



(12) 发明专利

(10) 授权公告号 CN 112309501 B
(45) 授权公告日 2025. 01. 28

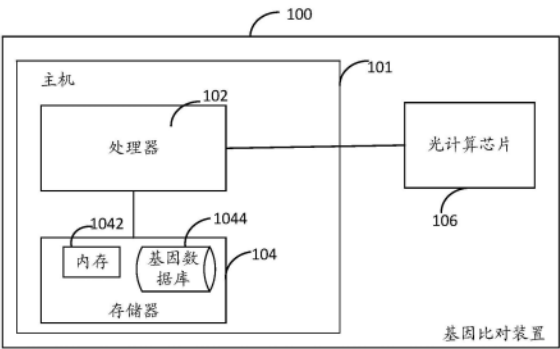
(21) 申请号 201911046513.5
(22) 申请日 2019.10.30
(65) 同一申请的已公布的文献号
 申请公布号 CN 112309501 A
(43) 申请公布日 2021.02.02
(66) 本国优先权数据
 201910713689.5 2019.08.02 CN
(73) 专利权人 华为技术有限公司
 地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72) 发明人 方涛 陈夏捷 董晓文
(74) 专利代理机构 北京同立钧成知识产权代理有限公司 11205
 专利代理师 杨泽 刘芳

(51) Int.Cl.
 G16B 30/10 (2019.01)
(56) 对比文件
 N. Brousseau等.Analysis of DNA sequences by an optical time-integrating correlator.APPLIED OPTICS.1992,第31卷(第23期),4802-4815.
 N. Brousseau等.Analysis of DNA sequences by an optical time-integrating correlator.APPLIED OPTICS.1992,第31卷(第23期),4802-4815.
 Fereshte Mozafari等.Speeding up DNA sequence alignment by optical correlator.Optics and Laser Technology.2018,第108卷124-135.
 审查员 韩沛

权利要求书4页 说明书18页 附图10页

(54) 发明名称
 基因比对技术

(57) 摘要
 一种基因比对技术。所述基因比对技术可以应用于包括光计算芯片的计算机系统中。在执行基因比对的过程中,可以先根据待测基因序列从基因数据库中获取第一组基因片段,所述第一组基因片段包括与所述待测基因序列的部分碱基匹配的多个参考基因片段。在获得所述第一组基因片段后,可以将所述待测基因序列与所述第一组基因片段中的多个参考基因片段输入所述光计算芯片进行光学比对。该技术可以大幅度提升基因比对速度,减少基因比对次数。



1. 一种基因比对方法,其特征在于,所述方法由包括光计算芯片的计算机系统执行,所述方法包括:

根据待测基因序列从基因数据库中获取第一组基因片段,其中,所述基因数据库中包
含有参考基因序列的多个参考基因片段,所述第一组基因片段包括与所述待测基因序列的
部分碱基匹配的多个参考基因片段;

将所述待测基因序列与所述第一组基因片段中的多个参考基因片段输入所述光计算
芯片进行光学比对;

根据所述光计算芯片的输出结果,确定所述待测基因序列与所述第一组基因片段中的
第一基因片段的相似度小于第一阈值且大于第二阈值;

根据待测基因序列的长度从所述参考基因序列获得多个子参考基因序列,其中,每个
子参考基因序列为所述参考基因序列的一部分;

将所述待测基因序列与所述多个子参考基因序列中的第二子参考基因序列输入所述
光计算芯片进行光学比对,获得所述待测基因序列与所述第二子参考基因序列的第四相似
度;

将所述待测基因序列与所述多个子参考基因序列中的第三子参考基因序列输入所述
光计算芯片进行光学比对,获得所述待测基因序列与所述第三子参考基因序列的第五相似
度,其中,所述第三子参考基因序列为与所述第二子参考基因序列连续的子参考基因序列;

确定所述第四相似度和所述第五相似度的和大于所述第一阈值;

根据所述第二子参考基因序列和所述第三子参考基因序列获得第四子参考基因序列,
其中,所述第四子参考基因序列包括所述第二子参考基因序列的部分碱基以及所述第三子
参考基因序列的部分碱基;

将所述待测基因序列和所述第四子参考基因序列输入所述光计算芯片进行光学比对,
获得所述待测基因序列和所述第四子参考基因序列的相似度;

其中,所述根据待测基因序列从基因数据库中获取第一组基因片段包括:

从所述基因数据库中获取与待测基因序列的前 m 个碱基以及后 n 个碱基相同的第一组
基因片段,其中, m 的值和 n 的值均大于0, m 与 n 的和小于所述待测基因序列中的碱基的数
量;

所述根据待测基因序列的长度从所述参考基因序列获得多个子参考基因序列,包括:

以所述待测基因序列的长度为窗口及滑动步长从所述参考基因序列中获得多个子参
考基因序列;或者,

按照所述待测基因序列的碱基长度将所述参考基因序列拆分成多个子参考基因序列。

2. 根据权利要求1所述的基因比对方法,其特征在于,所述方法还包括:

确定所述待测基因序列和所述第四子参考基因序列的相似度大于第三阈值且小于第
四阈值,其中,所述第四阈值不大于所述第一阈值;

响应上述确定,根据所述待测基因序列获得第一待测子基因序列和第二待测子基因序
列,其中,所述第一待测子基因序列和第二待测子基因序列的部分碱基相同;

将所述第一待测子基因序列与所述第四子参考基因序列输入所述光计算芯片进行光
学比对,获得第二相似度;

将所述第二待测子基因序列与所述第四子参考基因序列输入所述光计算芯片进行光

学比对,获得第三相似度。

3.根据权利要求2所述的基因比对方法,其特征在于,所述方法还包括:

当所述第二相似度大于所述第四阈值时,记录所述第四子参考基因序列在所述参考基因序列中的位置。

4.根据权利要求3所述的基因比对方法,其特征在于,所述方法还包括:

当所述第三相似度大于所述第三阈值且小于所述第四阈值时,根据所述第二待测子基因序列获得第一待测子基因序列单元和第二待测子基因序列单元,其中,所述第一待测子基因序列单元和所述第二待测子基因序列单元的部分碱基相同;

将所述第一待测子基因序列单元与所述第四子参考基因序列输入所述光计算芯片进行光学比对;

将所述第二待测子基因序列单元与所述第四子参考基因序列输入所述光计算芯片进行光学比对。

5.根据权利要求1所述的基因比对方法,其特征在于,所述方法还包括:

根据所述光计算芯片的输出结果确定所述第一组基因片段中的第二基因片段与所述待测基因序列的相似度;

当所述第一组基因片段中的第二基因片段与所述待测基因序列的相似度大于或等于第一阈值时,记录所述第二基因片段在所述参考基因序列中的位置。

6.根据权利要求1所述的基因比对方法,其特征在于,所述将所述待测基因序列与所述第一组基因片段中的多个参考基因片段输入所述光计算芯片进行光学比对包括:

分别将所述待测基因序列以及所述第一组基因片段中的所述多个参考基因片段进行光学编码;

将所述待测基因序列的光学编码与所述第一组基因片段中的所述多个参考基因片段的光学编码分别输入所述光计算芯片进行光学比对。

7.一种基因比对装置,包括:

处理器,用于根据待测基因序列从基因数据库中获取第一组基因片段,其中,所述数据库中包含有参考基因序列的多个参考基因片段,所述第一组基因片段包括与所述待测基因序列的部分碱基匹配的多个参考基因片段;

光计算芯片,连接所述处理器并用于对所述待测基因序列与所述第一组基因片段中的多个参考基因片段进行光学比对;

所述处理器还用于:

根据所述光计算芯片的输出结果,确定所述待测基因序列与所述第一组基因片段中的第一基因片段的相似度小于第一阈值且大于第二阈值;

根据待测基因序列的长度从所述参考基因序列获得多个子参考基因序列,其中,每个子参考基因序列为所述参考基因序列的一部分;

所述光计算芯片还用于:

对所述待测基因序列与所述多个子参考基因序列中的第二子参考基因序列进行光学比对;

对所述待测基因序列与所述多个子参考基因序列中的第三子参考基因序列进行光学比对,其中,所述第三子参考基因序列为与所述第二子参考基因序列连续的子参考基因序

列；

所述处理器还用于：

确定所述待测基因序列与所述第二子参考基因序列的第四相似度，以及所述待测基因序列与所述第三子参考基因序列的第五相似度的和大于所述第一阈值；

根据所述第二子参考基因序列和所述第三子参考基因序列获得第四子参考基因序列，其中，所述第四子参考基因序列包括所述第二子参考基因序列的部分碱基以及所述第三子参考基因序列的部分碱基；

将所述待测基因序列和所述第四子参考基因序列输入所述光计算芯片进行光学比对，以获得所述待测基因序列和所述第四子参考基因序列的相似度；

所述处理器，用于从所述基因数据库中获取与待测基因序列的前 m 个碱基以及后 n 个碱基相同的第一组基因片段，其中， m 的值和 n 的值均大于0， m 与 n 的和小于所述待测基因序列中的碱基的数量；

所述处理器，用于以所述待测基因序列的长度为窗口及滑动步长从所述参考基因序列中获得多个子参考基因序列；或者，按照所述待测基因序列的碱基长度将所述参考基因序列拆分成多个子参考基因序列。

8. 根据权利要求7所述的基因比对装置，其特征在于，所述处理器还用于：

确定所述待测基因序列和所述第四子参考基因序列的相似度大于第三阈值且小于第四阈值，其中所述第四阈值不大于所述第一阈值；

响应上述确定，根据所述待测基因序列获得第一待测子基因序列和第二待测子基因序列，其中，所述第一待测子基因序列和第二待测子基因序列的部分碱基相同；

所述光计算芯片还用于：

对所述第一待测子基因序列与所述第四子参考基因序列进行光学比对，获得第二相似度；以及

对所述第二待测子基因序列与所述第四子参考基因序列进行光学比对，获得第三相似度。

9. 根据权利要求8所述的基因比对装置，其特征在于，所述处理器还用于：

当所述第二相似度大于所述第四阈值时，记录所述第四子参考基因序列在所述参考基因序列中的位置。

10. 根据权利要求8所述的基因比对装置，其特征在于，所述处理器还用于：

当所述第三相似度大于所述第三阈值且小于所述第四阈值时，根据所述第二待测子基因序列获得第一待测子基因序列单元和第二待测子基因序列单元，其中，所述第一待测子基因序列单元和所述第二待测子基因序列单元的部分碱基相同；

所述光计算芯片还用于：

对所述第一待测子基因序列单元与所述第四子参考基因序列进行光学比对；

对所述第二待测子基因序列单元与所述第四子参考基因序列进行光学比对。

11. 根据权利要求7所述的基因比对装置，其特征在于，所述处理器还用于：

根据所述光计算芯片的输出结果确定所述第一组基因片段中的第二基因片段与所述待测基因序列的相似度；

当所述第一组基因片段中的第二基因片段与所述待测基因序列的相似度大于或等于

第一阈值时,记录所述第二基因片段在所述参考基因序列中的位置。

12.根据权利要求7所述的基因比对装置,其特征在于,所述处理器还用于:

分别将所述待测基因序列以及所述第一组基因片段中的所述多个参考基因片段进行光学编码;

将所述待测基因序列的光学编码与所述第一组基因片段中的所述多个参考基因片段的光学编码分别输入所述光计算芯片进行光学比对。

13.一种计算机程序产品,包括程序代码,所述程序代码包括的指令被计算机所执行以执行如权利要求1-6任意一项所述的基因比对方法。

14.一种计算机可读存储介质,包括计算机程序指令,当所述计算机程序指令在计算机上运行时,使得所述计算机执行如权利要求1-6任意一项所述的基因比对方法。

基因比对技术

技术领域

[0001] 本申请涉及光学技术领域,尤其涉及一种基因比对技术。

背景技术

[0002] 脱氧核糖核酸(Deoxyribonucleic acid,DNA)是染色体的主要化学成分,同时也是组成基因的材料。基因(Gene)是指携带有遗传信息的DNA序列,也称为遗传因子,是控制生物性状的遗传物质的基本结构单位和功能单位。基因通过指导蛋白质的合成来表达自己所携带的遗传信息,从而控制生物个体的性状表现。随着DNA测序技术的出现,直到人类基因组计划(Human Genome Project,HGP)完成,DNA序列数据的产生便呈现了指数级的增长。DNA序列比对是进行基因识别、信息分析、结构预测等问题的前提,通过多个DNA序列的比对,寻找它们相同和不同的位点、区域,从而帮助判断待测基因的同源性、变异点以及来源。

[0003] 随着新一代DNA测序技术的快速发展,DNA测序数据爆炸式积累的速度远远大于其被处理的速度。面对这些生物信息领域的大数据分析任务以及多种不同维度的数据整合,亟待需要一种快速、便捷的DNA比对方法。

发明内容

[0004] 本申请提供一种基因比对技术,能够提高DNA比对效率。

[0005] 第一方面,本发明实施例提供了一种基因比对方法,该方法应用于包括光计算芯片的计算机系统中。根据该方法,在实现基因比对的过程中,计算机系统的处理器可以根据待测基因序列从基因数据库中获取第一组基因片段,并将所述待测基因序列与所述第一组基因片段中的多个参考基因片段输入所述光计算芯片进行光学比对。其中,所述基因数据库中包含有参考基因序列的多个参考基因片段,所述第一组基因片段包括与所述待测基因序列的部分碱基匹配的多个参考基因片段。

