



(12) 发明专利

(10) 授权公告号 CN 107851135 B

(45) 授权公告日 2022.03.01

(21) 申请号 201680041161.2

(72) 发明人 H·林 S·卡玛拉卡兰

(22) 申请日 2016.07.11

(74) 专利代理机构 永新专利商标代理有限公司

(65) 同一申请的已公布的文献号

申请公布号 CN 107851135 A

(43) 申请公布日 2018.03.27

(51) Int.CI.

G16B 20/00 (2019.01)

(30) 优先权数据

G16B 20/50 (2019.01)

62/191,684 2015.07.13 US

G16B 10/00 (2019.01)

(85) PCT国际申请进入国家阶段日

G16B 20/20 (2019.01)

2018.01.12

G16H 50/30 (2018.01)

(86) PCT国际申请的申请数据

(56) 对比文件

PCT/IB2016/054139 2016.07.11

CN 104010570 A, 2014.08.27

(87) PCT国际申请的公布数据

US 2008206767 A1, 2008.08.28

W02017/009770 EN 2017.01.19

CN 1612936 A, 2005.05.04

(73) 专利权人 皇家飞利浦有限公司

CN 103975328 A, 2014.08.06

地址 荷兰艾恩德霍芬

审查员 王敏

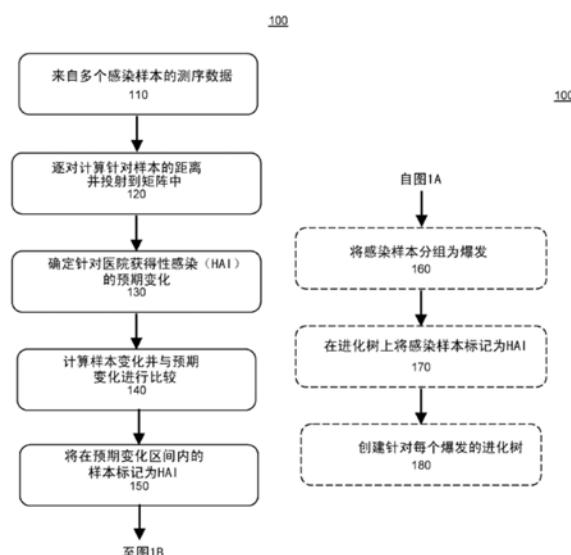
权利要求书2页 说明书6页 附图3页

(54) 发明名称

追踪医院环境中的感染

(57) 摘要

本发明提供了一种用于识别医院获得性感染的方法。所述方法包括：计算感染样本与多个感染样本的至少一个子集之间的随时间的变化数目，并且确定所述随时间的变化数目是否在预期变化数目的区间内。如果是这样，则将该感染样本标记为医院获得性感染。



1. 一种用于追踪感染的扩散的方法,所述方法包括:

在多个感染样本中,针对每个感染样本对数据进行测序;

通过使用包含针对不同样本随时间的变化的样本数据并且计算预测区间来确定针对病原生物的随时间的预期变化范围,其中,所述预测区间界定感染传播的所述预期变化范围;

针对至少一个感染样本,计算该样本与至少一个额外的感染样本之间的随时间的突变变化数目;

针对计算出的随时间的突变变化数目中的每个,确定样本之间的所述随时间的突变变化数目是否在所述预测区间内;并且

针对在所述预测区间内的所述计算出的随时间的突变变化数目中的每个,将相关联的样本对标记为潜在的传播。

2. 根据权利要求1所述的方法,其中,所述预期变化范围是通过以下操作来确定的:

构建测序样本的系统发育以识别密切相关的样本;并且

针对所述多个感染样本中的至少一个,计算所述样本与如由所述系统发育确定的至少一个紧密相关的样本之间的随时间的变化数目。

3. 根据权利要求1所述的方法,其中,所述预期变化范围是根据随时间从同一患者取得的多个样本来确定的,或者是根据从相互之间可能存在传播的两个患者取得的多个样本来确定的。

4. 根据权利要求2所述的方法,其中,构建所述系统发育包括逐对计算每对测序样本之间的距离并使用距离矩阵来构建进化树。

5. 根据权利要求1所述的方法,其中,感染样本测序数据是通过全基因组测序或靶向测序获得的。

6. 根据权利要求1所述的方法,其中,计算出的样本之间的突变变化数目通过除以两个样本都具有碱基调用的基因组位置的总数而被归一化。

7. 根据权利要求4所述的方法,还包括:

通过对所述样本的所述进化树进行广度优先搜索遍历变化数目在所述预期变化范围内的边来确定所述感染样本中的哪些感染样本来自相同的爆发。

8. 根据权利要求7所述的方法,还包括:

在经测序的感染样本的显示的进化树上标记被识别为潜在的传播的感染样本对。

9. 根据权利要求8所述的方法,还包括:

创建多个进化树,每个树由被标记为来自相同的爆发的样本构建。

10. 一种用于追踪感染的扩散的系统,所述系统包括:

接收器,其用于接收感染样本测序数据;

存储器,其被配置为存储所接收的感染样本测序数据;以及

处理器,其被配置为:

通过使用包含针对不同样本随时间的变化的样本数据并且计算预测区间来确定针对病原生物的随时间的预期变化范围,其中,所述预测区间界定感染传播的所述预期变化范围;

针对至少一个感染样本,计算该样本与至少一个额外的感染样本之间的随时间的突变

变化数目；

针对计算出的随时间的突变变化数目中的每个，确定样本之间的所述随时间的突变变化数目是否在所述预测区间内；并且

针对在所述预测区间内的所述计算出的随时间的突变变化数目中的每个，将相关联的样本对标记为潜在的传播。

11. 根据权利要求10所述的系统，其中，所配置的处理器通过以下操作来确定所述预期变化范围：

构建测序样本的系统发育以识别密切相关的样本；并且

针对多个感染样本中的至少一个，计算所述样本与由所述系统发育确定的至少一个紧密相关的样本之间的随时间的变化数目。

12. 根据权利要求11所述的系统，其中，所配置的处理器通过逐对计算每对测序样本之间的距离并使用距离矩阵构建进化树来构建所述系统发育。

13. 根据权利要求12所述的系统，其中，所述处理器还被配置为：

通过对所述样本的所述进化树进行广度优先搜索遍历变化数目在所述预期变化范围内的边来确定所述感染样本中的哪些感染样本来自相同的爆发。

追踪医院环境中的感染

技术领域

[0001] 本文描述的各个实施例涉及与医疗保健相关联的感染，并且更具体地但不排他地涉及用于追踪医院获得性感染的扩散的方法和装置。

背景技术

[0002] 与医疗保健相关联的感染(HAI)是患者在针对另一种病症的医疗保健治疗期间获得的感染。医学文献中的HAI通常被称为医院感染。HAI可能是致命的，并且在医院中经常发生。它们常常是因为细菌或真菌。每20名住院患者中就有近1人将患HAI。2002年进行的最新的一项研究估计有近170万人受到HAI。这170万受到HAI的人每年造成或导致近99000例的死亡。欧洲已经发现了类似的趋势，那就是每年25000例的死亡中，三分之二的人受到HAI感染。

[0003] 医院感染会导致严重的肺炎和尿路、血流和身体其他部位的感染。许多类型难以用抗生素治疗，并且抗生素耐药性正在扩散到能够感染医院环境外的人群的革兰氏阴性菌。在美国，全院最常见的感染类型是尿路感染(36%)，其次是手术部位感染(20%)、血流感染和肺炎(二者均为11%)。

[0004] 根据20世纪70年代中期进行的一项研究，“医院感染控制效力研究”(SENIC)，在1992年确定了HAI的显著经济后果。在公布时，医疗保健的直接成本估计为66.5亿美元(使用2007年的CPI调整通货膨胀)。然而，最近公布的证据显示，今天的直接成本在284亿美元到338亿美元之间。这些费用大部分与患者住院时间较长、隔离医院部分以及发现并根除感染源有关。

发明内容

[0005] 根据前述内容，期望提供用于识别和追踪HAI以促进对其进行控制和根除的方法和装置。

[0006] 应当理解，本文使用的措辞和术语是为了描述而不是限制的目的。

[0007] 本文公开的各个实施例总体上涉及与医疗保健相关联的感染，并且更具体地涉及用于追踪医院获得性感染在医院环境中的扩散的方法、装置以及存储指令的非瞬态机器可读介质，所述指令由处理器运行以用于追踪医院获得性感染在医院环境中的扩散。为此，一些实施例追踪从细菌和真菌获得的与医院相关联的感染，并且以进化的方式将其与其他样本进行比较以帮助对传播的溯源，识别相关的病例等。

