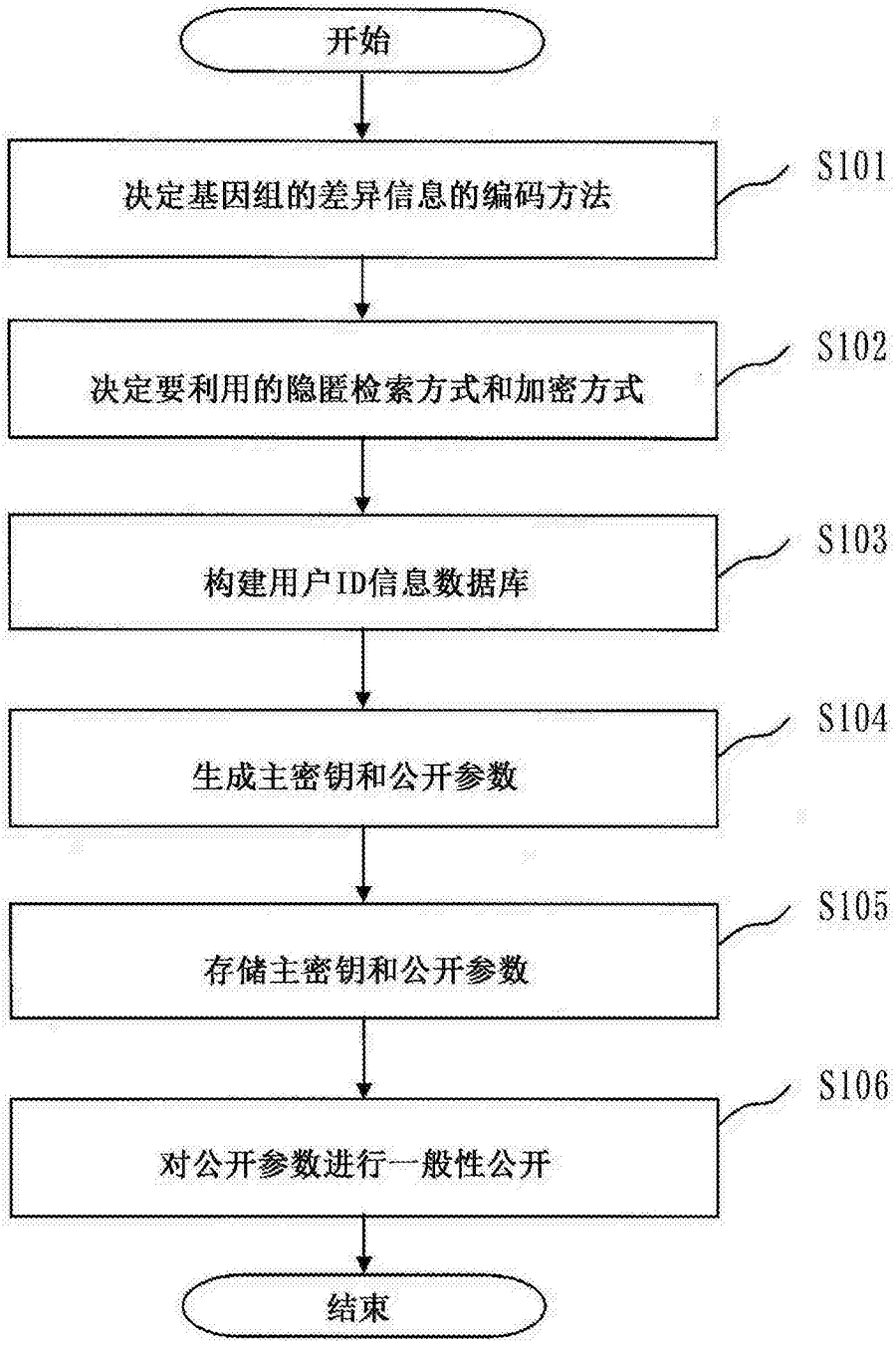

开始
决定基因组的差异信息的编码方法
S101
决定要利用的隐匿检索方式和加密方式
S102
构建用户ID信息数据库
S103
生成主密钥和公开参数
S104
存储主密钥和公开参数
S105
对公开参数进行一般性公开
S106
结束