[0006] 本发明实施例提供的基因比对方法,结合了数据库查找和光学自相关比对两种方式,通过构建的基因数据库与所述待测基因序列进行初次匹配,从而筛选出与所述待测基因序列可能匹配的第一组参考基因片段。通过本发明实施例提供的基因数据库对待比对的基因片段进行筛选后,可以大幅度减少需要详细比对的参考基因片段的数量。并且,在本发明实施例中,在获得第一组参考基因片段后,进一步通过光计算芯片对所述待测基因序列和所述第一组参考基因片段中的多个参考基因片段进行光学比对,由于光计算芯片进行光学比对,与通过电学方式进行基因比对的方法相比,比对速度更快。从而本发明实施例提供的基因比对方法,也极大的提高了比对效率。

[0007] 实际应用中,处理器可以根据待测基因序列的部分碱基从所述数据库中获得第一组基因片段。例如,根据待测基因序列的前 m 个碱基以及后 n 个碱基从所述数据库中获得第一组基因片段,其中, m 的值和 n 的值均大于0, m 与 n 的和小于所述待测基因序列中的碱基的数量。通常, m 和 n 的取值可以根据待测基因序列的长度、参考基因序列的长度等因素来确定。

[0008] 在一种可能的实现方式中,所述数据库可以为键-值(key-value)数据库,其中,key值为所述参考基因序列的多个参考基因片段的部分碱基,value值为所述多个参考基因片段在所述参考基因序列中的位置。

[0009] 在一种可能的实现方式下,所述方法还包括当根据所述光计算芯片的输出结果,确定所述待测基因序列与所述第一组基因片段中的第一基因片段的相似度小于第一阈值且大于第二阈值时,从所述参考基因序列获得多个子参考基因序列,并且,将所述待测基因序列与所述多个子参考基因序列中的第一子参考基因序列输入所述光计算芯片进行光学比对,获得所述待测基因序列与所述第一子参考基因序列的第一相似度,其中,每个子参考基因序列为所述参考基因序列的一部分。

[0010] 在本发明实施例中,当所述待测基因序列与所述第一组基因片段中的至少一个基因片段的相似度小于第一阈值且大于第二阈值时,说明待测基因序列很有可能在参考基因序列中找到匹配的参考基因片段,需要进一步的比对。因此,可以将所述待测基因序列和参考基因序列的多个子参考基因序列进一步进行光学比对,以便能够快速找到与所述待测基因序列的至少一部分片段匹配的参考基因片段。

[0011] 在又一种可能的实现方式中,所述方法还可以包括确定所述第一相似度大于第三阈值且小于第四阈值,并响应上述确定,根据所述待测基因序列获得第一待测子基因序列和第二待测子基因序列,其中,所述第四阈值不大于所述第一阈值,所述第一待测子基因序列和第二待测子基因序列的部分碱基相同。进一步的,将所述第一待测子基因序列与所述第一子参考基因序列输入所述光计算芯片进行光学比对以获得第二相似度,以及将所述第二待测子基因序列与所述第一子参考基因序列输入所述光计算芯片进行光学比对以获得第三相似度。根据这种方式,当经过待测基因序列和第一子参考基因序列的相似度满足预设条件时,可以将待测基因序列进一步进行拆分,并将拆分后的第一待测子基因序列和第二待测子基因序列分别与所述第一子参考基因序列进行比对,从而能够尽快定位出所述待测基因序列中与所述第一子参考基因序列匹配的部分片段。并且,由于这种最大相似匹配的方法能够容忍碱基的缺失现象,从而能够实现对待测基因序列中的缺失部分或变异部分的精确定位。实际应用中,第一待测子基因序列可以包括从所述待测基因序列的头部向尾部方向获取的第一预设长度的碱基,所述第二待测子基因序列可以包括从所述待测基因序列的尾部向头部方向获取的所述第一预设长度的碱基,且所述第一待测子基因序列和所述第二待测子基因序列的部分碱基重合。

[0012] 在又一种可能的实现方式中,所述方法还包括当所述第二相似度大于所述第四阈值时,记录所述第一子参考基因序列在所述参考基因序列中的位置。根据这种方式,当所述第一待测子基因序列与所述第一子参考基因序列的第二相似度大于所述第四阈值时,可以确认所述第一待测子基因序列与所述第一子参考基因序列最大相似匹配,从而,可以记录所述第一子参考基因序列在所述参考基因序列中的位置,获得所述第一待测子基因序列的最大相似匹配片段。

[0013] 在又一种可能的实现方式中,所述方法还包括:当所述第三相似度大于所述第三阈值且小于所述第四阈值时,根据所述第二待测子基因序列获得第一待测子基因序列单元和第二待测子基因序列单元,并将所述第一待测子基因序列单元与所述第一子参考基因序列输入所述光计算芯片进行光学比对,以及将所述第二待测子基因序列单元与所述第一子

参考基因序列输入所述光计算芯片进行光学比对。其中,所述第一待测子基因序列单元和所述第二待测子基因序列单元的部分碱基相同。根据这种方式,若所述第二待测子基因序列和所述第一子参考基因序列的匹配结果仍然没有达到最大相似匹配标准,则可以继续对所述第二待测子基因序列进行拆分比对,从而,根据这种递归查找的方式,能够快速定位出所述第二待测基因序列的至少部分片段的最大相似匹配片段。由于这种最大相似匹配的方法能够容忍碱基的缺失现象,从而实现对基因缺失和基因变异点的精确定位。

[0014] 在又一种可能的实现方式中,所述方法还包括:将所述待测基因序列与所述多个子参考基因序列中的第二子参考基因序列输入所述光计算芯片进行光学比对,获得所述待测基因序列与所述第二子参考基因序列的第四相似度,以及将所述待测基因序列与所述多个子参考基因序列中的第三子参考基因序列输入所述光计算芯片进行光学比对,获得所述待测基因序列与所述第三子参考基因序列的第五相似度,其中,所述第三子参考基因序列为与所述第二子参考基因序列连续的子参考基因序列。在确定所述第四相似度和所述第五相似度的和大于所述第一阈值时,根据所述第二子参考基因序列和所述第三子参考基因序列获得第四子参考基因序列,并将所述待测基因序列和所述第四子参考基因序列输入所述光计算芯片进行光学比对。其中,所述第四子参考基因序列包括所述第二子参考基因序列的部分碱基以及所述第三子参考基因序列的部分碱基。

[0015] 根据这种方式,在确定所述待测基因序列与第二子参考基因序列的相似度的值不满足需要继续与所述第二子参考基因序列进行进一步匹配的条件,而所述待测基因序列与所述第二子参考基因序列以及所述第三子参考基因序列的相似度的和大于所述第一阈值的情况下,可以及时调整子参考基因序列的位置,从所述第二子参考基因序列和所述第三子参考基因序列中取连续的部分获得所述第四子参考基因序列,从而能够尽快从所述第四子参考基因序列中查找到所述待测基因序列的最大相似匹配片段,而无需将所述待测基因片段与所述第三子参考基因序列后面的子参考基因序列继续比对。这种根据部分比对结果及时调整子参考基因序列的方式,能够提高获得最大相似基因片段的概率和速度,减少比对次数。

[0016] 可以理解的是,实际应用中,可以根据所述第四相似度和所述第五相似度的比例分别从所述第二子参考基因序列和第三子参考基因序列获取一部分参考基因片段组成所述第四子参考基因序列。

[0017] 在又一种可能的实现方式中,所述方法还包括根据所述光计算芯片的输出结果确定所述第一组基因片段中的第二基因片段与所述待测基因序列匹配,并记录所述第二基因片段在所述参考基因序列中的位置。

[0018] 在又一种可能的实现方式中,所述将所述待测基因序列与所述第一组基因片段中的多个参考基因片段输入所述光计算芯片进行光学比对包括:分别将所述待测基因序列以及所述第一组基因片段中的所述多个参考基因片段进行光学编码,将所述待测基因序列的光学编码与所述第一组基因序列中的所述多个基因片段的光学编码分别输入所述光计算芯片进行光学比对。实际应用中,可以根据光的强度信息和/或光的空间信息对所述待测基因序列与所述多个参考基因片段进行光学编码。

[0019] 第二方面,本发明实施例提高一种基因比对装置,包括处理器和光计算芯片。所述处理器用于根据待测基因序列从数据库中获取第一组基因片段,其中,所述数据库系

统中包含有参考基因序列的多个参考基因片段,所述第一组基因片段包括与所述待测基因序列的部分碱基匹配的多个参考基因片段。所述光计算芯片连接所述处理器并用于对所述待测基因序列与所述第一组基因片段中的多个参考基因片段进行光学比对。

[0020] 在一种可能的实现方式中,处理器可以根据待测基因序列的部分碱基从所述数据库中获取第一组基因片段。例如,根据待测基因序列的前 m 个碱基以及后 n 个碱基从所述数据库中获取第一组基因片段,其中, m 的值和 n 的值均大于0, m 与 n 的和小于所述待测基因序列中的碱基的数量。具体的,所述数据库可以为键-值(key-value)数据库,其中,key值为所述参考基因序列的多个参考基因片段的部分碱基,value值为所述多个参考基因片段在所述参考基因序列中的位置。

[0021] 在一种可能的实现方式中,所述处理器还用于根据所述光计算芯片的输出结果,确定所述待测基因序列与所述第一组基因片段中的第一基因片段的相似度小于第一阈值且大于第二阈值,并从所述参考基因序列获得多个子参考基因序列,其中,每个子参考基因序列为所述参考基因序列的一部分。所述光计算芯片还用于对所述待测基因序列与所述多个子参考基因序列中的第一子参考基因序列进行光学比对,获得所述待测基因序列与所述第一子参考基因序列的第一相似度。

[0022] 在又一种可能的实现方式中,所述处理器还用于确定所述第一相似度大于第三阈值且小于第四阈值,其中所述第四阈值不大于所述第一阈值,并且响应上述确定,根据所述待测基因序列获得第一待测子基因序列和第二待测子基因序列,其中,所述第一待测子基因序列和第二待测子基因序列的部分碱基相同。所述光计算芯片还用于对所述第一待测子基因序列与所述第一子参考基因序列进行光学比对,获得第二相似度;以及对所述第二待测子基因序列与所述第一子参考基因序列进行光学比对,获得第三相似度。

[0023] 在又一种可能的实现方式中,所述处理器还用于当所述第二相似度大于所述第四阈值时,记录所述第一子参考基因序列在所述参考基因序列中的位置。

[0024] 在又一种可能的实现方式中,所述处理器还用于当所述第三相似度大于所述第三阈值且小于所述第四阈值时,根据所述第二待测子基因序列获得第一待测子基因序列单元和第二待测子基因序列单元,其中,所述第一待测子基因序列单元和所述第二待测子基因序列单元的部分碱基相同。所述光计算芯片还用于对所述第一待测子基因序列单元与所述第一子参考基因序列进行光学比对,以及对所述第二待测子基因序列单元与所述第一子参考基因序列进行光学比对。

[0025] 在又一种可能的实现方式中,所述光计算芯片还用于对所述待测基因序列与所述多个子参考基因序列中的第二子参考基因序列进行光学比对;对所述待测基因序列与所述多个子参考基因序列中的第三子参考基因序列进行光学比对,其中,所述第三子参考基因序列为与所述第二子参考基因序列连续的子参考基因序列。所述处理器还用于:确定所述待测基因序列与所述第二子参考基因序列的第四相似度与所述待测基因序列与所述第三子参考基因序列的第五相似度的和大于所述第一阈值;根据所述第二子参考基因序列和所述第三子参考基因序列获得第四子参考基因序列;并将所述待测基因序列和所述第四子参考基因序列输入所述光计算芯片进行光学比对。其中,所述第四子参考基因序列包括所述第二子参考基因序列的部分碱基以及所述第三子参考基因序列的部分碱基。

[0026] 在又一种可能的实现方式中,所述处理器还用于根据所述光计算芯片的输出结果

确定所述第一组基因片段中的第二基因片段与所述待测基因序列匹配,并记录所述第二基因片段在所述参考基因序列中的位置。

[0027] 在又一种可能的实现方式中,所述处理器还用于分别将所述待测基因序列以及所述第一组基因片段中的所述多个参考基因片段进行光学编码,并将所述待测基因序列的光学编码与所述第一组基因序列中的所述多个基因片段的光学编码分别输入所述光计算芯片进行光学比对。

[0028] 第三方面,本发明实施例提供一种比对装置,包括处理器和光计算芯片。所述处理器用于根据待匹配的第一对象从数据库中获取第一组参考对象,其中,所述第一组参考对象中包括与所述第一对象的部分特征相同的多个参考对象。所述光计算芯片连接所述处理器并用于对所述第一对象以及所述多个参考对象进行光学比对。

[0029] 本发明实施例提供的比对装置,结合了数据库查找及光学比对两种方式,通过数据库对待比对的参考对象进行筛选后,可以大幅度减少需要详细比对的参考对象的数量。并且,采用光计算芯片进行比对,能够大幅度提升比对速度。本发明实施提供的比对装置不仅能够用于对基因检测场景,还可以应用于各种需要进行海量数据比对的场景中。

[0030] 在一种可能的实现方式中,所述处理器还用于根据所述光计算芯片的输出结果,确定所述第一对象与所述第一组参考对象中的第一参考对象的相似度小于第一阈值其而大于第二阈值,根据标准对象获得多个子参考对象,其中,每个子参考对象为所述参考对象的一部分。所述光计算芯片还用于对所述第一对象与所述多个子参考对象中的第一子参考对象进行光学比对,获得所述第一对象与所述第一子参考对象的第一相似度。

