



(12)发明专利

(10)授权公告号 CN 105431553 B

(45)授权公告日 2020.02.07

(21)申请号 201480038087.X
 (22)申请日 2014.05.29
 (65)同一申请的已公布的文献号
 申请公布号 CN 105431553 A
 (43)申请公布日 2016.03.23
 (30)优先权数据
 61/828,582 2013.05.29 US
 (85)PCT国际申请进入国家阶段日
 2015.12.31
 (86)PCT国际申请的申请数据
 PCT/US2014/040082 2014.05.29
 (87)PCT国际申请的公布数据
 W02014/194131 EN 2014.12.04
 (73)专利权人 生物辐射实验室股份有限公司
 地址 美国加利福尼亚州
 (72)发明人 J·希利
 (74)专利代理机构 上海专利商标事务所有限公
 司 31100
 代理人 余颖 沈端

(51)Int.Cl.
 C12Q 1/6869(2018.01)
 C40B 30/10(2006.01)
 C40B 40/06(2006.01)
 (56)对比文件
 WO 2012078710 A1,2012.06.14,
 WO 0040758 A2,2000.07.13,
 CN 101946010 A,2011.01.12,
 US 2008166720 A1,2008.07.10,
 Calero O et al..Apolipoprotein E
 genotyping method by Real Time PCR, a
 fast and cost-effective alternative to
 the TaqMan® and FRET assays.《Journal of
 neuroscience methods》.2009,第183卷(第2
 期),摘要,材料和方法,表1.
 Calero O et al..Apolipoprotein E
 genotyping method by Real Time PCR, a
 fast and cost-effective alternative to
 the TaqMan® and FRET assays.《Journal of
 neuroscience methods》.2009,第183卷(第2
 期),摘要,材料和方法,表1.

审查员 王胜佳

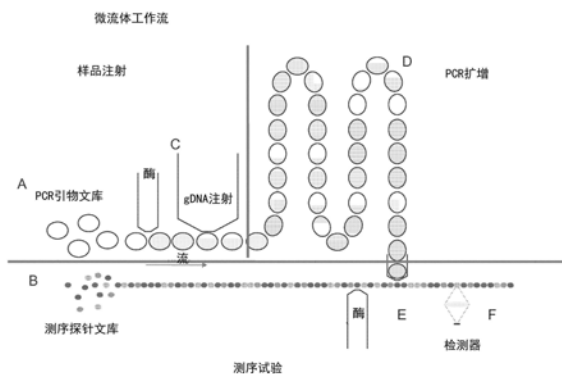
权利要求书2页 说明书43页 附图27页

(54)发明名称

用于在基于乳液的微流体中测序的系统和
方法

(57)摘要

提供了用于核苷酸测序的方法、文库和试剂
盒。



1. 一种确定靶核酸中核苷酸序列的方法,所述方法依赖于不同引物序列与靶核酸的杂交或缺乏杂交,所述方法包括,

划分包含靶核酸的多个拷贝的混合物,从而生成多个靶划分产物,所述靶划分产物是液滴;

以多个引物划分产物提供一系列被设计成与靶核酸杂交的引物,至少大部分所述引物划分产物包含两种或更多种不同的引物,所述两种或更多种不同的引物具有至少4个指定核苷酸,所述引物挑选为依据核酸序列中存在的变体与靶核酸序列杂交,其中,各引物出现在多个不同引物划分产物中,并且,任两个引物组的交集都不含超过一种共同的引物,各引物组含有已知数目和特征的引物,所述一系列被设计成与靶核酸杂交的引物包含指定核苷酸的引物序列的可能组合的至少40%;

以一比一合并 (i) 所述靶划分产物的至少一部分和 (ii) 引物划分产物以形成多个反应划分产物;

在完全互补的引物与所述靶核酸序列杂交且不完全互补的引物不与所述靶核酸杂交的条件下在所述反应划分产物中使所述靶核酸与所述引物杂交;

确定所述系列中的何种引物与所述靶核酸杂交,当且仅当划分产物中含有的所有引物都不杂交时,划分产物生成试验阴性应答;所述靶核酸包含荧光部分且退火至包含淬灭剂的淬灭寡核苷酸,所述淬灭剂退火至所述靶核酸淬灭所述荧光部分的荧光;所述确定包括使退火至淬灭剂寡核苷酸的所述靶核酸接触引物依赖性聚合酶,如果杂交,所述引物的延伸导致所述淬灭剂寡核苷酸被置换,从而生成荧光信号;所述确定还包括检测是否存在所述荧光信号,存在所述荧光信号表示所述多种引物之一与所述靶核酸杂交,且缺少所述荧光信号表示所述多种引物无一与所述靶核酸杂交;以及

基于何种引物与所述靶核酸杂交以及何种引物与所述靶核酸不杂交来确定靶核酸中的核苷酸序列;

所述引物划分产物含有一种或多种与划分产物中具体引物一致性关联的分光光度物质,由此可通过检测所述划分产物的分光光度特征来确定所述划分产物中引物的序列,且所述确定还包括检测分光光度特征并将所述分光光度特征与所述引物的序列关联。

2. 如权利要求1所述的方法,所述一系列引物包含指定核苷酸数目不同的至少2种引物,指定核苷酸多于4个核苷酸。

3. 如权利要求1所述的方法,所述一系列引物包含指定核苷酸数目不同的至少3种引物,指定核苷酸多于4个核苷酸。

4. 如权利要求1所述的方法,所述一系列引物包含指定核苷酸数目不同的至少4种引物,指定核苷酸多于4个核苷酸。

5. 如权利要求1所述的方法,其中,生成至少100份部分所述靶划分产物;且所述合并包括将这100份部分靶划分产物各自分别与不同的引物划分产物合并。

6. 如权利要求1所述的方法,所述靶核酸是扩增子。

7. 如权利要求1所述的方法,所述方法还包括基于所述荧光信号的存在或不存在和所述引物序列反卷积所述靶核酸的核苷酸序列。

8. 如权利要求1所述的方法,所述一系列引物包含n组引物,不同的划分产物含有不同的引物组,不同的引物组具有2-20种不同的独特引物,且不同的引物划分产物之间具有不

超过一种共有引物,其中,n是1000-300000。

9.如权利要求8所述的方法,其中,n组中每一组的划分产物各含有一种或多种分光光度物质,由此使得n组中的每一组都可通过分光光度特征区分。

10.如权利要求8所述的方法,所述组中的至少大部分引物出现在两个不同的引物组中。

11.如权利要求1所述的方法,其中,一个引物划分产物中任两种引物的序列没有超过两个核苷酸的重叠。

12.如权利要求1所述的方法,所述系列中的一些引物具有6-18个指定核苷酸。

13.如权利要求1或12所述的方法,所述指定核苷酸是连续的。

14.如权利要求1或12所述的方法,其中,所述指定核苷酸至少一些是不连续的,由此使得所述指定核苷酸中至少两个被至少一个简并核苷酸位置或通用核苷酸隔开。

15.如权利要求1所述的方法,所述系列具有400-16000种不同序列的引物。

16.如权利要求1所述的方法,所述靶核酸的长度为50-1000个核苷酸。

17.如权利要求1所述的方法,所述靶核酸包含3' 茎环。

18.如权利要求1所述的方法,所述液滴包含在乳液中。

用于在基于乳液的微流体中测序的系统和方法

[0001] 相关申请的交叉参考

[0002] 本申请要求2013年5月29日提交的美国临时专利申请号61/828,582的优先权,该文通过引用纳入本文。

背景技术

[0003] DNA测序是确定DNA分子中核苷酸的精确顺序的过程,例如DNA链中四种碱基(腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶)的顺序。快速DNA测序方法的出现极大地加速了生物和药理学研究和发现。

[0004] DNA序列的知识已用于基础生物学研究和多种应用领域,例如诊断、生物技术、法医生物学和生物系统学。使用现代DNA测序技术获得快速测序已被用于完整DNA序列或多种类型和物种的基因组(包括人基因组和许多动物、植物和微生物物种的其他完整DNA序列)的测序。

发明内容

[0005] 本发明提供了确定靶核酸中核苷酸序列的方法。在一些实施方式中,该方法包括,划分包含多个靶核酸拷贝的混合物,从而生成多个靶划分产物;在多个引物划分产物中提供一系列引物,至少大部分引物划分产物包含两种或更多种引物,所述引物具有至少4个指定核苷酸,这些引物的其他核苷酸位置(如果有)是简并位置或通用核苷酸;以一比一的基础合并(i)至少一部分靶划分产物和(ii)引物划分产物以形成多个反应划分产物;在完全互补的引物与靶核酸杂交且不完全互补的引物不与靶核酸杂交的条件下在反应划分产物中使靶核酸杂交引物;确定该系列中哪些引物结合靶核酸;以及基于结合靶核酸的引物确定靶核酸中的核苷酸顺序。

[0006] 在一些实施方式中,该系列的引物包含指定核苷酸数目不同的至少2、3或4种引物,指定核苷酸多于4、5、6、7、8、9或10个核苷酸,例如一些引物具有一种数目(如 $X > 4$)的指定核苷酸且该系列中的其他引物具有不同数目(如 $Y > 4$)。

[0007] 在一些实施方式中,生成至少100、500、1000或10000份部分的靶划分产物;且所述合并包括将100、500、1000或10000份部分中的每一个与不同的引物划分产物合并。

[0008] 在一些实施方式中,该靶核酸是扩增子。

[0009] 在一些实施方式中,该靶核酸包含荧光部分且退火至包含淬灭剂的淬灭寡核苷酸,其中,淬灭剂退火至靶核酸淬灭荧光部分的荧光。在一些实施方式中,所述确定包括使退火至淬灭剂寡核苷酸的靶核酸接触引物依赖性聚合酶,如果杂交,该引物的延伸导致淬灭剂寡核苷酸的置换,从而生成荧光信号。在一些实施方式中,分成的各份引物包含具有不同序列的多种引物,且所述确定包括:检测是否存在荧光信号,存在荧光信号表示多种引物之一杂交至靶核酸而缺少荧光信号表示多种引物无一杂交至靶核酸;且该方法还包括基于是否存在荧光信号和引物序列反卷积(deconvolute)靶核酸的核苷酸序列。

[0010] 在一些实施方式中,这些引物划分产物含有一种或多种与划分产物中具体引物具

有一致性关联的分光光度物质,从而可通过检测划分产物的分光光度特征确定划分产物中引物的序列,且所述确定还包括检测分光光度特征并将分光光度特征与引物序列关联。

[0011] 在一些实施方式中,该系列引物包含n组引物,具有含有不同组引物的不同划分产物,不同的引物组具有2-20种不同的独特引物且不同的引物划分产物之间具有不超过一种共有引物,其中n是1000-300000。在一些实施方式中,n组中每一组的划分产物含有一种或多种分光光度物质,使得n组中的每一组都可通过分光光度特征区分。在一些实施方式中,该组中的至少大部分引物出现在两个不同的引物组中。

[0012] 在一些实施方式中,引物划分产物中任两种引物的序列没有超过两个核苷酸的重叠。

[0013] 在一些实施方式中,这些系列中的一些引物具有6-18个(例如6、7、8、9、10、11、12、13、14、15、16、17或18个)指定核苷酸。

[0014] 在一些实施方式中,这些指定核苷酸是连续的。在一些实施方式中,这些指定核苷酸至少一些是不连续的,使得指定核苷酸中至少两个被至少一个简并核苷酸位置或通用核苷酸隔开。

[0015] 在一些实施方式中,这些系列具有400-16000种不同序列的引物。

[0016] 在一些实施方式中,该靶核酸的长度为50-1000个核苷酸。

[0017] 在一些实施方式中,该靶核酸包含3' 茎环。

[0018] 在一些实施方式中,这些划分产物是液滴。在一些实施方式中,这些液滴包含在乳液中。

[0019] 还提供了至少n种不同引物划分产物的文库,这些引物划分产物包含n组引物,不同划分产物含有不同引物组,各组引物具有2-20种不同的独特引物且不同的引物划分产物之间具有不超过一种共有引物。在一些实施方式中,n是1000-300000。在一些实施方式中,不同的引物具有至少4个指定核苷酸,引物的其他核苷酸位置(如果有)是简并位置或通用核苷酸。在一些实施方式中,引物划分产物包含n组引物,不同划分产物含有不同引物组,各引物组具有8-12种不同的独特引物且不同的引物划分产物之间具有不超过一种共有引物。在一些实施方式中,n组中每一组的划分产物含有一种或多种分光光度物质,使得n组中的每一组都可通过分光光度特征区分。

[0020] 在一些实施方式中,该组中的至少大部分引物出现在两个不同的引物组中。

[0021] 在一些实施方式中,引物划分产物中任两种引物的序列没有超过一个、两个或三个核苷酸的重叠。

[0022] 在一些实施方式中,一些(或大部分)引物具有6-18个指定核苷酸。

[0023] 在一些实施方式中,一些、大部分或全部指定核苷酸是连续的。

[0024] 在一些实施方式中,这些指定核苷酸至少一些是不连续的,使得指定核苷酸中至少两个被至少一个简并核苷酸位置或通用核苷酸隔开。

[0025] 在一些实施方式中,文库包含400-16000种不同序列的引物。

[0026] 在一些实施方式中,这些划分产物是液滴。在一些实施方式中,这些液滴包含在乳液中。

[0027] 还提供了包含上述或本文其它地方所述的文库的容器。在一些实施方式中,该容器包含一个或多个与文库流体连通的微流体通道。

[0028] 还提供了用于生成靶核酸的核苷酸序列的系统。在一些实施方式中,该系统包含第一容器、第一微流体通道和检测器,所述第一容器包含多个引物液滴中的一系列引物,其中至少大部分引物液滴包含两种或更多种不同引物,所述引物具有至少4个指定核苷酸,引物的其他核苷酸位置(如果有)是简并位置或通用核苷酸,所述第一微流体通道提供第一容器和检测器之间的流体连通。

[0029] 在一些实施方式中,该系统还包含含有用于扩增靶核酸的引物对的第二容器;提供第二容器与第一微流体通道之间流体连通的第二微流体通道;以及与第二微流体通道流体连通的样品核酸容器。

[0030] 在一些实施方式中,该系统还包含第一液滴注射器和第二液滴注射器,所述第一液滴注射器构设成将核酸从样品核酸容器中注入包含来自第二容器的引物对且位于第二微流体通道中的液滴内以形成混合物液滴,所述第二液滴注射器构设成将混合物液滴分成的各份注入向下游通过第一微流体通道的引物液滴中。

[0031] 在一些实施方式中,这些液滴包含在乳液中。

[0032] 在一些实施方式中,该第二微流体通道的一部分是蜿蜒的。

[0033] 在一些实施方式中,该第一容器包含至少100、500、1000或10000个所述引物液滴。

[0034] 在一些实施方式中,该系统还包含一个或多个泵,用于推动液滴通过一个或多个微流体通道。

[0035] 在一些实施方式中,该组中的至少大部分引物出现在两个不同的引物组中。在一些实施方式中,引物划分产物中仍两种引物的序列没有超过一个、两个或三个核苷酸的重叠。

[0036] 在一些实施方式中,这些系列中的一些、大部分或全部引物具有6-18个指定核苷酸。

[0037] 在一些实施方式中,这些指定核苷酸是连续的。

[0038] 在一些实施方式中,这些指定核苷酸至少一些是不连续的,使得指定核苷酸中至少两个被至少一个简并核苷酸位置或通用核苷酸隔开。

[0039] 在一些实施方式中,这些系列具有400-16000种不同序列的引物。

[0040] 还提供了基于引物杂交确定靶核酸的核苷酸序列的方法。在一些实施方式中,该方法包括:从涉及多种引物与靶核酸杂交的实验中接收数据,该数据包括阳性杂交和阴性引物杂交结果;基于阳性杂交引物组装支架核苷酸序列,该支架包含基于阳性杂交引物的明确的核苷酸;进行以下两个步骤之一或全部两个步骤:(a) 通过移除或改变支架序列中与阴性引物杂交结果不一致的一个或多个核苷酸来校准支架序列;(b) 确定基于阳性杂交引物而言模糊的支架位置的模糊选项;并基于阴性杂交引物辨析至少一个所述模糊选项,从而基于引物杂交确定靶核酸的核苷酸序列。

[0041] 还提供了细分代表多个反应划分产物中杂交的恒定数据信号流的方法,所述信号流生成自使靶核酸划分产物的各部分与引物划分产物合并。在一些实施方式中,该方法包括获得恒定的数据流;确定跨固定长度滑动窗的时间的函数形式的反应划分产物中自相矛盾的比率;监测该比率的最小值;并使用该最小值作为边界细分数据,边界勾画了来自靶核酸划分产物的信号端。

[0042] 还提供了基于引物杂交确定靶核酸的核苷酸序列的方法。在一些实施方式中,该

方法包括通过计算机系统从涉及多种引物与靶核酸杂交的实验中接收数据,该数据包括阳性杂交和阴性引物杂交结果;通过计算机系统分析数据以鉴定与靶核酸阳性杂交的第一引物集合和与靶核酸杂交的第二引物集合;通过计算机系统基于第一引物集合组装支架核苷酸序列,该支架核苷酸序列包含其中具体核苷酸已确定的明确位置和其中具体核苷酸未确定的模糊位置;通过计算机系统基于第二引物集合辨析至少一个所述模糊位置。

[0043] 在一些实施方式中,辨析第一模糊位置包括:当以下情形时确定第一集合的第一引物不与第一模糊位置对齐,即当与对齐一致的是对齐第二集合的第二引物也与第一模糊位置对齐。

[0044] 在一些实施方式中,辨析第一模糊位置包括:基于第二引物集合通过排除明确引物的比对位置鉴定第一集合的一种或多种明确引物;使用第一集合的一种或多种明确引物来确定一个或多个模糊位置处的一个或多个核苷酸,从而获得更新的支架核苷酸序列。

[0045] 在一些实施方式中,该方法还包括:鉴定模糊引物第一子集,该第一子集是第一引物集合的子集,模糊引物与更新的支架核苷酸序列中的多个位置对齐;基于第二引物集合排除第一子集的第一模糊引物的对齐位置,从而使第一模糊引物成为第一明确引物;以及使用第一明确引物辨析更新的支架核苷酸序列中的第二模糊位置。

[0046] 在一些实施方式中,组装支架核苷酸序列还基于第二引物集合。在一些实施方式中,组装支架核苷酸序列可包括当以下情形时确定第一集合的第一引物不与第一模糊位置对齐,即当与对齐一致的是对齐第二集合的第二引物也与第一模糊位置对齐。在一些实施方式中,该方法还包括:基于确定第一集合的第一引物不与支架核苷酸序列的第一位置对齐来将第一位置鉴定为模糊位置。在一些实施方式中,第一引物集合包括在核苷酸之间有缺口的引物。

[0047] 在一些实施方式中,该方法还包括通过比较支架与参考序列来辨析支架中的模糊选项。

[0048] 在一些实施方式中,来自涉及多种引物与靶核酸杂交的实验的数据代表的不同引物的杂交,所述引物具有至少4个指定核苷酸,引物的其他核苷酸位置(如果有)是简并位置或通用核苷酸。

[0049] 在一些实施方式中,来自涉及多种引物与靶核酸杂交的实验的数据代表引物划分产物内引物的杂交,这些引物划分产物包含n组引物,不同划分产物含有不同引物组,各引物集合具有8-12种不同的独特引物且不同的引物划分产物之间具有不超过一种共有引物。

[0050] 在一些实施方式中,所述数据还包含来自n组中每一组的划分产物的数据,所述划分产物含有一种或多种分光光度物质,使得n组中的每一组都可通过分光光度特征区分。

[0051] 在一些实施方式中,该集合中的至少大部分引物出现在两个不同的引物组中。在一些实施方式中,引物划分产物中任两种引物的序列没有超过两个核苷酸的重叠。在一些实施方式中,一些(或大部分)引物具有6-18个指定核苷酸。在一些实施方式中,这些指定核苷酸是连续的。

[0052] 在一些实施方式中,这些指定核苷酸至少一些(或大部分)是不连续的,使得指定核苷酸中至少两个被至少一个简并核苷酸位置或通用核苷酸隔开。

[0053] 在一些实施方式中,多种序列具有400-16000种不同序列的引物。

[0054] 还提供了鉴定引物/靶核酸划分产物生成自相同靶核酸划分产物混合物液滴划分

产物的方法。在一些实施方式中,该方法包括:接收一段时间上检测器获得的数据信号,该数据信号包括来自生成自多个混合物液滴的多个反应划分产物的信号,各混合物液滴对应于一个划分产物集合且包括至少一个靶核酸的拷贝,各划分产物包含一种或多种引物,该数据信号包含关于划分产物中至少一种引物是否与靶核酸杂交的数据;基于各划分产物的各信号鉴定各划分产物的杂交状态,划分产物的杂交状态显示划分产物中的引物是否与靶核酸杂交,特定划分产物的信号对应于特定时间;对于时间段中多个特定时间中的每一个:对于围绕特定时间的窗口:计算具有矛盾杂交状态且包含相同引物的划分产物量,从而获得时间函数;鉴定该时间函数中的极值;以及确定时间函数中对应于相同混合物液滴的相应极值之间出现的连续划分产物集合。

[0055] 在一些实施方式中,这些极值是峰之间的最小值。在一些实施方式中,具有矛盾杂交状态的划分产物的量对应于具有矛盾杂交状态的划分产物中引物的量。

[0056] 在一些实施方式中,用时间窗口中不同划分产物的数目对划分产物的量进行标定,当两个划分产物包含至少一种不同的引物时这两个划分产物是不同的。

[0057] 在一些实施方式中,该时间窗口被描述成划分产物数目,且该时间窗口围绕特定时刻的划分产物。在一些实施方式中,基于创建自混合物液滴的划分产物数目来选择时间窗口中的划分产物数目。

[0058] 在一些实施方式中,该方法还包括:鉴定第一连续划分产物集合对应于包含第一靶核酸的第一混合物液滴;基于第一集合的信号确定第一集合中的各引物是否与第一靶核酸杂交;使用第一集合中的引物和第一集合中引物的杂交状态来确定第一靶核酸的核苷酸序列。

[0059] 还提供了一种鉴定混合物液滴包含靶核酸的方法。在一些实施方式中,该方法包括:接收多个划分产物中每一个的杂交状态,各划分产物包含一种或多种引物,多个混合物液滴各自被划分且各自包含一种或多种靶核酸的拷贝,划分产物的杂交状态显示该划分产物中至少一种引物是否与靶核酸杂交;鉴定第一划分产物集合对应于第一混合物液滴;对第一划分产物集合建立第一位向量,位向量中的各值对应于第一集合的各划分产物的杂交状态;比较第一位向量与多个参考位向量以获得相对于多个参考位向量中每一个的差值,各参考位向量对应于不同的参考核酸且包含第一集合中各划分产物的关于参考核酸的杂交状态的值;以及基于这些差值鉴定第一混合物液滴中的第一靶核酸。

[0060] 在一个实施方式中,比较第一位向量与第一参考位向量包括:对具有不同于第一参考位向量的杂交状态的第一集合的各划分产物增加一个筹数。在一些实施方式中,具有最低筹数的参考核酸被选为第一靶核酸。在一些实施方式中,具有最低筹数的参考核酸被选为第一靶核酸,前提是最低的筹数比下一最高筹数少至少预定量。

[0061] 在一些实施方式中,各参考核酸具有已知的序列。该方法还包括:对各参考核酸,通过确定第一集合中各划分产物的预期杂交状态建立参考位向量,基于划分产物中的引物确定划分产物的预期杂交状态。

[0062] 在一些实施方式中,这些参考核酸对应于其他混合物液滴中的核酸。在一些实施方式中,该方法还包括:对多个混合物液滴的位向量进行聚类;以及基于第一簇的位向量的类似性鉴定包含第一靶核酸的第一混合物液滴簇。

[0063] 在一些实施方式中,该方法还包括:使用第一划分产物集合中的引物确定第一靶

核酸的核苷酸序列。

[0064] 在一些实施方式中,各划分产物包含多种引物,该方法还包括:鉴定第一划分产物集合中与第一靶核酸阳性杂交的引物;组装鉴定的引物以确定第一靶核酸的核苷酸序列。

[0065] 还提供了确定混合物液滴的引物的杂交状态的方法。在一些实施方式中,该方法包括:接收一段时间上检测器获得的数据信号,该数据信号包括来自多个混合物液滴的多个划分产物的信号,各混合物液滴对应于划分产物集合且包含至少一种靶核酸的拷贝,各划分产物包含多种引物;鉴定对应于第一混合物液滴的第一划分产物集合;确定第一划分产物集合对应于第一靶核酸;基于各划分产物的各信号确定第一集合的各划分产物的杂交状态,划分产物的杂交状态显示是否该划分产物中至少一种引物与第一靶核酸杂交;鉴定第一集合的各划分产物中的引物,第一集合的多个划分产物包含第一引物;确定第一引物是否位于第一集合任何具有阴性杂交状态的划分产物中;以及基于第一引物是否位于第一集合任何具有阴性杂交状态的划分产物中来确定第一引物是否与第一靶核酸杂交。

[0066] 在一些实施方式中,当第一引物处于具有阴性杂交状态的一个或多个划分产物中时,确定该第一引物未与第一靶核酸杂交。

[0067] 在一些实施方式中,当相对于具有阳性杂交状态的划分产物,第一引物处于具有阴性杂交状态的至少特定比例的划分产物中时,确定该第一引物未与第一靶核酸杂交。

[0068] 在一些实施方式中,该方法还包括:确定多个其他划分产物组对应于第一靶核酸;鉴定其他划分产物组中每一组中的引物;确定第一引物是否位于其他组任何具有阴性杂交状态的划分产物中;并基于第一引物是否位于其他组任何具有阴性杂交状态的划分产物中来确定第一引物是否与第一靶核酸杂交。