图 8

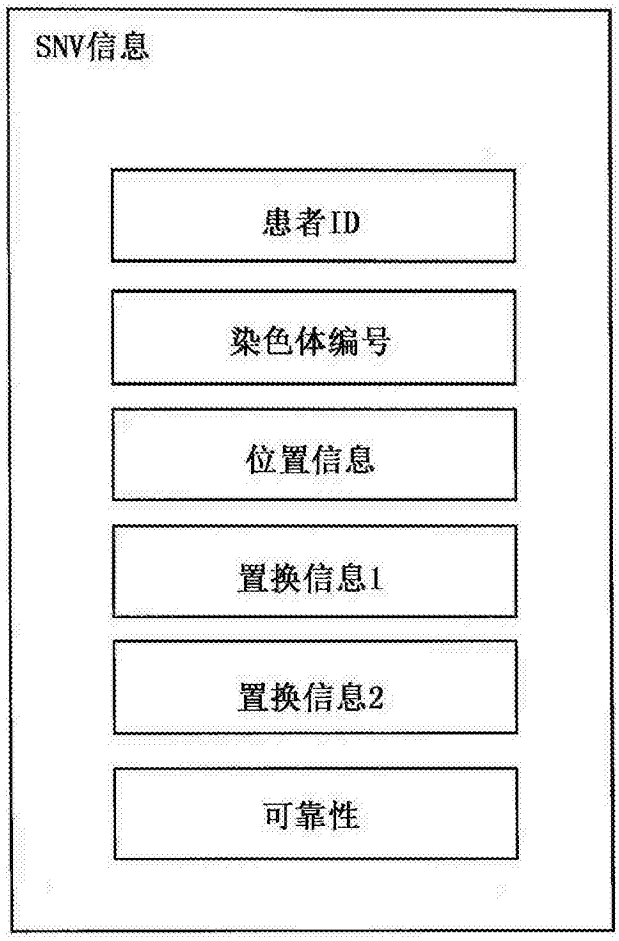

SNV信息
患者ID
染色体编号
位置信息
置换信息1
置换信息2
可靠性

图 9

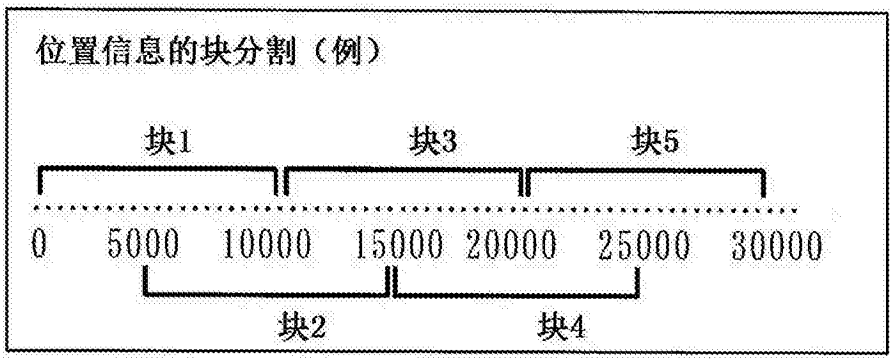

位置信息的块分割（例）
块1
块3
块5
0
5000
10000
15000
20000
25000
30000
块2
块4

图 10

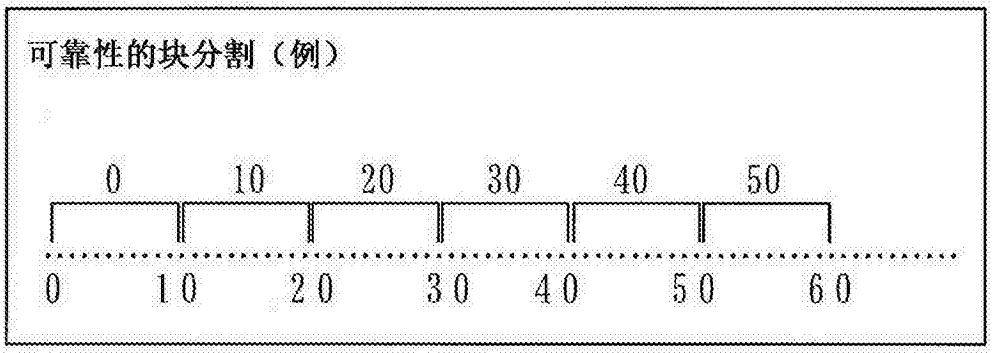