[0031] 在又一种可能的实现方式中,所述处理器还用于确定所述第一相似度大于第三阈值且小于第四阈值,并响应上述确定,根据所述第一对象获得第一子对象和第二子对象。其中,所述第四阈值不大于所述第一阈值,所述第一子对象和所述第二子对象的部分数据相同。所述光计算芯片还用于对所述第一子对象和所述第一子参考对象进行光学比对,获得第二相似度,以及对所述第二子对象和所述第一子参考对象进行光学比对,获得第三相似度。

[0032] 在又一种可能的实现方式中,所述处理器还用于当所述第二相似度大于所述第四阈值时,记录所述第一子参考对象在所述标准对象中的位置。

[0033] 第四方面,本申请还提供了一种比对装置,包括获取模块、比对模块、结果处理模块、判断模块等用于实现第一方面或第一方面的任意一种可能的实现方式的功能模块。

[0034] 第五方面,本申请还提供了一种计算机程序产品,包括程序代码,所述程序代码包括的指令被计算机所执行,以实现所述第一方面以及所述第一方面的任意一种可能的实现方式中所述的基因比对方法。

[0035] 第六方面,本申请还提供了一种计算机可读存储介质,所述计算机可读存储介质用于存储程序代码,所述程序代码包括的指令被计算机所执行,以实现前述第一方面以及所述第一方面的任意一种可能的实现方式中所述的基因比对方法。

附图说明

[0036] 为了更清楚的说明本发明实施例或现有技术中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些

实施例。

- [0037] 图1为本发明实施例提供的一种基因比对装置的结构示意图；
- [0038] 图2A为本发明实施例提供的一种基因数据库的示意图；
- [0039] 图2B为本发明实施例提供的一种光学编码示意图；
- [0040] 图3A为本发明实施例提供的一种光计算芯片的结构示意图；
- [0041] 图3B为本发明实施例提供的另一种光计算芯片的结构示意图；
- [0042] 图3C为本发明实施例提供的光学比对的原理示意图；
- [0043] 图4为本发明实施例提供的一种基因比对方法流程图；
- [0044] 图5A、图5B、图5C和图5D为本发明实施例提供的光学编码的示例；
- [0045] 图6为本发明实施例提供的另一种基因比对方法流程图；
- [0046] 图7为本发明实施例提供的子参考基因序列和待测子基因序列的示意；
- [0047] 图8为本发明实施例提供的另一种基因比对方法流程图；
- [0048] 图9为本发明实施例提供的一种比对装置的结构示意图；
- [0049] 图10为本发明实施例提供的另一种比对装置的结构示意图。

具体实施方式

[0050] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚的描述。显然,所描述的实施例仅仅是本发明一部分的实施例,而不是全部的实施例。

[0051] 如前所述,由于脱氧核糖核酸(Deoxyribonucleic acid,DNA)测序技术的快速发展,DNA测序数据爆炸式增长。因此,如何提高DNA比对的速度是一个亟需解决的技术问题。现有技术中,通常是通过在计算机系统中对参考基因序列构建索引的方式来加速查找速率。索引的本质是通过优化数据结构来提高查找效率。然后索引优化本身存在瓶颈,并且同时创建很多的负责索引会耗费大量的实际。因此这种基因比对方式效率难以承受DNA测序数据的大量增长。本发明实施例提供的基因比对方案,能够大幅提供基因比对速度,即使在面对海量基因测序数据时,也能够快速实现基因比对。

[0052] 为了更清楚的理解本方案,先对本发明实施例涉及的几个技术术语进行描述。

[0053] 基因:是指控制生物性状的遗传信息,通常由DNA序列来承载。基因也可视作基本遗传单位,亦即一段具有功能性的DNA或核糖核酸(Ribonucleic acid,RNA)序列。弄清其序列本身的过程叫基因测序。

[0054] 待测基因序列:也可以被称为reads,是一小段测序片段,是一种由高通量测序平台产生的测序数据。在对整个基因组进行测序的过程中,就会产生成百上千万的reads,然后将这些reads拼接起来就能获得基因组的全序列。

[0055] 参考基因序列(也可以被称为reference sequence):是一种经过验证和编辑的标准序列。参考基因序列可以为人类基因组的功能注解提供一个基础。为突变分析、基因表达研究和多态发现提供一个稳定的参考点。

[0056] 碱基对:是形成DNA、RNA单体以及编码遗传信息的化学结构。组成碱基对的碱基包括腺嘌呤A、鸟嘌呤G、胸腺嘧啶T、胞嘧啶C、尿嘧啶U。严格地说,碱基对是一对相互匹配的碱基(即A-T,G-C,A-U相互作用)被氢键连接起来。它常被用来衡量DNA和RNA的长度(尽管RNA

是单链)。

[0057] 下面将对本发明实施例进行详细介绍。图1为本发明实施例提供的一种利用光学系统实现基因比对的示意图。如图所示,基因对比装置100可以包括处理器102、存储器104以及光计算芯片106。其中,处理器102和存储器104可以看成是主机101的一部分。光计算芯片106可以通过主机接口与主机101连接。主机接口可以包括标准的主机接口以及网络接口(network interface)。例如,主机接口可以包括快捷外设互联标准(Peripheral Component Interconnect Express,PCIE)接口。数据可以通过主机接口发送给光计算芯片106,光计算芯片106处理后的数据也可以通过主机接口发送给处理器102。处理器102也可以通过主机接口监测光计算芯片106的工作状态。实际应用中,处理器102和存储器104也可以不作为主机的一部分,处理器102、存储器104和光计算芯片106可以是系统芯片(System on a Chip,SOC)的一部分。

[0058] 处理器(Processor)102是基因比对装置100的运算核心和控制核心(Control Unit)。处理器102中可以包括多个处理器核(core)。处理器102可以是一块超大规模的集成电路。在处理器102中安装有操作系统和其他软件程序,从而处理器102能够实现对内存1042、缓存、磁盘及外设设备(如图1中的光计算芯片106)的访问。可以理解的是,在本发明实施例中,处理器102中的Core例如可以是中央处理器(Central Processing unit,CPU)、图形处理器(Graphics Processing Unit,GPU)、现场可编程逻辑门阵列(Field Programmable Gate Array,FPGA)、还可以是其他特定集成电路(Application Specific Integrated Circuit,ASIC)等。

[0059] 存储器104用于存储数据。存储器104可以包括内存1042、磁盘等其他存储数据的存储器。内存1042是主机101的主存。内存1042可以通过双倍速率(double data rate,DDR)总线和处理器102相连。内存1042通常用来存放操作系统中各种正在运行的软件、输入和输出数据以及与外存交换的信息等。为了提高处理器102的访问速度,内存1042需要具备访问速度快的优点。实际应用中,通常采用动态随机存取存储器(Dynamic Random Access Memory,DRAM)作为内存1042。处理器102能够通过内存控制器(图1中未示出)高速访问内存1042,对内存1042中的任意一个存储单元进行读操作和写操作。

[0060] 在本发明实施例中,存储器104可以用于存储基因数据库1044。基因数据库1044可以根据参考序列建立的键-值(key-value)数据库。其中,key值可以根据基因片段的部分碱基获得。value值可以包括key值对应的参考基因片段在存储器中的位置,还可以包括key值对应的参考基因片段在参考基因序列中的位置。

[0061] 在本发明实施例中,可以通过将参考基因序列的部分碱基作为key值,例如,可以将预设长度的参考基因片段的前m个碱基以及后n个碱基作为key值。m和n可以相同也可以不相同,在此不做限定。通过遍历参考基因序列,定位出所有满足该key值的参考基因片段,并将所有参考基因片段的位置信息作为所述key值对应的value值进行记录。图2A为本发明实施例提供的一种基因数据库的示意图。如图2A所示,基因数据库1044可以包括键(key)1044_1以及值(value)1044_2。其中,键1044_1部分以10个碱基为例,具体的,可以分别取参考片段头部的5个碱基和尾部的5个碱基为key值。在本发明实施例中,以150个碱基为参考基因片段的长度为例对如何建立基因数据库1044进行描述。具体的,首先构建一个空集的索引表(只具备key值),其行数为 4^{5+5} ,key的排序组合即为AAAAAAAAAA到TTTTTTTTTT的字母

表排序。映射方式如图2B所示。具体的,以头部碱基作为高位,尾部碱基作为低位的顺序进行排列。每单位上的碱基以A、C、G、T向前递进,满T则进一位单位碱基。待尾部碱基全为TTTTT时向头部碱基部分进一位碱基。通过这种方式,可以如下顺序的碱基:AAAAAAAAAA、AAAAAAAAAAC、AAAAAAAAAAG、AAAAAAAAAAT、AAAAAAAAACA、AAAAAAAAACC、AAAAAAAAACG、AAAAAAAAACT等等。由此,可以获得如图2A中所示的键1044_1。

[0062] 在建立键值索引表后,依次以预设碱基长度作为单位窗口,以步长为单位碱基(即1个碱基)在参考基因序列上进行滑动,从而能够获得多个参考基因片段。在获得每一个参考基因片段的过程中,可以分别根据该参考基因片段头部的5个碱基和尾部的5个碱基,获得该参考基因片段的key值,并在与该key值对应的value 1044_2中记录该参考基因片段在参考基因序列中的位置。例如,可以记录所述参考基因片段的第一个碱基所在的位置。按照这种方式,一直滑动到参考基因序列的末端,获得所述参考基因序列的所有参考基因片段的value值(即,参考基因片段的位置信息)。从而可以建立如图2A所示的基因数据库1044。

[0063] 实际应用中,key值映射的方式取决于排列组合的形式。假设令前n后m个碱基的序列片段分别为Seq₁、Seq₂,key值的映射定义为:

$$[0064] \quad Key_{mapping} = \left(\sum_{i=1}^n Seq_1[i-1] \times 4^{i-1} \right) \times 4^n + \sum_{j=1}^m Seq_2[j-1] \times 4^{j-1}$$

[0065] 例如,如果存在一条DNA序列为GTGGA……CGAGC,令A、C、G、T的值分为0、1、2、3,则此条序列对应的Key值为:

$$[0066] \quad Key_{GTG \dots AGC}$$

$$[0067] \quad = (Seq_1[4] \times 4^4 + Seq_1[3] \times 4^3 + Seq_1[2] \times 4^2 + Seq_1[1] \times 4^1 + Seq_1[0] \times 4^0) \times 4^5 + Seq_2[4] \times 4^4 + Seq_2[3] \times 4^3 + Seq_2[2] \times 4^2 + Seq_2[1] \times 4^1 + Seq_2[0] \times 4^0$$

$$[0068] \quad \times 4^1 + Seq_1[0] \times 4^0) \times 4^5 + Seq_2[4] \times 4^4 + Seq_2[3] \times 4^3 + Seq_2[2] \times 4^2 + Seq_2[1] \times 4^1 + Seq_2[0] \times 4^0$$

$$[0069] \quad \times 4^3 + Seq_2[2] \times 4^2 + Seq_2[1] \times 4^1 + Seq_2[0] \times 4^0$$

$$[0070] \quad = 728 \times 4^5 + 393 = 745865$$

[0071] 可以理解的是,n和m碱基数目的选择直接影响到算法本身的效率,n和m的增大会导致key值存储的value(即位置信息)减少。如果不考虑硬件因素,平均每增加单位碱基对于每一条待测基因序列的寻址速率会提高4倍。然而,由于测序误差以及基因突变限制了n和m不可无限增加,增加n和m可能会降低Key值可靠性。因此,m和n的值可以根据需要自行确定,参考基因片段的长度也可以根据实际需要进行设定。通常,m和n的取值可以根据待测基因序列的长度、参考基因序列的长度等因素来确定。实际应用中,参考基因片段的长度通常与待测基因序列的碱基长度相同。

[0072] 光计算芯片106可以是一种片上光计算系统。图3A为本发明实施例提供的一种光计算芯片的结构示意图。如图3A所示,该光计算芯片106可以包括光源阵列202、调制器阵列204、探测器阵列206、第一凹面镜208和第二凹面镜210。其中,光源阵列202位于第一凹面镜208的物面焦平面上。调制器阵列204位于第一凹面镜208的像面焦平面上,并且,调制器阵列204也位于第二凹面镜210的物面焦平面上。探测器阵列206位于第二凹面镜210的像面焦平面上。

[0073] 光源阵列202用于数据的调制和发送,作为光计算芯片106的数据输入单元。光源阵列202可以根据输入数据生成多个不同光强的光信号。第一凹面镜208用于对光源阵列

202发送的数据光信号实现标准的傅里叶变换。调制器阵列204有两种工作模式：记录模式和调制模式。其中，记录模式用于获得光源阵列202发送的数据光信号经过第一凹面镜208后频谱面的像。调制模式用于将光源阵列202发送的数据光信号频谱面像调制在调制器阵列204上。第二凹面镜210用于对经过调制器阵列204后的光信号实现标准的傅里叶逆变换。探测器阵列206用于光强度信号探测，作为光计算芯片106的结果输出单元。

[0074] 图3B为本发明实施例提供的又一种光计算芯片的结构示意图。与图3A提供的片上集成光计算芯片不同的是，图3B所示的光计算芯片中，将光源阵列202和探测器阵列206布置在芯片的同一侧，使得整个计算芯片的结构更加紧凑，可以降低芯片尺寸。如图3B所示，与图3A所示的光计算芯片相比，第一凹面镜208、第二凹面镜210、以及调制器阵列204的位置没有改变，光源阵列202、调制器阵列204和探测器阵列206分别相对于第一凹面镜208和第二凹面镜210的焦距位置也没有改变。图3B所示的各个器件的实现可以参考对图3A所示的光计算芯片中各个器件的描述。在此不再赘述。