[0069] 在一些实施方式中,该方法还包括:确定与第一靶核酸杂交的一群引物;以及使用该群引物的序列组装第一靶核酸的核苷酸序列。

[0070] 在一些实施方式中,组装第一靶核酸的核苷酸序列包括使用第一靶核酸的参考序列。

[0071] 本文他处描述本发明的其他方面。

[0072] 定义

[0073] 术语“核酸扩增”或“扩增反应”指用于倍增核酸靶序列拷贝的任何体外方法。这类方法包括但不限于聚合酶链反应(PCR)、DNA连接酶链反应(参见美国专利号4,683,195和4,683,202;《PCR方案:方法和应用指南》(Innis等编,1990))、(LCR)、QBeta RNA复制酶、和基于RNA转录(如TAS和3SR)的扩增反应以及本领域技术人员已知的其它反应。

[0074] “扩增”指将溶液置于足以扩增多核苷酸的条件下的步骤。扩增反应的组分包括,例如,引物、多核苷酸模板、聚合酶、核苷酸等。术语“扩增”一般是指靶核酸的“指数型”增长。然而,本文所用的扩增也可指核酸的选择靶序列数量的线性增长,如由循环测序所得。

[0075] “聚合酶链反应”或“PCR”是指靶双链DNA的特定区段或子序列得以几何级数式扩增的一种方法。PCR是本领域技术人员所熟知的;参见例如,美国专利号4,683,195和4,683,202;和《PCR方案:方法和应用指南》,Innis等编,1990。示例性PCR反应条件一般包括两步或三步式循环。两步骤循环具有变性步骤,之后是杂交/延长步骤。三步骤循环包括变性步骤,之后是杂交步骤,之后是独立的延长步骤。PCR可以终点PCR(即仅在终点处监测)或定量PCR(“实时”监测)的方式进行。

[0076] “引物”指与靶核酸上的序列杂交并且用作核酸合成的起始点的多核苷酸序列。引物可以是各种长度的但通常长度小于50个核苷酸,例如长度为5-30个核苷酸。可基于本领域技术人员已知的原理设计用于PCR的引物的长度和序列,参见例如Innis等(同上)。

[0077] “模板”指包含待扩增的多核苷酸、其侧或为一对引物杂交位点的多核苷酸序列。因此,“靶模板”包含侧接“正向”引物和“反向”引物的杂交位点的靶多核苷酸序列。

[0078] 本文所用的“核酸”表示DNA、RNA、单链、双链、或更高度聚集的杂交基序及其任意化学修饰。修饰包括但不限于,提供整合入其它电荷、极化性、氢键、静电相互作用、与核酸配体碱基或核酸配体整体的连接点和作用点的化学基团的那些修饰。这类修饰包括但不限于,肽核酸(PNA)、磷酸二酯基团修饰(例如,硫代磷酸酯、甲基磷酸酯)、2'-位糖修饰、5-位嘧啶修饰、8-位嘌呤修饰、环外胺处的修饰、4-硫尿核苷的取代、5-溴或5-碘-尿嘧啶的取代、骨架修饰、甲基化、不常见的碱基配对组合如异碱基(isobases)、异胞苷和异胍(isoguanidine)等。核酸也可包含非天然碱基,如硝基咪唑。修饰还可包括3'和5'修饰,包括但不限于用荧光团(例如,量子点)或其他部分加帽。

[0079] 术语“多肽”、“肽”和“蛋白质”在本文中互换使用,指氨基酸残基的聚合物。该术语可用于表示其中一个或多个氨基酸残基是相应天然产生氨基酸的人造化学模拟物的氨基酸聚合物,以及天然产生的氨基酸聚合物和非天然产生的氨基酸聚合物。

[0080] “聚合酶”是指进行模板引导的多核苷酸(例如,DNA和/或RNA)合成的酶。该术语同时包括全长多肽和具有聚合酶活性的结构域。DNA聚合酶是本领域技术人员熟知的,包括但不限于从激烈火球菌(*Pyrococcus furiosus*)、滨海嗜热球菌(*Thermococcus litoralis*)和海栖热袍菌(*Thermotoga maritime*)分离或衍生的DNA聚合酶或其修饰版本。市售的聚合酶的其他示例包括,但不限于:克列诺片段(新英格兰生物实验室公司(New England Biolabs® Inc.))、Taq DNA聚合酶(凯杰公司(QIAGEN))、9°N™ DNA聚合酶(新英格兰生物实验室公司)、Deep Vent™ DNA聚合酶(新英格兰生物实验室公司)、Manta DNA聚合酶(Enzymatics公司)、Bst DNA聚合酶(新英格兰生物实验室公司)、和phi29 DNA聚合酶(新英格兰生物实验室公司)。聚合酶包括DNA-依赖聚合酶和RNA-依赖聚合酶,如逆转录酶。已知至少5个DNA-依赖DNA聚合酶家族,虽然大多数落入A、B和C家族。各家族之间很少或没有序列相似性。大多数A家族聚合酶是可含有多重酶促功能(包括聚合酶、3'到5'外切核酸酶活性和5'到3'外切核酸酶活性)的单链蛋白质。B家族聚合酶通常有具有聚合酶和3'到5'外切核酸酶活性的单个催化结构域,以及辅助因子。C家族聚合酶通常是具有聚合和3'到5'外切核酸酶活性的多亚基蛋白质。在大肠杆菌中,已经发现了3种类型的DNA聚合酶,DNA聚合酶I(A家族)、DNA聚合酶II(B家族)和DNA聚合酶III(C家族)。在真核细胞中,核复制中涉及3种不同的B家族聚合酶,DNA聚合酶 α 、 δ 和 ϵ ,并且A家族聚合酶,聚合酶 γ 用于线粒体DNA复制。其它类型的DNA聚合酶包括噬菌体聚合酶。相似地,RNA聚合酶通常包括真核RNA聚合酶I、II和III,和细菌RNA聚合酶以及噬菌体和病毒聚合酶。RNA聚合酶可以是DNA-依赖的和RNA-依赖的。

[0081] 术语“标记物”、“可检测标记物”、“可检测部分”和类似术语指可通过光谱、光化学、生物化学、免疫化学、化学或其它物理手段检测的组合物。例如,可用的标记物包括荧光染料(荧光团)、发光剂、高电子密度试剂、酶(例如,ELISA中常用)、生物素、地高辛、³²P和其它同位素、半抗原、以及可被检测的蛋白质(例如通过将放射性标记物整合至肽中或或用于

检测与肽特异性反应的抗体)。该术语包括单一标记试剂的组合,例如,提供独特可检测特征(例如,特定波长或波长组合下)的荧光团的组合。可以采用本领域已知的用于将标记物偶联到需要的试剂的任何方法,例如,使用如下文献中所述的方法:Hermanson,《生物偶联技术》(Bioconjugate Techniques)1996,圣迭戈的学术出版社有限公司(Academic Press, Inc.)。

[0082] “茎环”,也称为“发夹”或“发夹环”,是指当线性链的第一部分中的互补碱基与相同链的第二部分中的碱基杂交时由单链寡核苷酸形成的二级结构。序列的第二部分中的序列与第一部分序列反向互补,从而能够杂交。

[0083] 本文所用术语“划分”或“划分的”指将样品分为多个部分或多个“划分产物”。划分产物可以是固体或流体。在一些实施方式中,划分产物是固体划分产物,例如微通道。在一些实施方式中,划分产物是流体划分产物,例如液滴。在一些实施方式中,液体划分产物(如液滴)是不互溶的液体(如水和油)的混合物。在一些实施方式中,流体划分产物(如液滴)是水性液滴,其被不互溶的运载体流体(如油)包围。

[0084] 附图简要说明

[0085] 图1:微流体 workflow。

[0086] 图1是用于测序的微流体装置的高水平示意图。存在两个储器,其在进行测序实验前预加载有乳化试剂。储器A含有进行PCR反应所需的乳化试剂,包括至少一对PCR引物。储器B含有测序试验所需的乳化试剂,包括至少一种测序引物寡聚体。在注射点C处将小部分基因组DNA样品注入各PCR划分产物中。部分D中的蜿蜒通道流过不同的热区中至少两个且作为在线的热循环仪。各PCR划分产物都结束于第二注射点E,其中含有的扩增的材料被注入一系列较小的测序引物划分产物中。测序引物划分产物从注射点E的下游流至点F处的光学检测器。F处的光学检测器读取来自各引物划分产物的荧光信号。

[0087] 图2:软件结构

[0088] 图2显示哪种软件模块在过程的哪个阶段被执行且在哪个时间粒度处进行和在何种输入物上运行。例如,信号加工模块在每个引物划分产物被光学检测器测量后执行一次。

[0089] 图3A:四维染料概况簇

[0090] 该图中的各点代表单个引物划分产物测量。图中的各簇是来自对应于同一划分产物的引物划分产物的染料强度读数的集合。在该实验中,各划分产物对应于单个引物寡聚体序列。这些簇根据引物种类着色。

[0091] 图3B:四维染料概况簇-根据试验强度着色

[0092] 图3B显示与图3A相同的数据,但根据各划分产物测得的试验染料强度着色。较暗的簇表示较亮的试验荧光。

[0093] 图4A:根据序列位置的试验强度簇

[0094] 图4A是散点图,显示使用本发明对已知含有突变的样品中BRAF基因的外显子15测序的测序试验的试验度数据。图中的各点表示光学检测器读取的单个引物划分产物的试验荧光强度。各点根据引物划分产物的类别着色。在该实验中,划分产物特性和其中含有的单个引物是相同的。该散点图中的各点以含有特定引物的划分产物中观察到的平均试验强度的函数形式自然地聚集在y轴中。相对于靶标的最终一致性序列组装,通过其相应引物的位置,标记为组A的点簇在x轴周围是故意有序的。组A含有对应于被称为对测序试验呈阳性的

引物划分产物的所有点簇。组B含有对应于被称为对测序试验呈阴性的引物划分产物的所有点簇。组B中的簇在X轴周围是随机排序的且仅代表来自实验的全部试验阴性液滴的随机样品。区域C代表我们的“无应答区 (no call zone)”，其描述了对引物划分产物没有试验应答的试验强度范围。

[0095] 图4B: 含有3个简并碱基的引物的试验强度

[0096] 图4B与图4A相同, 不同之处在于围绕所有点簇的圈, 其代表含有具有其制造过程中使用的至少3个简并碱基的引物寡聚体的引物划分产物。

[0097] 图5: BRAF靶标的参考辅助组装

[0098] 图5描述了针对代表BRAF基因的外显子15的扩增子通过组装软件进行的参考辅助的组装过程。野生型序列“D”被用作对比组装过程的参考。该图中, 仅描述了全部参考序列的子序列。生成阳性试验应答的引物组 (组“A”和“E”) 被用于相对于野生型初始化潜在突变的假设空间。阳性引物“G”被发现在1的编辑距离内在靠近组“E”的位置处与参考序列对齐, 而以较大的2的编辑距离在靠近组“A”的位置处对齐。然而, 通过软件发现组“A”的一致性序列生成的假设正确的可能性大得多。该软件联用阳性组“E”与组“H”的阴性结合引物以在参考物的区域中形成一致性序列, 该区域匹配野生型且与引物“G”假设的G/A错配相比正确的可能性高得多。对该靶标应答的最终一致性序列列于序列“I”且含有来自样品的正确应答的2-核苷酸取代突变, 如桑格测序方法所示。这显示参考辅助组装的基本步骤; 在一些编辑距离阈值内对各引物寻找针对参考序列的对齐, 使用与其他阳性和阴性引物的重叠关系来排列候选物对齐并对参考序列中的每一个碱基位置都如此操作。随后在参考序列的各位置处选择最大可能的假设。如果任何引物都被排除在赢得假设性公认序列的最终集合外, 则其可与其他引物比较以尝试根据参考物进行潜在插入物的无参考组装 (该微组装步骤在本文中未显示)。

[0099] 图6A: 多重化引物划分的实施例

[0100] 图6A显示多重化引物划分产物的两个实施例。两种划分产物 (A和B) 都确切地含有10种不同的引物。这些引物由其鉴定物编码; 未以任何方式描述其寡聚序列。鉴定为编号1的引物出现在全部两个划分产物中且是两个划分产物共有的唯一引物。

[0101] 图6B: 多重划分产物的正面和负面结果

[0102] 图6B显示多重划分产物可生成阳性或阴性试验应答的方式。图6B添加了多种靶分子 (或扩增子) 的细节, 其假定先前已注入划分产物中。例如, 划分产物A显示扩增子分子的三种情况, 其结合引物编号1的寡核苷酸的三种情况。划分产物B显示三种不同的引物, 其各自结合三分之一的可用靶分子。划分产物A和B都生成单试验阳性应答。划分产物C显示划分产物可生成试验阴性应答的唯一方法: 当且仅当其中含有的所有引物都不杂交时。

[0103] 图7A: 5重实验的结果

[0104] 试验结果见图7A。结果显示为与错配试验引物信号的标准偏差距离 ($\frac{[{\text{荧光}}] - [\text{错配试验引物的平均荧光}]}{[\text{标准偏差错配试验引物}]}$)。合并试验引物不干扰试验性能且结果在整个引物试验组中保持恒定。含有匹配试验引物的组在所有组合中都生成恒定的高信号且含有错配试验引物的组生成恒定的低信号, 与合并的组中包含的试验引物数目无关。

[0105] 图7B: 10重实验的结果

[0106] 试验结果见图7B。如同合并试验引物的5试验引物组实验那样, 不干扰试验性能且

结果在整个引物试验组中保持恒定,其中含有匹配试验引物的组在所有组合中都生成恒定的高信号且含有错配试验引物的组生成恒定的低信号,与合并的组中包含的试验引物数目无关。

[0107] 图8:引物寡核苷酸中的简并碱基

[0108] 图8显示向引物中加入简并碱基的一些选择。

[0109] 图9:基于矛盾率的时间分割

[0110] 图9显示在固定时间窗口内观察到超过一次的引物划分产物在试验状态方面自身矛盾的比率。与矛盾率中局部最小值交叉的纵条标记代表不同的PCR划分产物的引物划分产物测量值之间的时间边界。

[0111] 图10:参考序列的划分产物向量

[0112] 图10显示用于预计算单个参考序列的划分产物向量的软件机制。划分产物向量储存相对于相应野生型参考序列的引物划分产物文库中各划分产物的预期试验结果。A描述参考序列,其对应于PCR划分产物文库中PCR引物对之一所捕获的一些感兴趣区域的野生型。划分产物C、D和E各自描述了测序文库中实际多重引物划分产物的子集。C、D和E各自用“+”或“-”标记在右上角以表示相对于参考序列A的该划分产物的预期试验结果。表B描述参考序列1000的划分产物向量。表1010中的左栏含有序列文库中各划分产物的标识符。表1010中的右栏显示位的值根据对于相应划分产物的预期试验结果设定。划分产物1020、1030和1040具有箭头,指向对应于其标识符的表1010中的各行。针对该野生型进行测试时,划分产物1020和1040预期生成阳性试验应答。划分产物1030预期是试验阴性的。在阳性划分产物1020和1040中,作为阳性信号来源的引物高亮显示为蓝色并斜体化。具有匹配来自划分产物1020和1040的阳性引物序列的参考序列1000的子序列也高亮显示为蓝色并斜体化。

[0113] 图11:计算映射评分的过程

[0114] 图11显示用于计算引物划分产物集合的映射评分的软件方法。标记为组“1110”的划分产物集合被假定是先前刚与时间片段过程相关的组。组1110中的各圆圈含有“+”或“-”以代表该划分产物试验应答。表“1120”是长度等于测序文库中划分产物总数的位向量。“1110”中的划分产物排列为对齐划分产物性质顺序。表1140显示划分产物标识符。表1130代表实验期间进行的测序实验的所有参考序列的划分产物向量的集合。柱“1132”显示“1130”中各参考向量与“1120”中集合向量之间的汉明距离(Hamming distance)。应注意“1120”中的一些柱是空白的。这表明一个划分产物碰巧没有在集合内被观察到。集合向量“1120”和1130中所有参考向量内的位位置在计算汉明距离前被掩蔽。矩形“1134”选择最匹配的参考序列,因为其具有最小的汉明距离。

[0115] 图12:集合向量的聚类

[0116] 图12显示在参考序列不可用或未发现充分强匹配的情况中对集合向量进行聚类的软件过程。“1250”标记一批未聚类的集合向量。1210、1220、1230和1240标记在一对集合向量之间直接做出的汉明距离计算。对“1250”中的所有集合向量对做出该距离计算。这些集合随后凝聚地聚类并产生“1270”中的聚类的向量组。“1260”仅标记在两个向量中都观察到划分产物的划分产物向量中的位,且这些位是距离计算中使用的仅有的位。

[0117] 图13:集合内的否定

[0118] 图13显示利用来自集合的试验阴性划分产物数据“1310”来否定来自同一集合的试验阳性划分产物数据“1300”所得人工假阳性引物的软件过程。该软件预计算所有划分产物对之间的组交集。其用于否定同一组内阳性和阴性划分产物中都发现的任何引物。被用于否定其阳性对应物的引物的阴性实施方式斜体化。通过否定过程的阳性划分产物中的引物未从图中划掉。

[0119] 图14:集合之间的否定

[0120] 图14显示由软件进行的另一否定步骤,其中,来自一个集合的阴性划分产物被用于否定不同但相关集合中的假阳性引物。该过程与图13相同,不同之处在于必须首先通过将其聚类在一起或直接或间接通过映射到相同参考序列来建立集合之间的相关性。在该图中,如果否定同一集合内的阳性结果则阴性引物斜体化,或如果其否定其他集合内的阳性结果则阴性引物带下划线。在该图中,两个集合“1420”和“1440”通过映射到相同的参考序列“1400”而相关。

[0121] 图15:显示具有缺口的引物

[0122] 图15显示具有缺口的探针的若干示例。部分A列举三种不同的缺口方案,其跨长度为6个碱基对至9个碱基对的序列。部分A中的各缺口方案都确切使用6个指定碱基,导致4096种不同指定六聚体模式的总可寻址复杂性。部分B显示使用两种互补缺口方案。具体而言,其允许仅具有6个指定碱基的单个探针跨越该情况下缺口寡聚体本身的合并长度长达9个核苷酸的串联重复和同源聚合物区域。

[0123] 图16:显示组装过程

[0124] 图16显示从头组装的过程。在该图中,假定了含有遵循缺口方案混合物的引物序列的引物划分产物文库的总体实施方式。在该过程的若干阶段利用阴性证据来删除重叠图中的假分支,消除组装支架内的假补充物以及消除假阳性突变假设。

[0125] 图17:使用阳性和阴性证据显示基本微组装。

[0126] (1710) 阳性结合的探针为蓝色。(1705) 假拟的C/T多态性被对试验测试为阳性的若干重叠六聚体探针所证实。(1712) 证实C/T多态性的探针子集也支持1712插入多态性的替代性假设。该1712插入多态性与其他阳性探针矛盾,这些其他阳性探针与参考序列的相同区域良好地对齐。(1722) 该C插入假设还与若干探针矛盾,这些探针在试验中测试为阴性,且似乎精确匹配该假设的参考内容中的假拟C插入。在这一方面,阴性证据与假阳性C插入假设相对立。这以简化的方式表明阴性证据的一般效用。此外,在(1724)中,我们看到阴性证据与参考序列本身相对立,其进一步证实了C/T多态性假设。这表明阴性证据通过否定或权衡野生型或“参考”等位基因来促进真阳性假设的一般能力。

[0127] 图18是根据本发明的实施方式用于确定靶核酸的核苷酸序列的方法1800的流程图。

[0128] 图19是鉴定引物/靶核酸划分产物生成自相同靶核酸划分产物混合物液滴的划分产物方法1900的流程图。

[0129] 图20是鉴定混合物液滴包含靶核酸的方法的流程图。

[0130] 图21是确定混合物液滴的引物的杂交状态的方法的流程图。

[0131] 图22是根据本发明的实施方式基于引物杂交确定靶核酸的核苷酸序列的方法2200的流程图。

[0132] 图23显示一种计算机系统。

[0133] 发明详述

[0134] 引言

[0135] 本发明提供了基于引物与靶核酸的杂交确定靶核酸的核苷酸序列的方法,其中,杂交显示存在与引物互补的核苷酸序列且缺少引物杂交显示不存在引物的精确互补。如本发明所述,该方法在划分产物中进行(例如在乳液中的液滴中),且具体包括生成(i)包含靶核酸的划分产物(“靶划分产物”)和(ii)包含一种或多种引物的划分产物(“引物划分产物”)并以一比一的基础合并靶划分产物的各部分与引物划分产物以在合并的划分产物中生成杂交反应(“反应划分产物”),可随后测试其杂交。

[0136] 在一些方面中,该方法包括合并并在相同的引物划分产物中具有不同序列的多个引物,所述引物划分产物随后与一部分靶划分产物合并。随后确定杂交反应中是否存在引物中任一种的杂交。合并的引物划分产物可经设计,从而可基于多个引物组的杂交对多种划分产物组合的杂交进行反卷积以生成靶标的逻辑序列。多个引物组的设计在下文中详细描述。

[0137] 还提供了本发明所述引物划分产物的文库以及用于进行这些方法和分析结果的系统。

[0138] 虽然在一些情况中靶核酸核苷酸序列完全是未知的,但在许多实施方式中,靶核酸序列的总体结构和序列是已知的,但精确的核苷酸序列是未知的。例如,特定的遗传生物标记物序列是已知的,但个体携带的确切的遗传变体是未知的。本发明所述方法可特别用于确定样品(例如来自个体的样品)中已知遗传序列的精确遗传变体。

[0139] 通过杂交测序概述

[0140] 本发明所述方法依赖于大量不同引物序列与靶核酸的杂交或缺乏杂交。通过检测重叠引物的杂交和类似但不同序列引物的杂交不存在,可预测靶序列的核苷酸序列。通过杂交进行测序的一个实施例示于图17,其显示预测的靶核酸(粗体)和一系列不同的变体引物。仅完全互补的引物与靶核酸结合而不完全互补的引物不结合。在图17中,已在试验中将预测的靶序列上方的引物序列鉴定为与靶核酸杂交而将靶序列下方的引物序列鉴定为不与靶核酸杂交。虽然图17中的靶核酸被描述为与引物取向相同的单链序列,应理解事实上这些引物将在靶标的另一链上与其互补序列杂交。

[0141] 划分产物中的引物可用于确定核酸的序列,参见PCT公开号W02012/078710。例如,该试验可包括将(例如该检测器可包含荧光标记物)整合至靶核酸并将该靶核酸添加至引物划分产物。此外,这些引物划分产物可包含聚合酶(例如Klenow片段(新英格兰生物实验室公司(New England Biolabs® Inc.))、Taq DNA聚合酶(凯杰公司(QIAGEN))、9°N™ DNA聚合酶(新英格兰生物实验室公司)、Deep Vent™ DNA聚合酶(新英格兰生物实验室公司)、Manta DNA聚合酶(酶学公司(Enzymatics®))、Bst DNA聚合酶(新英格兰生物实验室公司),和phi29 DNA聚合酶(新英格兰生物实验室公司)、逆转录酶(新英格兰生物实验室公司)或另一种市售可得的聚合酶)、dNTP、核酸抑制剂(例如该抑制剂可包含接合匹配荧光标记的靶核酸的核酸序列的淬灭剂)、染料标记物和聚合酶反应所需的盐缓冲剂(所需的盐缓冲剂是聚合酶特异性的且包含诸如KCL、Tris-HCl和MgCl的组分)。

[0142] 使用的引物组合可设计为使得其在单个试验中协同作用。例如,引物的长度可变

化,以使其在解链温度上匹配。例如,以下引物:GACTGTCA、AGGCGTT、ATTGAACTT、ATTTAACTT在长度上不同且具有类似的解链温度,使得其都可在相同温度和试验条件下与核酸靶标杂交。在其他实施方式中,使用通用碱基向引物序列中导入缺口。在其他实施方式中,通过混合的方法导入缺口,包括但不限于简并碱基和通用碱基。

[0143] 用于确定靶核酸序列的不同引物序列的精确数目将是靶序列长度和复杂性,以及引物中指定(如下文所述)核苷酸和其他位置的函数。一旦引物的杂交结果已知,可使用反卷积方法和软件来基于存在或缺失何种序列的知识从逻辑上衍生靶核酸序列,该知识衍生自杂交。

[0144] 已对这些探针组的优化选择、其合并的序列复杂性和优化的表面条件给予了很大关注以最大化杂交信号以及最大化靶分子的可辨析尺寸且该信息可应用于本发明。参见例如R.Drmanac等,Science 260:1649-1652(1993);R.Drmanac等,J.Biomol.Struct.Dyn.5:1085(1991);PEVZNER等,J.BIOMOLECULARSTRUCTURE&DYNAMICS 9(2):399-410(1991);B.HUDSON:"An Experimental Study of SBH with Gapped Probes(使用有缺口探针的SBH的实验性研究)"TECHNICAL REPORTCS-99-07,布朗大学计算机科学系,1999年4月;PCT专利公开号WO 2000/022171。

[0145] 通常,当“引物”据说存在于划分产物或反应混合物中时,其表明该引物的多个拷贝被包括在该划分产物中,除非本文中明确说明了拷贝数目。例如,如果两种引物被描述为存在于划分产物中,实际上划分产物中包含两种引物中每一种的大量拷贝。

[0146] 引物组的生成和应用

[0147] 各引物杂交反应都在一个划分产物内确定。在一些实施方式中,各引物划分产物中包含多种不同(即具有不同序列)的引物。杂交确定为是否存在杂交。

[0148] 在一些实施方式中,多种不同的引物(“组”)置于各划分产物内。可选择引物组,使得这些组是不随机的,但各组含有已知数目和特性的引物。引物组的数目通常是至少50、100、200、500、1000、10000、50000、100000、200000或更多。例如,可提供1024种不同的引物组,各组含有10种不同的引物。各组可,但非必须,包含相同数目的引物。在一些实施方式中,各引物出现在多个组中。其实施例描述于图2,其中显示两类多重乳液,各自含有10种不同探针。这些探针被编号,这只是为了将其在组内彼此区分并作为图中跨划分产物的鉴定物。