可靠性的块分割（例）
0
10
20
30
40
50
0
10
20
30
40
50
60

图 11

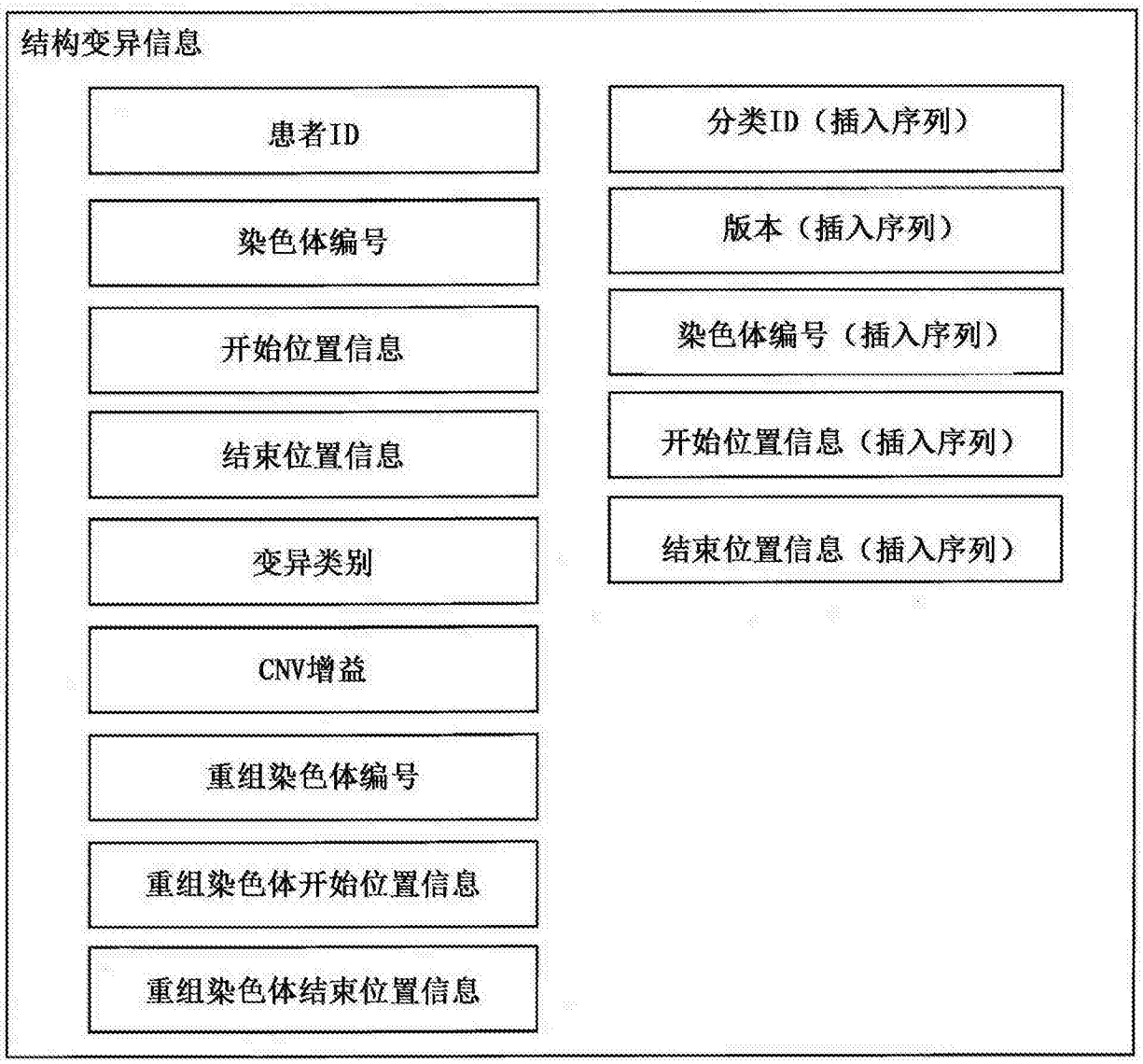

结构变异信息
患者ID
染色体编号
开始位置信息
结束位置信息
变异类别
CNV增益
重组染色体编号
重组染色体开始位置信息
重组染色体结束位置信息
分类ID（插入序列）
版本（插入序列）
染色体编号（插入序列）
开始位置信息（插入序列）
结束位置信息（插入序列）