[0075] 图3A和图3B仅仅是本发明实施例提供的光计算芯片的结构示意图，实际应用中，并不对光计算芯片106的具体结构进行限制，还可以采用其他结构的光计算芯片。例如，光计算芯片106还可以是利用4F光计算系统原理实现的其他结构的光计算芯片。图3C为4F光计算系统原理示意图。如图3C所示，第一调制器302位于第一凸透镜304的物面焦点位置。第二调制器306位于所述第一凸透镜304的像面焦点位置，且位于第二凸透镜308的物面焦点位置。所述第一凸透镜304和第二凸透镜308之间的间隔为所述两个凸透镜(304和308)的焦距之和。探测器310在第二凸透镜308的像面焦点位置，整个系统长度为4倍焦距。在利用图3C所示的4F光学系统进行数据比对时，可以分别将待比对的第一数据加载在第一调制器302上，将经过翻转后的第二数据的频谱数据加载在第二调制器306上，从而根据第一数据产生的光信号穿过第一凸透镜304后发生傅里叶变换在第二调制器306位置处变成频谱光信号，与第二调制器306上的翻转后的第二数据的频谱数据在光空间完成了乘法操作。本质上改变了第一数据的频谱光信号在光空间的光场能量分布。相乘后的频谱光信号经过第二凸透镜308发生反傅里叶变换又变回时域光信号。探测器310可以根据探测穿过第二凸透镜308的时域光信号的强度获得两个数据的自相关结果。需要说明的是，上述加载在光计算芯片上的第一数据和第二数据可以均为向量。

[0076] 可以理解的是，上述图3A-3C的光计算芯片实现数据比对的过程是通过探测两个数据的光信号在光空间的自相关结果获得的。本领域技术人员可以知道，自相关也叫序列相关，是一个信号于其自身在不同时间点的互相关。换一种表达方式，自相关是两次观察之间的相似度对它们之间的时间差的函数。自相关是一种找出随机变量序列重复模式的数学工具。在实际进行序列识别时，通过使用自相关运算，能保证在待测序列和目标序列相同时，其自相关结果中会出现一个明显的最大值位置，监测最大值的出现就可以较为容易地实现序列的比对。

[0077] 下面将具体介绍如何采用图1所示的基因比对装置实现基因比对，提高基因比对速度。图4为本发明实施例提供的一种基因比对方法流程图。下面将结合图1对图4所示的方法进行具体介绍。为描述清楚、简便，本发明实施例以检测一个待测基因序列为例进行描述。可以理解的是，即使实际应用中会一次检测多个待测基因序列，但对每一个待测基因序列都可以参考本发明实施例进行比对。如4所示，该方法包括下述步骤。

[0078] 在步骤402中,处理器102根据待测基因序列的部分碱基从数据库中获取第一组基因片段。具体的,可以按照基因数据库1044的key 1044_1的获取方式,获取待测基因序列的key值。例如,可以将待测基因序列的头部的5个碱基和尾部的5个碱基作为待测基因序列的key值。根据所述待测基因序列的key值查找所述基因数据库1044,获得与所述key值匹配的多个value值,该多个value值用于指示待测基因序列在参考基因序列上的可能位置。由于基因数据库1044中与某个key值对应的value值指示的是对应的参考基因片段在参考基因序列中的位置信息,因此,可以根据匹配的多个value值获得多个参考基因片段。在本发明实施例中,将与待测基因序列的key值匹配的多个参考基因片段称为第一组基因片段。

[0079] 在步骤404中,通过光计算芯片106对所述待测基因序列与所述第一组基因片段中的多个参考基因片段进行光学比对。具体的,处理器102可以分别对待测基因序列和所述多个参考基因片段进行光学编码,并将待测基因序列的光学编码以及所述多个参考基因片段的光学编码加载到所述光计算芯片进行比对。在对待测基因序列和参考基因片段进行光学编码的过程中,可以分别对待测基因序列和参考基因片段中的碱基字符串进行编码。例如,以4个点光源作为单碱基的单位簇,以其不同的明暗程度(0表示光源灭,1表示光源亮)表示四种不同碱基,A、C、G、T的编码方案为0001、0010、0100、1000,如图5A所示。根据单碱基A、C、G、T的编码方式,可以获得待测基因序列和第一组基因片段中的多个参考基因片段的光学编码。从而可以将获得的待测基因序列和第一组基因片段中的多个参考基因片段的光学编码发送给光计算芯片106进行光学比对。

[0080] 实际应用中,由于不同的编码方式直接影响解码难度以及自相关结果输出的可信度。又一种情形下,还可以在编码过程中包括光的强度信息和/或光的空间信息。在本发明实施例中,可以将利用光的强度信息进行编码的方式称为强度编码方式,将利用光的空间信息进行编码的方式称为空间编码方式。实际应用中,还可以将两种编码方式进行结合,可以将这种结合的方式称为混合编码方式。强度编码方式可以利用不同的电压幅度调制光强,通过不同强度的光信号表示四种不同碱基。强度编码方式可以如5B所示。空间编码方式则可以利用多个点光源作为单碱基的单位簇,以不同的明暗程度(例如,0表示光源灭,1表示光源亮)表示四种不同碱基。空间编码方式可以如图5C所示,可以采用多个具有相同电压且具有不同光强的光信号表示不同的碱基。混合编码方式则可以是结合强度编码和空间编码的方式,例如可以如图5D所示,可以采用多个具体不同电压以及不同光强的光信号结合表示不同的碱基。在本发明实施例中不对具体的编码方式进行限定。

[0081] 在光计算芯片106进行基因比对的过程中,光源阵列202可以先根据翻转后的待测基因序列的编码发送第一光信号,第一光信号经过第一凹面镜208反射后发生傅里叶变换变成频谱光信号,调制器阵列204接收第一光信号的反射频谱光信号,并将所述第一光信号的反射频谱光信号调制在调制器阵列204上。然后,光源阵列202根据所述第一组参考基因片段中的多个参考基因片段的光学编码分别发送多个光信号,从而根据参考基因片段的光学编码发送的光信号穿过第一凹面镜208在调制器阵列204位置处变成频谱光信号后,与所述第一光信号的反射信号在光空间完成乘法操作。调制器阵列204输出的频谱光信号经过第二凹面镜210进行反傅里叶变换成为时域光信号,最后探测器阵列206通过探测第二凹面镜210输出的时域光信号的光强,能够分别获得第一光信号和所述多个参考基因片段的光信号的匹配结果。本领域技术人员可以知道,频谱数据相乘后经过反傅里叶变换后就是两

个数据的自相关结果。

[0082] 在步骤406中,处理器102根据所述光计算芯片的输出结果确定所述待测基因序列与所述多个参考基因片段的相似度。实际应用中,探测器阵列206获得匹配结果后,光计算芯片106可以将匹配结果发送给处理器102。例如,可以通过一些外围电路采集探测器阵列206探测获得的光强信号,将采集的光强信号转换为电信号的,并将电信号转换为数字信号后发送给处理器102,从而,处理器102能够获得光计算芯片106对待测基因序列以及参考基因片段的比对结果。可以理解的是,实际应用中,探测器阵列206可以每获得一比对结果就产生反馈,也可以在相似度达到预设阈值时产生反馈。需要说明的是,本发明实施例的相似度用于指示待测基因序列和参考基因片段的匹配程度。

[0083] 在步骤408中,处理器102确定所述待测基因片段与所述多个参考基因片段中的第一参考基因片段的相似度是否大于或等于第一阈值,如果大于或等于第一阈值,则进入步骤410,当确定待测基因片段与所述第一参考基因片段的相似度小于所述第一阈值时,该方法进入步骤412。在本步骤中,处理器102在获得比对结果后,可以和设置的阈值进行比较。对于待测基因序列与任意一个参考基因片段的匹配结果,都可以与设置的阈值进行比较。本发明实施例以将待测基因序列与第一组参考基因片段中的第一参考基因片段为例进行描述,其中,第一参考基因片段为第一组参考基因片段中的任意一个参考基因片段。当所述待测基因片段与第一参考基因片段的相似度是否大于或等于第一阈值,该方法进入步骤410,否则该方法进入步骤412。

[0084] 在步骤410中,处理器102记录所述第一参考基因片段在参考基因序列中的位置,结束对所述待测基因序列的匹配。在本发明实施例中,可以将相似度大于或等于第一阈值的匹配结果认为匹配成功。当处理器102确定所述待测基因序列与所述第一参考基因片段匹配成功时,可以记录所述第一基因片段在所述参考基因序列中的位置。结束对所述待测基因序列的匹配,匹配过程结束。可以理解的是,在本发明实施例中,相似度用于指示待测基因序列与参考基因片段的匹配程度。第一阈值用于指示是否达到需求的匹配标准。实际应用中,第一阈值可以用于指示完全匹配,也可以用于指示最大相似度匹配。如果相似度大于或等于设置的第一阈值,则可以认为所述待测基因序列与所述参考基因序列匹配或最大相似匹配。例如,第一阈值可以为100%,也可以为95%,在此不进行限定。

[0085] 若在步骤408中,处理器确定所述待测基因片段与所述第一基因片段的相似度小于所述第一阈值,则在步骤412中,处理器102会进一步判断所述待测基因片段与所述第一基因片段的相似度是否大于第二阈值,当所述待测基因片段与所述第一基因片段的相似度大于第二阈值时,该方法进入步骤414,进入最大相似度匹配的流程。否则该方法进入步骤416,确认所述待测基因序列与所述第一参考基因片段不匹配,结束所述待测基因片段与所述第一基因片段的匹配。在本发明实施例中,可以将第二阈值设置为50%。当待测基因片段与所述第一参考基因片段的相似度小于第一阈值且大于第二阈值时,说明所述待测基因序列与所述参考基因序列能够匹配的可能性较大,或者说,所述待测基因序列中的部分片段可能与所述参考基因序列进行匹配。因此需要进一步的对所述待测基因序列与所述参考基因序列进行比对,该方法进入最大相似度匹配流程。

[0086] 可以理解的是,图4中的步骤408至步骤416是以待测基因序列与第一参考基因片段的匹配为例进行描述。实际应用中,可以通过步骤404和步骤406获得所述待测基因序列

与多个参考基因片段的相似度后,再分别根据所述待测基因序列与每个参考基因片段的相似度按照步骤408和步骤416进行处理。当然,在获得第一组参考基因片段后,也可以依次对待测基因序列与第一组参考基因片段中每个参考基因片段执行步骤404到步骤416的操作。在此,不对具体的实现方式进行限定。

[0087] 本发明实施例提供的基因比对方法,通过构建的基因数据库与所述待测基因序列进行初次匹配,从而筛选出与所述待测基因序列可能匹配的第一组参考基因片段。本领域技术人员知道,以人的参考基因片段为例,人的参考基因片段有30亿个碱基,如果直接将待测基因片段与参考基因片段一一比对,会花费很多时间。而通过本发明实施例提供的基因数据库对待比对的基因片段进行筛选后,可以将需要比对的参考基因片段从30亿降低到几百次,从而大幅度的减少了需要进行比对的参考基因片段的数量。并且,在本发明实施例中,在获得第一组参考基因片段后,进一步通过光计算芯片对所述待测基因序列和所述第一组参考基因片段中的多个参考基因片段进行光学比对,由于光计算芯片进行光学比对,与通过电学方式进行基因比对的方法相比,比对速度更快。从而本发明实施例提供的基因比对方法,也极大的提高了比对效率。

[0088] 需要说明的是,在本发明实施例中,只要待测基因序列与所述第一组参考基因片段中的任意一个参考基因片段的相似度小于所述第一阈值且大于所述第二阈值,则可以按照图6所示的最大相似度匹配方法对所述待测基因序列进一步的进行比对。图6为本发明实施例提供的又一种基因比对方法流程图。图6所示的方法仍然由基因匹配装置100来执行。如图6所示,该方法可以包括下述步骤。

[0089] 在步骤602中,处理器102从参考基因序列中获得多个子参考基因序列。具体的,处理器102根据待测基因序列的长度从参考基因序列中获得多个子参考基因序列。例如,可以以所述待测基因序列的长度为窗口及滑动步长,从所述参考基因序列中获得多个子参考基因序列。也可以按照所述待测基因序列的碱基长度将所述参考基因序列拆分成多个子参考基因序列。例如,如图7所示,可以按照所述待测基因序列702的长度根据所述参考基因序列700获得多个子参考基因序列。以参考基因序列有30亿个碱基为例,如果待测基因序列为150个碱基,则可以获得0.2亿个子参考基因序列。

[0090] 在步骤604中,将所述待测基因序列与步骤602中获得的第*i*个子参考基因序列输入光计算芯片106进行光学比对。*i*的初始值为1,且*i*的值不大于步骤602中获得的子参考基因序列的数量。具体的,处理器102可以分别对所述待测基因序列和所述第*i*子参考基因序列进行光学编码,并将所述待测基因序列和所述第*i*子参考基因序列的光学编码加载到光计算芯片106中进行光学比对,以获得待测基因序列与第*i*子参考基因序列的相似度,光计算芯片106会将比对的结果发送给处理器102。在本发明实施例中,可以将所述待测基因序列与所述多个子参考基因序列中的第一子参考基因序列的相似度称为第一相似度。