[0149] 向引物划分产物中加入的引物可具有任何核酸序列。例如,这些核酸序列可具有4-15个核苷酸。在另一个实施例中,在引物划分产物中将多种引物合并在一起。例如,引物组合可包含共有的核酸序列和可变的核酸序列。引物组合的一个示例是NACTTCA,其中N表示以下引物组合的由A、G、C或T组成的可变序列:AACTTCA、CACTTCA、TACTTCA和GACTTCA。其他引物组合可包含任何类型的多种可变序列,例如NNNATGCT、CTNGGN或GTCVTGC,其中V是A、C或G。在另一个实施例中,长度和/或序列不同的引物可合并在一起,例如但不限于:AACTTCAGG、GTCGC和GGTCACT。在另一个实施例中,引物组合可包含含有天然核苷酸和化学改性的核苷酸的引物。例如,引物可包含锁核酸、荧光标记物(如Cy5、Cy3、FAM、MAX、TAMRA等)、生物素、硝基吡啶、脱氧肌苷、脱氧尿苷或其他修饰。

[0150] 在一些实施方式中,任何两个引物组的交集都不含有超过一种引物。这种类型的关系描述于图2,其中划分产物A和B仅共有鉴定为编号“1”的探针。

[0151] 当针对靶核酸测试大量不同引物时,通常大部分(例如大于20%、大于30%、大于40%、大于50%、大于60%、大于70%、大于80%)将不与反应划分产物中存在的一种或多种靶核酸杂交。这至少部分是因为给定靶序列的复杂性比测试的一系列引物的复杂性小得多。因此,如果在单个划分产物中测试各引物,大多数划分产物将仅提供有限量的信息*(一种引物不杂交)。然而,通过在划分产物中合并多种引物并在划分产物中同时测试这些引物,可以对单个划分产物获得显著较多的信息(例如2、3、4、5、6、7、8、9种等不同引物不与靶标杂交)。

[0152] 在一些实施方式中,该系列引物包含指定核苷酸的引物序列的可能组合的至少40、50、60、70、80、90或100%。例如,存在4096种不同的六聚体组合,其中所有核苷酸都是指定的。

[0153] 划分产物中的引物数目可以是相同的(例如每个划分产物1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多种)或者引物数目可以在划分产物间变化。如下文所述,可通过在划分产物中纳入一种或多种特定的分光光度特征来确定划分产物中一种或多种特定引物的特性,其中,各引物组由不同的性质代表。

[0154] 在一些实施方式中,组中所有的引物都具有相同数目的指定核苷酸(例如各引物都具有六个指定核苷酸)。更频繁地,需要引物组包含不同数目指定核苷酸的引物(例如,一些引物具有5个指定核苷酸,一些引物具有6个指定核苷酸,一些引物具有7个指定核苷酸)。

[0155] 杂交确定为划分产物中是否存在杂交。因此,在一个实施例中,当划分产物含有10种引物时,如果例如引物划分产物中十个引物中的一个或九个与靶核酸杂交,则该划分产物被评价为杂交,而如果十种引物中无一杂交,该划分产物被评价为不杂交。可测量的聚集体结果类型的示例描述于图3。在情况A中,液滴内仅一种引物结合靶分子,因此划分产物整体显示为阳性。在情况B中,引物结合靶标,且同样乳液整体显示为阳性。仅在情况C中,无一引物结合靶标,划分产物显示为阴性状态。可以使用阴性结果(缺少杂交),例如在情况C中,以对抗例如情况A和B中出现的假阳性。此外,可基于具有已知的引物重叠的两个划分产物之间不同的重叠来验证阳性探针,例如如图6A所示。阴性结果也有助于排除其它逻辑上可用的序列。例如,在图17的1722部分中,阴性引物(如CACTTT)显示靶序列的所示“-”中的C的替代事实上不存在(否则CATTT也会杂交)。

[0156] 在一些实施方式中,各划分产物中使用的一种或多种引物是“缺口”引物,其中引物的一些位置具有指定核苷酸且其他位置(即缺口)是简并的或允许在该位置与任何碱基杂交。因此,具有缺口的引物将包含指定核苷酸的区域(其在杂交条件下通过A-T或G-C配对等选择性碱基配对)和通用核苷酸的区域(其显示简并性(即在该位置处具有A、C、T和G的引物混合物),或在核苷酸之间基本没有选择性)。示例性通用核苷酸包括但不限于5-硝基肌苷、3-硝基吡咯和脱氧次黄苷。具有缺口的引物的设计描述于,例如,美国专利公开号2003/0064382。

[0157] 除上文描述的引物序列外,可向引物序列中加入其他核苷酸,通常在5'端,以调节引物的解链温度(T_m)。在一些实施方式中,使用的引物组将是 T_m “平衡的”,即通过在引物的5'端加入一个或多个核苷酸(例如简并、通用和指定)来上调组中 T_m 低于平均 T_m 的单个引物,使得所得引物组具有比初始引物组更窄的 T_m 范围。这允许在组中全部引物都用于试验的杂交条件下。在一些实施方式中,组中引物的 T_m 的不超过0%、1%、5%、10%、15%或20%

与该组引物的平均 T_m 相差超过1、2、3、4、5、7、10、15或 20°C 。

[0158] 图15显示若干可能的示例性缺口模式,其中引物序列中六个核苷酸位置已指定,即是固定的。在一些实施方式中,该系列引物包含具有缺口的引物和无缺口引物的混合物。在一些实施方式中,一系列中至少10、20、30、40、50、60、70、80或90%的引物是带缺口的。

[0159] 引物组中缺口模式的混合物可实现以下一种或多种:

[0160] 组装过程可跨较长,或重复的核苷酸序延伸;

[0161] 可明确地组装较长的靶序列,因为引物序列组的复杂性总体升高(即任何给定引物在靶核酸内杂交和独特的可能性增加);

[0162] 无需增加靶划分产物数目以补偿引物文库复杂性的增加。

[0163] 引物的划分产物,分光光度特征

[0164] 本发明所述划分产物可含有其他标志物以鉴定特定划分产物内的试剂(如引物)。例如,在一些实施方式中,可将一种或多种标志物试剂插入各不同的引物划分产物中,使得各引物组由基于引物划分产物中的一种或多种标志物试剂的预定且已知的独特信号代表。通过允许针对各使用的引物组的独特可检测性质,可因此确定哪个引物组导致杂交。例如,在一些实施方式中,对各反应划分产物检测存在/不存在杂交(如本文他处所述)和标志物性质,该标志物性质表示反应中引物的特性。在一个简单的示例中,如果使用两个引物组且引物组A包含标志物性质X且引物组B包含标志物性质X,则“杂交和性质X”的结果显示引物组A的引物杂交而“无杂交和性质B”的结果显示引物组B中的引物无一杂交。可随后查找引物组A和B的精确引物序列,因为标志物性质是预定且独特的。

[0165] 在一些实施方式中,通过存在一种或多种分光光度物质来生成标志物性质。在一些实施方式中,该分光光度物质包含一种或多种选择性吸收分子。本文所用“选择性吸收分子”指,在将宽频带光源导向分子时,吸收光的某些特征性颜色或波长同时允许光的其他颜色或波长通过或经由分子传输的分子。本领域技术人员应了解和理解,可以使用的多种选择性吸收分子包括本发明所述选择性吸收物质/组分,包括但不限于可购自艾克斯顿公司(Exciton)(俄亥俄州丹顿)和QCR溶液公司(QCR Solutions, Corp.)(佛罗里达州圣露西港)的那些。

[0166] 在该实施方式的另一个方面中,该分光光度物质包含一种或多种荧光分子。本文所用“荧光分子”指“荧光材料”或“荧光标记物”或“荧光团”或“荧光染料”,其在本文中使用时各自可以是荧光分子、荧光半导体纳米粒(称为“量子点”)或螯合的镧系稀土元素(lanthanide)或镧系元素(lanthanoid),其能够吸收来自特定波长光的能量,随后以具有另一特定波长(其是特定分子或量子点的特征)的荧光的形式发射该能量。在这种情况下,荧光团将有利于最终试验的读出值,所述读出值显示样品中是否存在感兴趣的特定靶标。

[0167] 使用的特定荧光团对于本发明而言并不重要。荧光团是本领域已知的且描述于例如, Marras, “Selection of Fluorophore and Quencher Pairs for Fluorescent Nucleic Acid Hybridization Probes (用于荧光核酸杂交探针的荧光团和淬灭剂对的选择)”, 刊于V. Didenko编, 2006, Fluorescent Energy Transfer Nucleic Acid Probes: Designs and Protocols (《荧光能量转移核酸探针: 设计和实验方案》) (Methods in Molecular Biology (《分子生物学方法》), 第335卷), 新泽西州: 胡马纳出版公司 (Humana Press Inc.), 第3-16页。可用于本发明的荧光团的示例包括但不限于由Marras 2006描述

和下文详细描述的那些。本领域技术人员应理解存在多种荧光染料,其可用作荧光分子并用于本发明并可从多个销售商处获得。

[0168] 可使用多种荧光染料来标记引物划分产物。荧光染料可购自多个供应商。一些示例包括但不限于:花青染料(多个供应商,如GE医疗公司(GE Healthcare)、AAT生物公司(AAT Bioquest)、塞默飞世尔科学公司(Thermo Scientific))例如Cy3、Cy5、Cy5.5、Cy7; DyLight染料(塞默飞世尔科学公司)例如DyLight 550、DyLight 594、DyLight 633、DyLight 650、DyLight 680;Atto染料(奥拓科技公司(Atto-Tec))例如Atto 610、Atto 647、Atto 680;IFluor染料(AAT生物公司)例如iFluor 633、iFluor 647、iFluor 750、iFluor 790;藻红蛋白和PerCP荧光蛋白(多个供应商,如生命技术公司(Life technologies)、AS公司(AnaSpec)、哥伦比亚生物科学公司(Columbia Biosciences));荧光素及其衍生物(例如异硫氰酸荧光素(FITC)、羧基荧光素(FAM)、四氯荧光素(TET)、2',7'-二氟荧光素(OregonGreen® 488)、OregonGreen® 514羧酸和具有氯和甲氧基取代基的荧光素(JOE和6-JOE));若丹明衍生物(例如四甲基若丹明(TAMRA)、异硫氰酸四甲基若丹明(TRITC)、四甲基若丹明(TMR)、羧基-X-罗丹明(ROX)、德克萨斯红(异构的磺酰氯和磺基若丹明的混合物;英杰公司(Invitrogen™)和德克萨斯红-X(德克萨斯红琥珀酰亚胺基酯,其在荧光团及其反应基团之间含有额外的七原子氨基己酰间隔子("X";英杰公司),和若丹明X);花青衍生物(如吡啶羧花青(Quasar® 570、Quasar® 670和Quasar® 705)、OregonGreen®异硫氰酸酯和伊红异硫氰酸酯(EITC));N-羟基琥珀酰亚胺基1-萘丁酸酯(PYB);N-羟基琥珀酰亚胺基1-苯乙烯磺酸酯(PYS);(5-(2'-氨基乙基)氨基萘(EDANS);CAL Fluor® 金540、CAL Fluor® 橙560、Fluor® 红590、CAL Fluor® 红610和CAL Fluor® 红635(可购自生物检索技术公司(Biosearch Technologies, Inc.)的专有荧光团);VIC®; HEX®(6-异构体亚磷酰胺);以及NED®。

[0169] 标记引物划分产物所需的荧光染料的浓度取决于使用的检测系统。例如,一些实施方式中引物划分产物中染料浓度的范围是10nM至1000nM(对于所列实施例中的非蛋白质染料)和0.1ug/ml至50ug/ml(对于与荧光蛋白偶联的一种或多种蛋白质染料)。与提供的示例相比,不同的检测系统可能需要较高或较低的染料浓度。

[0170] 使用的特定量子点(QD)对于本发明而言并不重要。量子点是本领域已知的且描述于例如Han等“Quantum-dot-tagged Microbeads for Multiplexed Optical Coding of Biomolecules(用于生物分子的多重光学编码的量子点标记的微珠)”,Nat Biotechnol(2001年7月)第19卷,第631-635页。本领域技术人员应理解存在多种量子点,其可用作荧光标记物并用于本发明并可从多个销售商处获得。可用于本发明的量子点(QD)的示例包括但不限于以下物质:硒化镉(CdSe)量子点纳米颗粒(例如CdSe量子点核心,480-640nm发射光谱,西格玛-奥德里奇公司(Sigma-Aldrich®));硫化镉(CdS)量子点纳米颗粒(例如CdS量子点核心,380-480nm发射光谱,西格玛-奥德里奇公司);硫化锌加帽的硒化镉(ZnS加帽的CdSe)纳米晶体(例如CdSe/ZnS Lumidots™和CdSe/ZnS NanoDots™,480-640nm发射光谱,西格玛-奥德里奇公司);以及无镉的量子点(例如CFQD™,400-650nm发射光谱,西格玛-奥德里奇公司)。

[0171] 使用的特定螯合的镧系稀土元素或镧系元素对于本发明而言并不重要。镧系稀土

元素和镧系元素是本领域已知的,其包括原子序数为57-71从镧(La)至镥(Lu)的15种金属化学元素。可用于本发明的螯合形式的镧系稀土元素或镧系元素的示例包括但不限于以下物质:镧(La)、铈(Ce)、镨(Pr)、钕(Nd)、钷(Pm)、钐(Sm)、铕(Eu)、钆(Gd)、铽(Tb)、镝(Dy)、钬(Ho)、铒(Er)、铥(Tm)、镱(Yb)和镥(Lu)。

[0172] 可使用本领域技术人员了解和理解的任何注射方法在液滴形成时或液滴形成后将染料整合至载剂(如液滴)中。可通过以液流(fluid stream)的形式流动或涌动所需染料复合物来在液滴形成期间将染料整合至液滴制造设计(droplet-maker design)。液滴制造设计和方法包括但不限于国际专利公开WO 2004/002627和WO 2006/096571中描述的那些,其各自全文纳入本文。

[0173] 根据本发明的方法,可分析待测试样品的各分光光度物质的分光光度强度测量,其中参考分光光度物质的分光光度强度测量可用于校正一种或多种样品分光光度物质的分光光度强度测量。根据应用,分光光度特征可包括:对待测试样品照射后从待测试样品散射的光;通过待测试样品内化学作用以化学发光形式发射的光;将宽频带光源导向待测试样品后由待测试样品选择性吸收的光;或者激发待测试样品后以荧光形式从待测试样品发射的光。

[0174] 可通过本领域技术人员了解和理解的用于分光光度分析的任何方法来测量分光光度物质的分光光度强度和波长。可用于本发明的分光光度方法包括但不限于:本领域技术人员已知的激光和光检测器对系统或较复杂的光学器件,其中光束的路径与分光光度物质的路径相交且分光光度物质的激发或照射被光程捕捉,所述光程包括一个或多个物镜、反光镜和/或透镜(len),以将光导向光电倍增管(PMT)或光敏照相机。用于本发明的本领域技术人员了解和理解的已知的荧光检测方法是使用流式细胞术仪器。例如,通过以八种不同的染料浓度提供四种不同的染料,可生成 $8^4 (=4096)$ 种不同的独特鉴定物,其各自均可用于鉴定独特的引物组。

[0175] 分光光度强度测量可包括一种或多种方法,包括但不限于光散射、吸收、化学发光、荧光强度、辐射衰减计数、比色等。将待测试样品置于激发能量源(如光源)的路径中,所述光源选自但不限于激光器、发光二极管(LED)、弧光灯、宽频带光源和高强度灯泡。待测试样品中的分光光度物质以波长基本不同于光源的波长的光的形式散射、吸收、化学发光(chemiluminesce)或发荧光(在本文中也称作“信号”)。随后通过检测器或传感器捕获该来自待测试样品的光,所述检测器或传感器选自但不限于照相机、电荷偶联装置(CCD)、互补金属氧化物半导体(CMOS)(或者称作互补对称金属氧化物半导体(COS-MOS))、一种或多种单一光电二极管、光电二极管阵列(PDA)、雪崩光电二极管(APD)、雪崩光电二极管阵列、光电倍增管(PMT)或光电倍增管阵列。

[0176] 已知的光学或电子方法可任选地用于扩增来自光源的光和/或来自待测试样品的光和/或将一种或全部两种分离为其组分波长。针对特定待测试样品选择参考分光光度物质和一种或多种样品分光光度物质,使得各分光光度物质在基本不同的波长下散射光、选择性吸收光、以化学放光或荧光的形式发射光(取决于分光光度物质和特定应用),以更容易地分离相应波长。参考分光光度物质的预测值与测量值之间的差异可用于对输出值定量“噪音”的贡献,前提是参考分光光度物质和一种或多种样品分光光度物质处于相同测量条件(例如光源的功率、检测器或传感器噪音、湿度、热量、待测试样品的pH和包含待测试样品

本身的载剂)。“噪音”对参考分光光度物质信号的贡献应与噪音对一种或多种样品分光光度物质信号的贡献相关。该相关性可以且通常是正比例的,但可线性地、指数地或以其他方式或函数的形式变化。

[0177] 生成用于标志单个引物划分产物的独特信号的其他信息可参见例如W02012/135327。

[0178] 靶核酸

[0179] 靶核酸可以是任何天然或合成的核酸,其可涉及“沃森-克里克”碱基配对。在许多实施方式中,该靶核酸将是DNA或RNA。这些核酸可来源于任何生物体。在一些实施方式中,本发明中使用的靶核酸可获自一种或多种真核或原核细胞。在一些实施方式中,细胞是动物细胞,包括但不限于人或非人哺乳动物细胞。非人哺乳动物细胞包括但不限于灵长类细胞、小鼠细胞、大鼠细胞、猪细胞和牛细胞。在一些实施方式中,这些细胞是非哺乳动物细胞,例如禽、爬行动物或其他细胞。在一些实施方式中,这些细胞是植物细胞。细胞可以是,例如,培养的原代细胞、永生化培养细胞或者可来自活检或组织样品的细胞,任选地经培养和刺激以在试验之前分裂。在透化和/或DNA修饰步骤之前和/或期间,培养的细胞可处于悬浮或粘附状态。在一些实施方式中,这些细胞可来自肿瘤活检样品或其他患病组织。

[0180] 这些靶核酸可以是双链或单链的。这些靶核酸可以是需要的任何长度。通常,较长的靶核酸将需要更复杂的反卷积(因为需要辨析的逻辑替代物的数目增加)。在一些实施方式中,该靶核酸是50-1000bp、100-500bp或100-250bp。

[0181] 在一些实施方式中,该靶核酸包含或是疾病、预后或病征的遗传生物标志物的一部分。作为一个示例,在一些实施方式中,该靶核酸的基因型与特定癌症或糖尿病表型相关。在一些实施方式中,该生物标志物可用于预测对用于治疗病症(包括但不限于癌症)的药物的反应。

[0182] 在一些实施方式中,该靶核酸是扩增子,即通过扩增生成。使用PCR反应扩增DNA基因座是熟知的(参见美国专利号4,683,195和4,683,202;《PCR方案:方法和应用指南》(PCR PROTOCOLS:A GUIDE TO METHODS AND APPLICATIONS)(Innis等编,1990))。通常,使用PCR来扩增DNA模板。然而,已经描述并也可采用扩增的替代性方法。在一些实施方式中,双链扩增子可通过例如以下所述的方法转化成单链核酸:Mitsis等,Nucleic Acids Res(1999),第27卷,第15期,第3057-3063页;Sanchez等,PNAS(2004),第101卷,第7期,第1933-1938页;Chen等,“异步PCR(Asynchronous PCR)”,刊于D.Park编,2011.PCR方案(《分子生物学方法》Methods in Molecular Biology),第687卷),新泽西州:胡马纳出版公司(Humana Press Inc.),第231-243页。

[0183] 在一些实施方式中,在划分产物中生成靶核酸,例如使用PCR反应。包含靶核酸的划分产物有时在本文中称作“小块(slug)”。如图1所示,可提供一种PCR引物文库,其扩增一种或多种靶核酸。包含适当引物以扩增靶核酸的划分产物与模板多核苷酸(例如样品基因组DNA、cDNA、线粒体DNA、RNA等)合并并随后在热循环条件下反应多轮(例如5、10、15、20、25、30或更多)以生成扩增子。如图1所示,使划分产物处于热循环条件下的一种方法是使划分产物通过蜿蜒的微流体通道,使得蜿蜒的不同区域暴露于热循环反应的不同温度(例如引物延伸温度、引物退火温度等)。

[0184] 虽然上文所述方法已讨论了单数形式的靶核酸(例如,靶核酸划分产物中的一种

靶核酸接触引物组以检测杂交), 在一些实施方式中, 这些靶核酸划分产物包含不同序列的超过一种靶核酸。例如, 在一些实施方式中, 这些靶核酸划分产物含有两种或三种或四种不同的靶核酸。本文中“不同的核酸”指不同序列且非等位基因的靶核酸(即其来自不同的基因组区域且不是简单的来自杂合子个体的两个等位基因)。不同的核酸各自可出现在多个拷贝中。多种靶核酸的应用大部分应用于总体靶核酸序列已知(例如两种已知的人基因序列)但个体中精确的等位基因未知的情况下。发明人发现, 由于对任何特定靶序列而言都存在许多阴性杂交的划分产物, 可对两种或三种靶核酸的组合同时进行检测和反卷积。事先知道基因的总结构可辅助序列反卷积, 其将某些引物的杂交“锚定”于特定靶核酸且因此允许反卷积过程基于这些锚建立单独的靶序列。可生成多种靶核酸, 例如通过多重化样品核酸的扩增以生成多种靶扩增子。或者, 可生成并随后混合单独的扩增子。在另一个实施方式中, 不同的靶核酸划分产物将含有不同的靶核酸序列但所得引物杂交将在不知道哪种靶核酸位于特定划分产物的情况下确定。随后的反卷积可继而逻辑上确定哪种杂交来自于哪种靶核酸。

[0185] 包含靶核酸的划分产物(靶划分产物)可随后与引物划分产物合并。鉴于待与靶核酸合并的引物划分产物的数目, 在一些实施方式中, 一部分靶核酸划分产物可与引物划分产物合并, 使得所有需要的引物划分产物(其具有不同的引物组)各自接触靶核酸划分产物的不同部分, 从而细分靶划分产物并将各细分物注入不同的引物划分产物。将液滴注入其他液滴的方法描述于例如US2012/0132288。在一些实施方式中, 靶核酸划分产物可被分为至少50、100、200、300、400或更多个(例如50-1000、50-500、50-5000个)部分, 各部分随后使用不同的引物划分产物注射(混合)以形成反应划分产物。应注意, 虽然需要合并一部分靶核酸划分产物与各种不同类型的引物划分产物, 通常不需要获得精确的序列。实际上, 在一些实施方式中, 常见情况是组中5%、10%、20%或更多的引物不与靶核酸反应。

[0186] 引物杂交试验

[0187] 如上文所述, 该方法包括检测反应划分产物中是否存在一种或多种引物的杂交。其可按需要和便利实现。在一个实施方式中, 如WO 2012/078710中所述检测杂交, 其通过引用全文纳入本文。简言之, 该方法可包括生成包含荧光标记物或其他可检测物质的靶核酸(例如与荧光标记物共价连接的核酸序列)并退火至包含淬灭剂的抑制剂多核苷酸, 使得抑制剂多核苷酸与靶核酸的杂交导致荧光标记物信号的淬灭。该检测器核酸(即与抑制剂多核苷酸杂交的靶标相关的序列)可以是靶序列的一部分或被添加至靶核酸序列。测试引物(即反应划分产物中的一种或多种引物)可与含链置换聚合酶的靶核酸/抑制剂多核苷酸双链体合并划分产物, 使得如果引物退火至靶核酸, 则聚合物延伸引物并置换抑制剂多核苷酸, 从而生成荧光信号, 显示引物已杂交。如果引物未杂交, 淬灭剂不置换且检测不到信号(或检测到降低的信号)。应注意, 在替代性配置中, 淬灭剂和荧光标记物还可分别与靶核酸和抑制剂多核苷酸连接。链置换试验可在等温下发生且因此无需热循环。在一些实施方式中, 该靶核酸将以扩增子生成, 具有5' 荧光标记物且任选具有3' 茎(即由寡核苷酸与3' 端杂交所形成的双链端)或茎环。

[0188] 该检测器核酸是整合至靶核酸的寡核苷酸, 其功能为反应抑制剂的结合位点。在一个实施方式中, 该检测器核酸通过适体连接整合至靶核酸中。在该实施方式的一个示例中, 这些适体是相对于彼此类似的两个寡核苷酸。在该实施例中, 这些适体将检测器接合至

靶核酸。在另一个实施方式中,使用PCR引物将检测器核酸整合至核酸样品。在该实施方式的一个示例中,这些PCR引物包含靶标特异性序列(在引物的3'端上)、设计为在下游步骤中与抑制剂杂交的通用核酸序列(在引物的5'端上)和检测器。参见例如WO 2013/122826。在任何实施方式中,该检测器核酸整合至靶核酸序列中并取向为靶核酸序列的5'。