图 12

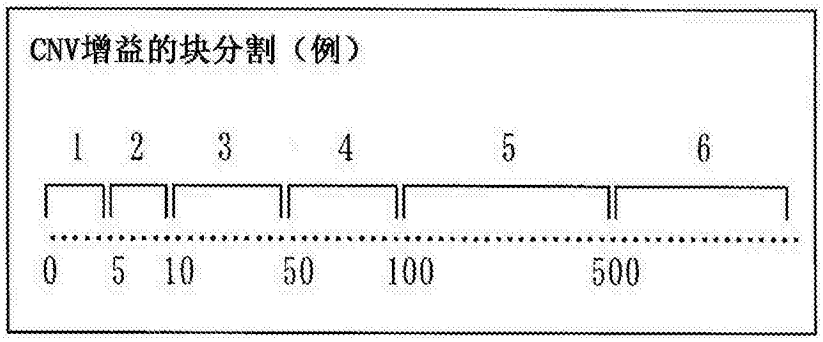

CNV增益的块分割（例）
1
2
3
4
5
6
0
5
10
50
100
500

图 13

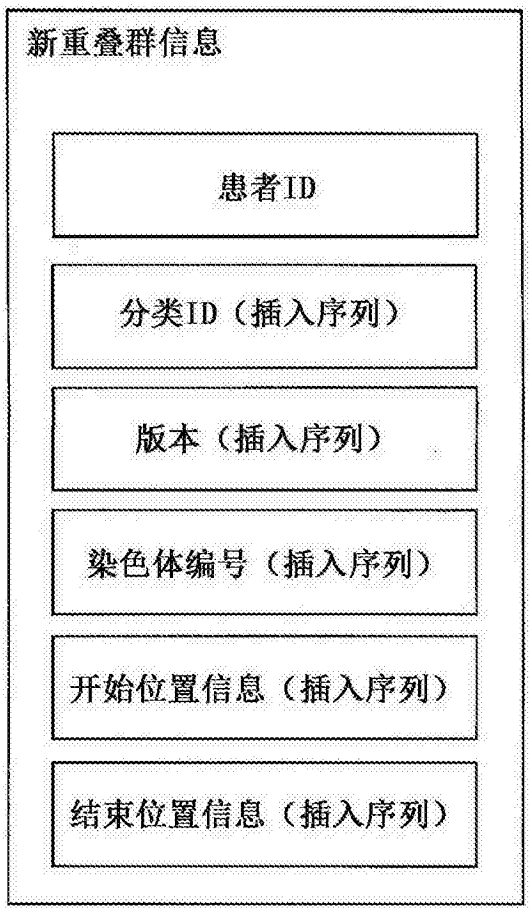

新重叠群信息
患者ID
分类ID（插入序列）
版本（插入序列）
染色体编号（插入序列）
开始位置信息（插入序列）
结束位置信息（插入序列）

图 14

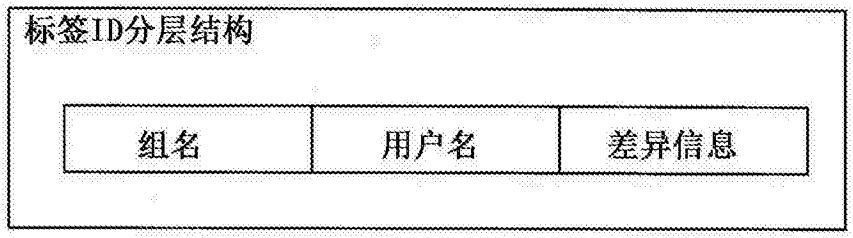


图 15

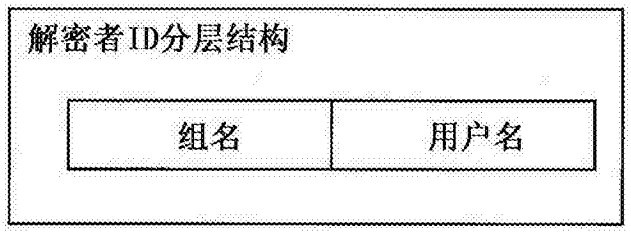


图 16

用户ID信息数据库

组名	用户名	部门信息	有效期间
A制药公司	田中太郎	研究所/ 感冒药部	2005/4～