[0008] 为了辅助感染控制流程，各个实施例还识别可能在单次爆发(而不是单独的不相关的感染)中扩散的感染样本。如果有多个同一类型感染的患者，那么重要的是要识别哪些感染可能在医院内扩散与哪些感染可能是社区获得性感染(即，在医院环境外传播的感染)。

[0009] 一般而言，公开了追踪感染的扩散的方法和系统。更一般地，来自感染样本的测序数据被用于确定存在的主要病原体(例如，细菌、病毒或真菌)。然后将测序读数映射到病原

体的参考序列以获得变体(例如,SNP和indel)的列表。利用测序样本的集合,逐对计算每个样本之间的距离以创建距离矩阵。这个距离矩阵被用于构建样本的进化树以显示感染爆发可能如何扩散。为了确定哪些感染可能是相关的,过去病原体如何进化的测量结果被用于确定随时间的预期变化率,在患者到患者的扩散的情况下预期会看到这种变化率。

[0010] 能够通过对样本数据使用线性回归来计算预期变化率,所述样本数据包含针对不同样本的随时间看到的变化数目。还计算了预测区间,所述预测区间是在HAI传播时预期看到的预期变化数目界定的区间。使用特定阈值(例如,95%)的预测区间,密切相关的样本能够被分类为在患者之间传播或不相关。在确定可能相关的样本对之后,能够通过各种方法在进化树上可视地标记这些样本,例如在相关样本周围标一个框。这些样本通常应该落在进化树的子树内。

[0011] 本文公开的各个实施例包括一种用于追踪感染的扩散的方法。所述方法包括:基于感染样本测序数据构建感染测序样本的系统发育,以识别密切相关的样本;确定针对病原生物的随时间的预期变化范围;针对至少一个感染样本,计算该样本与至少一个额外的样本之间的随时间的突变变化数目;针对计算出的随着时间的突变变化数目中的每个,确定样本之间的所述随时间的突变变化数目是否在所述预期变化范围内;并且针对在所述预期变化范围内的所述计算出的随时间的突变变化数目中的每个,将相关联的样本对标记为潜在的传播。

[0012] 在一些实施例中,公开了如上所述的追踪感染的扩散的方法,其中,所述预期变化范围是通过针对所述多个感染样本中的至少一个计算所述样本与由所述系统发育确定的至少一个密切相关的样本之间的随时间的变化数目来确定的。

[0013] 在各个实施例中,公开了如上所述的追踪感染的扩散的方法,其中,所述预期变化范围是根据随时间从同一患者取得的多个样本来确定的,或者是根据从相互之间可能存在传播的两个患者取得的多个样本来确定的。

[0014] 在各个实施例中,公开了如上所述的追踪感染的扩散的方法,其中,构建所述系统发育包括逐对计算每对测序样本之间的距离并利用诸如FastTree2或Kimura 2参数模型的方法使用距离矩阵来构建进化树。

[0015] 在各个实施例中,公开了如上所述的追踪感染的扩散的方法,其中,确定针对病原生物的随时间的预期变化范围包括对样本数据使用线性回归并计算预测区间,所述样本数据包含针对不同样本的随时间的变化。

[0016] 在各个实施例中,公开了如上所述的追踪感染的扩散的方法,其中,感染样本测序数据是通过全基因组测序或靶向测序获得的。

[0017] 在各个实施例中,公开了如上所述的追踪感染的扩散的方法,其中,计算出的样本之间的突变变化数目通过除以两个样本都具有碱基调用的基因组位置的总数而被归一化。

[0018] 在各个实施例中,如上所述的追踪感染的扩散的方法还包括通过对所述样本的所述进化树进行广度优先搜索遍历变化数目在所述预期变化范围内的边来确定哪些感染样本来自相同的爆发。