[0189] 在一个实施方式中,该检测器核酸偶联于荧光团。荧光团是一种特定分子,其能够从特定波长的光吸收能量然后将该能量以另一特定波长特征的荧光发出。在这种情况下,荧光团将有利于最终试验的读出值,所述读出值显示是否存在靶核酸。使用的特定荧光团对于本发明而言并不重要。荧光团是本领域已知的且描述于Marras,“Selection of Fluorophore and Quencher Pairs for Fluorescent Nucleic Acid Hybridization Probes (用于荧光核酸杂交探针的荧光团和淬灭剂对的选择)”,刊于V.Didenko编,2006, Fluorescent Energy Transfer Nucleic Acid Probes: Designs and Protocols (《荧光能量转移核酸探针:设计和实验方案》) (Methods in Molecular Biology (《分子生物学方法》),第335卷),新泽西州:胡马纳出版公司(Humana Press Inc.),第3-16页。可用于本发明的荧光团的示例包括但不限于由Marras 2006描述和下文详细描述的那些。与检测器核酸相关的荧光团的具体位置对于本发明而言并不重要。该荧光团可沿检测器核酸的任意位置处接合,包括5'端、3'端或沿检测器核酸内部的任意位置。

[0190] 该抑制剂是与检测器核酸杂交的寡核苷酸。该抑制剂的功能是仅在寡核苷酸探针匹配靶核酸时允许信号被检测。抑制剂与检测器核酸的杂交发生于标准反应缓冲液中,例如,在DNA聚合酶反应缓冲液中,从而在适当温度下在缓冲液中混合检测器核酸和抑制剂。在一个实施例中,反应物可加热至95℃持续30秒,随后冷却至抑制剂退火温度下5℃。

[0191] 在一个实施方式中,该抑制剂多核苷酸偶联于淬灭剂。该淬灭剂是一种特定分子,其功能是降低(即淬灭)荧光强度,具体方式为将能量从第一荧光团转移至第二荧光团或非荧光分子。使用的特定淬灭剂对于本发明而言并不重要。淬灭剂是本领域已知的且描述于例如Marras 2006。可用于本发明的淬灭剂的示例包括但不限于由Marras 2006描述和下文详细描述的那些。与抑制剂相关的淬灭剂的具体位置对于本发明而言并不重要。该淬灭剂可沿抑制剂多核苷酸的任意位置处接合,包括5'端、Y端或沿抑制剂内部的任意位置。

[0192] 在一个替代性实施方式中,该检测器核酸偶联于淬灭剂。

[0193] 可使用任何置换聚合酶在杂交引物存在情况下实现抑制剂的置换。市售可得的聚合酶的例子包括但不限于:克列诺片段(新英格兰生物实验室公司)、Taq DNA聚合酶(凯杰公司(QIAGEN))、9°N™ DNA聚合酶(新英格兰生物实验室公司)、Deep Vent™ DNA聚合酶(新英格兰生物实验室公司)、Manta DNA聚合酶(酶学公司)、Bst DNA聚合酶(新英格兰生物实验室公司)、和phi29DNA聚合酶(新英格兰生物实验室公司)。

[0194] 如上文所述,在一些方面中,生成具有3'茎或茎环的靶核酸。茎环部分将包含形成前半茎的序列,之后是形成环的多个核苷酸,之后是前半茎的反向互补物。前半茎和后半茎通常会有完全相同的长度,但在一些实施方式中,一半茎可比另一半茎多一个核苷酸,使得当两半退火时,一半茎的一个核苷酸不退火。各半的茎可以是所需的任意尺寸,例如,5-15个核苷酸长,例如,5、6、7、8、9、10、11、12、13、14或15个核苷酸。环将具有至少一个核苷酸,并且在一些实施方式中,将具有2-6个,例如,2、3、4、5或6个核苷酸。在一些实施方式中,茎将具有至少40℃、45℃、50℃、55℃、60℃或65℃的溶解温度(T_m)。在一些实施方

式中,例如, T_m 比引物延伸试验所采用的温度高至少5度或10度。可凭经验确定环的 T_m ,并且在一些实施方式中,例如,使用F.Baldino, Jr, M.-F.Chesselet和M.E.Lewis, *Methods in Enzymology* 168:766 (1989) 中所述的公式估计环的 T_m 。在一些实施方式中,茎在37°C下具有-8.5、-3或更低的最小自由能(ΔG)。理想地,多核苷酸3'末端序列将允许很少或没有替代性构象(例如,其它替代性二级结构)。可通过本领域普通技术人员已知的软件分析这些方面。这类软件的示例包括可从mfold.rna.albany.edu得到的UNAFold软件。或者,可形成更复杂的(锤头装)二级结构,只要该结构减弱了3'末端的可及性以降低假启动。在一些实施方式中,多核苷酸的最后3'末端与茎的3'端(即,茎结构的最后核苷酸)一致。或者,在一些实施方式中,在3'末端处,1、2、3、4、5或6个不是茎环的部分的核苷酸可存在于5'末端处,由于在茎的5'方向上缺少与附近的核苷酸的互补性。

[0195] 示例性工作流

[0196] 下文提供了一种示例而非旨在限制要求保护的发明。

[0197] 根据图1,在划分产物中提供靶引物和包含聚合酶的其他PCR试剂(图1A)。将来自样品的基因组DNA(图1B)注入划分产物中并通过压力驱动其通过蜿蜒的微流体通道使其处于热循环条件下(图1C),从而生成扩增子,其是靶核酸。扩增反应还将荧光分子导入靶核酸的5'端。

[0198] 同时,在容器中提供提供的引物划分产物文库。该引物划分产物包含多个引物(一个组),不同的划分产物包含不同的引物组。作为示例,各可能的序列(4096)的连续(或缺口)六聚体的文库提供于10个引物的组中,各组包含一种也出现在至少一个其他组中的引物。这些引物划分产物还包含若干种可能浓度之一的一种或多种不同染料,使得不同染料的浓度显示特定引物划分产物中引物组的性质。向引物划分产物中加入链置换聚合酶,这一操作在其与靶核酸划分产物合并之前或之后进行。向引物划分产物中注射靶核酸划分产物的各部分(使得向许多引物划分产物中注射一种靶核酸)以形成反应划分产物(图1D)。

[0199] 包含置换聚合酶和杂交淬灭剂多核苷酸的反应划分产物在允许淬灭剂多核苷酸置换的条件下通过微流体通道,前提是至少一个来自引物划分产物的引物与靶核酸杂交。

[0200] 可连续生成反应划分产物的信号。可由一个或多个检测器检测信号(图1F)。检测的信号包括是否存在来自靶核酸上荧光分子的荧光(或其差异水平),从而确定杂交。还检测了划分产物中多种染料的水平和性质,从而提供划分产物中特定引物的性质。本发明的若干实施方式使用图6A所述的软件架构。图2显示哪种软件模块在过程的哪个阶段被执行且在哪个时间粒度处进行和在何种输入物上运行。例如,信号加工模块在每个引物划分产物被光学检测器测量后执行一次。

[0201] 随后加工所得信号。例如,使用多种染料的性质的水平(例如各自以不同波长检测)来确定引物组的性质,并因此确定反应划分产物中各引物的序列。随后可将数据分选为不杂交的引物组和杂交的引物组。存在或不存在特定引物的杂交可从各组的杂交模式中通过逻辑确定(例如经由计算机软件)。例如,阳性杂交划分产物最初显示潜在地划分产物中的所有引物都杂交。然而,通过观察阴性杂交的划分产物,可确定阴性和阳性杂交的划分产物之间的重叠引物实际上不杂交。随后可使用引物杂交的所得信息(再次经由计算机软件)以及靶核酸总体结构的知识来基于是否存在引物杂交生成预测的杂交序列。

[0202] 此外,来自反应的信号可经连续监测以确定来源于不同靶核酸划分产物的信号的

边缘,即当来自第一靶核酸划分产物的各部分被耗尽且正检测来自第二靶核酸划分产物(与引物划分产物混合)的划分产物时。该方法特别适用于在系统中测试多个靶核酸时(例如在其中不同的引物对被用于生成不同的靶核酸划分产物的实施方式中,或在等位基因变异或样品变异的情况中),至少部分是因为可以随后基于划分产物中靶核酸的预测性质来预测并分选来自划分产物的信号。例如,如果系统中存在两种靶标,且各靶标的野生型序列是已知的,则可预测哪个引物划分产物将与野生型参考序列杂交和哪个不会。随后可将该引物划分产物杂交模式与引物划分产物杂交的实际模式相比较以确定正在具体靶核酸划分产物(“小块”)中测试的是哪一种靶标。该信息可在之后组装序列时有帮助,因为在组装靶序列前将已知例如与靶标2相比施加于靶标1的杂交信号。

[0203] 图16显示一种示例性组装过程,通过该过程基于是否存在一系列引物的杂交来确定靶核酸序列。作为初始阶段,杂交引物以明确重叠模式排列以生成初始靶支架序列。在许多实施方式中,该支架将是不完整的,在序列中具有缺口,其无法由阳性杂交单独确定。例如,在一些情况中,可将杂交引物逻辑上置于两个或更多个位置中的支架中的阵列中。在该情况中,阴性(不)杂交的类似但不同序列的引物可被应用于辨析模糊位置。该方面还示于图17,其中带圆圈的阳性杂交引物可逻辑上出现在支架中的任一位置。阴性引物1722和1724可随后用于辨析带圆圈的引物属于位置C。在一些实施方式中,对于使用阴性引物数据辨析某些阳性引物,可随后将其他先前模糊的阳性引物序列置于支架中。任选地,参考序列(即靶序列的总体结构)可与支架比较以辨析模糊处。作为示例,如果p53的一部分是靶序列,数据库参考p53序列可用作用于比较的支架的参考。

[0204] 划分产物

[0205] 划分产物可包括多种类型的划分产物中的任一种,包括固体划分产物(如孔或管)和流体划分产物(如油相内的水性液滴)。在一些实施方式中,这些划分产物是液滴。在一些实施方式中,这些划分产物是微通道。划分样品的方法和组合物描述于,例如,公开的专利申请WO 2010/036352、US 2010/0173394、US 2011/0092373、US 2011/0092376、US2012/0222748;WO2013/09573;和US 2011/0218123,其全部内容各自通过引用纳入本文。

[0206] 在一些实施方式中,将样品划分成多个液滴。在一些实施方式中,液滴包含乳液组合物,即不互溶的流体(如水和油)的混合物。在一些实施方式中,液滴是水性液滴,其被不互溶的载体流体(如油)包围。在一些实施方式中,液滴是油性液滴,其被不互溶的载体流体(如水性溶液)包围。在一些实施方式中,本文所述液滴是相对稳定的并在两种或更多种液滴之间具有最小聚结。在一些实施方式中,由样品生成的液滴中少于0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%与其他液滴聚结。这些乳液还可具有有限的絮凝,一种分散相从薄片悬浮液产生的过程。

[0207] 在一些实施方式中,使油相流过包含待检测标记物的水性溶液,从而形成液滴。在一些实施方式中,包含待检测标记物的水性样品包括缓冲溶液和检测标记物的试剂。用于油相的油可经合成或天然存在。在一些实施方式中,所述油包括碳和/或硅。在一些实施方式中,所述油包括氢和/或氟。示例性的油包括但不限于硅油、矿物油、氟碳油、植物油或其组合。

[0208] 该油相可包含氟化基础油,其可通过与氟化表面活性剂(如全氟聚醚)联用而进一

步稳定。在一些实施方式中,该基础油包括以下一种或多种:HFE7500、FC-40、FC-43、FC-70或另一种常见氟化油。在一些实施方式中,该油相包含阴离子含氟表面活性剂。在一些实施方式中,该阴离子含氟表面活性剂是Ammonium Krytox (Krytox-AS)、Krytox FSH的铵盐或Krytox FSH的吗啉基衍生物。Krytox-AS的浓度可以是约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1.0%、2.0%、3.0%或4.0% (w/w)。在一些实施方式中,Krytox-AS的浓度是约1.8%。在一些实施方式中,Krytox-AS的浓度是约1.62%。Krytox FSH的吗啉基衍生物的浓度可以是约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1.0%、2.0%、3.0%或4.0% (w/w)。在一些实施方式中,Krytox FSH的吗啉基衍生物的浓度是约1.8%。在一些实施方式中,Krytox FSH的吗啉基衍生物的浓度是约1.62%。

[0209] 在一些实施方式中,该油相还包含用于调节油性质(如蒸气压、粘性或表面张力)的添加剂。非限制性示例包括全氟辛醇和1H,1H,2H,2H-全氟癸醇。在一些实施方式中,1H,1H,2H,2H-全氟癸醇以约0.05%、0.06%、0.07%、0.08%、0.09%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1.0%、1.25%、1.50%、1.75%、2.0%、2.25%、2.5%、2.75%或3.0% (w/w)的浓度添加。在一些实施方式中,1H,1H,2H,2H-全氟癸醇以约0.18% (w/w)的浓度添加。

[0210] 在一些实施方式中,该乳液配制为生成具有类液界面膜的高度单分散液滴,其可通过加热转化为具有类固界面膜的微胶囊;这类微胶囊可作为生物反应器以通过一段时间的孵育保持其含量。转化为微胶囊形式可在一经加热后即发生。例如,这类转化可发生于大于约40°、50°、60°、70°、80°、90°或95°C的温度下。加热过程期间,可使用流体或矿物油覆盖物来阻止蒸发。可在加热前除去或不除去过量的连续相油。这些生物相容性胶囊可在大范围的热和机械处理下抗聚结和/或絮凝。

[0211] 转化后,这些微胶囊可储存于约-70°、-20°、0°、3°、4°、5°、6°、7°、8°、9°、10°、15°、20°、25°、30°、35°或40°C下。在一些实施方式中,这些胶囊可用于生物医学应用,例如稳定、数字化的大分子包封,特别是包含目标分子(如核酸、蛋白质或上述两者)的混合物的水性生物液体;药物和疫苗递送;生物分子文库;临床成像应用;等。

[0212] 这些微胶囊划分产物可含有本文所述的一种或多种亲和试剂并可抗聚结,特别是在高温下。因此,这些胶囊可在非常高的密度(例如每单位体积的划分产物数)下孵育。在一些实施方式中,可每毫升孵育超过100,000、500,000、1,000,000、1,500,000、2,000,000、2,500,000、5,000,000或10,000,000个划分产物。在一些实施方式中,样品-探针孵育发生在单个孔中,例如微量滴定板的孔,此时各划分产物之间不具有内部混合。这些微胶囊还可含有孵育所需的其他组分。

[0213] 在一些实施方式中,生成的液滴在形状和/或尺寸方面基本均匀。例如,在一些实施方式中,这些液滴在平均直径方面基本均匀。在一些实施方式中,生成的液滴的平均直径为约0.001微米、约0.005微米、约0.01微米、约0.05微米、约0.1微米、约0.5微米、约1微米、约5微米、约10微米、约20微米、约30微米、约40微米、约50微米、约60微米、约70微米、约80微米、约90微米、约100微米、约150微米、约200微米、约300微米、约400微米、约500微米、约600微米、约700微米、约800微米、约900微米,或约1000微米。在一些实施方式中,生成的液滴的平均直径为小于约1000微米、小于约900微米、小于约800微米、小于约700微米、小于约600

微米、小于约500微米、小于约400微米、小于约300微米、小于约200微米、小于约100微米、小于约50微米,或小于约25微米。在一些实施方式中,生成的液滴在形状和/或尺寸方面是不均匀的。

[0214] 在一些实施方式中,生成的液滴在体积上基本均匀。例如,在一些实施方式中,生成的液滴的平均体积为约0.001nL、约0.005nL、约0.01nL、约0.02nL、约0.03nL、约0.04nL、约0.05nL、约0.06nL、约0.07nL、约0.08nL、约0.09nL、约0.1nL、约0.2nL、约0.3nL、约0.4nL、约0.5nL、约0.6nL、约0.7nL、约0.8nL、约0.9nL、约1nL、约1.5nL、约2nL、约2.5nL、约3nL、约3.5nL、约4nL、约4.5nL、约5nL、约5.5nL、约6nL、约6.5nL、约7nL、约7.5nL、约8nL、约8.5nL、约9nL、约9.5nL、约10nL、约11nL、约12nL、约13nL、约14nL、约15nL、约16nL、约17nL、约18nL、约19nL、约20nL、约25nL、约30nL、约35nL、约40nL、约45nL或约50nL。

[0215] 文库

[0216] 还提供了本发明所述引物划分产物的文库。例如,在一些实施方式中,提供了划分产物的文库,这些划分产物包含多种不同的引物且这些划分产物中提供至少1000、10000、100000或更多种不同的独特引物组划分产物,每个划分产物一个引物组。在一些实施方式中,这些组将含有2-20种(例如5-15、5-10、7-12种)不同引物。在一些实施方式中,这些引物将包含至少4、5、6、7、8、9、10或更多个指定核苷酸以及任选的1、2、3、4、5或更多个“缺口”简并位置。

[0217] 在一些实施方式中,划分产物的文库将还包含一种或多种分光光度可区分的分子以确定划分产物中引物的性质。因此,例如,如果各划分产物中存在n组不同的引物,则将存在n种不同的分光光度可区分的特性,各不同的组分配给一种具体特性。如上文所述,可例如通过提供1、2、3、4、5或更多种不同的可检测染料来实现,其各自具有1或2、3、4、5、6、7、8、9、10或更多种浓度,使得各染料的浓度组合提供针对各引物组的独特的分光光度上可区分的特性。用于监测分光光度强度组合的方法和系统描述于例如WO 2012/135327。

[0218] 系统

[0219] 还提供了进行本发明所述方法的系统。在一些实施方式中,这些系统包括第一容器,其包含上文所述多个引物划分产物(如液滴)中的一系列引物。在一些实施方式中,该系统还包含第一微流体通道和检测器,所述第一微流体通道提供第一容器和检测器之间的流体连通。在一些情况中,该系统还包含含有用于扩增靶核酸的引物对的第二容器;提供第二容器与第一微流体通道之间流体连通的第二微流体通道;以及与第二微流体通道流体连通的样品核酸容器。参见例如图1。在一些实施方式中,该第一通道包含蜿蜒部分,其经设置使得蜿蜒通道的不同部分可设置为不同的温度(例如具有珀耳帖元件或其他温度控制元件且任选具有自动调温器)。在一些实施方式中,所有上述组间都以单个筒匣的一部分提供。在一些实施方式中,该筒匣可继而插入歧管中以接合一个或多个泵,这些泵设置成将推动液滴通过微流体通道。

[0220] 在一些实施方式中,该系统还包含一种或多种液滴注射器。在一些实施方式中,该系统包含第一液滴注射器和第二液滴注射器,所述第一液滴注射器构设成将核酸从样品核酸容器中注入包含来自第二容器的引物对且位于第二微流体通道中的液滴内以形成混合物液滴,所述第二液滴注射器构设成将混合物液滴分成的各份注入向下游通过第一微流体通道的引物液滴中。液滴注射器描述于例如US 2012/0132288。

[0221] 示例性系统组件描述于,例如,US2011/0267457、US2011/0151578、US2011/0218123、US2012/0222748、US2011/0218123、2012/0222748、WO2012/135201、WO2012/135259、WO2014/043388、WO 2012/135327。

[0222] 本发明所述的检测器可检测来自以下的一种或全部两种信号:(i) 杂交试验或(ii) 引物划分产物中的染料以鉴定引物。在一些实施方式中,乳液流中的液滴流过微流体通道,其通过测量来自试验的荧光信号的光学检测器。在一些实施方式中,随时间生成和聚集相同靶分子信号的多组测量值。

[0223] 可通过本领域技术人员了解和理解的用于分光光度分析的任何方法来测量分光光度物质的分光光度强度和波长。可用于本发明的分光光度方法包括但不限于:本领域技术人员已知的激光和光检测器对系统或更复杂的光学器件,其中光束的路径与分光光度物质的路径相交且分光光度物质的激发或照射被光程捕捉,所述光程包括一个或多个物镜、反光镜和/或透镜,以将光导向光电倍增管(PMT)或光敏照相机。用于本发明的本领域技术人员了解和理解的已知的荧光检测方法是使用流式细胞术仪器。

[0224] 分光光度强度测量可包括一种或多种方法,包括但不限于光散射、吸收、化学发光、荧光强度、辐射衰减计数、比色等。将待测试样品置于激发能量源(如光源)的路径中,所述光源选自但不限于激光器、发光二极管(LED)、弧光灯、宽频带光源和高强度灯泡。待测试样品中的分光光度物质以波长基本不同于光源的波长的光的形式散射、吸收、化学发光(chemiluminesce)或发荧光(在本文中称作“信号”)。随后通过检测器或传感器捕获该来自待测试样品的光,所述检测器或传感器选自但不限于照相机、电荷偶联装置(CCD)、互补金属氧化物半导体(CMOS)(或者称作互补对称金属氧化物半导体(COS-MOS))、一种或多种单一光电二极管、光电二极管阵列(PDA)、雪崩光电二极管(APD)、雪崩光电二极管阵列、光电倍增管(PMT)或光电倍增管阵列。

[0225] 已知的光学或电子方法可任选地用于扩增来自光源的光和/或来自待测试样品的光和/或将一种或全部两种分离为其组分波长。针对特定待测试样品选择参考分光光度物质和一种或多种样品分光光度物质,使得各分光光度物质在基本不同的波长下散射光、选择性吸收光、以化学放光或荧光的形式发射光(取决于分光光度物质和特定应用),以更容易地分离相应波长。参考分光光度物质的预测值与测量值之间的差异可用于对输出值定量“噪音”的贡献,前提是参考分光光度物质和一种或多种样品分光光度物质处于相同测量条件(例如光源的功率、检测器或传感器噪音、湿度、热量、待测试样品的pH和包含待测试样品本身的载剂)。“噪音”对参考分光光度物质信号的贡献应与噪音对一种或多种样品分光光度物质信号的贡献相关。该相关性可以且通常是正比例的,但可线性地、指数地或以其他方式或函数的形式变化。

[0226] 在一些但不是所有实施方式中,本发明所述系统和方法的所有组件都是微流体的。本文中,“微流体”指含至少一个流体通道的装置、仪器或系统包,该流体通道的截面尺寸小于1mm,且长度与垂直于通道的最大截面尺寸比为至少约3:1。本文所用的“微流体通道”是符合这些标准的通道。

[0227] 可提供能够导致两个或更多个液滴融合或聚结成一个液滴的微流体系统,例如在两个或更多个液滴通常无法融合或聚结的情况中,例如由于组成、表面张力、液滴尺寸等,如本领域普通技术人员已知的那样。上文已经描述了其中两个或更多个液滴融合的实施方

式的示例。可使用任何合适的技术将流体液滴融合在一起,例如,参见Link等2005年10月7日提交的美国专利申请系列号11/246,911(标题为“Formation and Control of Fluidic Species (流体物质的形成和控制)”),其在2006年7月27日公开为美国专利申请公开号2006/0163385;或Link等2006年2月23日提交的美国专利申请系列号11/360,845(标题为“Electronic Control of Fluidic Species (流体物质的电子控制)”),其在2007年1月4日公开为美国专利申请公开号2007/0003442,其各自通过引用纳入本文。例如,在微流体系统中,相对于液滴尺寸的液滴表面张力可阻止发生液滴的融合或聚结。在一个实施方式中,对两个液滴给予相反的电荷(即正电荷和负电荷,无需相同幅度),其可提高两个液滴的电相互作用从而发生液滴的融合或聚结。通过使用泰勒锥或通过任何其他合适的技术将电荷(正或负)施加到液滴上。例如,可对含有液滴的反应器施加电场,这些液滴可通过电容器,可发生化学反应导致液滴带电荷,在具有相反湿润性质的区域上流动液滴,等。

[0228] 通道的“截面尺寸”以垂直于流体流动方向测量。本发明的组件中的大多数流体通道的最大截面尺寸小于约2mm,且在一些情况中小于约1mm。在一组实施方式中,所有含有本发明的实施方式的流体通道都是微流体或最大截面尺寸不超过约2mm或约1mm。在另一个实施方式中,这些流体通道可部分通过单一组件形成(例如蚀刻的底物或模塑单元)。当然,较大的通道、管道、腔体、储器等可用于储存大量流体和将流体递送至本发明的组件。在一组实施方式中,包含本发明的实施方式的通道的最大截面尺寸小于约500微米、小于约200微米、小于约100微米、小于约50微米或小于约25微米。

[0229] 本文中,“通道”指至少部分引导流体流的制品(物质)上方或内部的特征。该通道可具有任何截面形状(圆形、椭圆形、三角形、不规则形状、方形或矩形等)并可以是被覆盖的或未覆盖的。在其被完全覆盖的实施方式中,至少一部分通道可具有完全闭合的截面,或整个通道可沿其整个长度完全闭合(其入口和出口除外)。通道还可具有至少约2:1的纵横比(长度与平均截面尺寸的比率),更通常是至少约3:1、至少约5:1或至少约10:1或更多。开放的通道通常将包括促进对流体运输控制的特性,例如结构特性(延长的压痕)和/或物理或化学特性(疏水性对比亲水性)或可对流体施加作用力(如包含力(containing force))的其他特性。通道内的流体可部分或全部充满通道。在一些情况中,当使用开放的通道时,流体可维持在通道内,例如使用表面张力(即凸形或者凹形弯月面)。

[0230] 该通道可以是任何尺寸,例如,与流体流垂直的最大尺寸小于约5mm或约2mm,或小于约1mm,或小于约500微米,小于约200微米,小于约100微米,小于约60微米,小于约50微米,小于约40微米,小于约30微米,小于约25微米,小于约10微米,小于约3微米,小于约1微米,小于约300nm,小于约100nm,小于约30nm,或小于约10nm。在一些情况中,可选择通道的尺寸使得流体能够自由地流过制品或物质。还可选择通道的尺寸以例如允许通道内流体有特定体积或线性流速。当然,可通过本领域普通技术人员已知的任何方法改变通道的数目和通道的形状。在一些情况中,可以使用超过一条通道或毛细管。例如,可以使用两条或更多条通道,其中其位于彼此内部,彼此相邻,彼此相交等。