[0091] 在步骤606中,处理器102判断所述待测基因序列与所述第*i*个子参考基因序列的相似度是否大于设置的第三阈值。若不大于所述第三阈值,则所述待测基因序列与第*i*个子参考基因序列不匹配,该方法进入步骤608,令*i*=*i*+1,并返回步骤604,继续将所述待测基因序列与下一个子参考基因序列进行比对,直到将所述待测基因序列与在步骤602中获得的所有子参考基因序列通过光计算芯片106完成光学比对。如在步骤606中,处理器102判断所述待测基因序列与所述第*i*个子参考基因序列的相似度大于所述第三阈值,则该方法进

入步骤610。在本发明实施例中,为了尽可能的找到与所述待测基因序列的至少部分片段匹配的参考基因片段,可以将第三阈值设置为低于50%的相似度,例如,可以所述第三阈值可以设置为20%。可以理解的是,实际应用中,第三阈值也可以与第二阈值相同,在此不做限定。

[0092] 若所述待测基因序列与所述第i子参考基因序列的相似度大于所述第三阈值,则在步骤610中,处理器102进一步判断所述待测基因序列与所述第i子参考基因序列的相似度是否大于第四阈值。若所述待测基因序列与所述第i子参考基因序列的相似度大于所述第四阈值,该方法进入步骤612。在本发明实施例中,第四阈值不大于所述第一阈值,第一阈值可以用于指示完全匹配设置的阈值,第四阈值为用于指示最大相似度匹配的阈值。通常第一阈值可以设置为100%,第四阈值可以设置为95%。可以理解的是,实际应用中,第四阈值也可以与第一阈值相同,例如,第一阈值和第四阈值都可以设置为95%,用于指示达到最大相似度匹配的阈值。在此不做限定。在步骤612中,处理器102确定第i子参考基因序列为与所述待测基因序列具有最大相似度的基因片段,记录所述第i子参考基因序列在参考基因序列中的位置,结束对所述待测基因序列的比对流程。若所述待测基因序列与所述第i子参考基因序列的相似度不大于所述第四阈值,该方法进入步骤614。

[0093] 在步骤614中,处理器102根据所述待测基因序列获得第一待测子基因序列和第二待测子基因序列。继续参考图7,在本步骤中,处理器102可以根据待测基因序列702获得第一待测子基因序列7022和第二待测子基因序列7024。其中,第一待测子基因序列7022和第二待测子基因序列7024的部分碱基相同。例如,第一待测子基因序列7022可以包括从所述待测基因序列702的头部开始向尾部方向获取的第一预设长度的碱基,第二待测子基因序列7024可以包括从所述待测基因序列702的尾部开始向头部方向获取的第一预设长度的碱基,第一待测子基因序列7022和第二待测子基因序列7024的部分碱基相同。该方法进入步骤616。

[0094] 在步骤616中,通过光计算芯片106将所述第j待测子基因序列与所述第i子参考基因序列进行光学比对。其中,j的初始值为1,且j的值可以不大于待测子基因序列的数量。由于本发明实施例中从待测基因序列获得了两个待测子基因序列,因此,在本发明实施例中,j的值不大于2。可以理解的是,如果实际应用中,需要获得p(p大于2)个待测子基因序列,则j的值不大于p即可。在本步骤中,处理器102也需要先对第j待测子基因序列单元进行光学编码,然后将所述第j待测子基因序列单元的光学编码与所述第i子参考基因序列的光学编码加载到光计算芯片106进行光学比对,以获得第j待测子基因序列与所述第i子参考基因序列的相似度。该方法进入步骤618。在步骤618中,处理器102确定所述第j待测子基因序列与所述第i子参考基因序列的相似度是否大于所述第三阈值,如果不大于所述第三阈值,则该方法进入步骤620,令 $j = j + 1$,并进入步骤616,将所述第j+1待测子基因序列与所述第i子参考基因序列进行光学比对,以获得第j+1待测子基因序列与所述第i子参考基因序列的相似度。如果在步骤618中,处理器102确定所述第j待测子基因序列与所述第i子参考基因序列的相似度大于所述第三阈值,则该方法进入步骤622,进一步判断所述第j待测子基因序列与所述第i子参考基因序列的相似度是否大于所述第四阈值。在本发明实施例中,为了描述清楚方便,可以将所述光计算芯片对第一待测子基因序列与所述第一子参考基因序列的匹配结果称为第二相似度,将所述光计算芯片对第二待测子基因序列与所述第一子参考基

因序列的匹配结果称为第三相似度。

[0095] 若在步骤622中,处理器102确定所述第j待测子基因序列与所述第i子参考基因序列的相似度大于所述第四阈值,则该方法进入步骤624,记录所述第i子参考基因序列中与所述第j待测子基因序列匹配的参考基因片段在所述参考基因序列中的位置,结束对所述待测基因序列的匹配。需要说明的是,实际应用中,若已经确定第j待测子基因序列与所述第i子参考基因序列的部分片段相似度大于所述第四阈值的情况下,为了提高匹配速度,也可以不继续将所述第j+1待测子基因序列与所述第i子参考基因序列进行匹配,而直接结束对所述待测基因序列的比对流程。当然,可以理解的是,实际应用中,也可以根据需要,继续对第j+1待测子基因序列与所述第i子参考基因序列进行光学比对。

[0096] 若在步骤622中,处理器102确定所述第j待测子基因序列与所述第i子参考基因序列的相似度不大于所述第四阈值,则该方法进入步骤626。在步骤626中,处理器102获取所述第j待测子基因序列的第一待测基因序列单元和第二待测基因序列单元,其中所述第一待测基因序列单元和第二待测基因序列单元的部分碱基相同。具体的,可以参考步骤614中从所述待测基因序列中获取第一待测子基因序列和所述第二待测子基因序列的方法。例如,所述第一待测基因序列单元可以包括从所述第j待测子基因序列的头部向尾部方向获取的第二预设长度的碱基,所述第二待测基因序列单元可以包括从所述j待测子基因序列的尾部向头部方向获取的第二预设长度的碱基。

[0097] 在步骤628中,通过光计算芯片106将第k个待测基因序列单元与所述第i子参考基因序列进行光学比对。其中,k的初始值为1,且k的值不大于待测基因序列单元的数量。在本发明实施例中,由于以根据第j待测子基因序列获得两个待测基因序列单元为例,因此,k的值不大于2。具体的,在步骤628中,处理器102可以对第k个待测基因序列单元进行光学编码,并分别将所述第k个待测基因序列单元的光学编码与所述第i子参考基因序列的光学编码加载到光计算芯片106上进行光学比对。该方法进入步骤630。在步骤630中,处理器102判断所述第k个待测基因序列单元与所述第i子参考基因序列的相似度是否大于所述第三阈值。如果不大于所述第三阈值,则进入步骤632,令 $k=k+1$,并进入步骤628,通过所述光计算芯片106将第二待测基因序列单元与所述第i子参考基因序列进行光学比对。

[0098] 若在步骤630中,处理器102判断所述第k个待测基因序列单元与所述第i子参考基因序列的相似度大于所述第三阈值,则该方法进入步骤634,判断所述第k个待测基因序列单元与所述第i子参考基因序列的相似度是否大于第四阈值,如果大于所述第四阈值,则该方法进入步骤636,记录所述第i子参考基因序列中与所述第k待测基因序列单元匹配的基因片段在所述参考基因序列的位置,并结束匹配。具体的,一种情形下,为了提高匹配速度,在获得最大相似度的基因片段后,可以结束对所述待测基因序列的匹配。另一种情形下,也可以结束对所述第j待测子基因序列的匹配,或结束对所述第k待测基因序列单元的匹配,而继续对所述k+1待测基因序列单元的匹配或继续对第j+1待测子基因序列进行匹配。

[0099] 若在步骤634中,处理器102确定所述第k个待测基因序列单元与所述第i子参考基因序列的相似度不大于所述第四阈值,则该方法进入步骤638,按照递归的方式继续对所述第k待测基因序列单元进行拆分,并对所述第k待测基因序列单元的子单元与所述第i子参考基因序列进行光学比对,直到找到与所述第i子参考基因序列的相似度大于所述第四阈值的待测基因片段为止。在本发明实施例中,可以将子参考基因片段中与所述待测基因序

列中的部分待测基因片段的相似度大于所述第四阈值的参考基因片段称为最大相似基因片段。

[0100] 本发明实施例提供的基因比对方法,对于通过图4未能精确匹配的待测基因片段,能够继续通过图6所示的基因比对方法进一步对待测基因片段进行最大相似度匹配。由于图6所示的方法可以允许所述待测基因与获得的最大相似基因片段不完全一致,待测基因序列中可能有部分碱基的缺失或与参考基因片段不同,从而能够实现对待测基因序列中缺失的基因或变异基因的精确定位。

[0101] 在又一种情况下,本发明实施例提供的基因比对方法还可以包括图8所示的方法流程。图8所示的方法可以在图6所示的步骤604之后。如图8所示,该方法可以包括下述步骤。在步骤802中,处理器102确定所述待测基因序列与所述第 i 子参考基因序列的第一相似度小于第三阈值。并且,在步骤804中,处理器102进一步确定所述待测基因序列与所述第 $i+1$ 子参考基因序列的第二相似度大于所述第三阈值时,该方法进入步骤806。需要说明的是,步骤802和步骤804的描述可以参考图6中关于步骤606的描述,第三阈值可以与步骤606中设置的第三阈值相同,例如可以为50%。

[0102] 在步骤806中,处理器进一步判断所述第一相似度与所述第二相似度的和是否大于100%。如果所述第一相似度与所述第二相似度的和不大于100%,则该方法进入步骤808,通过光计算芯片106将所述待测基因序列与所述第 $i+2$ 子参考基因序列进行光学比对。如果所述第一相似度与所述第二相似度的和大于100%,则该方法进入步骤810。在步骤810中,所述处理器102根据所述第 i 子参考基因序列和所述第 $i+1$ 子参考基因序列获得新的子参考基因序列。在步骤810中,可以根据第一相似度和第二相似度的比例分别从所述第 i 子参考基因序列和第 $i+1$ 子参考基因序列获取一部分参考基因片段组成新的子参考基因序列。例如,如果第一相似度为40%,第二相似度为80%,一个参考基因序列的长度是150个碱基对,则可以将所述第 i 子参考序列尾部的50个碱基对与所述第 $i+1$ 头部的100个碱基对组成一个连续的长度为150碱基对的新的子参考序列。在获得新的子参考序列后,该方法进入步骤812,通过光计算芯片106将所述待测基因序列与获得的所述新的子参考序列进行光学比对,具体的光学比对方法可以参见图6中步骤604的描述。并且,在将所述待测基因序列与获得的所述新的子参考序列比对的过程中,可以参见图6中将所述待测基因序列与第 i 子参考序列进行比对的过程。根据这种方式,如果所述待测基因序列与所述新的子参考基因序列的相似度大于所述第三阈值,则可以继续参见图6中步骤610至步骤638的方法,从所述新的子参考序列中查找与所述待测基因序列的相似度大于所述第四阈值的参考基因片段。在本发明实施例中,可以将按照图6和图8所示的比对方法在参考基因序列中查找到的与所述待测基因序列的相似度大于所述第四阈值的参考基因片段均称为最大相似基因片段。

[0103] 图8所示的方法可以与图6所示的方法结合使用。例如,当确定所述待测基因序列与所述第 i 子参考基因序列的相似度较低,而与第 $i+1$ 子参考基因序列的相似度较高时,则可以转而执行图8所示的方法,从而能够通过从第 i 子参考基因序列和第 $i+1$ 子参考基因序列中获得新的参考基因序列与所述待测基因序列进行比对。这种根据部分比对结果及时调整子参考基因序列的方式,能够提高获得最大相似基因片段的概率和速度,减少比对次数。可以理解的是,实际应用中,也可以先按照图6所示的方法将所述待测基因序列与步骤602中获得的多个子参考基因序列比对完成后,再执行图8所示的方法,调整子参考基因序列后

进行比对。在本发明实施例中,不对具体的执行方式进行限定。

[0104] 需要说明的是,图8是以所述待测基因序列与第*i*子参考基因序列为例进行描述,实际应用中,第*i*子参考基因序列可以是所述多个子参考基因序列中的任意一个子参考基因序列。例如,在步骤802中,处理器可以以将所述待测基因序列与所述多个子参考基因序列中第二子参考基因序列的比对情况为例,若所述待测基因序列与所述第二子参考基因序列的相似度为第四相似度,且第四相似度小于所述第三阈值。在步骤804中,处理器102确定所述待测基因序列与所述多个子参考基因序列中第三子参考基因序列的相似度为第五相似度,第五相似度大于所述第三阈值。若在步骤806中,处理器进一步判断所述第四相似度与所述第五相似度的和大于100%,则处理器102可以按照图8所示的方法根据第二子参考基因序列和第三子参考基因序列获得新的子参考基因序列。

[0105] 在本发明实施例中,在通过图6和图8的方法查找到所述待测基因序列的最大相似基因片段后,还可以通过Smith-Waterman局部比对算法将所述最大相似基因片段在所述待测基因序列和所述参考序列上进行扩展,以便能得到更长的最大相似基因片段,从而便于更好的对待测基因片段做进一步的基因分析工作。

[0106] 可以理解的是,上述实施例所示的方法是以待测基因序列与所述多个子参考基因序列中一个子参考基因序列的比对为例进行描述。实际应用中,可以与多个子参考基因序列分别进行比对,在此不进行限定。本申请实施例提及“第一”、“第二”等序数词用于对多个对象进行区分,不用于限定多个对象的顺序、时序、优先级或者重要程度。