⋮	⋮	⋮	⋮

图 17

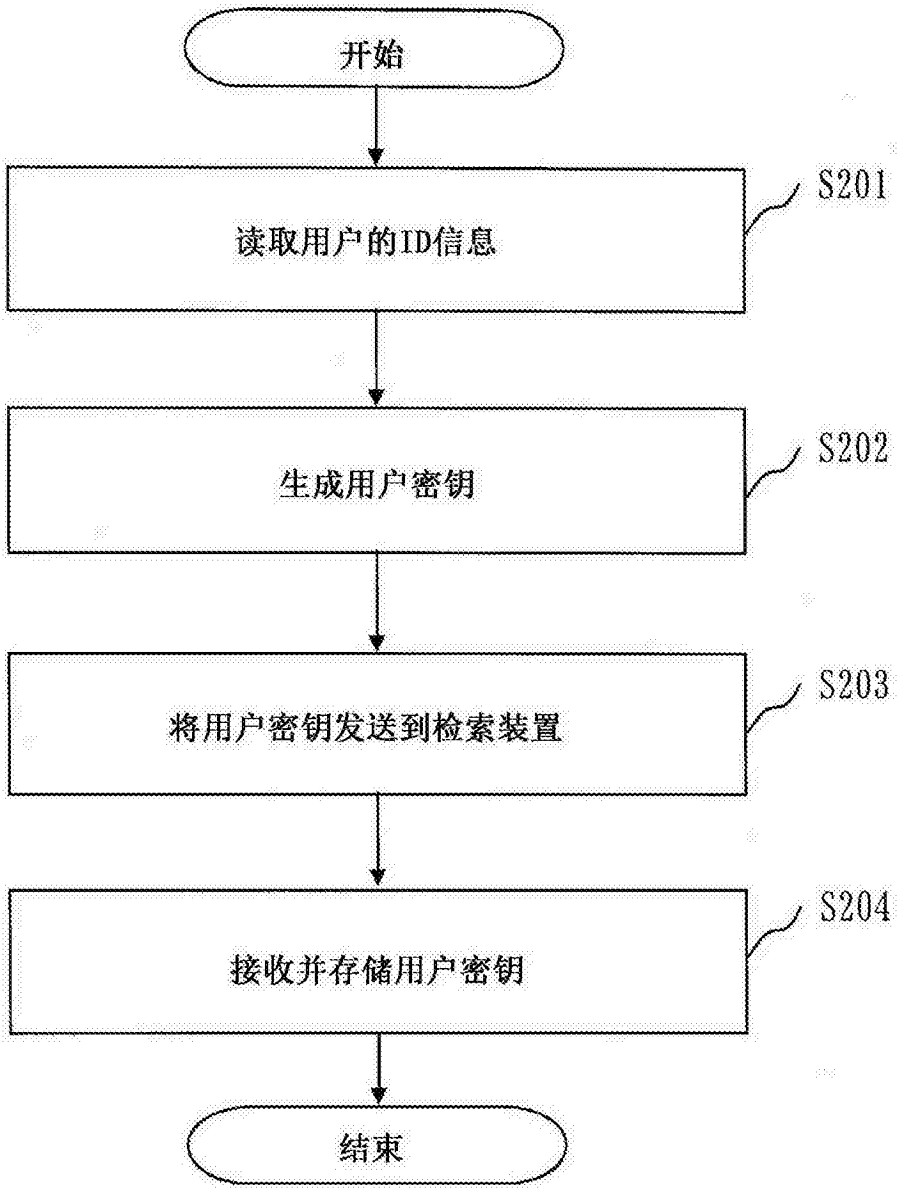

开始
读取用户的ID信息
S201
生成用户密钥
S202
将用户密钥发送到检索装置
S203
接收并存储用户密钥
S204
结束

图 18

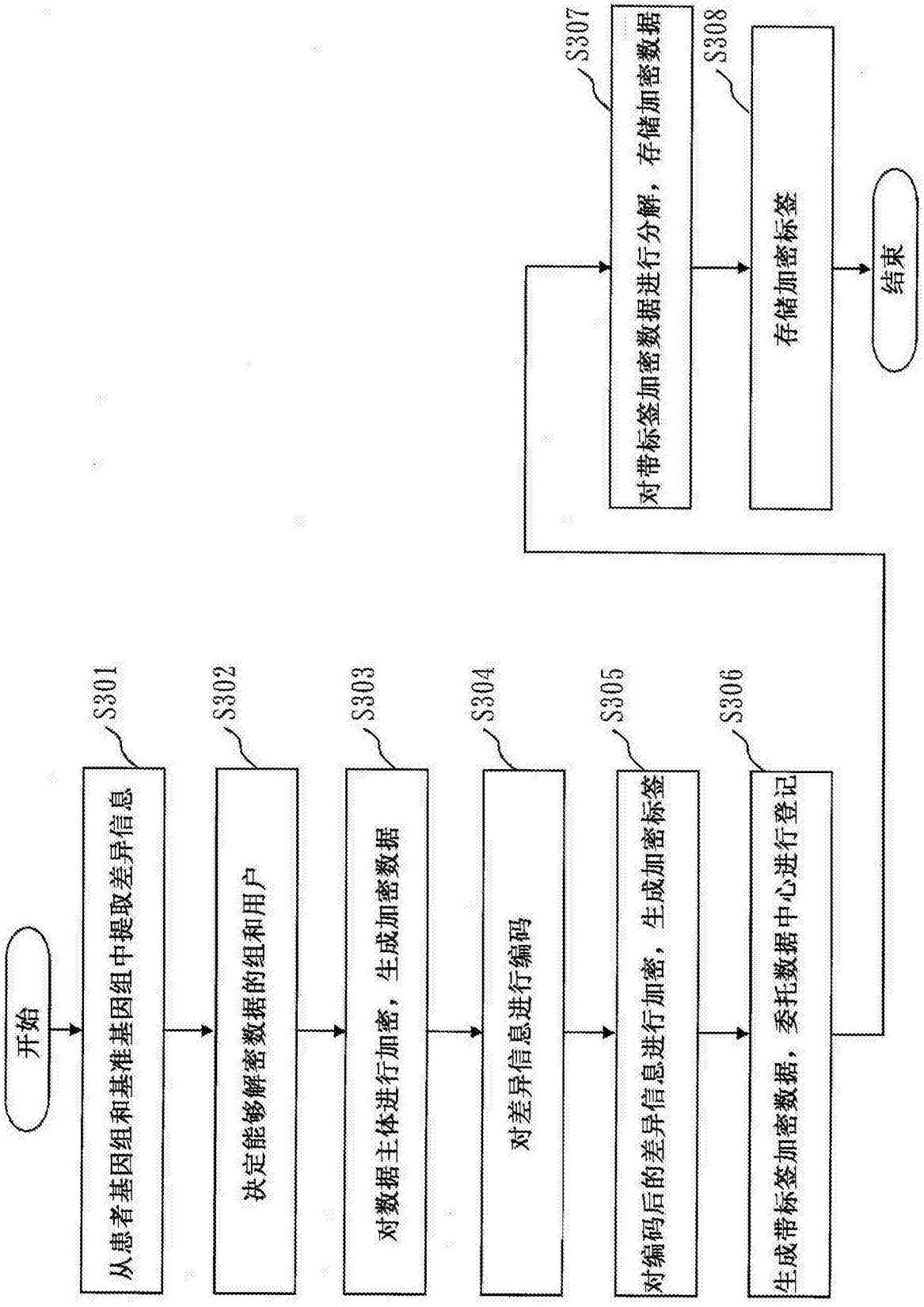

开始
S301
从患者基因组和基准基因组中提取差异信息
S302
决定能够解密数据的组和用户
S303
对数据主体进行加密，生成加密数据
S304
对差异信息进行编码
S305
对编码后的差异信息进行加密，生成加密标签
S306
生成带标签加密数据，委托数据中心进行登记
S307
对带标签加密数据进行分解，存储加密数据
S308
存储加密标签
结束