[0019] 在各个实施例中,如上所述的追踪感染的扩散的方法还包括在经测序的感染样本的显示的进化树上标记被识别为潜在的传播的感染样本对。

[0020] 在各个实施例中,如上所述的追踪感染的扩散的方法还包括创建多个进化树,每

个树由被标记为来自相同的爆发的样本构建。

[0021] 各个实施例涉及一种用于追踪感染的扩散的系统。所述系统包括：接收器，其用于接收感染样本测序数据；存储器，其被配置为存储所接收的感染样本测序数据；以及处理器，其被配置为：构建测序样本的系统发育以识别紧密相关的样本；确定针对病原生物的随时间的预期变化范围；针对至少一个感染样本，计算该样本与至少一个额外的感染样本之间的随时间的突变变化数目；针对计算出的随时间的突变变化数目中的每个，确定样本之间的所述随时间的突变变化数目是否在所述预期变化范围内；并且针对在所述预期变化范围内的所述计算出的随时间的突变变化数目中的每个，将相关联的样本对标记为潜在的传播。

[0022] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所配置的处理器通过针对多个感染样本中的至少一个计算所述样本与由所述系统发育确定的至少一个紧密相关的样本之间的随时间的变化数目来确定所述预期变化范围。

[0023] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所配置的处理器根据随时间从同一患者取得的多个样本来确定的预期变化范围，或者根据从相互之间可能存在传播的两个患者取得的多个样本来确定的预期变化范围。

[0024] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所配置的处理器通过逐对计算每对测序样本之间的距离并使用距离矩阵构建进化树来构建所述系统发育。

[0025] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所配置的处理器通过对样本数据使用线性回归并计算预测区间来确定针对病原生物的随时间的预期变化范围，所述样本数据包含针对不同样本的随时间的变化。

[0026] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，感染样本测序数据是从全基因组测序数据或靶向测序数据的源接收的。

[0027] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，计算出的样本之间的突变变化数目通过除以两个样本都具有碱基调用的基因组位置的总数而被归一化。

[0028] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所述处理器还被配置为通过对所述样本的所述进化树进行广度优先搜索遍历变化数目在所述预期变化范围内的边来确定哪些感染样本来自相同的爆发。

[0029] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所述处理器还被配置为在经测序的感染样本的显示的进化树上标记被识别为潜在的传播的感染样本对。

[0030] 在各个实施例中，公开了如上所述的追踪感染的扩散的系统，其中，所述处理器还被配置为创建多个进化树，每个树由被标记为来自相同的爆发的样本构建。

[0031] 在阅读了以上描述之后，各种备选实施例对于本领域技术人员来说将是明显的。这些实施例被认为在主题公开的范围和精神内。

附图说明

[0032] 在附图中，贯穿不同的视图，相同的附图标记通常指的是相同的部件。在以下描述中，参考以下附图来描述各个实施例，在附图中：

[0033] 图1A-B是追踪感染的方法的实施例的示意图；并且

[0034] 图2是追踪感染的系统的实施例的示意图。

具体实施方式

[0035] 本文呈现的描述和附图示出了各种原理。将意识到，本领域技术人员将能够设计出各种布置，尽管这些布置在本文中没有被明确描述或示出，但是体现了这些原理并且被包括在本公开内容的范围内。除非另外指明（例如，“或其他”或“或在备选方案”中），否则本文所使用的术语“或”指的是非排他性的或（即，和/或）。额外地，本文描述的各个实施例不一定是相互排斥的，并且可以被组合以产生结合本文描述的原理的额外的实施例。

[0036] 本公开内容的各个实施例满足上面讨论的需要。实施例一般涉及与医疗保健相关联的感染，并且更具体地涉及用于追踪医院获得性感染的方法和装置。

[0037] 参考图1，图示了根据各个实施例的追踪医院获得性感染的方法的示意图100。在步骤110中，对多个感染样本中的每个感染样本进行测序。能够使用各种测序技术执行测序，例如，诸如Illumina HiSeq或MiSeq的新一代测序技术或Pacific Biosciences。实施例能够利用来自生物体的全基因组测序或靶向测序的数据。

[0038] 对测序数据进行处理并将测序数据与针对测序生物体的一个或多个参考序列进行比较。能够使用例如现有知识或MLST分型来识别合适的参考序列。一旦选择了参考序列，就能够使用诸如BWA、Bowtie等众所周知的方法将测序数据与参考序列进行对齐，并且能够通过各种方法（例如，samtools或GATK）来识别测序数据与参考序列之间的差异以调用变体。基因组中的变异可以被注释以确定哪些突变发生在重要基因内或发生在基因组的较不重要的区域内。