[0107] 可以理解的是,本发明实施例的比对方法仅仅是以基因比对为例。实际应用中还可以将本发明实施例提供的将基于数据库实现的电学比对方法和基于光计算芯片的光学比对相结合的比对方法应用于各种其他的应用场景。图9为本发明实施例提供的一种比对的示意图。该比对装置可以用于实现包括基因比对在内的各种数据比对场景。

[0108] 如图9所示,该比对装置900可以包括处理器902、存储器904和光计算芯片906。其中,处理器902用于根据待匹配的第一对象从存储器904存储的数据库中获取第一组参考对象,其中,所述第一组参考对象中包括与所述第一对象的部分特征相同的多个参考对象。光计算芯片906用于连接所述处理器并用于对所述第一对象以及所述多个参考对象进行光学比对。处理器902还可以用于根据所述光计算芯片的输出结果确定所述第一对象与所述多个参考对象的相似度。

[0109] 在又一种情形下,处理器902还可以用于根据所述光计算芯片的输出结果,确定所述第一对象与所述第一组参考对象中的第一参考对象的相似度小于第一阈值其而大于第二阈值,并根据标准对象获得多个子参考对象,其中,每个子参考对象为所述参考对象的一部分。所述光计算芯片906还可以用于对所述第一对象与所述多个子参考对象中的第一子参考对象进行光学比对,获得所述第一对象与所述第一子参考对象的第一相似度。

[0110] 在又一种情形下,所述处理器902还可以用于确定所述第一相似度大于第三阈值且小于第四阈值,并且响应上述确定,根据所述第一对象获得第一子对象和第二子对象,其中,所述第四阈值不大于所述第一阈值,所述第一子对象和所述第二子对象的部分数据相同。所述光计算芯片906还可以用于对所述第一子对象和所述第一子参考对象进行光学比对,获得第二相似度,并且,对所述第二子对象和所述第一子参考对象进行光学比对,获得第三相似度。实际应用中,处理器902还可以用于当所述第二相似度大于所述第四阈值时,

记录所述第一子参考对象在所述标准对象中的位置。

[0111] 可以理解的是,图9所示的比对装置可以用于实现图1所示的比对装置的功能,对图9的比对装置的描述可以本发明实施例中前述图1-图8的描述。图9所示的比对装置均可以应用于包括基因比对在内的各种需要进行数据比对或特征比对的场景。可以说,图1所示的基因比对装置是图9所示的比对装置的一种具体应用。需要说明的是,图9所示的比对装置以及本方面实施例提供的比对方法还可以用于进行图片比对、以图搜图、序列比对、模糊匹配等场景,在此不进行限定。

[0112] 图10为本发明实施例提供的另一种比对装置的示意图。如图10所示,该比对装置1000可以包括获取模块1002、比对模块1004和结果处理模块1006。获取模块1002用于根据待测基因序列从数据库中获得第一组基因片段,其中,所述数据库系统中包含有参考基因序列的多个参考基因片段,所述第一组基因片段包括与所述待测基因序列的部分碱基匹配的多个参考基因片段。比对模块1004用于对所述待测基因序列与所述第一组基因片段中的多个参考基因片段进行光学比对。结果处理模块1006用于根据所述比对模块1004的输出结果确定所述待测基因序列与所述第一组基因片段中的所述多个参考基因片段的相似度。

[0113] 在又一种情形下,比对装置1000还可以包括判断模块1008。判断模块1008用于根据所述比对模块1004的输出结果确定所述待测基因序列与所述第一组基因片段中的第一基因片段的相似度小于第一阈值且大于第二阈值。所述获取模块1002还用于当判断模块1008判断所述待测基因序列与所述第一组基因片段中的第一基因片段的相似度小于第一阈值且大于第二阈值时,从所述参考基因序列获得多个子参考基因序列,其中,每个子参考基因序列为所述参考基因序列的一部分。所述比对模块1004还用于对所述待测基因序列与所述多个子参考基因序列中的第一子参考基因序列进行光学比对。所述结果处理模块1006还用于根据所述光计算芯片的输出结果获得所述待测基因序列与所述第一子参考基因序列的第一相似度。

[0114] 在又一种情形下,所述判断模块1008还用于确定所述第一相似度大于第三阈值且小于第四阈值,其中,所述第四阈值不大于所述第一阈值。所述获取模块1002还用于响应于所述判断模块1008的判断,根据所述待测基因序列获得第一待测子基因序列和第二待测子基因序列,其中,所述第一待测子基因序列和第二待测子基因序列的部分碱基相同。所述比对模块1004还用于将所述第一待测子基因序列与所述第一子参考基因序列进行光学比对,获得第二相似度,以及将所述第二待测子基因序列与所述第一子参考基因序列进行光学比对,获得第三相似度。

[0115] 在又一种情形下,所述结果处理模块1006还用于当所述第二相似度大于所述第四阈值时,记录所述第一子参考基因序列在所述参考基因序列中的位置。

[0116] 在又一种情形下,获取模块1002还用于当判断模块1008判断所述第三相似度大于所述第三阈值且小于所述第四阈值时,根据所述第二待测子基因序列获得第一待测子基因序列单元和第二待测子基因序列单元,其中,所述第一待测子基因序列单元和所述第二待测子基因序列单元的部分碱基相同。所述比对模块1004还用于将所述第一待测子基因序列单元与所述第一子参考基因序列进行光学比对,以及将所述第二待测子基因序列单元与所述第一子参考基因序列进行光学比对。

[0117] 在又一种情形下,所述比对模块1004还用于将所述待测基因序列与所述多个子参

考基因序列中的第二子参考基因序列进行光学比对,获得所述待测基因序列与所述第二子参考基因序列的第四相似度;并且,将所述待测基因序列与所述多个子参考基因序列中的第三子参考基因序列进行光学比对,获得所述待测基因序列与所述第三子参考基因序列的第五相似度,其中,所述第三子参考基因序列为与所述第二子参考基因序列连续的子参考基因序列。所述判断模块1008在确定所述第四相似度和所述第五相似度的和大于所述第一阈值时,所述获取模块1002还用于根据所述第二子参考基因序列和所述第三子参考基因序列获得第四子参考基因序列,其中,所述第四子参考基因序列包括所述第二子参考基因序列的部分碱基以及所述第三子参考基因序列的部分碱基。所述比对模块1004还用于将所述待测基因序列和所述第四子参考基因序列输入所述光计算芯片进行光学比对。

[0118] 在又一种情形下,所述结果处理模块1006还用于根据所述光计算芯片的输出结果确定所述第一组基因片段中的第二基因片段与所述待测基因序列匹配,并记录所述第二基因片段在所述参考基因序列中的位置。

[0119] 可以理解的是,图10所示的比对装置可以用于实现图1所示的基因比对装置的功能。具体可以参见前面对图1相关模块的功能的描述。在此不在赘述。可以理解的是,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。例如,多个模块或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,上述实施例所讨论的模块相互之间的连接可以是电性、机械或其他形式。所述作为分离部件说明的模块可以是物理上分开的,也可以不是物理上分开的。作为模块显示的部件可以是物理模块或者也可以不是物理模块。另外,在申请实施例各个实施例中的各功能模块可以独立存在,也可以集成在一个处理模块中。

[0120] 本发明实施例还提供一种用于实现基因比对的计算机程序产品,包括存储了程序代码的计算机可读存储介质,所述程序代码包括的指令用于执行前述任意一个方法实施例所述的方法流程。本领域普通技术人员可以理解,前述的存储介质包括:U盘、移动硬盘、磁碟、光盘、随机存储器(random-access memory, RAM)、固态硬盘(solid state disk, SSD)或者非易失性存储器(non-volatile memory)等各种可以存储程序代码的非短暂性的(non-transitory)机器可读介质。

[0121] 需要说明的是,本申请所提供的实施例仅仅是示意性的。所属领域的技术人员可以清楚的了解到,为了描述的方便和简洁,在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其他实施例的相关描述。在本发明实施例、权利要求以及附图中揭示的特征可以独立存在也可以组合存在。在本发明实施例中以硬件形式描述的特征可以通过软件来执行,反之亦然。在此不做限定。