[0231] 可用于本发明的微流体系统的非限制性示例公开于:2005年10月7日提交标题为“Formation and Control of Fluidic Species (流体物质的形成和控制)”的美国专利申请系列号11/246,911,其在2006年7月27日公开为美国专利申请公开号2006/0163385;2004年12月28日提交标题为“Method and Apparatus for Fluid Dispersion (流体分散的方法

和设备)”的美国专利申请系列号11/024,228,其在2005年8月11日公开为美国专利申请公开号2005/0172476;2006年2月23日提交标题为“Electronic Control of Fluidic Species (流体物质的电子控制)”的美国专利申请系列号11/360,845,其在2007年1月4日公开为美国专利申请公开号2007/000342;2006年3月3日提交的标题为“Method and Apparatus for Forming Multiple Emulsions (形成多种乳液的方法和设备)”的国际专利申请号PCT/US2006/007772,其在2006年9月14日公开为W0 2006/096571;2006年3月3日提交的标题为“System and Method for Forming Particles (形成颗粒的系统和方法)”的美国专利申请系列号11/368,263,其在2007年3月8日公开为美国专利申请公开号2007/0054119;2007年3月28日提交的标题为“Multiple Emulsions and Techniques for Formation (多种乳液及其形成技术)”的美国临时专利申请系列号60/920,574;以及2006年1月20日提交的标题为“Systems and Methods for Forming Fluidic proplets Encapsulated in Particles Such as Colloidal Particles (用于形成包封在诸如胶体颗粒的颗粒中的流体的液滴的系统和方法)”的国际专利申请号PCT/US2006/001938,其在2006年7月27日公开为W0 2006/078841,其各自通过引用全文纳入本文。

[0232] 软件

[0233] 如上文所述,本发明所述方法生成涉及阳性杂交引物以及阴性杂交引物的数据。因此,在一些实施方式中,提供了使用阳性和阴性数据的辨析靶核酸序列的方法。在一些实施方式中,杂交引物以明确重叠模式排列以生成初始靶支架序列。在该步骤中,仅使用明确的对齐。在许多实施方式中,该支架将是不完整的,在序列中具有缺口,其无法由阳性杂交单独确定。例如,在一些情况中,可将一种或多种杂交引物逻辑上置于两个或更多个位置中支架中的阵列中。在该情况中,可仅基于阳性杂交将阴性(不)杂交的类似但不同序列的引物应用于辨析模糊位置。在一些实施方式中,对于使用阴性引物数据辨析某些阳性引物,可随后将其他先前模糊的阳性引物序列置于支架中。任选地,参考序列(即靶序列的总体结构)可与支架比较以辨析模糊处。

[0234] 在靶核酸划分产物的各部分(在这里称作“小块”或“混合物液滴”)与不同的引物划分产物部分合并以形成反应划分产物的实施方式中,需要确定哪个反应信号来源于哪个靶核酸划分产物。例如,在简单的实施例中,第一靶核酸划分产物的各部分与引物划分产物合并以生成第一靶核酸反应划分产物,其生成信号。一旦第一靶核酸划分产物被消耗,该系统立即生成来自第二靶核酸划分产物的部分并将那些部分与其他引物划分产物一比一合并。因此需要在来源于一个混合物液滴的数据(例如来自第一靶核酸划分产物)结束且来源于另一混合物液滴的数据(例如来自第二靶核酸划分产物)开始时确定生成的信号,特别是在其中不同的靶核酸划分产物可含有不同的等位基因、扩增子等的情况中。

[0235] 图18是根据本发明的实施方式用于确定靶核酸的核苷酸序列的方法1800的流程图。方法1800可完全或部分由计算机系统进行。

[0236] 在模块1850处,接收数据信号。该数据信号对应于包括多种引物与靶核酸杂交的实验,例如如上文所述。可随时间通过检测器获得数据信号。该数据信号可由各种信号组成,其各自针对由一部分混合物液滴与引物划分产物合并生成的各反应划分产物。可使用多种混合物液滴,且各混合物液滴可包含一种或多种靶核酸的拷贝。各反应划分产物包含一种或多种引物,且所有划分产物都可包含多种引物。

[0237] 在模块1820处,确定哪种划分产物对应于哪种混合物液滴。该数据信号可以由检测器收集的连续数据流。随着各新划分产物通过检测器,接收额外的数据。由于各混合物液滴被划分为多个反应划分产物且连续划分产物流通过检测器划分产物,因此难以知道一种混合物液滴结束的划分产物位置和下一混合物液滴开始的划分产物位置。对应于同一混合物液滴的划分产物在本文中称作一个集合。

[0238] 在一个实施方式中,可通过分析划分产物的杂交数据来确定集合的起始(先前集合的结束)。如果划分产物集合对应于同一混合物液滴,则那些划分产物的杂交应是类似的。例如,同一引物应与靶核酸杂交而无需考虑引物处于集合的哪个划分产物中。可追踪杂交数据中的矛盾,且时间中特定点处矛盾数据量的最小值可表示一个集合的结束和下一集合的开始。

[0239] 在方框1830处,对各划分产物集合确定相应的靶核酸。因为在多个混合物液滴上总共可存在许多靶核酸,所以难以知道混合物液滴内含有何种靶核酸。此外,混合物液滴可包含超过一种靶核酸的事实也使问题变得复杂。

[0240] 在一个实施方式中,可针对特定的划分产物集合建立杂交概况(如位向量)。该杂交概况可与参考杂交概况比较,其各自对应于不同的参考(如野生型)靶核酸。且可鉴定最接近的参考杂交概况。一个集合可被鉴定为对应于具有最接近杂交概况的靶核酸。多个集合可对应于同一靶核酸。并且,一个集合可对应于多个靶核酸(例如当混合物液滴包含多种靶核酸时)。

[0241] 在方框1840处,确定引物对靶核酸的杂交状态。引物的杂交状态表明引物是否与特定靶核酸杂交。划分产物的杂交状态表明该划分引物中的至少一种引物与特定靶核酸杂交。可基于来自划分产物的信号确定该划分产物的杂交状态。例如,根据由划分产物发射并通过检测器检测的特定颜色,该信号可以是二进制值。

[0242] 由于一个划分产物可包含多种引物,难以从划分产物的杂交状态中了解划分产物中引物的杂交状态。然而,同一引物包含在多个划分产物中。并且,当划分产物数目较大时,任意划分产物内的具体引物将变化。在一个实施方式中,可以比较具有相同引物的划分产物的杂交状态以鉴定矛盾状态。这类分析可显示相同引物的杂交状态。

[0243] 在方框650处,与靶核酸杂交的引物被用于组装靶核酸的核苷酸序列。例如,可通过确定或重叠彼此一致的核苷酸来组装这些引物的序列。还可使用不与靶核酸杂交的引物的知识。例如,可以使用多个选项(其中引物可与靶核酸对齐),且阴性引物(即不杂交的引物)可帮助排除这些选项之一。

[0244] A. 确定集合

[0245] 集合的确定可视作引物划分产物读数的时间分割。可以多种方式使用引物不杂交的信息。由于一个划分产物内的引物无法消除彼此的阳性,可观察运行期间不同时间处测量的其他划分产物。这包括使各划分产物彼此相关。在一些实施方式中,一种相关性是来自同一混合物液滴的划分产物之间的相关性,即集合内。由于集合内的划分产物可能包含同一靶核酸,来自相同混合物液滴的划分产物之间共有的引物应反映相同的与混合物液滴的靶核酸的杂交。

[0246] 图9的图900显示本发明的实施方式所述的集合划分产物边缘。横轴对应于划分产物数目,其对应于特定时间。纵轴对应于矛盾值。

[0247] 矛盾率910显示为划分产物零和划分产物50000之间。使用划分产物上的滑动窗口计算矛盾率。在多个实施方式中,各滑动窗口的中心或起始处可提供该划分产物的数据点。例如,10000处的矛盾值可由划分产物10000之后的2000个划分产物(可能包括划分产物10000)来确定。

[0248] 当所有划分产物都对应于相同混合物液滴时,矛盾的量将最小,前提是杂交状态测量与相同靶核酸的杂交。然而,当这些划分产物来自不同的混合物液滴时,不同的划分产物具有针对不同靶核酸测量的杂交状态;且因此矛盾率最大。在图9中,混合物液滴(小块)之间的边缘920显示为峰之间矛盾率910的最小值。在其他实施方式中,这些边缘可鉴定为最大值,这取决于如何定义滑动窗口,例如当滑动窗口位于特定时间中央时。

[0249] 矛盾率910可确定为比率或原始数。对于比率,分子可对应于显示矛盾杂交状态的划分产物量。该量可计数为显示矛盾数据的引物数目或划分产物数目。例如,两个划分产物可具有相同引物但具有不同的杂交状态,这可视为矛盾数据。在多个实施方式中,可对矛盾的划分产物数目进行计数,或可对矛盾的引物数目进行计数。

[0250] 图19是鉴定生成自相同靶核酸划分产物混合物液滴的引物/靶核酸划分产物的方法1900的流程图。

[0251] 在方框1910处,接收一段时间上检测器获得的数据信号。该数据信号可包括来自生成自多个混合物液滴的多个反应划分产物的信号。各混合物液滴可对应于划分产物集合且包括至少一种靶核酸的拷贝。各划分产物可包含一种或多种引物。该数据信号包括涉及划分产物中至少一种引物是否与靶核酸杂交的数据。

[0252] 在方框1920处,基于各划分产物的各信号鉴定各划分产物的杂交状态。划分产物的杂交状态可显示该划分产物中的引物是否与靶核酸杂交。特定划分产物的信号对应于特定时间。

[0253] 在方框1930处,计算具有矛盾杂交状态且包含相同引物的划分产物的量,从而获得时间函数。对时间段中多个特定时间的每一个确定量。由对应于特定时间的时间窗口内的划分产物确定特定时间处的量。

[0254] 在一个实施方式中,具有矛盾杂交状态的划分产物的量对应于具有矛盾杂交状态的划分产物中引物的量。例如,该量可以是在一个划分产物中具有阳性杂交状态和在一个划分产物中具有阴性杂交状态的引物的计数。在另一个实施方式中,该量可以是具有相同引物但具有矛盾杂交状态的划分产物的计数。

[0255] 可用时间窗口中不同划分产物的数目来标定划分产物的量。当两个划分产物包含至少一种不同的引物时,两个划分产物可以是不同的。

[0256] 在一个实施方式中,该时间窗口可被描述成多个划分产物。该时间窗口可起始于特定时间处的划分产物或围绕特定时刻的划分产物。可基于创建自混合物液滴的划分产物数目来选择时间窗口中的划分产物数目。

[0257] 在方框1940处,可在时间函数中鉴定极值。该极值可以是最大值或最小值。该极值是局部极值。可使用任何合适的确定极值的技术。例如,两个峰之间的最低值可用作最小值。

[0258] 在方框1950处,时间函数中对应极值之间出现的连续划分产物的集合被确定为对应于相同的混合物液滴。该组可一并进行分析。在一些实施方式中,当第一连续划分产物集

合对应于包含第一靶核酸的第一混合物液滴时,可基于第一集合的信号确定第一集合中的各引物是否与第一靶核酸杂交。下文提供了该确定的具体细节。随后,第一集合中的引物和第一集合中引物的杂交状态可用于确定第一靶核酸的核苷酸序列。

[0259] B. 确定集合的靶标

[0260] 图10显示根据本发明的实施方式确定混合物液滴的杂交概况。图10显示靶核酸1000的序列。测试的引物的确切序列可以是未知的,但可获得相同区域的参考序列。出于说明目的,显示了靶核酸1000的序列。

[0261] 位向量1010生成杂交概况。位向量定义为指定对应于混合物液滴的集合中各划分产物的杂交状态的任何方法。当杂交状态为真或假时,可通过位指定各值。这些位值可以是0或1,或两个数目的任意组合。仅显示了位向量1010的一部分,即集合的最初10个划分产物的值。在该实施例中,划分产物0具有阳性杂交状态而划分产物9为阳性杂交状态。

[0262] 通过显示多个划分产物中的引物来进一步显示杂交状态。引物集合1020包含CGTAGG,其存在于靶核酸900的序列中,且因此杂交状态为阳性。引物集合1030不存在于靶核酸1000的序列中。引物集合1040包含GATGCT,其存在于靶核酸1000的序列中。

[0263] 图11显示,根据本发明的实施方式,与参考杂交概况相比混合物液滴的杂交概况的比较,所述参考杂交概况与参考核酸相关。行1110显示针对集合的各划分产物测量的杂交状态。各符合对应不同的划分产物。行1140提供划分产物编号。行1120显示二进制值代表的特定划分产物的杂交状态。一些划分产物在该集合中显示,且因此不具有对应的测量值。

[0264] 阵列1130对应于已知参考序列的杂交概况。阵列1130中的各行对应于不同的参考序列,且因此是不同的杂交概况。由于各划分产物的引物是已知的,可确定某一划分产物的预期位值。可针对所有可能的划分产物确定参考序列的位向量,且因此位值不从这些杂交概况中缺失。

[0265] 列1132对应于位向量1120与参考位向量之间的差异。例如,位向量1120在三个划分产物处(0、7和8)不同于第一参考位向量。因此,该差异是3。可使用各向量之间的其他距离。参考位向量1134与位向量1120最类似;且实际上这两个位向量在集合内包含的划分产物方面是相同的。

[0266] 多种标准可用于选择一种或多种位置描述的参考物。例如,需要最低的类似性。最低的类似性可以是原始值,例如小于一个或两个差异。另一个标准可以是足以与下一最低差值区分的最低差值,例如这两个值必须相差至少N,N是预定值。

[0267] 图12显示根据本发明的实施方式混合物液滴彼此间杂交概况的比较。显示多种混合物液滴的杂交概况的集合2150。显示了集合1250的两种特定的杂交概况。行1210和1240显示划分产物的杂交状态。行1220和1230显示相应的位值。混合物液滴的位向量可针对杂交概况中常见的划分产物1260彼此进行比较。

[0268] 杂交概况之间的差异可用于聚类杂交概况。可根据对应于同一靶核酸选取给定簇的杂交概况。在该实施例中,参考杂交概况(位向量)对应于其他混合物液滴的杂交概况。

[0269] 图20是鉴定包含靶核酸的混合物液滴的方法的流程图。方法2000可使用来自本发明所述其他方法的特征。方法2000可完整或部分由计算机系统进行。

[0270] 在方框2010处,接收多个划分产物的每一个的杂交状态。各划分产物包含一种或多种引物,且各自可包含多种引物。可划分多个混合物液滴中的每一个且各自包含一种或

多种靶核酸的拷贝。划分产物的杂交状态可显示该划分产物中的至少一种引物是否与靶核酸杂交。

[0271] 在方框2020处,可根据对应于第一混合物液滴来鉴定第一划分产物集合。方法1900可用于确定第一集合。

[0272] 在方框2030处,可针对第一划分产物集合创建第一位向量。该位向量中的各值可对应于第一集合的各划分产物的杂交状态。这些杂交状态可以是二进制值或具有更多值。

[0273] 在方框2040处,该第一位向量可与多个参考位向量比较以获得相对于多个参考位向量中每一个的差值。各参考位向量可对应于不同的参考核酸且包含第一集合中各划分产物的关于参考核酸的杂交状态值。

[0274] 在一个实施方式中,对于具有不同于第一参考位向量的杂交状态的第一集合的各划分产物,筹数增加。该筹数可确定两个位向量之间的距离(例如汉明距离)。

[0275] 在一个实施方式中,各参考核酸具有已知的序列。例如,这些参考序列可对应于参考基因组。可通过确定第一集合中各划分产物的预期杂交状态来建立参考位向量。可基于划分产物中的引物来确定该划分产物的预期杂交状态。例如,第一集合中的引物是已知的,且这些引物可与参考序列相比较以确定其中之一是否杂交。

[0276] 在另一个实施方式中,这些参考序列对应于与相同靶核酸相关联的其他集合。因此,这些参考核酸可对应于其他混合物液滴中的核酸。这些位向量可以是簇。例如,可对多个混合物液滴的位向量进行聚类。可以基于第一簇的位向量的类似性鉴定包含第一靶核酸的混合物液滴的第一簇。

[0277] 在方框2050处,基于差值鉴定第一混合物液滴中的第一靶核酸。在一个实施方式中,具有最低筹数的参考核酸被选为第一靶核酸。在另一个实施方式中,当最低的筹数比下一最高筹数少至少预定量时,具有最低筹数的参考核酸可被选为第一靶核酸。例如,最低值必须比下一最低值低至少N。N的示例是2-10或10-50。

[0278] 一旦鉴定到第一靶核酸,各实施方式可在第一划分产物集合中使用这些引物以确定第一靶核酸的核苷酸序列。例如,各划分产物可包含多种引物。随后,可鉴定与第一靶核酸阳性杂交的第一划分产物集合中的引物。鉴定的引物可经组装以确定第一靶核酸的核苷酸序列。

[0279] C. 引物的杂交状态

[0280] 图13显示根据本发明的实施方式通过比较同一集合的划分产物的杂交状态来确定引物的杂交状态。划分产物1300和划分产物1310是同一集合的一部分。划分产物1300具有阳性杂交状态,且各自具有4种引物。划分产物1310具有阴性杂交状态,且各自具有4种引物。

[0281] 选取划分产物1300本身,各划分产物中四种引物中哪一种实际上与靶核酸杂交是未知的。但是,可交叉参考划分产物1310中引物的阴性状态以排除划分产物1300中阳性杂交的引物。例如,引物1315可被鉴定为具有阴性状态(即不与靶核酸杂交),因为引物1315处于具有阴性杂交状态的划分产物中。引物1315也在划分产物1300之一中,但可因为引物1315具有阴性杂交状态而被排除。显示在划分产物1300中删除具有阴性杂交状态的引物,从而将引物1303和1306鉴定为具有阳性组织状态。

[0282] 图14显示根据本发明的实施方式通过比较不同集合的划分产物的杂交状态来确

定引物的杂交状态。图14显示靶核酸1400的序列。划分产物1420来自与靶核酸1400相关的第一集合,且划分产物1440来自也与靶核酸1400相关的第二集合。

[0283] 在该实施例中,具有阳性杂交状态的划分产物中的某些引物可基于这些引物在具有阴性杂交状态的划分产物中而被排除。由于两个集合对应于同一靶核酸1400,这些引物的杂交应显示相同或相似的结果。例如,引物1445位于具有阴性杂交状态的第二集合的划分产物内,且因此引物1445可被鉴定为具有阴性杂交状态,即使引物1445出现在具有阳性杂交状态的第一集合的划分产物中时也是如此。

[0284] 图21是确定混合物液滴的引物的杂交状态的方法的流程图。方法2100使用来自本发明所述其他方法的特征。方法2100可完整或部分由计算机系统进行。

[0285] 在方框2110处,通过检测器获得数据信号。该数据信号包括来自多个混合物液滴的多个划分产物的信号。各混合物液滴可对应于划分产物集合且包括至少一种靶核酸的拷贝。各划分产物可包含多种引物。

[0286] 在方框2120处,鉴定第一混合物液滴对应于第一划分产物集合。这可如本发明所述进行。

[0287] 在方框2130处,经确定第一划分产物集合对应于第一靶核酸。这可如本发明所述进行。

[0288] 在方框2140处,基于各划分产物的各信号确定第一集合的各划分产物的杂交状态。划分产物的杂交状态可显示该划分产物中的至少一种引物是否与第一靶核酸杂交。

[0289] 在方框2150处,鉴定第一集合的各划分产物中的引物。第一集合的多个划分产物包含第一引物。

[0290] 在方框2160处,确定第一引物是否位于第一集合任何具有阴性杂交状态的划分产物中。在一个实施方式中,当第一引物处于一个或多个具有阴性杂交状态的划分产物中时,可确定该第一引物未与第一靶核酸杂交。在另一个实施方式中,当相对于具有阳性杂交状态的划分产物,第一引物处于至少特定比例的具有阴性杂交状态的划分产物中时,可确定该第一引物未与第一靶核酸杂交。该比例可以是百分比。例如,至少5%或10%的划分产物可被确定为具有阴性杂交状态。

[0291] 在方框2170处,基于第一引物是否位于具有阴性杂交状态的第一集合的任意划分产物中来确定第一引物是否与第一靶核酸杂交。其他划分产物组可被确定为对应于第一靶核酸且这些引物可用于确定第一靶核酸的核苷酸序列。

[0292] D. 使用阴性引物组装

[0293] 一旦已知何种引物与靶核酸杂交,可组装该第一引物组以确定靶核酸的核苷酸序列。在一些实施方式中,何种引物不杂交的知识也可用于组装过程中。例如,不杂交的第二引物组可用于辨析第一引物组可对齐处的模糊性。因此,第一引物(其杂交)可确定为不对齐核苷酸序列中的特定位置,其基于第二引物(其不杂交)在被置于特定位置时将对齐第一引物的所得序列。

[0294] 图16显示根据本发明的实施方式的组装过程各阶段的各阶段。在阶段1中,在重叠图中确定明确路径。与靶核酸杂交的多种引物显示彼此重叠。这些引物以明确的方式彼此重叠。

[0295] 在阶段2中,从重叠图中生成初始支架。可以看到,引物的位置及其重叠对应于初始支架。初始支架中的短划线表示模糊位置,其中具体的核苷酸是未知的。

[0296] 作为确定初始支架的一部分,可使用种序列(seed)开始并添加与种序列重叠的引物。随着种序列延伸,可鉴定不同的可能的核苷酸,其可对应于模糊位置。当所得核苷酸序列包含不杂交的引物时,可排除某些核苷酸。在这种方式中,可辨析模糊位置。可重复该过程以填充标记为短划线的模糊位置。

[0297] 在阶段2后,可保留模糊引物。这些模糊引物是与靶核酸杂交的引物,但可与初始支架中的超过一个位置对齐。这些摇摆引物可填充模糊位置,但哪种对齐位置是正确的仍是未知的。不杂交的引物(阴性引物)可用于辨析模糊引物的对齐位置。

[0298] 在阶段3中,阴性引物可用于排除模糊引物的某些对齐位置。在显示的实施例中(其不对应于阶段2中的初始支架),且模糊引物#1显示为T-A-A-AGA,其可与引物#2 TGATAA重叠以形成TGATAAAGA,但随后引物GATAAA将与靶核酸杂交。然而,GATAAA不与靶核酸杂交,且因此模糊引物#1不与引物#2重叠(即不对齐该特定位置)。因此,可排除该模糊引物的具体对齐,从而在留下正确的对齐位置中排除所有不正确的对齐位置。

[0299] 在阶段4中,可使用明确对齐缺口的阳性引物来填充该缺口(模糊位置的示例)。因此可获得更新的支架。在阶段5中,可重复阶段3和4,其中在各迭代处获得更新的支架。

[0300] 在阶段6中,比较最终组装物与参考序列。在阶段7中,可使用阴性引物来相对于参考序列确认最终组装物中的变体应答。

[0301] 图17显示使用阴性引物确认变体应答的示意图。显示对应于靶核酸的参考核酸1700。第一引物组1710是与靶核酸杂交的阳性引物。第二引物组1720是不与靶核酸杂交的阴性引物。

[0302] 这些阳性引物似乎在位置1730处显示相对于参考序列的变化。然而,引物1710的部分1712也可与模糊位置1740对齐。然而,如果两个引物1712不与模糊位置1740对齐,则摇摆位置(improbable position)1740将具有C。但是,如果模糊位置1740数据具有C,则阴性引物1722将与模糊位置1740对齐。因此,可确定两种引物1712无法与模糊位置1740对齐。因此,两种引物1712被确认与位置1730对齐,且增加了更多的位置1730处存在变异的确认性数据。

[0303] 图22是根据本发明的实施方式基于引物杂交确定靶核酸的核苷酸序列的方法2200的流程图。方法2200可使用来自本发明所述其他方法的特征。方法2200可完整或部分由计算机系统进行。

[0304] 在方框2210处,接收来自涉及多个引物与靶核酸杂交的实验的数据。该数据可包含阳性杂交和阴性引物杂交结果。例如,一些引物可与靶核酸杂交且一些引物不杂交。

[0305] 在方框2220处,可对数据进行分析以鉴定与靶核酸阳性杂交的第一引物集合和与靶核酸杂交的第二引物集合。

[0306] 在方框2230处,可基于第一引物集合组装支架核苷酸序列。该支架核苷酸序列可包含明确位置(其中列出了具体的核苷酸)和模糊位置(其中未列出具体的核苷酸)。可基于确定第一集合的第一引物不与支架核苷酸序列的第一位置对齐将第一位置鉴定为模糊位置。

[0307] 在方框2240处,可基于第二引物集合辨析至少一个所述模糊位置。在一个实施方式中,辨析第一模糊位置可包括:当以下情形时确定第一集合的第一引物不与第一模糊位置对齐,即当与对齐一致的是对齐第二集合的第二引物也与第一模糊位置对齐。例如,如果

将阳性引物与某一位置对齐将产生阴性引物也对齐的序列,则该阳性引物的对齐可被确定为错误的。

[0308] 在另一个实施方式中,辨析第一模糊位置可包括鉴定第一集合的一种或多种明确引物,具体方法为基于第二引物集合排除明确引物的对齐位置。第一集合的一种或多种明确引物可用于确定一个或多个模糊位置处的一个或多个核苷酸,从而获得更新的支架核苷酸序列。

[0309] 在一些实施方式中,组装支架核苷酸序列可包括当以下情形时确定第一集合的第一引物不与第一模糊位置对齐,即当与对齐一致的是对齐第二集合的第二引物也与第一模糊位置对齐。

[0310] 在一个实施方式中,可鉴定第一模糊引物子集。该第一子集可以是第一引物集合的子集。模糊引物可与更新的支架核苷酸序列中的多个位置对齐。可基于第二引物集合排除第一子集的第一模糊引物的对齐位置,从而将第一模糊引物转变为第一明确引物。第一明确引物可用于辨析更新的支架核苷酸序列中的第二模糊位置。