图 19

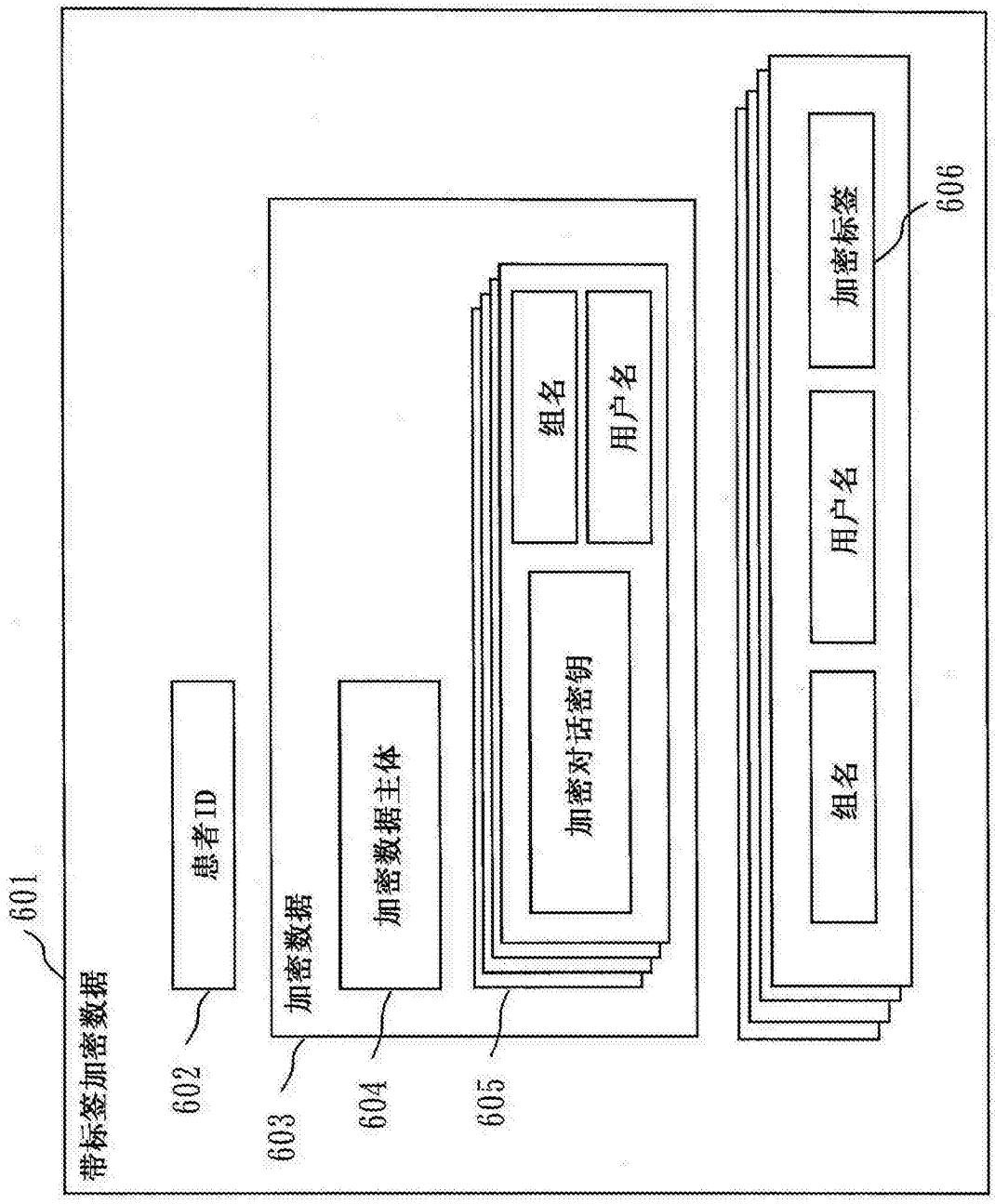

601
带标签加密数据
602
患者ID
603
加密数据
604
加密数据主体
605
加密对话密钥
组名
用户名
606
加密标签
用户名
组名

图 20

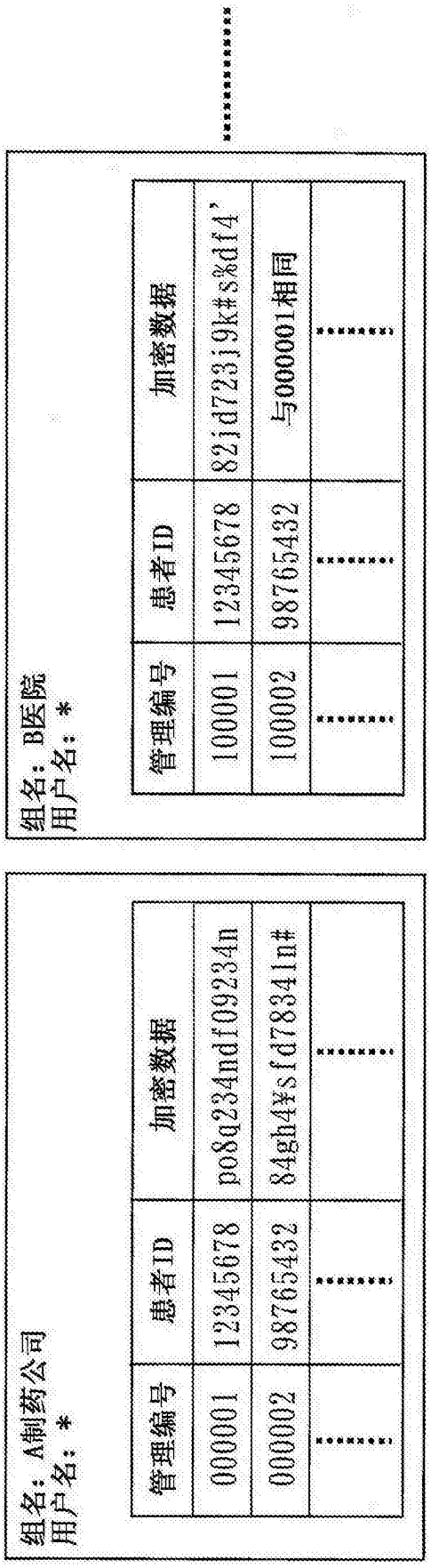

组名：A制药公司
用户名：*
管理编号
患者ID
加密数据
000001
12345678
po8q234ndf09234n
000002
98765432
84gh4¥sfd78341n#
组名：B医院
用户名：*
管理编号
患者ID
加密数据
100001
12345678
82jd723j9k#s%df4'
100002
98765432
与000001相同