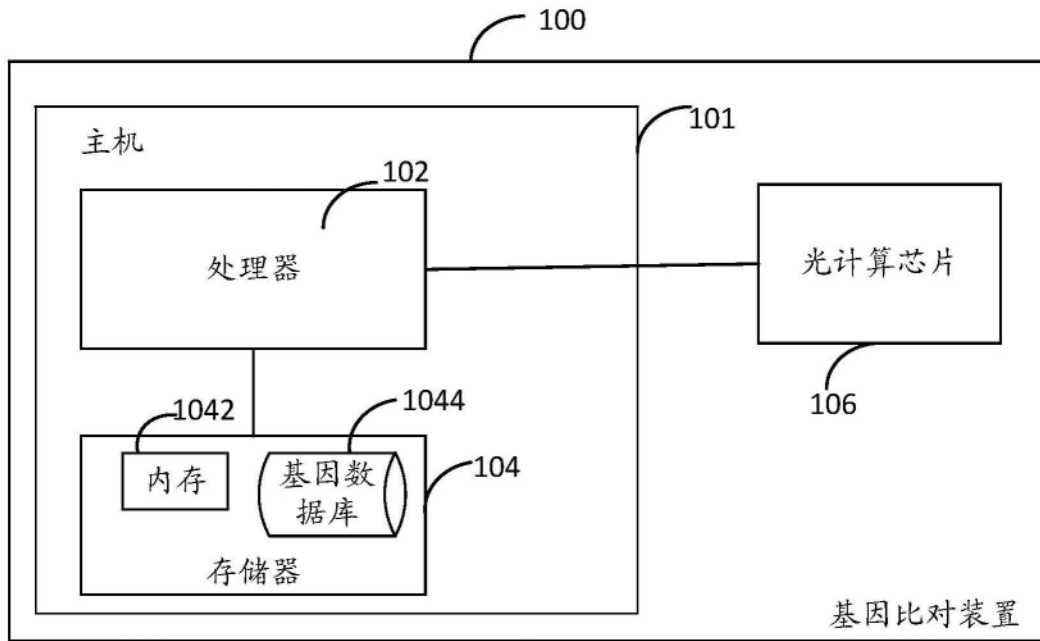


图1

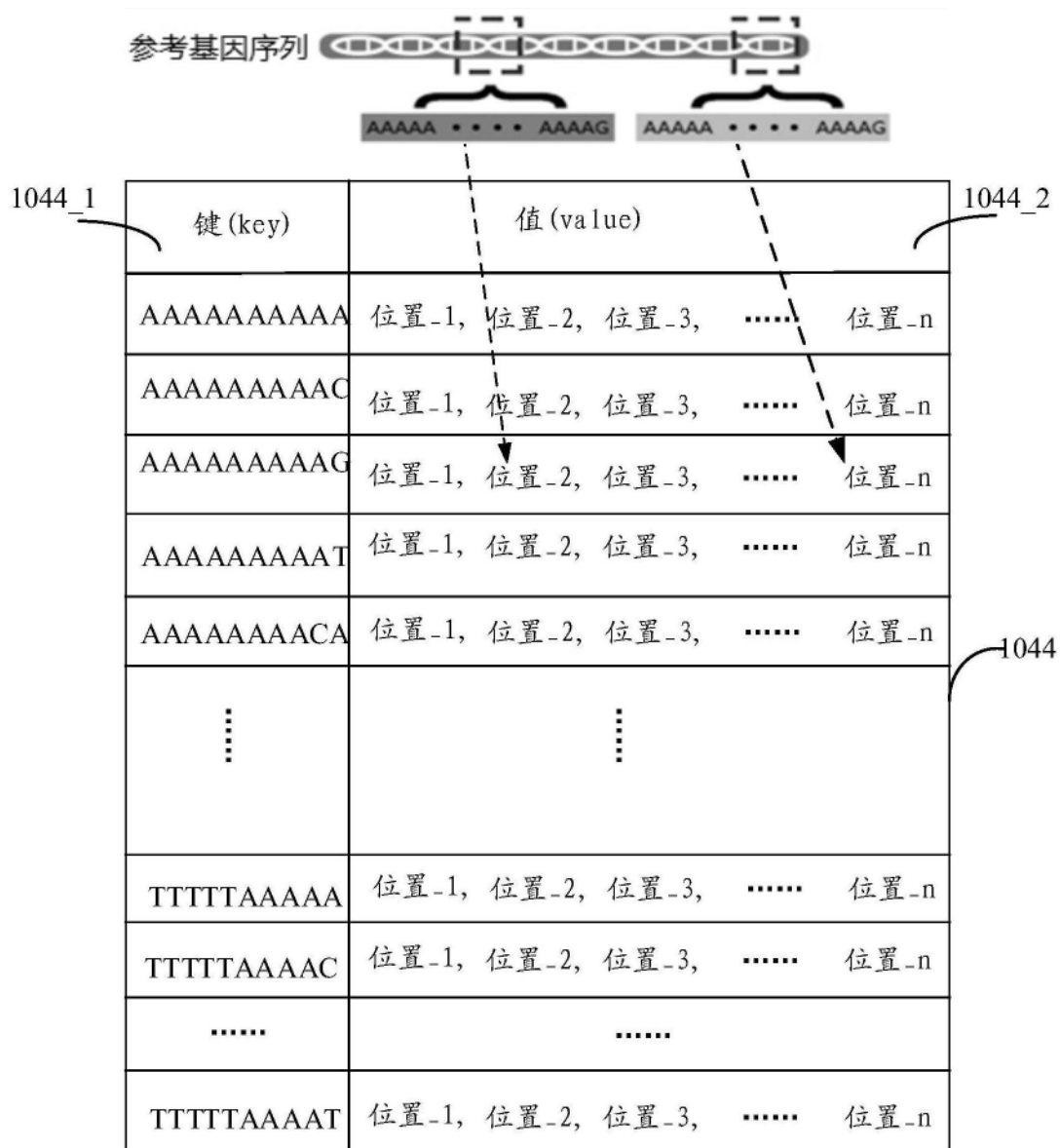


图2A

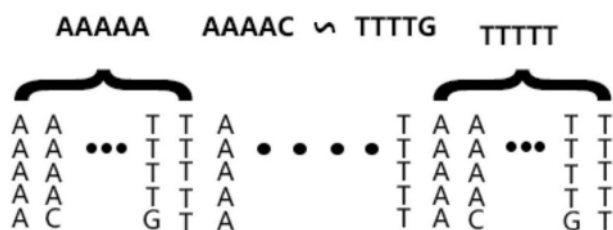


图2B

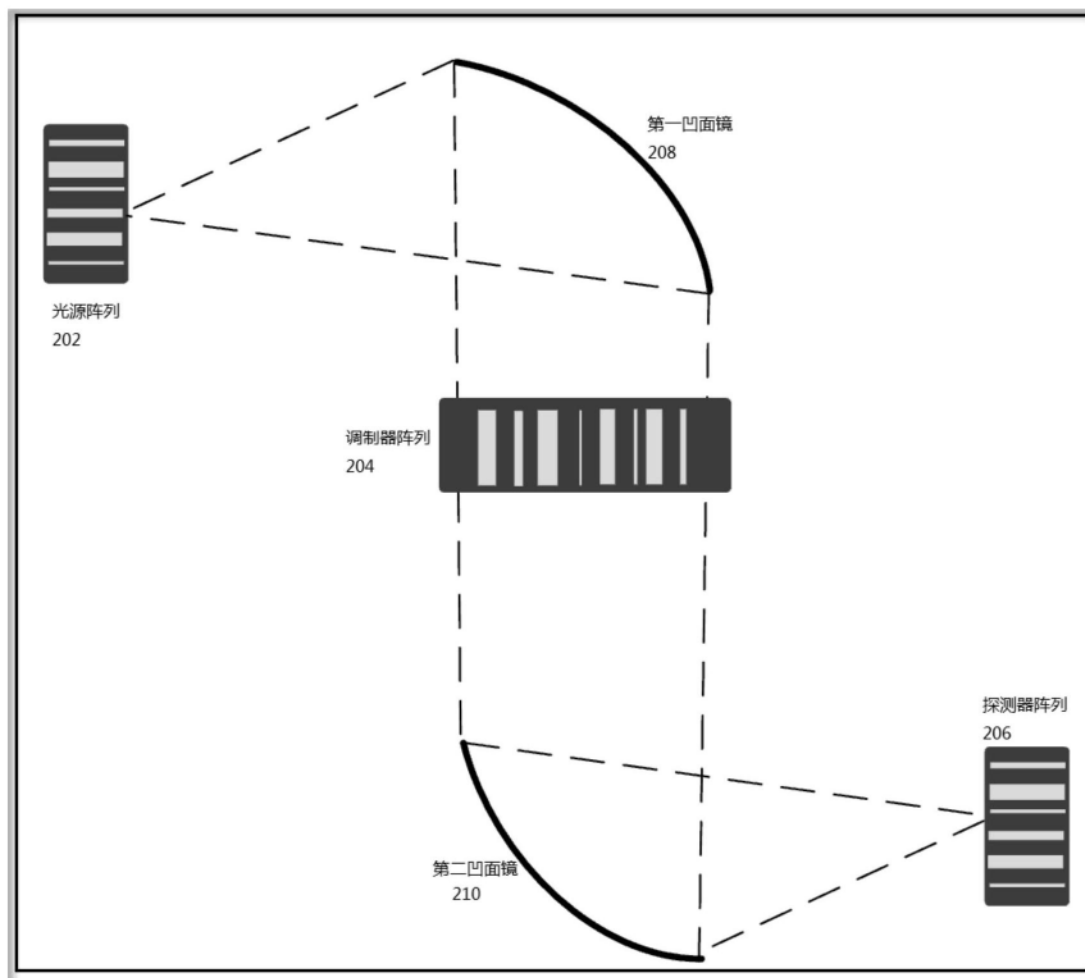


图3A

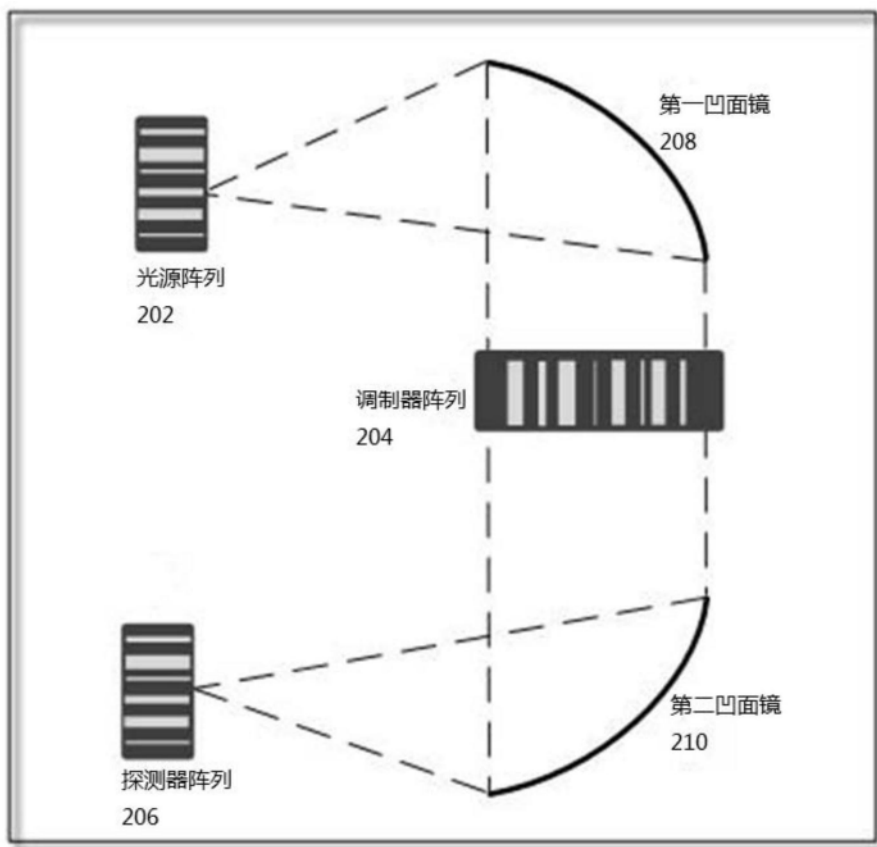


图3B

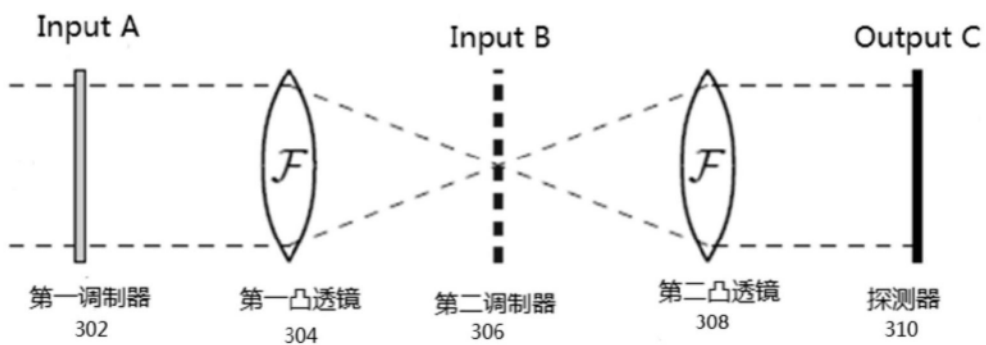


图3C

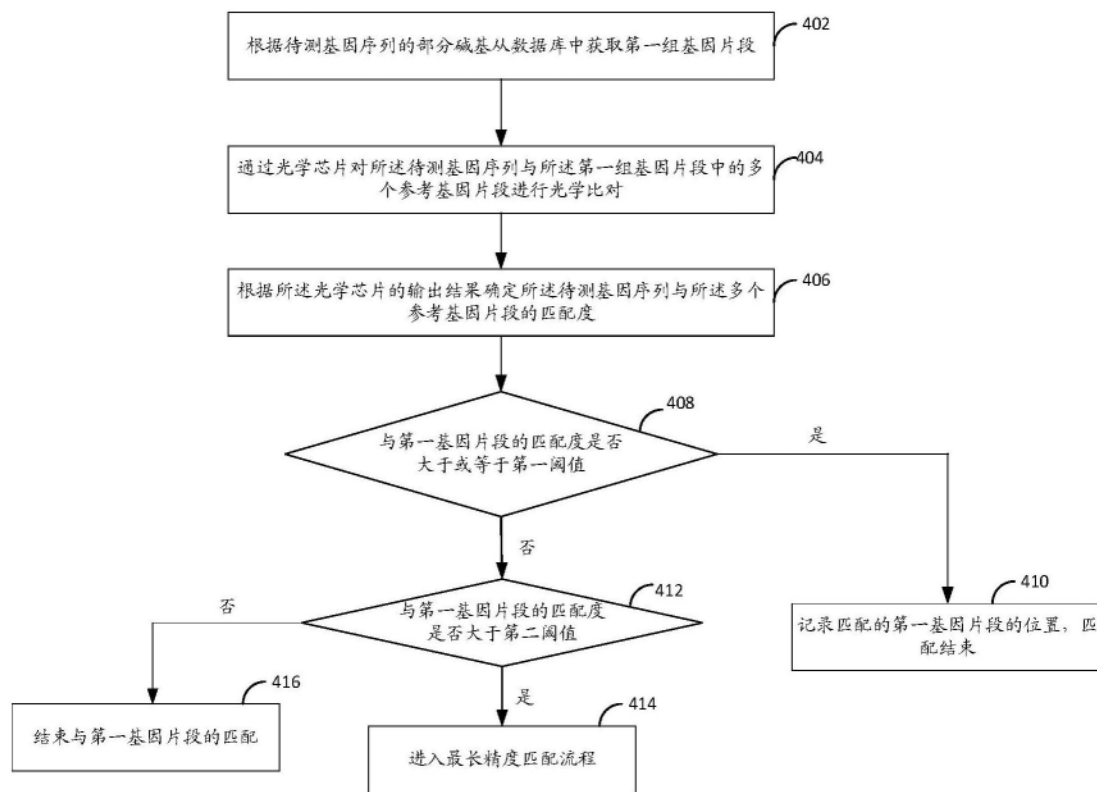


图4

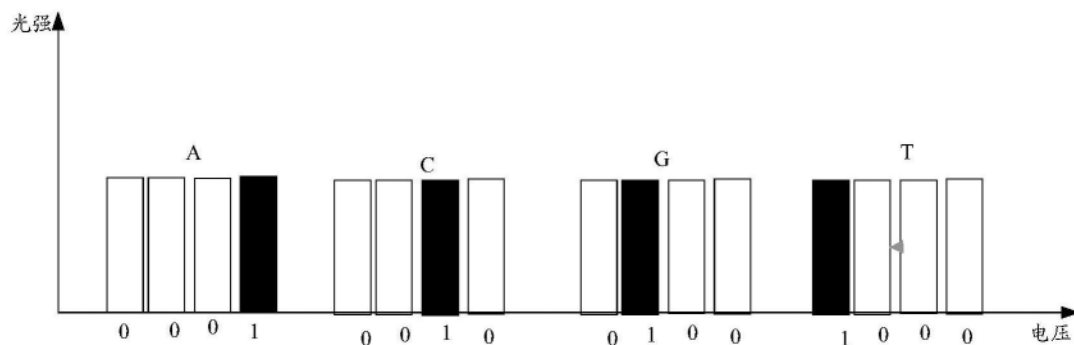