[0311] 计算机系统

[0312] 上述分析可在计算机上的软件或本发明所述的系统中进行。本发明提及的任何计算机系统都可利用任何适当数目的子系统。这类子系统的示例如图23中计算机设备10所示。在一些实施方式中,计算机系统包括单个计算机设备,其中子系统可以是该计算机设备的组件。在其他实施方式中,计算机系统可包括多个计算机设备,其各是一个子系统,具有内部组件。

[0313] 图23所示的子系统经由系统总线75互联。显示了其他子系统,如打印机74、键盘78、储存装置79、与显示适配器82偶联的监视器76等。与输入/输出(I/O)控制器71偶联的周边和I/O装置可通过任何数量的本领域已知方式(如输入/输出(I/O)端口977(如USB、FireWire[®]))连接至计算机系统。例如,I/O端口77或外部接口81(例如以太网、Wi-Fi等)可用于将计算机系统10连接至广域网(如因特网)、鼠标输入装置或扫描仪。经由系统总线75的互联允许中央处理器73与各子系统连通并控制来自系统内存72或储存装置79(例如固定磁盘,如硬盘或光盘)指令的执行以及子系统间信息的交换。系统存储器72和/或储存装置79可包含计算机可读介质。本文所述的任何数据都可从一种组件输出至另一种组件并可输出至用户。

[0314] 计算机系统可包括多种相同的组件或子系统,例如通过外部接口81或通过内部接口连接在一起。在一些实施方式中,计算机系统、子系统或设备可通过网络连通。在这种情况下,可将一台计算机作为客户端并将另一台计算机作为服务器,其中各计算机都可以是同一计算机系统的部分。客户端和服务器可各包括多个系统、子系统或组件。

[0315] 应理解,本发明的任意实施方式都可使用硬件(例如专用集成电路或现场可编程门阵列)以控制逻辑的形式和/或通过通常可编程的处理器使用计算机软件以模块化或集成化的方式来实施。本文中,处理器包括同一集成芯片上的多核处理器或者单个电路板或网络连接的多个处理单元。基于本发明的公开和教导,本领域普通技术人员应知晓并理解使用硬件以及硬件和软件的组合来实施本发明的实施方式的其他方式和/或方法。

[0316] 本申请中描述的任何软件组件或函数都可作为软件代码使用,以由处理器使用任何适当的计算机语言(如Java、C、C++、C#)或脚本语言(如Perl或Python)、使用例如常规或

面向对象的技术来执行。软件代码可作为一系列指令或命令储存于计算机可读介质上用于储存和/或传输,合适的介质包括随机存取存储器(RAM)、只读存储器(ROM)、磁性介质(如硬盘或软盘)、或光学介质(如光盘(CD)或DVD(数字多功能光盘))、闪速存储器等。计算机可读介质可以是这里储存或传输装置的任意组合。

[0317] 也可使用适用于传输的载波信号经由遵循多种协议的有线、光纤和/或无线网络(包括因特网)编码和传输这类程序。同样地,可使用这类程序编码的数据信号来建立本发明的一个实施方式所述的计算机可读介质。程序代码编码的计算机可读介质可与兼容性装置打包或由其他装置单独提供(例如通过因特网下载)。任何这类计算机可读介质可存在于单个计算机产品(例如硬盘、CD或整个计算机系统)之上或之内,且可存在于系统或网络中不同计算机产品之上或之内。计算机系统可包括监视器、打印机或将本发明所述任何结果提供给用户的其他合适显示装置。

[0318] 本文所述方法中任一种都可全部或部分地使用包括一个或多个处理器的计算机系统,可对其进行配置以完成步骤。因此,实施方式可针对经设置以进行本文所述任意方法的步骤的计算机系统,其中不同组分可能完成相应步骤或相应步骤组合。虽然以编号的步骤形式显示,但本文中方法的步骤可同时或以不同顺序进行。此外,这些步骤的部分可与来自其他方法的其他步骤的部分联用。同样,步骤的全部或部分可以是任选的。此外,任何方法的任何步骤都可使用模块、循环或用于进行这些步骤的其他手段进行。

[0319] 具体实施方式的特定细节可以任何合适方式合并而不背离本发明的实施方式的精神和范围但是,本发明的其它的实施方式可能涉及单独的方面、或者这些单独方面的特定组合的实施方式。

[0320] 以上本发明的示例性实施方式的描述是为了阐述和说明的目的。不应当离解为本发明仅限于所述的具体形式,通过阅读以上内容,可以进行许多的改良和变化。对这些实施方式进行选择 and 描述,以更好地解释本发明的原理及其实施,使得本领域其它技术人员能够更好地利用本发明的各种实施方式和各种改良,使其适用于预期的特定用途。

[0321] 除非另有明确的相反说明,否则“一个”、“一种”或“该”旨在表示“一个(种)或多个(种)”。除非另有明确的相反说明,否则使用“或”旨在表示“包括性或”而非“排除性或”。

实施例

[0322] 实施例1

[0323] 该系统的一个实施方式被用于测序具有已知突变状态的感兴趣区域中的人DNA样品。在该实验中,感兴趣的区域是BRAF基因的外显子15。

[0324] 引物混合物被设计为使得各混合物含有6个共有3'核苷酸(共有区域)和5'核苷酸的混合物。使用所有可能的6个核苷酸共有区域设计了总共4096种引物混合物。在各引物混合物中添加的5'核苷酸数目经优化以保持Tm在4096种混合物的整个组中保持恒定,如图8所示。将各引物混合物(2uM)与以下组分混合:Thermopol缓冲液1X(新英格兰生物实验室公司)、BSA 0.2mg/ml(新英格兰生物实验室公司)、抑制剂寡核苷酸0.4uM(3' Dabcyl标记的20nt寡核苷酸-IDT)、dTNP 0.2uM(新英格兰生物实验室公司)、DNA扩增子0.08uM(以生成扩增子BRaf基因的外显子15,其使用包含设计为与Dabcyl淬灭剂抑制剂杂交的5'序列的FAM标记的正向PCR引物和5'磷酸-反向PCR引物进行PCR扩增)、λ外切核酸酶0.3kU/u1(以促进

双链DNA扩增子至单链的转化)和BST 2.4kU/u1(新英格兰生物实验室公司)。此外,向4096种独特组合中的各试验引物混合物中加入以下染料(哥伦比亚生物科学公司)以促进结果读取后引物混合物的鉴定:藻红蛋白(R-PE)(在范围为0至6.7ug/ml的8种浓度之一中)、R-PE偶联的DyLight 594(在范围为0至9.6ug/ml的8种染料浓度之一中)、R-PE偶联的DyLight 633(在范围为0.4至12.5的8种浓度之一中)和R-PE偶联的Cy5.5(在范围为0至13.3ug/ml的8种染料浓度之一中)。随后将各引物混合物乳化为液滴(PCT专利公开号W02012/078710中所述的油包水乳液)并将来自4096种混合物的全部组的液滴混合在一起并在34℃下孵育30分钟。使用光学系统(参见PCT专利公开号W02012/135327和W02012/135201)读取试验结果。

[0325] 在若干实施方式中,这些引物划分产物含有超过一种不同的引物。一个示例性实施方式描述于图6A,且称作多重引物划分产物。在该实施方式中,该划分产物整体生成阳性或阴性结合信号,如同每个划分产物中单个引物类型的情况中那样。然而,差别是如果该划分产物内的所有引物都阴性结合则该划分产物仅生成阴性结果,而如果其中含有的任意引物阳性结合则划分产物整体生成阳性结果。该动态示于图6B中。

[0326] 试验引物多重化(每组多达5种引物)

[0327] 该系统的一个实施方式被用于测试多种多重化策略,包括其中各划分产物中存在5种引物的策略。在多孔板中将八种试验引物合并为19组(其范围为每组1-5种试验引物)(对添加的各试验引物使用2uM浓度,范围为对每组1-5种试验引物分别为总共2uM至10uM)。随后将各试验引物组与以下组分混合:Thermopol缓冲液1X(新英格兰生物实验室公司)、BSA 0.2mg/ml(新英格兰生物实验室公司)、抑制剂寡核苷酸0.4uM(3' Dabcyl标记的20nt寡核苷酸-IDT)、dTNP 0.2uM(新英格兰生物实验室公司)、DNA扩增子0.08uM(以生成扩增子KRas基因的外显子2,其使用包含设计与Dabcyl淬灭剂抑制剂杂交的5' 序列标签的FAM标记的正向PCR引物和5' 磷酸-反向PCR引物进行PCR扩增)、λ外切核酸酶0.3kU/u1(以促进双链DNA扩增子至单链的转化)和BST 2.4kU/u1(新英格兰生物实验室公司)。此外,向独特组合中的各试验引物组中加入以下染料(哥伦比亚生物科学公司)以促进结果读取后试验引物组的鉴定:PE偶联的DyLight 594(在范围为0至6.3ug/ml的针对各引物组的6种染料浓度之一中)和PE偶联的Cy5.5(在范围为0.8至6ug/ml的针对各引物组的4种染料浓度之一中)。随后将各试验引物组乳化为液滴(PCT公开号W02012/078710中所述的油包水乳液)并将来自19种混合物中每一种的液滴混合在一起并在37℃下孵育30分钟。使用光学系统(参见PCT专利公开号W02012/135327和W02012/135201)读取试验结果。八种引物中的三种被设计为与DNA靶标杂交(匹配试验引物)且其他五种在DNA靶标内不具有互补序列(错配试验引物)。

[0328] 结果:

[0329] 试验结果见图7A。结果显示为与错配试验引物信号的标准偏差距离($\frac{[荧光]-[错配试验引物的平均荧光]}{[标准偏差错配试验引物]}$)。合并试验引物不干扰试验性能且结果在整个引物试验组中保持恒定。含有匹配试验引物的组在所有组合中都生成恒定的高信号且含有错配试验引物的组生成恒定的低信号,与合并的组中包含的试验引物数目无关。

[0330] 试验引物多重化(每组多达10种引物)

[0331] 该系统的另一个实施方式用于测试其中各划分产物中含有10种引物的多重化策略。将17种试验引物合并成30组,范围为每组单种试验引物至10种合并的试验引物(对各试验引物使用2uM浓度,为对每组1种试验引物至10种试验引物分别为总共2uM至20uM)。进行

的实验与5引物组实验大致相同,不同之处在于:

[0332] 试验引物组各自包含多达10种试验引物。

[0333] 以八种水平之一(浓度范围为0至9.6ug/ml)向试验引物组中加入PE偶联的594染料并以六种水平之一(浓度范围为0.8至10.8ug/ml)向试验引物组中加入PE偶联的Cy5.5以生成包含在实验中的30种独特的特征染料“条码”。

[0334] 结果:

[0335] 试验结果见图7B。如同5试验引物组实验那样,合并试验引物不干扰试验性能且结果在整个引物试验组中保持恒定,其中含有匹配试验引物的组在所有组合中都生成恒定的高信号且含有错配试验引物的组生成恒定的低信号,与合并的组中包含的试验引物数目无关。

[0336] 讨论:

[0337] 上文图7A和7B中讨论的实施方式每个多重划分产物含有5或10种不同的引物。此外,那些实施方式中所有多重划分产物都含有相同数目的引物。

[0338] 其他实施方式允许少至2种引物/多重划分产物或多至20种引物/多重划分产物。其他实施方式还允许文库中的划分产物含有不同数目的引物且任何给定的划分产物都允许含有少至1种引物(且因此是单重性的)至多至20种引物。通常,这允许围绕引物审查过程的经验结果优化系统性能或回答关于样品的特定问题。例如,如果引物碰巧在多重化时生成较弱的信背比,我们可使用仅含有该引物的单重划分产物补充划分产物文库。

[0339] 上述5重和10重实验中使用的实施方式将各引物置于多个划分产物中的多种引物组合中。这么做主要是辅助下游反卷积过程(在下文中详述)。回到图6B所示的实施例,如果引物编码770(其对靶标为阴性)仅属于划分产物“B”,则由于该靶标的假阳性,我们不能将其排除,因为每次我们遇到该靶标时其都为阳性。需要给予引物770至少第二次机会以在靶标中生成其真实的阴性结果,从而使该结果从划分产物“B”中其姐妹引物的结果中被正确地反卷积。

[0340] 一些实施方式对于必须出现的多重组合的最小数目施加严格限制。该最小值在不同的实施方式中在2至20的范围间变化,且通常(但不必然)取决于文库中整体多重化程度。例如,一个实施方式包括一种多重化划分产物文库,其需要各引物出现在至少10种不同的组合中,所述组合各自包含10种引物。这类实施方式尝试平衡任何单个引物将不依赖于靶序列被正确反卷积的可能性。

[0341] 其他实施方式进一步限制划分产物文库的组合并需要不存在两个引物组合含有超过最大数目的共有引物。两个组合之间相交的最大数目可在不同实施方式之间在1至4之间变化。该最大阈值不依赖于文库中多重化程度而变化。

[0342] 其他实施方式允许随机选择划分产物内的引物组合。其他实施方式还允许在两个待随机的划分产物之间维持共有的引物数目。其他实施方式允许一些引物在同一划分产物种出现超过一次(有效地处于总浓度的较大百分比)。最后一种允许有助于补偿各引物之间结合效率的潜在差异。例如,一个实施方式允许10重划分产物中的一个引物包含该划分产物允许的总引物浓度的50%。这将为各10%浓度处的额外5种引物或各20%浓度处的2种其他引物留下足够的引物浓度。允许相同文库中每个划分产物中可变数目引物的前述实施方式中的一些实施方式也以该方式改变每个引物的浓度分配。分配给划分产物中任一种引物

的总浓度的百分比将在少至0.4%至多达100%之间变化。

[0343] 其他实施方式对于多重划分产物内含有的引物序列之间的关系施加限制。例如，一个实施方式不允许同一划分产物内的任意两种引物在其端部重叠超过一些最大数目的核苷酸。该最大数目在各实施方式中在1至8之间变化。这么做主要是避免任何背景偏见(contextual bias)(其是否是杂交动力学的函数,或由许多其他文库设计参数偶然导入,或任何其他不可预见的背景偏见来源)。许多实施方式对于划分产物内含有的引物没有这类限制。

[0344] 通过杂交处理测序的许多文献涉及靶标的从头组装。特别感兴趣的是可明确组装的靶标复杂性的最大化同时限制辨析序列所需杂交引物的长度和量。本发明经类似限制。如Pevzner等所示,可由K聚体的集合明确组装的序列长度的上限是2K。例如,设置K等于6,本发明需要含有4096种六聚体序列的完全互补物的引物文库明确地组装长度不超过64个核苷酸的靶标。

[0345] 然而,在上文详述的实施方式中,使用4096种六聚体引物序列对长度超过100个核苷酸的扩增子进行测序。这是可能的,因为其不是从头测序应用。PCR引物设计为针对感兴趣的区域捕获和富集。这些设计基于取自Genbank的野生型参考序列。该参考序列通过下游的组装和变体应答软件模块进一步应用以组装靶标的公认性序列(描述于下文)。

[0346] 该类型的参考辅助的组装还在文献(参考文献)中被研究且通常称作“再测序”且大部分所谓的下一代测序应用遵循该模型。参考辅助的组装允许使用给定长度的引物阐明复杂得多的靶标。其基于以下假设,即大部分靶标材料与参考序列类似。此时,标准变成引物与参考序列模糊对齐的可能性。换言之,在所需长度的靶标中给定长度的引物具有超过一种精确序列匹配的可能性。

[0347] 对于BRAF实验中使用的实施方式,六聚体探针的完全互补物的总体复杂性高于100bp靶标所需,因为其是参考辅助的。六聚体的完全互补物足够复杂以对长度最多1千碱基的靶标进行再测序。

[0348] 存在允许多个指定碱基的其他实施方式,所述指定碱基在少至4个核苷酸至多达24个核苷酸之间变化。这些实施方式中的一些具有引物,所述引物全部具有相同数目的指定碱基,如同BRAF实验中完成的那样。

[0349] 另一个实施方式允许引物划分产物文库中的引物在其指定碱基数目方面变化。例如,在一个实施方式中,含有六聚体的完全互补物的划分产物文库附加有具有较多数目的指定碱基的引物,这些引物具有跨较长的双核苷酸重复段(例如“ATATATATAT”)或均聚物段(例如“TTTTTTTTTT”)的能力。可如此操作以代替成本高得多但仍可行的实施方式,所述实施方式中的划分产物文库由11个核苷酸的指定碱基序列的完全互补物组成。在其引物之间存在指定碱基数目变化的实施方式可包括4至24个指定核苷酸的完全范围。

[0350] 一些实施方式还利用我们称作“ T_m 标准化”或“ T_m 平衡”的技术,其中向寡聚体的5'端添加简并碱基直至该寡聚体的预测解链温度落入可接受的范围内。如此操作的原因在于微流体工作流使所有引物划分产物在试验孵育期间处于同一热条件下。一些实施方式允许寡聚体设计物的预测 T_m 落入用于孵育的通用热条件外 $\pm 20\%$ 。其他实施方式更严格且允许少至 $\pm 5\%$ 。另一个实施方式尝试一些范围内的平衡,例如 $\pm 10\%$,但仅针对大部分寡聚体。例如,在一个实施方式中,至少70%的寡聚体的 T_m 必须落入靶范围的 $\pm 10\%$ 内,而剩

余寡聚体设计为尽可能接近靶范围。

[0351] 其他实施方式允许引物寡聚体除仅在5'端外还可在其他位置中具有简并碱基。这有效地导入缺口或“通配符”碱基,将其间插在寡聚体的简并碱基之间。这类“具有缺口的引物”的示例示于图15。这些具有缺口的引物允许组装和变体应答软件模块辨析更长的靶标而无需增加简并碱基的数目(下文中详述)。因此,被导入的缺口数目可在1至6之间变化并可4至24之间任何数目的简并碱基添加至寡聚体。一些实施方式含有同一划分产物文库中具有缺口的引物和无缺口的引物的混合物。一些实施方式还改变引物(带缺口的和无缺口的)中简并碱基的数目。这赋予在潜在测序环境中对所需应用调节含量偏好(或缺少含量偏好)的能力。一些应用需要从头合成的测序元件以及参考辅助的再测序的混合物且还可能需已知且单个事件的非常靶向的试验。

[0352] 实施例2

[0353] 参考图1中的微流体示意图,测序过程中软件获取其输入值的点位于微流体 workflow 中点“E”处的光学检测器处。软件的输入值的形式为与主要检测设备相连的CCD传感器获取的图像。该软件处理CCD图像以定位对应于流过检测器的引物划分产物的图像区域。在一个实施方式中,该软件使用分水岭分割方法以将图像分割为对应于引物划分产物的区域。替代性实施方式可使用不同的图像分割方法,例如边缘检测法、团点检测法等。其随后可从图像中的各划分产物区域中提取光谱概况。

[0354] 在一个实施方式中,该信号处理模块随后使用广义最小平方方法对划分产物荧光概况进行光谱分解。该步骤生成各划分产物内含有的各荧光染料组分的荧光强度。条形码分配模块随后使用这些染料强度来计算标准化的欧氏距离(Euclidean distance)以发现对应于正确划分产物性质的染料簇。在一些实施方式中染料簇的性质是在实验前已知的并使用简单的方格位置寻找。在其他实施方式中,使用若干种方格拟合方法之一在实验期间将性质方格拟合至染料簇。

[0355] 在上文详述的BRAF实验中,存在4096种不同的染料条码簇且其通过聚类方法发现。在该实施方式中,使用DBSCAN算法聚类染料强度向量。其他实施方式使用不同的基于强度的聚类方法的变体以及模板匹配方法,其中在运行前已建立染料簇质心的预期图谱。来自BRAF实验的染料簇示于图3A和3B。

[0356] 实施例3

[0357] 划分产物试验应答

[0358] 图4A显示对用于前述BRAF实验的引物划分产物进行测序所收集的原始试验强度。这些荧光强度未被标准化或以除通过广义最小平方分解从光谱概况中提取以外的其他方式转化。

[0359] 从图中可以清楚地看到,两个划分产物群体(试验阳性和试验阴性)是线性可分离的。可使用多种方法来划分这两种群体。在该实施方式中,使用朴素贝叶斯分类器来确定哪些划分产物是阳性的和哪些是阴性的。小、随机的原始试验强度子集被用作分类器的训练集(training set)。在该实施方式中,20%的划分产物被用作训练集。在其他实施方式中,该训练集代表不同百分比且可以少至5%或多达50%。

[0360] 图4B根据简并碱基判定试验的稳健性。该图含有与图4A相同的散点图。然而,在该图中,挑出对应于含有具有明确的3个简并碱基的引物的划分产物的簇。该散点图所捕获的

剩余划分产物含有少于3个简并碱基且具有0至2个简并碱基的引物。这表明,在最坏的情况下,在阳性或阴性结合状态下,简并碱基数目和所得引物试验强度之间都没有关系。

[0361] 时间分割

[0362] 从图2所示的条码鉴定器和试验应答器模块的图中,观察到连续的引物划分产物图像流。在时间上不存在明确的边界,所述边界划分引物划分产物的一个PCR划分产物价值的末端和下一个的起点。该时间关联必须来源于原始数据(即引物性质和试验应答的流)。

[0363] 图9显示被称作时间分割的过程。为此,在划分产物的滑动窗口内以时间的函数的形式追踪自我矛盾率。例如,在用于测序BRAF靶标的实施方式中,检查2400个引物划分产物的滑动窗口以观察有多少被观察到超过一次的引物划分产物实际上相对于其试验状态与其自身矛盾。该比率在图9中作图为时间的函数。该比率自然状态下显示一系列局部最小值和最大值。局部最小值是各引物测量值的PCR划分产物价值之间的时间边界的预测值。在图9中,垂直条标记局部最小值且因此以时间标记PCR划分产物边界。在一个实施方式中,一旦通过该模块使一组引物划分产物在时间上关联,其在剩余分析过程中保持不变。在其他实施方式中,该关联被允许打断,前提是可通过交换时间边界处的引物划分产物来生成更强的假设。

[0364] 实施例4

[0365] 划分产物组的聚类

[0366] 软件系统的设计是使其能够支持一组关联的假定,即取自特定PCR划分产物的任何给定的引物划分产物测量组:

[0367] a) 不保证是整个引物划分产物文库的完整取样

[0368] b) 不保证以任何特定顺序出现

[0369] c) 不保证在下一测量扩增子(即靶核酸)时确切地重复

[0370] d) 通常在靶向样品中同一感兴趣区域的PCR划分产物文库中存在多个PCR划分产物。

[0371] 基于这些假说,该软件设计能够生产完整的图,其不仅是扩增子的图,而且是待随时间累积的样品内等位基因变体的亚群体的图。这通过鉴定最可能的感兴趣的基因组区域来实现,所述感兴趣的基因组区域对应于手边的测量组。我们参考以映射形式鉴定对应于扩增子的参考序列的过程。

[0372] 该映射过程开始在测序轮次开始前,具体方式为每条被靶向的扩增子组中的参考序列使用一条条目来建立查找表。可通过简单的评分机制来进行鉴定过程,其中组中各感兴趣的区域各参考序列分配为一个评分向量,其中引物划分产物文库中的每个划分产物一个元素。各向量初始化为含有一(1)(其中对应的划分产物预期生成阳性结合事件)和零(0)(其中对应的划分产物预期生成阴性结合事件)。该表示于图10。

[0373] 当下一划分产物测量组通过软件系统时,该测量组被转化为向量,其结构与各参考序列所分配的那些相同。这些向量元素被类似地初始化为1(其中出现阳性)和0(相反情况)。

[0374] 随后将该测量向量针对各参考向量进行比较并确定评分。在一个实施方式中,计算当前时间分割的组中观察到的引物划分产物相对应的各向量中的位之间的汉明距离。该参考序列显示选择最大评分作为最可能的起始区域,且向其分配该测量组用于分析工作流

中的未来步骤。该评分过程示于图11。

[0375] 其他实施方式使用替代性参考序列查找表,其表条目包括参考序列中引物情况的计数。其他实施方式包括原始参考查找表的二级表和指数,其具有与参考表条目的特定子集的预计算距离度量。这些二级指数允许映射过程期间更快的查找时间。

[0376] 在另一个实施方式中,提供了相同鉴定方法的变体以在不存在参考序列的从头开始的应用中完成类似的任务。该方法变体示于图12。在该实施方式中,在一轮次的过程期间动态建立参考向量的表。这些向量通过与任意引物划分产物测量组相同的法则初始化。各新测量组与所有先前的测量组进行比较。评分度量可以是与上文所用相同的汉明距离,但在该情况中其用作测量值凝聚聚类目的的距离度量(凝聚聚类的参考)。

[0377] 我们选取与特定参考序列或测量质心相关的所有测量值作为“读取堆积(read stack)”,其借用自所谓下一代测序领域的术语。一旦从样品中提取所有测量值,可在读取堆积内进行进一步的聚类以鉴定不同的等位基因亚群体。该亚聚类过程使用与上文从头开始的情况中使用的凝聚聚类步骤相同的评分机制。

[0378] 从图中清晰可见,所有映射和聚类步骤都可发生而不必反卷积各多重化划分产物中含有的引物。这是一种有用的能力,因为根据应用可能存在对计算资源和运行时间的严格限制。从效率的角度来看,能够在反卷积之前进行尽可能多的分析 workflow 是高度有利的。

[0379] 实施例5

[0380] 多重化划分产物的反卷积和阴性证据

[0381] 在分析期间以若干方式通过分析软件利用阴性证据。已在上文计算适当映射评分的上下文中讨论了映射过程中阴性数据的使用。然而,在映射和聚类过程期间,出于效率的目的,阴性数据维持于划分产物而非特定引物的粒度处。存在阴性证据的其它应用,包括在各引物的较精细粒度处的那些。