[0039] 在步骤120中，使用例如R统计建模语言和可从<http://cran.r-project.org/web/packages/ape/index.html>获得的诸如APE的R程序包来逐对计算每对测序样本之间的距离，并用于形成距离矩阵。距离矩阵能够基于差异的绝对数目或者基于诸如Jukes-Cantor的概率模型。然后可以使用该距离矩阵来构建测序样本的进化树并针对测序样本中的每个识别任何紧密相关的样本。

[0040] 在步骤130中，确定针对传播的感染的随时间的预期变化范围。用于确定随时间的预期变化范围的一个流程涉及检查进化树指示密切相关的邻居（即，那些仅在它们之间具有有限数目的变化（例如，10个SNP）且被认为与进化相关的邻居）并测量其基因组之间的变化范围并按其初始感染的隔离之间的时间进行划分。线性回归也可以用于样本数据以确定每单位时间的平均变化范围。用于确定随时间的预期变化范围的备选流程可以包括使用受控方法，例如在已知的（一个或多个）时间段上从同一患者取样并且使用来自第三方来源或者出版物的先前确定的信息来测量观察到的变化，或者根据从已知的医院获得性感染取得的多个样本来确定预期变化范围。能够使用其他方法来计算预期能看到的随时间的变化范围上的预测区间，并且这些区间能够用于界定典型的预期变化范围。

[0041] 在步骤140中，针对至少一个感染样本，计算该样本与至少一个额外的感染样本之间的随时间的突变变化数目，并且确定这些样本对之间的随时间的变化数目是在预期变化范围的区间内还是不在预期变化范围的区间内。

[0042] 例如，可以基于样本之间的SNP差异的数目来逐对计算每对样本之间的距离。也可以通过对例如插入缺失（插入和缺失）、基因组重排（反转和易位）、拷贝数变化、基因的缺失

或存在或前述特征的某种组合进行计数来确定突变变化数目，并且可以在生物体的全基因组或部分基因组（例如，生物体的染色体或质粒）中测量这些变化。

[0043] 一旦逐对计算了距离，就可以将每个样本与每一个其他样本进行比较，以确定该对样本之间的差异数目是否落在预期差异范围内，我们预期从给定样本之间的时间差的传播的感染看到所述预期差异范围。可以使用两个样本都具有碱基调用的基因组位置的总数来对样本之间的突变变化的计算出的数目进行归一化，从而减小样本测序过程中在物理转导中产生的误差和噪声。

[0044] 在步骤150中，针对其中计算出的随时间的变化数目在预期变化范围内的区间内的每对样本，将相关联的样本对标记为潜在的传播的感染。例如，如果变化数目在基于先前数据预期变化数目的某个预测区间（例如，95%）内，则样本将被标记为潜在的传播的感染。

[0045] 在各个实施例中，一起公开了如上所述的用于识别传播的感染的方法以及确定哪些感染样本来自相同的爆发的步骤160。这能够通过在如下的图上进行广度优先搜索来实现：在所述图中，每个患者是节点并且如果上述步骤确定患者之间可能存在传播则在两个节点之间存在边。这样的搜索通过从一个患者开始并且注意到已经被标记为可能已经受到来自该患者的感染的其他患者而在根本上识别在相同的爆发中的患者。这个过程通过以下操作而迭代进行：识别有可能依次受到来自以下患者的感染的患者；所述患者受到来自初始患者的感染，以此类推，直到没有新的患者被识别出为止。

[0046] 在各个实施例中，一起公开了如上所述的用于识别传播的感染的方法以及将被识别为经测序的感染样本的进化树上的被传播的感染的感染样本进行标记的步骤170。进化树可以是与上面的传播确定相关地计算出的初始进化树，或者它可以是如以上所讨论的考虑传播的感染的重新计算的进化树。传播的感染的实际显示可以是例如在所显示的树上对样本的聚类或子树的突出显示。

[0047] 在各个实施例中，一起公开了如上所述的用于识别传播的感染的方法以及创建多个进化树的步骤180，每个树与单独的爆发相关联。

[0048] 参考图2，图示了根据各个实施例的识别传播的感染的系统的示意图200。系统200包括：接收器202，其用于接收感染样本测序数据204；存储器206，其被配置为存储所接收的感染样本测序数据204；以及处理器208。