图5A

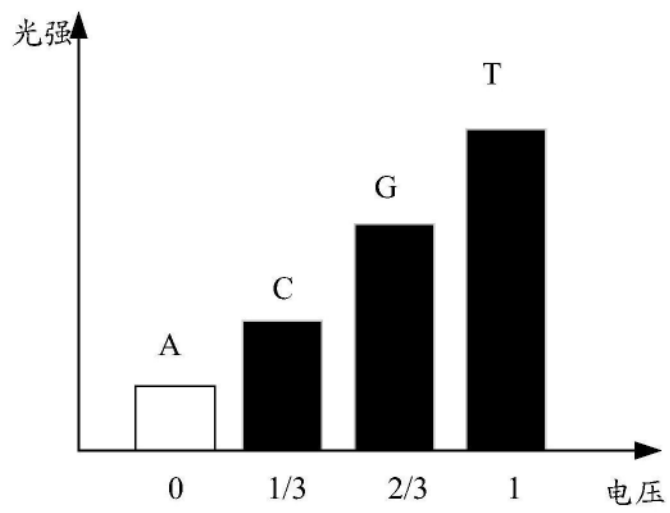


图5B

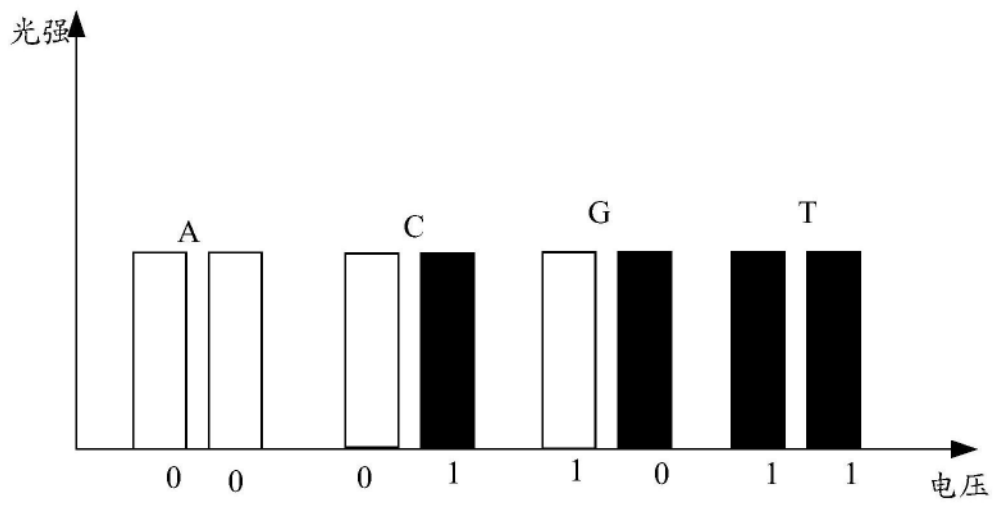


图5C

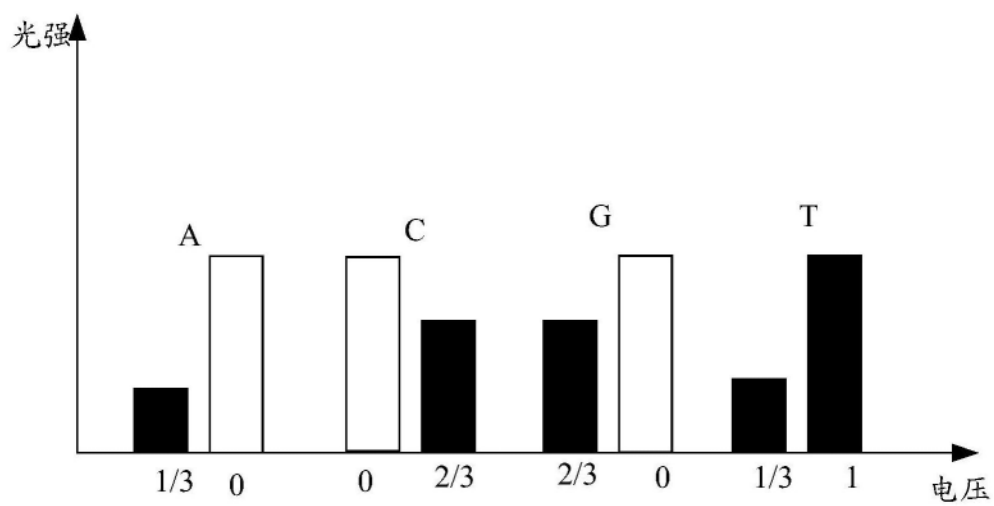


图5D

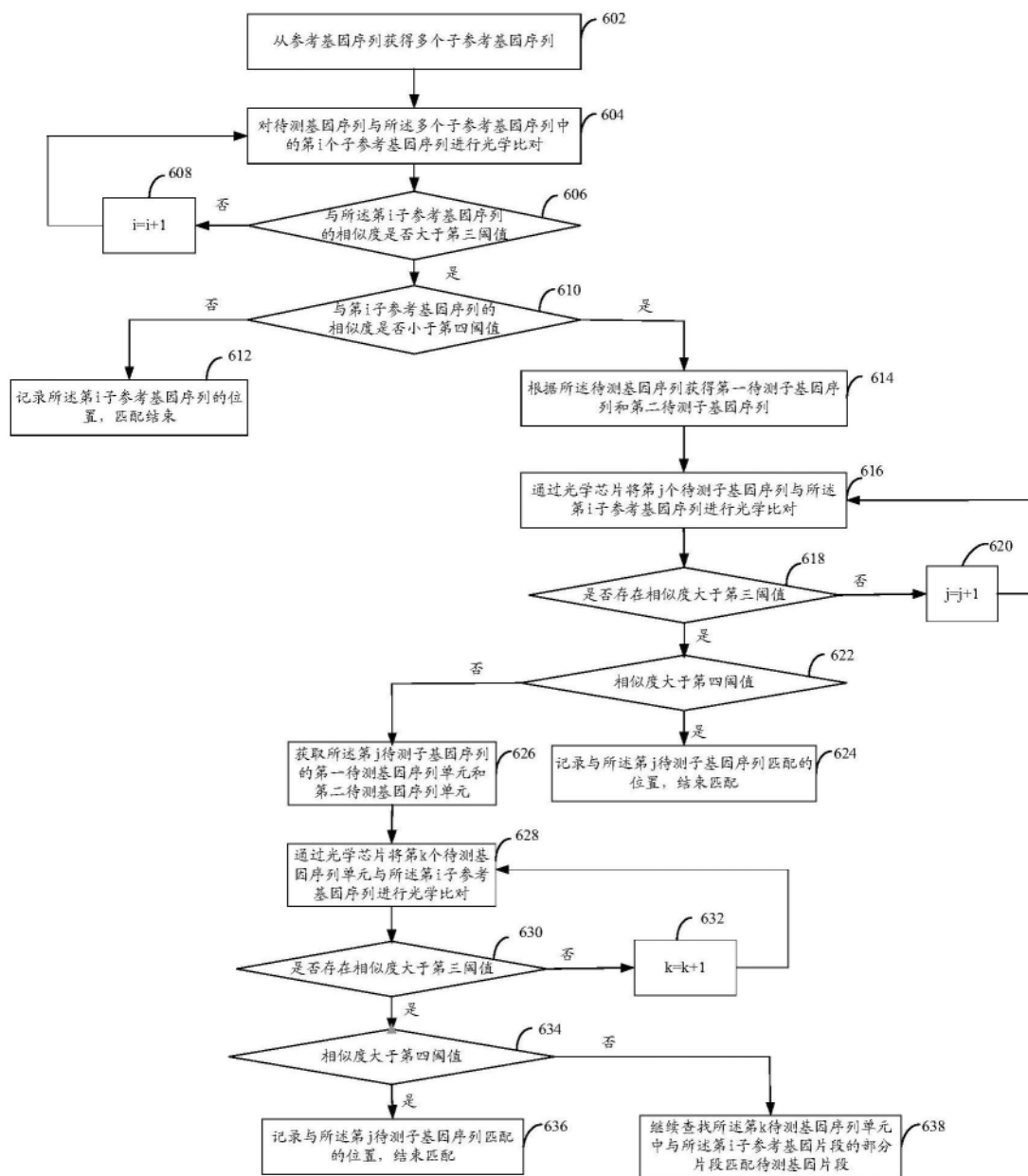


图6



图7

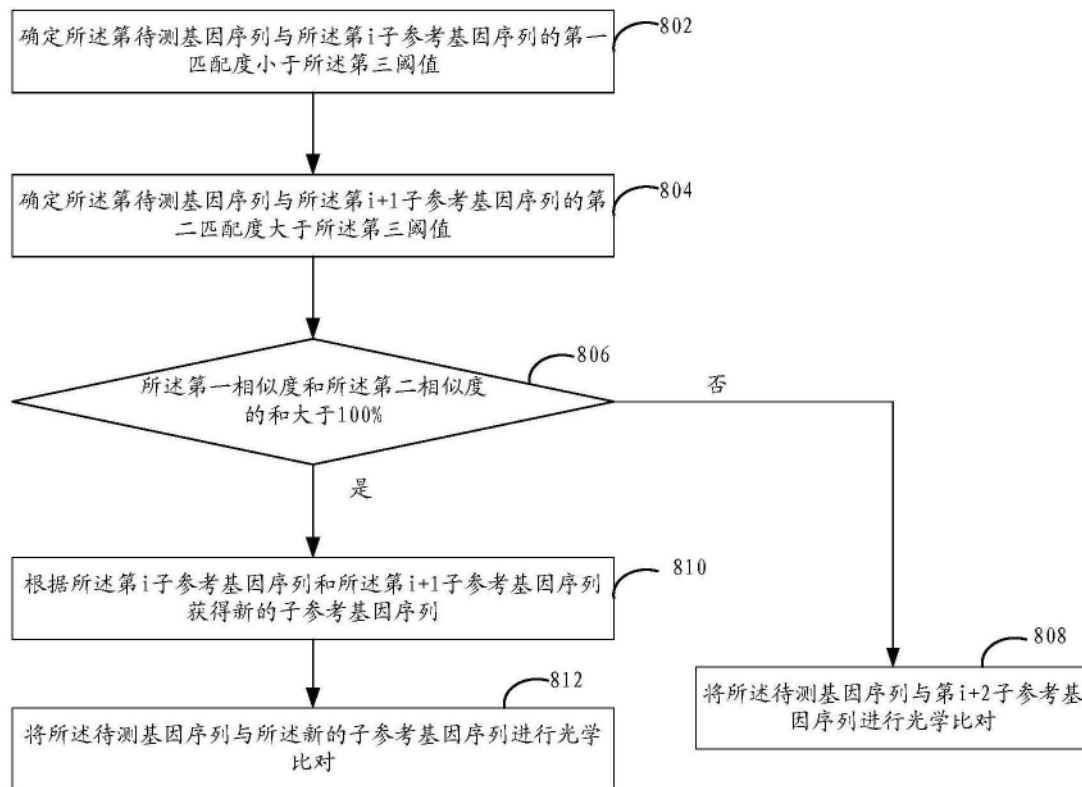


图8

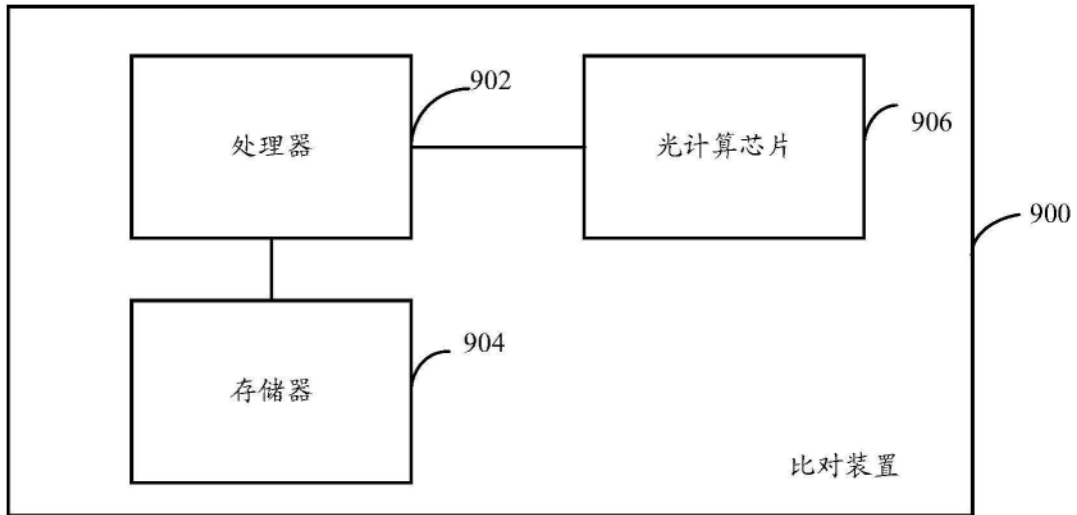


图9

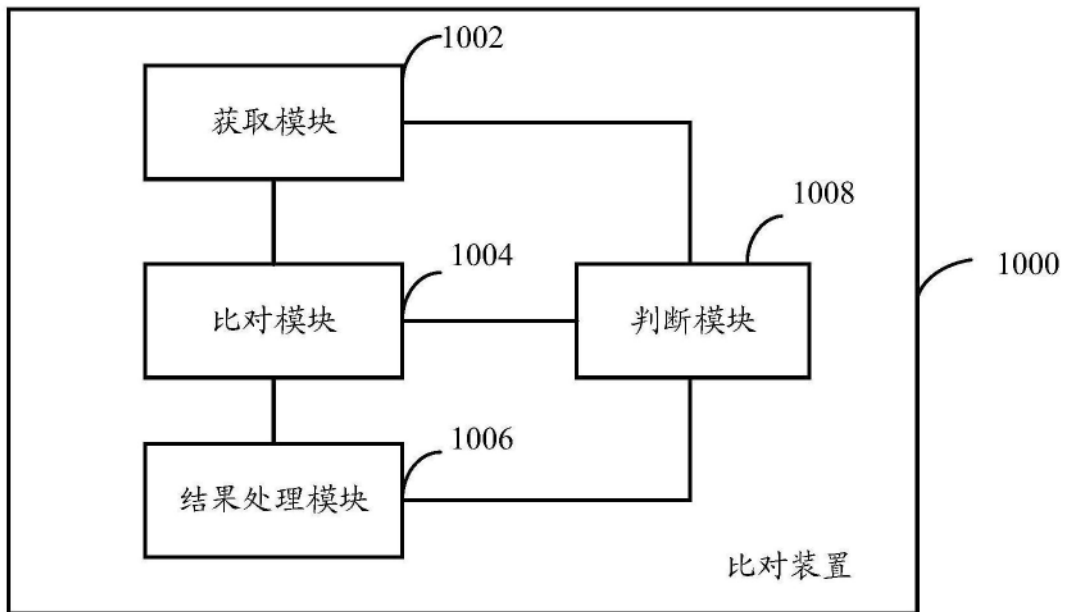


图10