[0382] 如上文引物选择过程期间所述的那样,划分产物内引物的多重化天然地(但仅暂时地)导入人工假阳性。使用来自图6B的实施例,在通过主要检测器测量划分产物“A”时导入了确切的9种假阳性引物。理想情况下,所有这9种假阳性都将被排除在考虑之外而不影响手中样品的最终变体应答和等位基因计数。这通过使用获自全阴性划分产物的真阴性证据来实现,即这些划分产物中的每条引物都与靶扩增子阴性杂交。真阴性证据用于排除假阳性,即实际上不杂交但处于具有至少一种杂交引物的划分产物中的引物。

[0383] 为能够正确地进行该排除,对应于样品中同一事件(即某一扩增子的同一克隆群体或同一等位基因)的全阳性和全阴性划分产物之间必须存在正确的关联。

[0384] 由于某一划分产物内的引物无法取消彼此的阳性,分析软件使用该轮期间不同时间处测量的其他划分产物。这需要该软件将划分产物测量与扩增子的同一内源性情况(即同一PCR划分产物中获得的测量)以及同一等位基因变体的情况相关联。

[0385] 第一种类型的关联(同一特定克隆群体的测量之间)通过图9所示前述时间分割过程进行。一旦对一组引物划分产物测量完成该过程,其形成在本文中称作“集合”的群体,其在整个剩余的分析过程期间维持不被打破的状态。该关联使我们能够使用已知存在于全阴性划分产物中的引物作为全阳性划分产物所导入假阳性的对抗手段。该过程示于图13。来自阴性划分产物的各引物都有效地排除阳性划分产物中发现的匹配引物的假阳性情况。

[0386] 第二种类型的关联(代表同一扩增子或等位基因的集合之间)通过图10、11或12所

示和上文所述的映射和聚类过程进行。排除过程与图13所示集合内类似物中的过程相同。然而,在这种情况下,允许从一组划分产物中获取的阴性引物排除其匹配情况(其是其他组中阳性引物划分产物的成员)。这种类型的排除示于图14。

[0387] 预期一个集合可代表较大划分产物文库的足够小的随机样品,从中我们无法保证观察到阳性划分产物导入的各假阳性的匹配阴性引物。在这类情况中,第一种关联类型生成针对该集合的不完整的反卷积。第二种关联类型(集合之间)可保证在一定程度上补偿该潜在的对抗手段缺少,其生成足够精确的变体应答。该方法通过增加代表任何感兴趣的区域中PCR划分产物文库中PCR划分产物数目来保证。也就是说,增加感兴趣的同一区域的重复测量的预期数目。这与文献中“覆盖深度”的概念类似。实际上,该系统通过将覆盖深度设置为适当水平来赋予其本身更多的机会观察先前假阳性测量的阴性对应物。该项能力简单地扩大我们可以提取的划分产物测量组的数目以进行图14所详述的组间消除。

[0388] 这是该系统的一种有用的能力:通过将引物划分产物测量组通过聚类与彼此直接关联(从头开始的情况)或通过映射至参考物(野生型序列)间接关联来将其自身假阳性测量的反卷积延迟至过程的较后阶段的能力。

[0389] 实施例6

[0390] 变体应答和组装

[0391] 在对任意多重划分产物进行反卷积并将所有引物划分产物组针对彼此或参考序列进行聚类后,该分析过程的下一阶段是调用组装软件模块和应答变体。

[0392] 在上文详述的BRAF实验中,相关实施方式靶向参考辅助的再测序应用技术。同样地,在组装靶序列的公认序列中使用参考序列。该过程示于图5。

[0393] 图5描述了针对代表BRAF基因的外显子15的扩增子通过组装软件进行的参考辅助的组装过程。野生型序列“D”被用作对比组装过程的参考。该图中,仅描述了全部参考序列的子序列。生成阳性试验应答的引物组(组“A”和“E”)被用于相对于野生型初始化潜在突变的假设空间。阳性引物“G”被发现在1的编辑距离内在靠近组“E”的位置处与参考序列对齐,而以较大的2的编辑距离在靠近组“A”的位置处对齐。然而,通过软件发现组“A”的一致性序列生成的假设正确的可能性大得多。该软件使用阳性组“E”与组“H”的阴性结合引物合并以在参考物的区域中形成一致性序列,该区域匹配野生型且与引物“G”假设的G/A错配相比正确的可能性高得多。对该靶标应答的最终一致性序列列于序列“I”且含有来自样品的正确应答的2-核苷酸取代,如桑格测序方法所示。这显示参考辅助组装的基本步骤;在一些编辑距离阈值内对各引物寻找针对参考序列的对齐,使用与其他阳性和阴性引物的重叠关系来排列候选物对齐并对参考序列中的每一个碱基位置都如此操作,随后在参考序列的各位置处选择最大可能性的假设。如果任何引物都被排除在赢得假设性公认序列的最终集合外,则其可与其他引物比较以尝试根据参考物进行潜在插入物的无参考组装。

[0394] 在重新进行的应用中,将聚类的引物划分产物组加入组装模块。该组装过程示于图16。在该图中,假定了含有遵循缺口方案混合物的引物序列的引物划分产物文库的总体实施方式。

[0395] 在一个实施方式中,该过程的第一阶段旨在发现在该过程的后续阶段中被填充的初始支架。该阶段使用德布鲁因图遍历方法(De Bruijn graph traversal method)实现,其类似于Velvet组装器使用的方法。具有阴性试验状态或“阴性证据”的引物被用于对抗该

阶段期间德布鲁因图中假的分支点或边缘。在一个实施方式中,仅在阴性证据中没有矛盾的边缘被允许保留在图中。在其他实施方式中,考虑阴性引物(获自试验应答模块)下试验应答的置信评分。

[0396] 在图16所述同一实施方式中,该过程的下一阶段尝试填充初始支架中的任何缺口。此外,利用阴性证据对抗任何潜在的假填充物。同样地,如果合并序列结果形成的一致性序列与任何阴性证据矛盾,则不允许任何引物填充间隙。该间隙填充过程是迭代的和顺序依赖的,表示所需的迭代数目取决于选择引物填充间隙的顺序。在一次迭代中模糊的引物可在展开性组装的后一阶段中变得明确。

[0397] 最后,一旦所有阳性引物贡献于组装或被阴性证据不可避免地否定,则在数据读数中报告最终组装。如果提供了参考序列,则使用成对对齐方法(如Smith-Waterman方法)比较最终组装物与参考序列以发现根据该参考序列的假设性变体。如果在这种情况下发现变体,则阴性证据可用作那些潜在变体的平衡物,正如其用于上述比较性再测序过程那样。

[0398] 应理解,本文所述的实施例和实施方式仅用于说明目的,本领域技术人员应了解据此作出的各种修饰或改变,且它们包括在本申请的主旨和范围以及所附权利要求书的范围内。本文引用的所有发表物、专利和专利申请通过引用全文纳入本文以用于所有目的。

微流体工作流

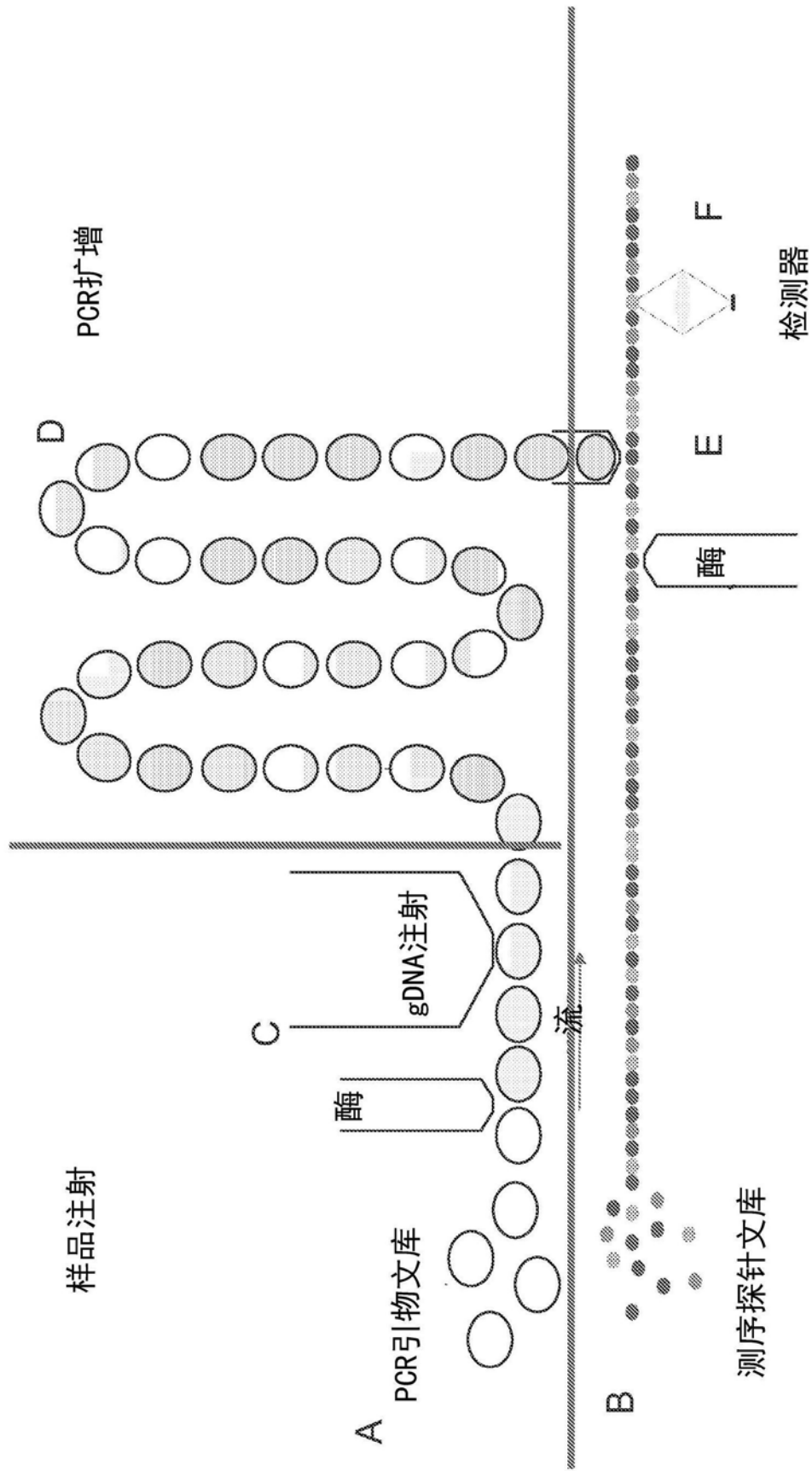


图1

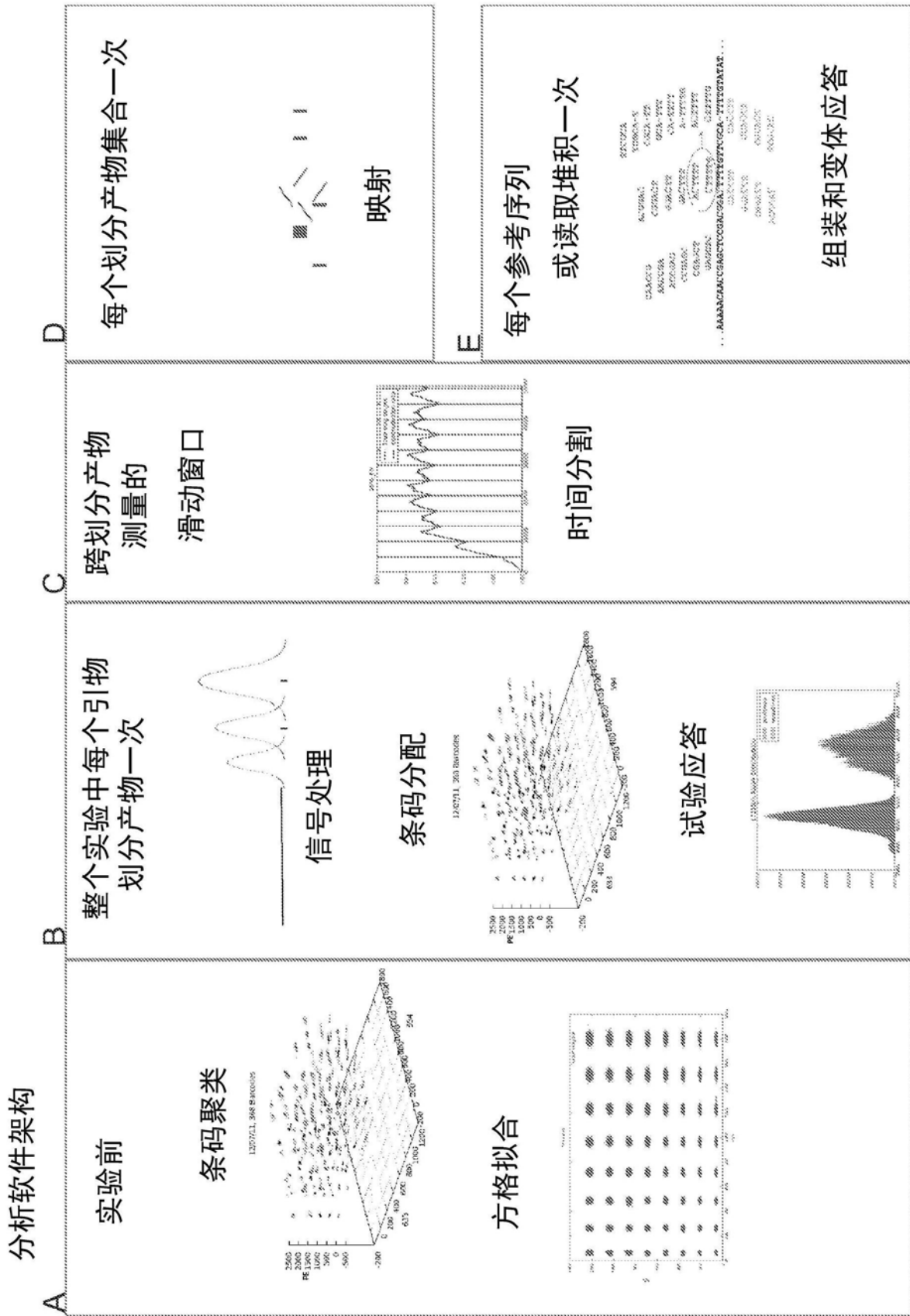


图2

四维染料概况聚类

BRAF全文库 - 染料比率

标记
 标记
 标记为 (行号)
 着色
 0
 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 ...
 所有值