图 21

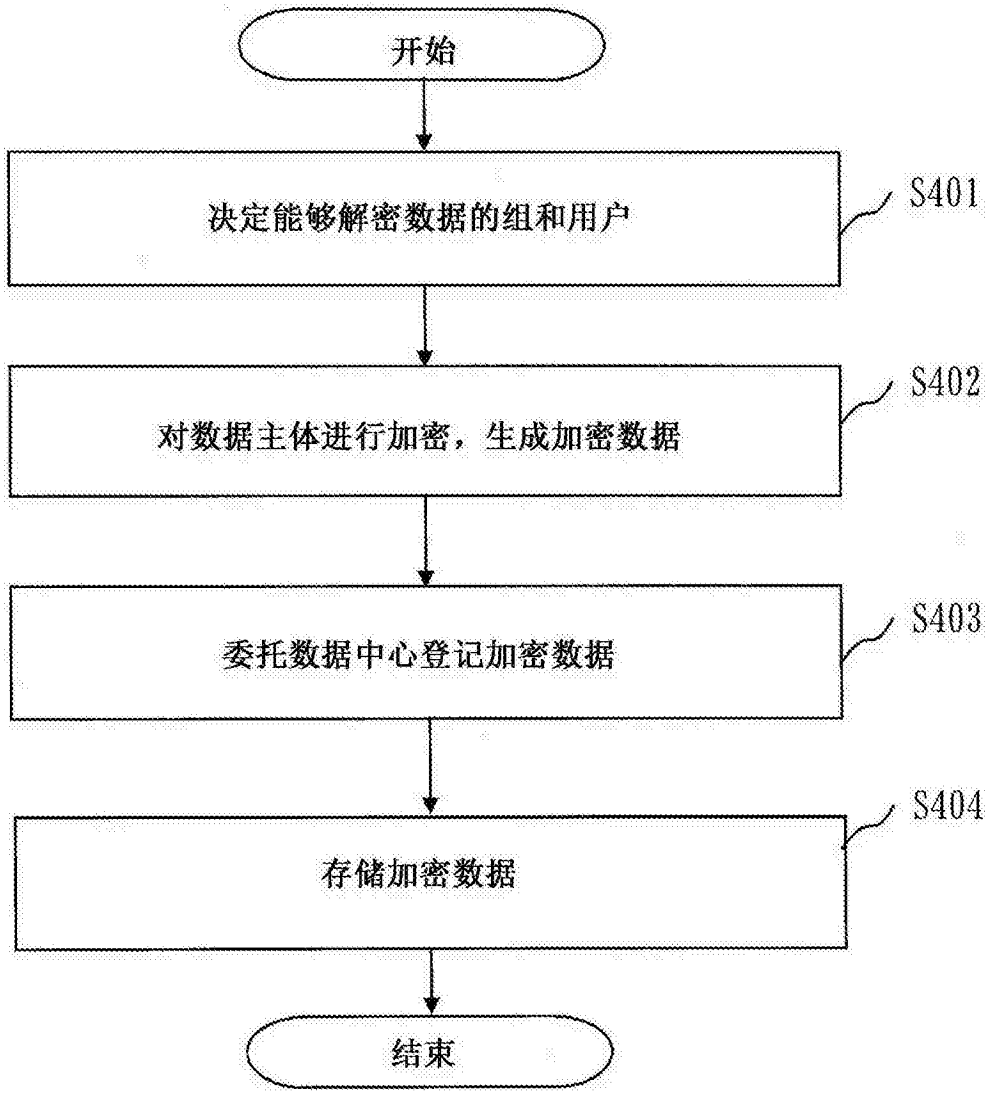

开始
决定能够解密数据的组和用户
S401
对数据主体进行加密，生成加密数据
S402
委托数据中心登记加密数据
S403
存储加密数据
S404
结束

图 22

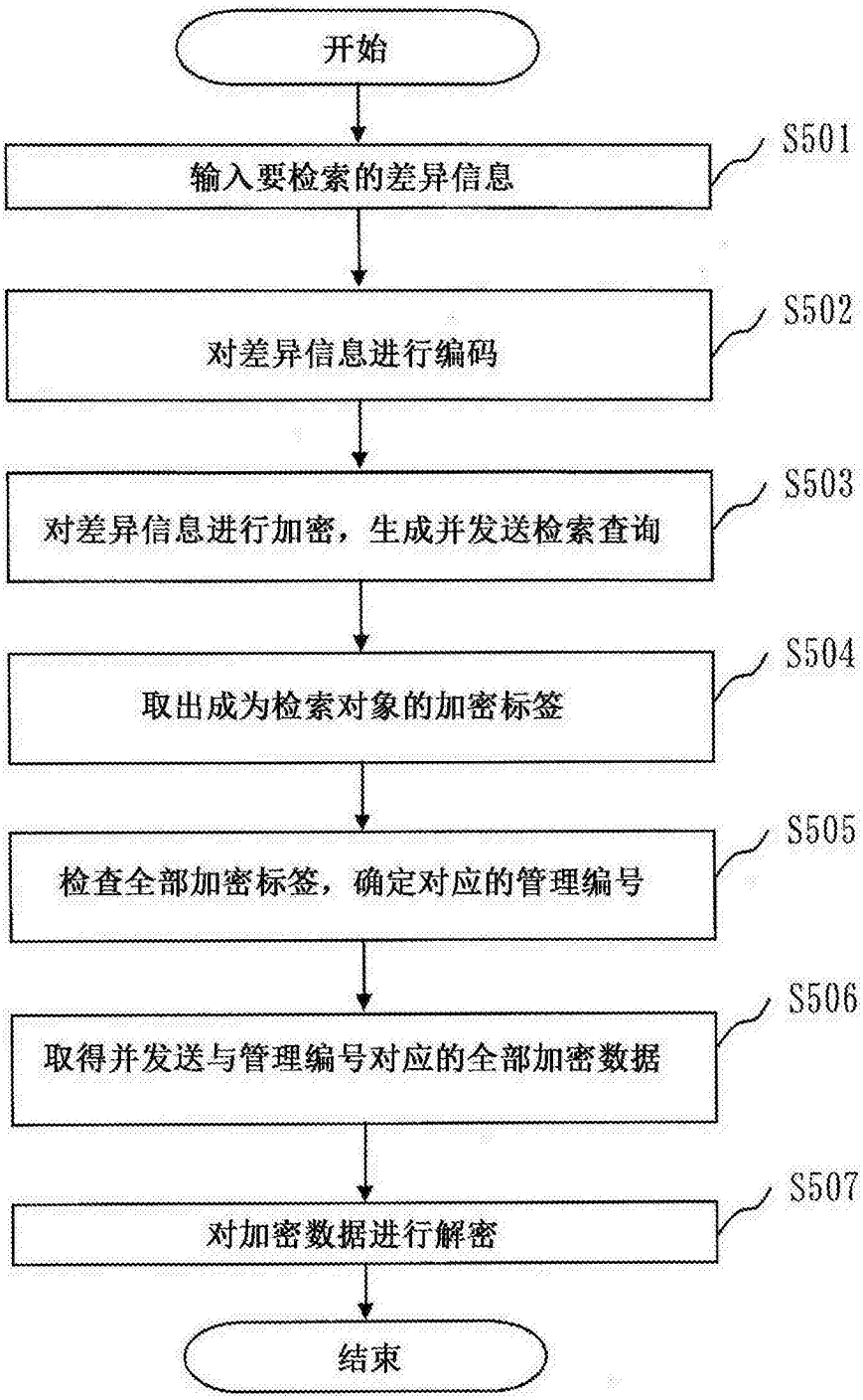

开始
输入要检索的差异信息
S501
对差异信息进行编码
S502
对差异信息进行加密，生成并发送检索查询
S503
取出成为检索对象的加密标签
S504
检查全部加密标签，确定对应的管理编号
S505
取得并发送与管理编号对应的全部加密数据
S506
对加密数据进行解密
S507
结束

图 23