[0049] 如以上所讨论的处理器208被配置为构建测序样本的系统发育以针对测序样本中的每个识别任何紧密相关的样本；确定针对传播的感染的随时间的预期变化范围；针对多个感染样本中的每个，计算出该样本与至少一个其他感染样本之间的随时间的变化数目；针对计算出的随着时间的变化数目中的每个，确定随着时间的变化数目是否在预期变化范围的区间内；并且针对在预期变化范围的区间内的计算出的随时间的变化数目中的每个，将相关联的样本对标记为传播的感染。

[0050] 处理器208可以包括一个或多个微处理器、现场可编程门阵列（FPGA）、专用集成电路（ASIC）、其他类似的设备或其组合。存储器206可以包括诸如高速缓存（L1/L2/L3高速缓存）、系统存储器或存储设备的各种存储器设备。本文所使用的术语“非瞬态机器可读介质”将被理解为包括易失性（例如，DRAM和SRAM）存储器和非易失性（例如，闪速、磁性和光学）存储器，但是排除瞬态信号。

[0051] 在各个实施例中，公开了如上所述的用于识别传播的感染的系统，其中，所配置的

处理器通过针对多个感染样本中的至少一个计算样本与由系统发育确定的至少一个紧密相关的样本之间的随时间的变化数目来确定预期变化范围。

[0052] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,所配置的处理器根据随时间从同一患者取得的多个样本来确定预期变化范围,或者根据从相互之间可能存在传播的两个患者取得的多个样本来确定预期变化范围。

[0053] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,所配置的处理器通过逐对计算每对测序样本之间的距离并使用距离矩阵构建进化树来构建系统发育。

[0054] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,所配置的处理器通过对样本数据使用线性回归并计算预测区间来确定针对病原生物的随时间的预期变化范围,所述样本数据包含针对不同样本的随时间的变化。

[0055] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,感染样本测序数据是从全基因组测序数据或靶向测序数据的源接收的。

[0056] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,计算出的样本之间的突变变化数目通过除以两个样本都具有碱基调用的基因组位置的总数而被归一化。

[0057] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,所述处理器还被配置为通过对样本的进化树进行广度优先搜索遍历变化数目在预期变化范围内的边来确定所述感染样本中的哪些感染样本来自相同的爆发。

[0058] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,所述处理器还被配置为在经测序的感染样本的显示的进化树上标记被识别为潜在的传播的感染样本对。

[0059] 在各个实施例中,公开了如上所述的用于识别传播的感染的系统,其中,所述处理器还被配置为创建多个进化树,每个树由被标记为来自相同的爆发的样本构建。

[0060] 尽管已经以某种程度的特殊性描述了实施例,但是能够表明,在不脱离本公开内容的精神和范围的情况下,可以在构造的细节和部件的布置方面做出许多改变。应当理解,本发明不限于本文出于示范目的而阐述的实施例。