访问权限管理表

访问者信息	患者ID	数据类别	允许的操作
A制药公司/田中	12345678	全部	检索/阅览
B医院/高桥	98765432	电子病历	阅览
⋮	⋮	⋮	⋮

图 24

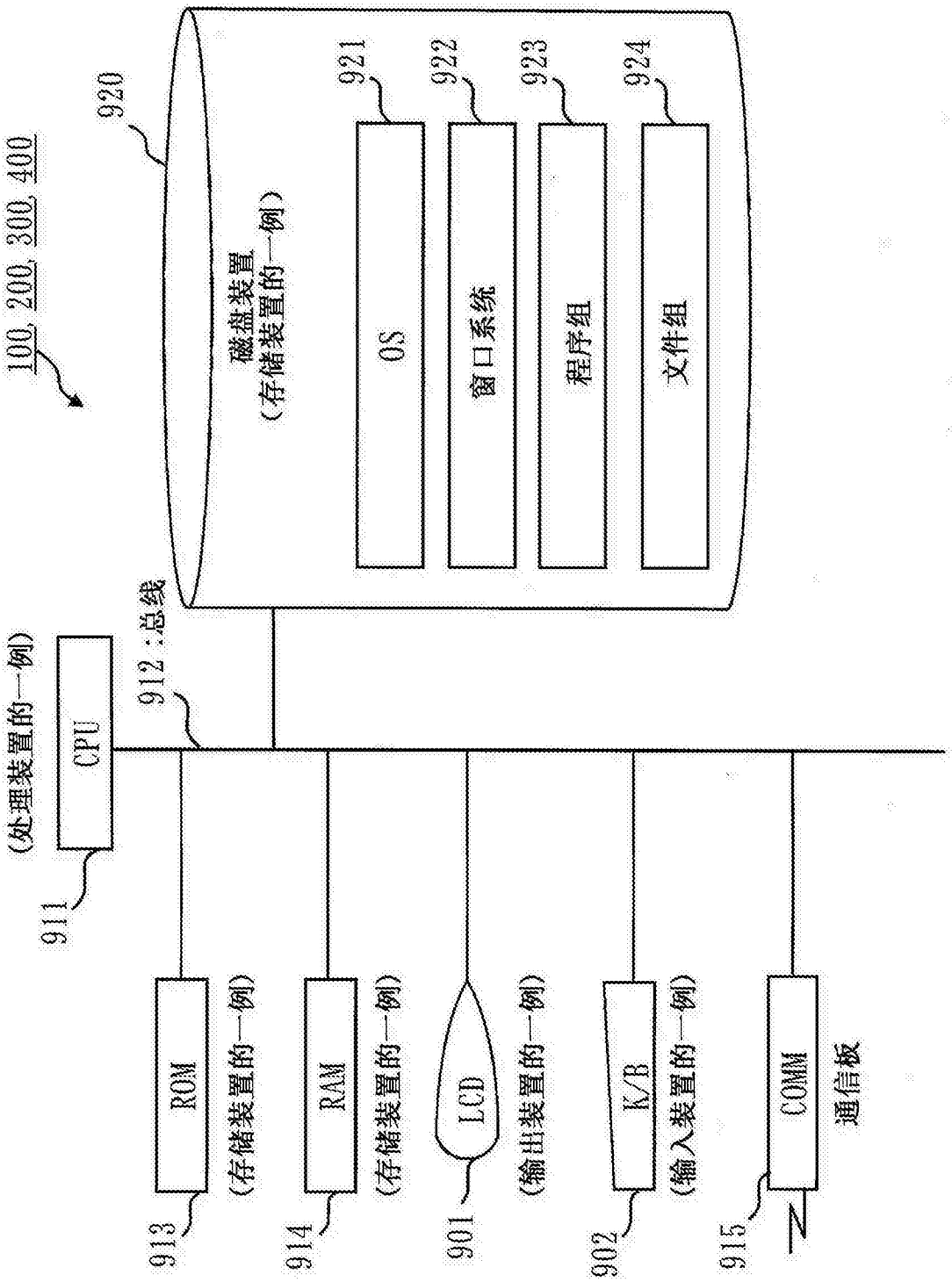

100, 200, 300, 400
920
磁盘装置
(存储装置的一例)
921
OS
922
窗口系统
923
程序组
924
文件组
912 :总线
(处理装置的一例)
CPU
911
ROM
(存储装置的一例)
913
RAM
(存储装置的一例)
914
LCD
(输出装置的一例)
901
K/B
(输入装置的一例)
902
COMM
通信板
915

图 25