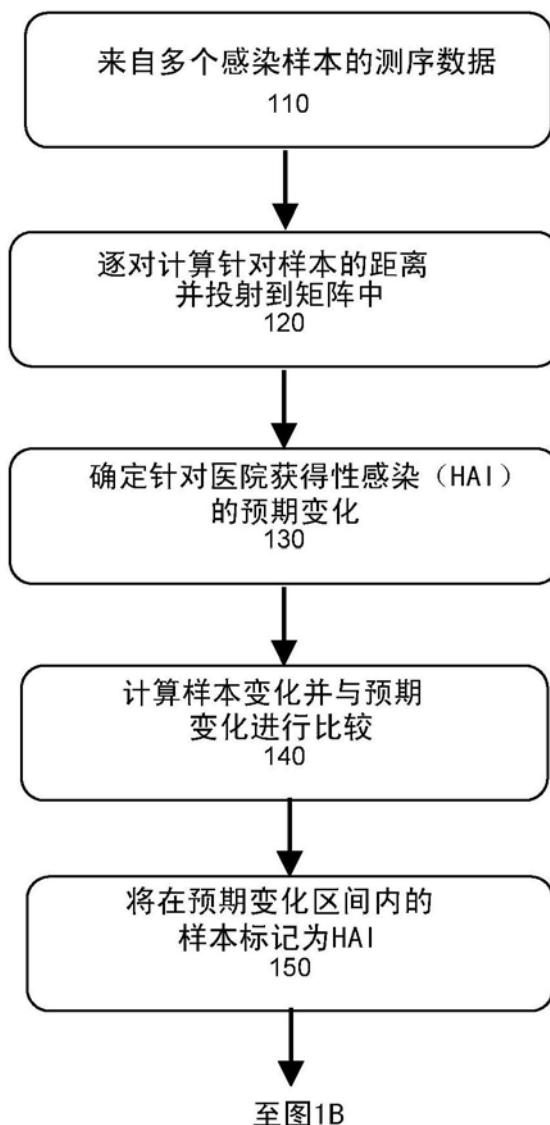
100

图1A

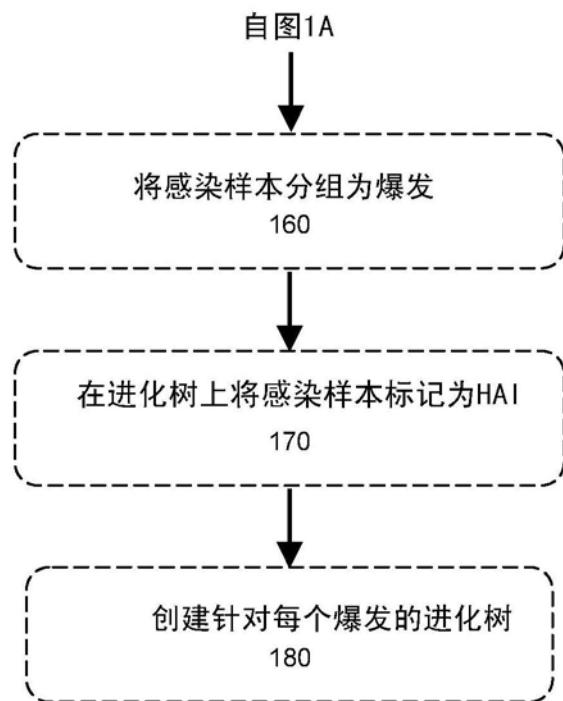
100

图1B

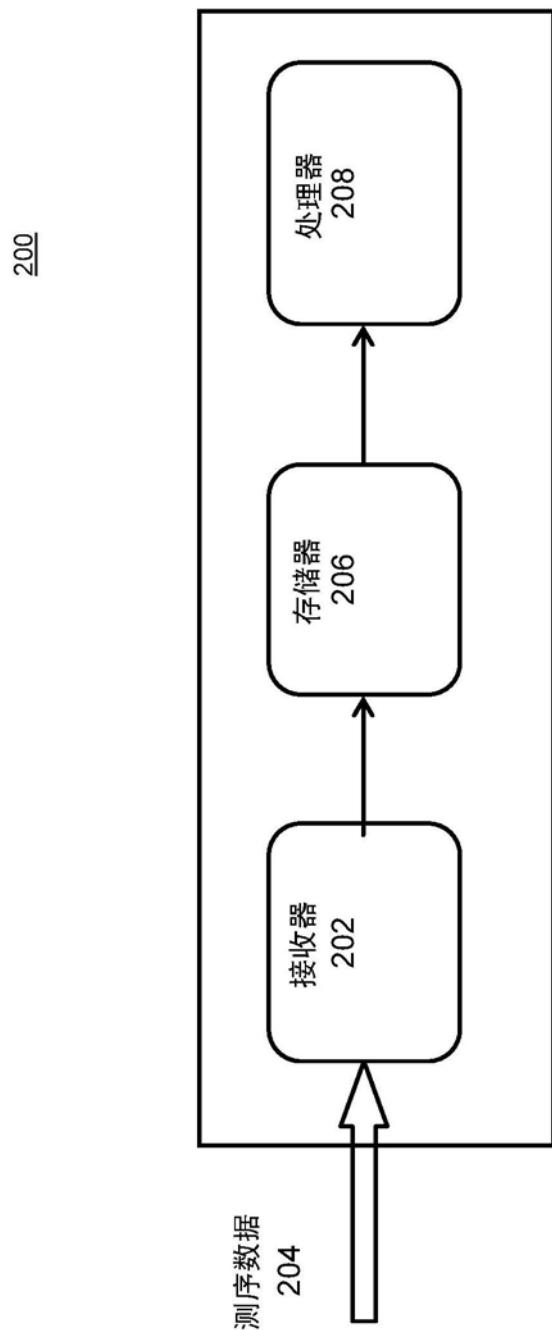


图2