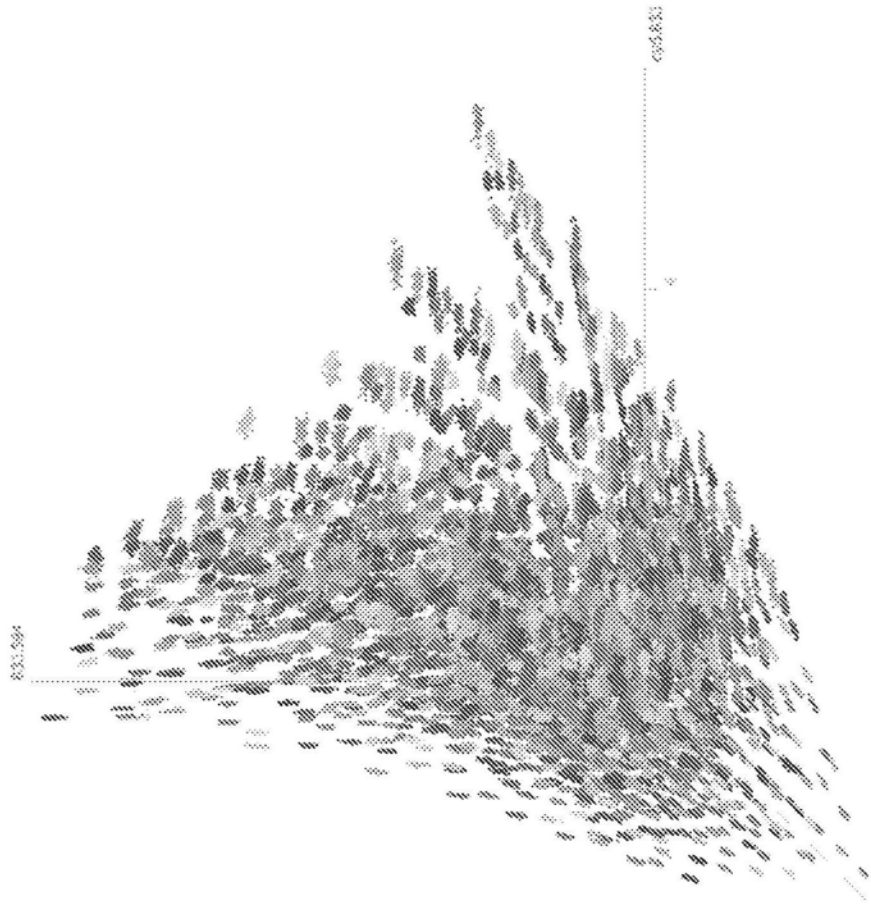


图3A

四维染料概况 - 根据试验强度着色

全文库 - 染料比率 + 试验

- 标记
- 标记
- 标记为 (行号)
- 着色
- 最大 (4711.40)
- 最小 (1864.86)
- 所有值

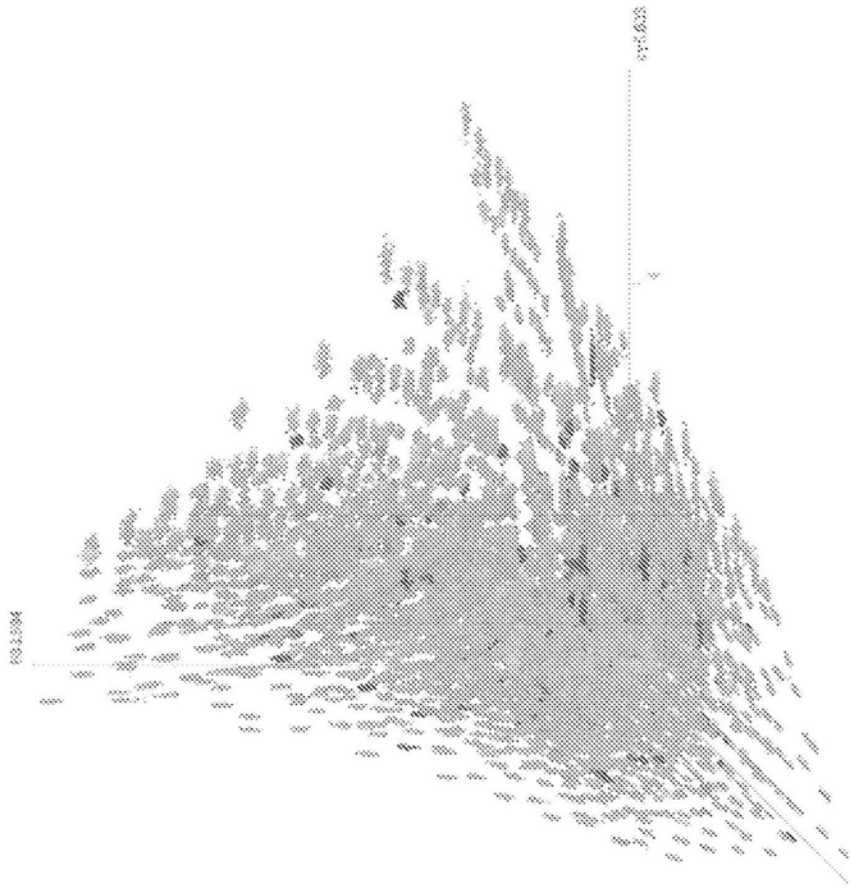


图3B

BRAF靶标的参考辅助组装

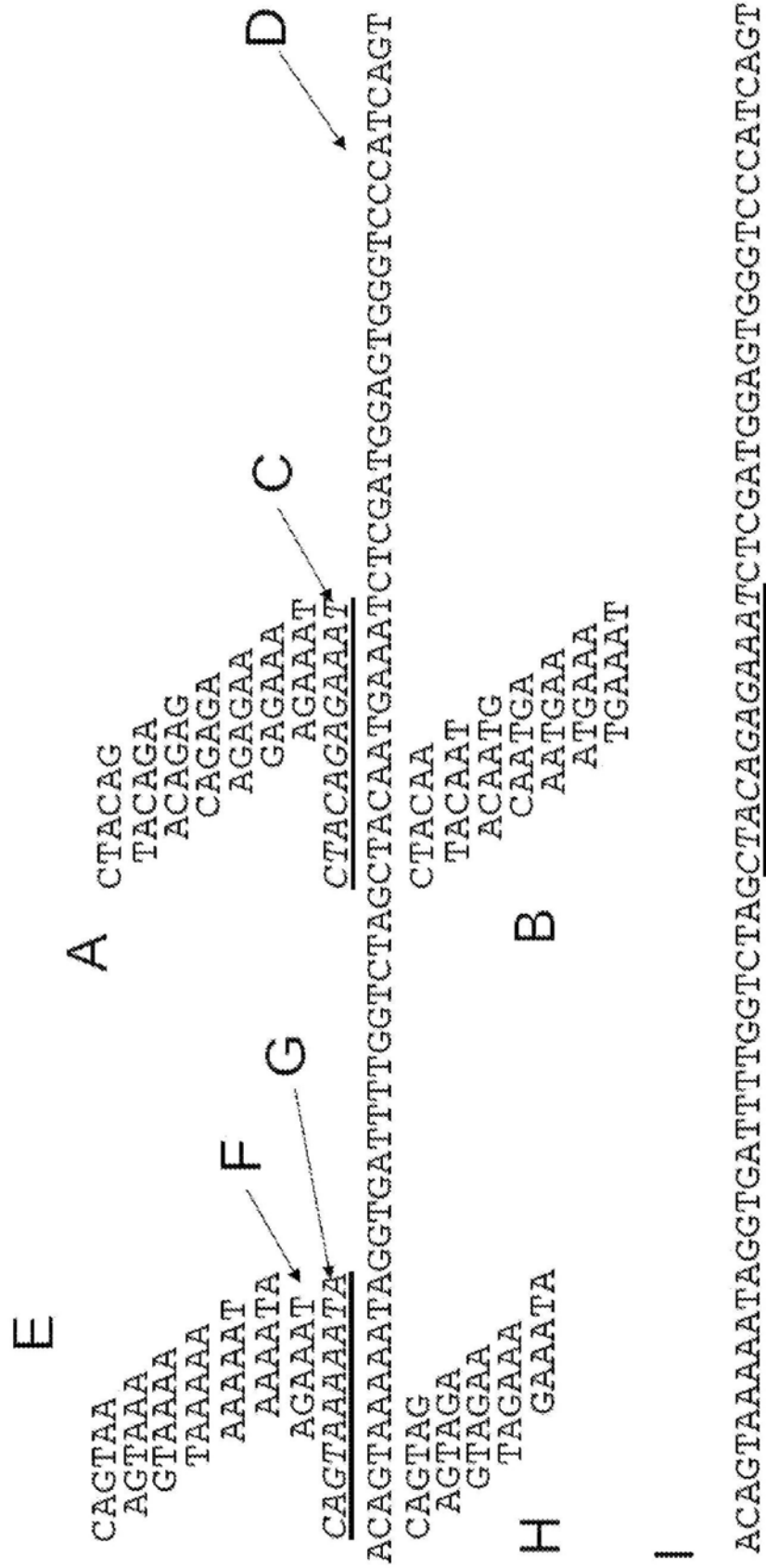


图5

多重化引物划分的实施例

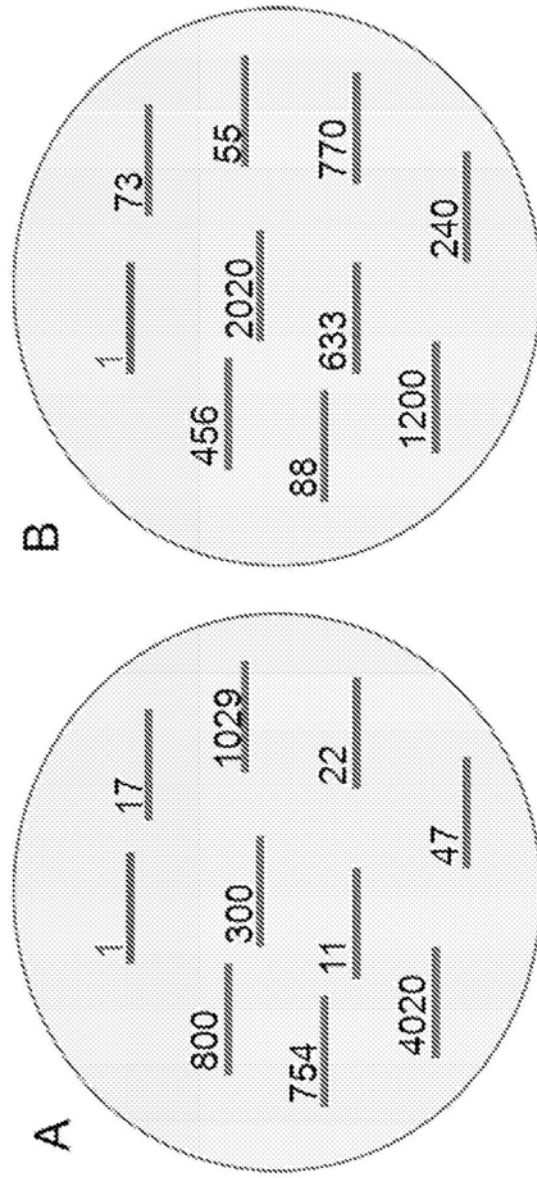


图6A

多重划分产物的阳性和阴性结果

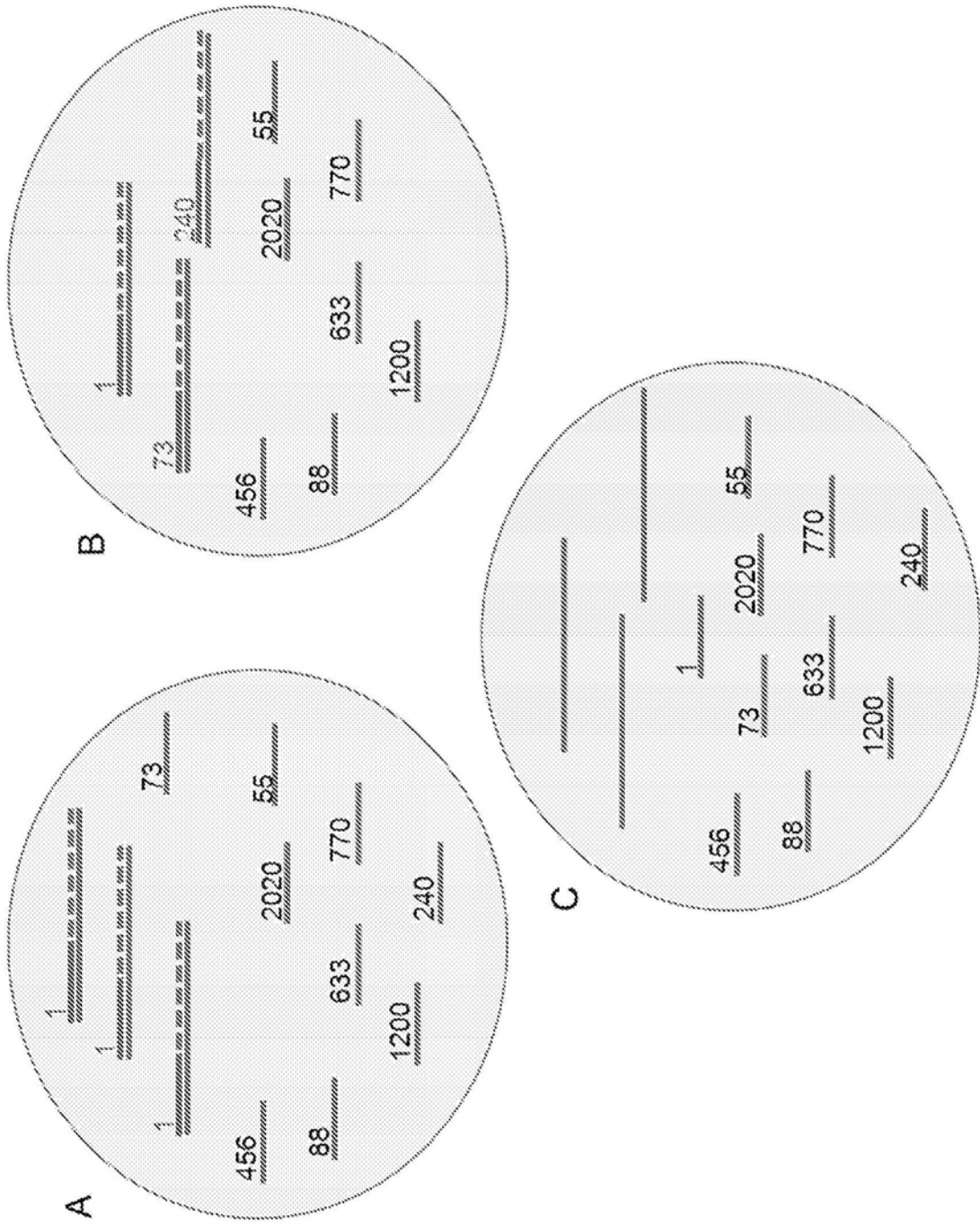


图6B

引物寡核苷酸中的简并碱基

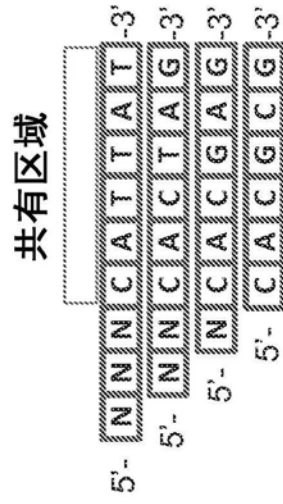


图8

基于矛盾率的时间分割

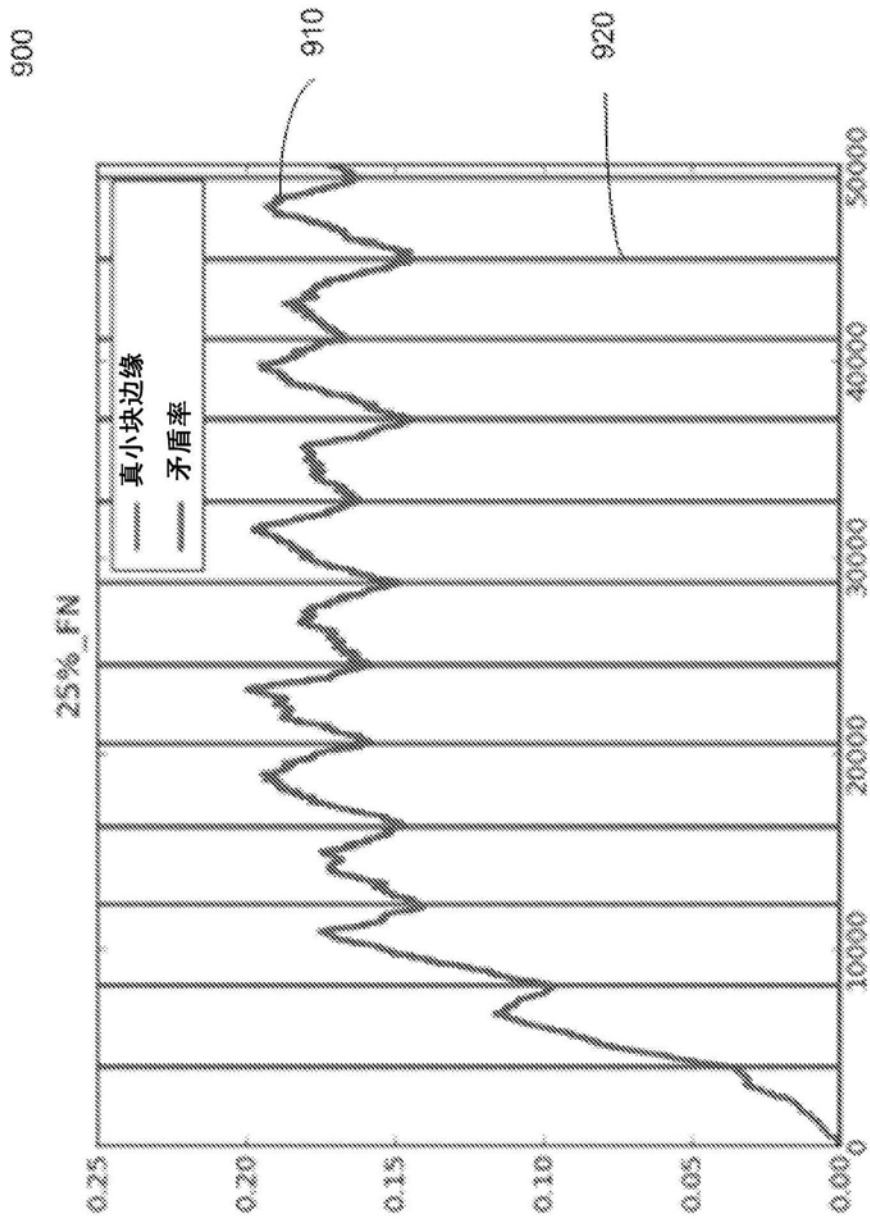


图9

参考序列的划分产物向量

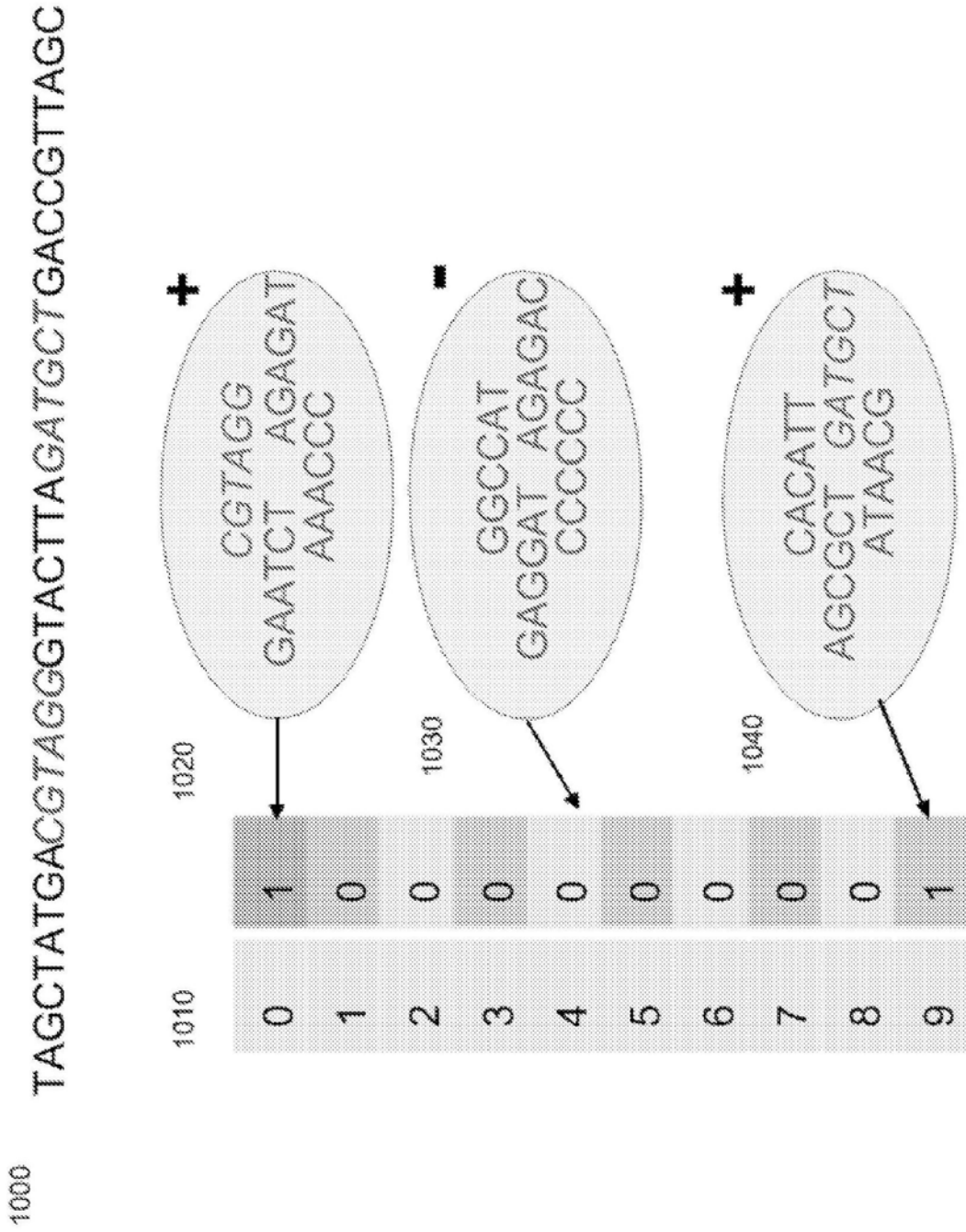


图10

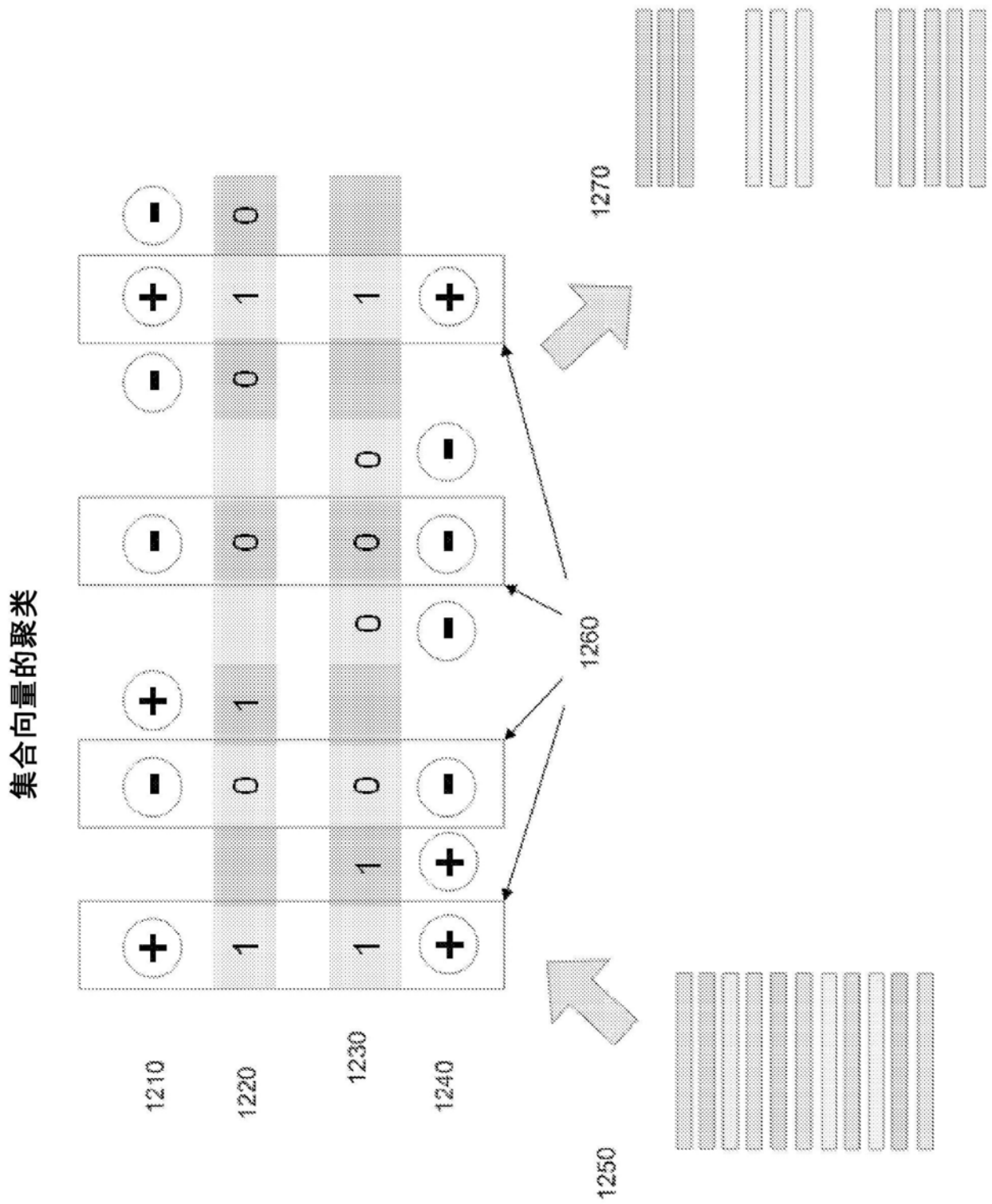


图12

集合内的否定

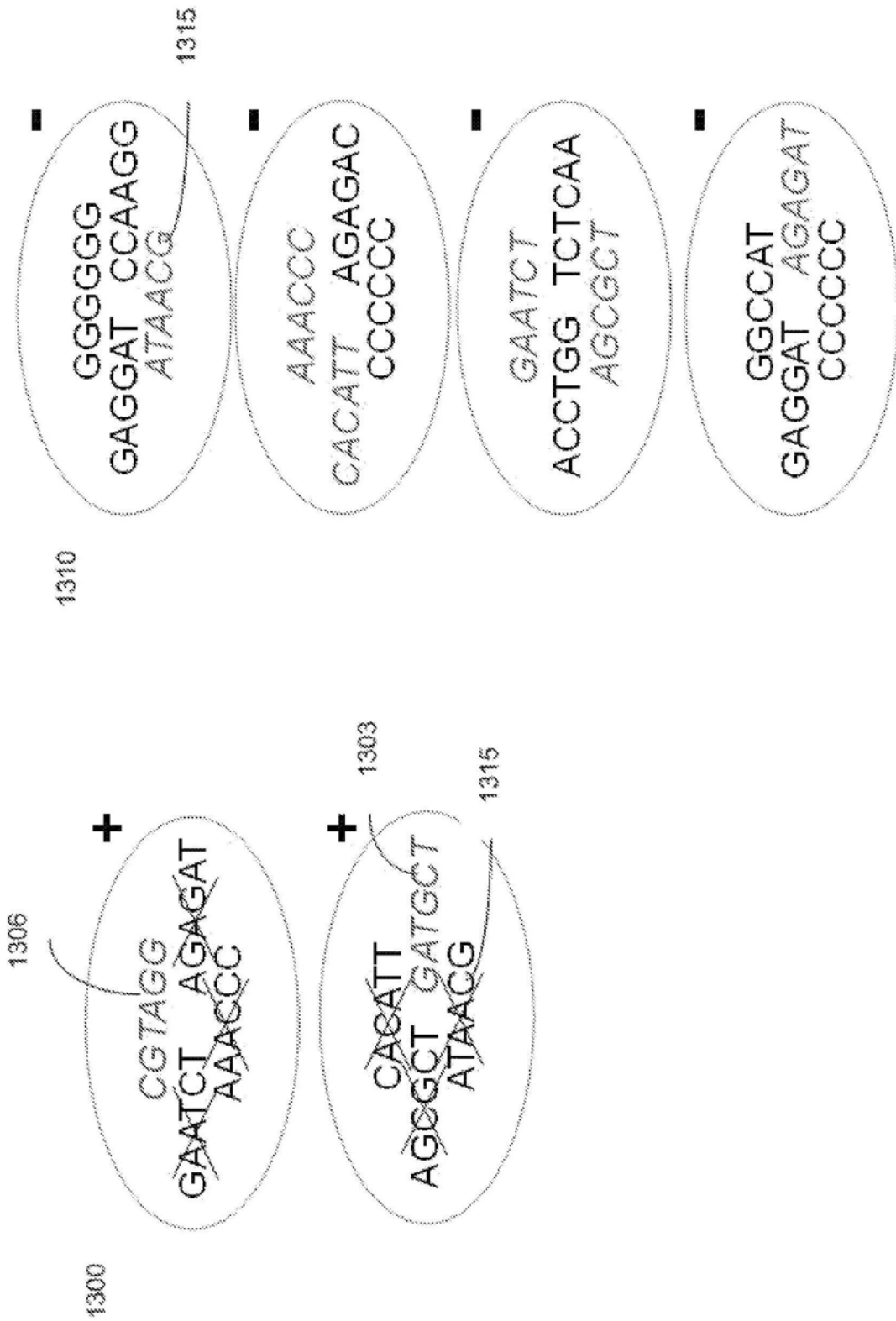


图13

显示具有缺口的引物

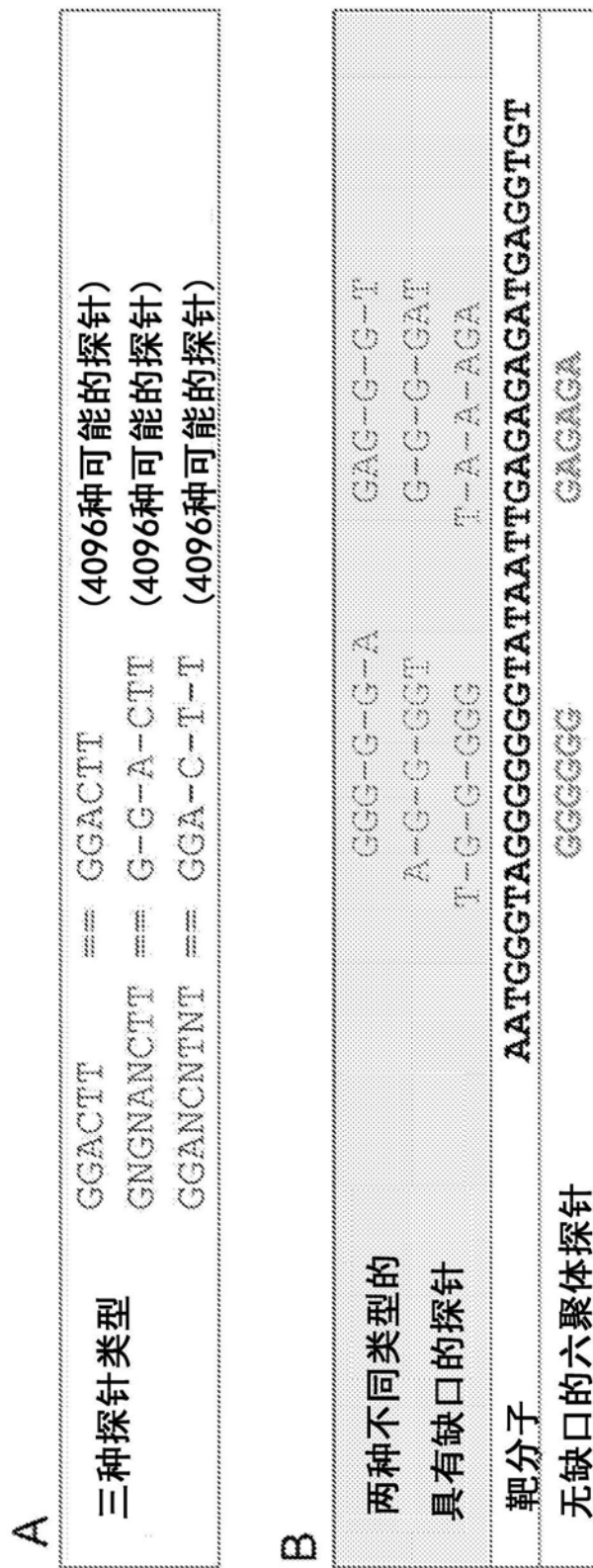


图15

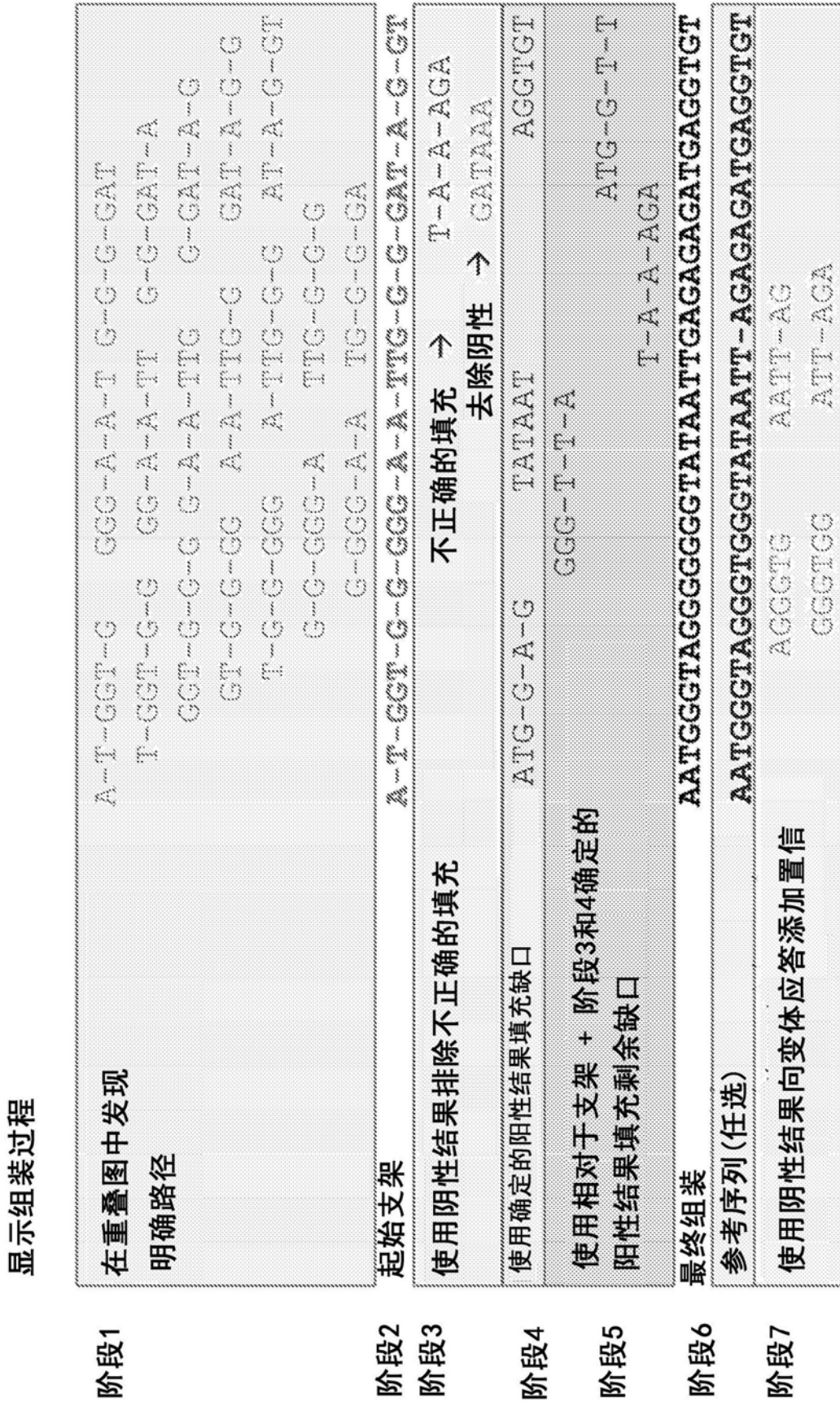


图16

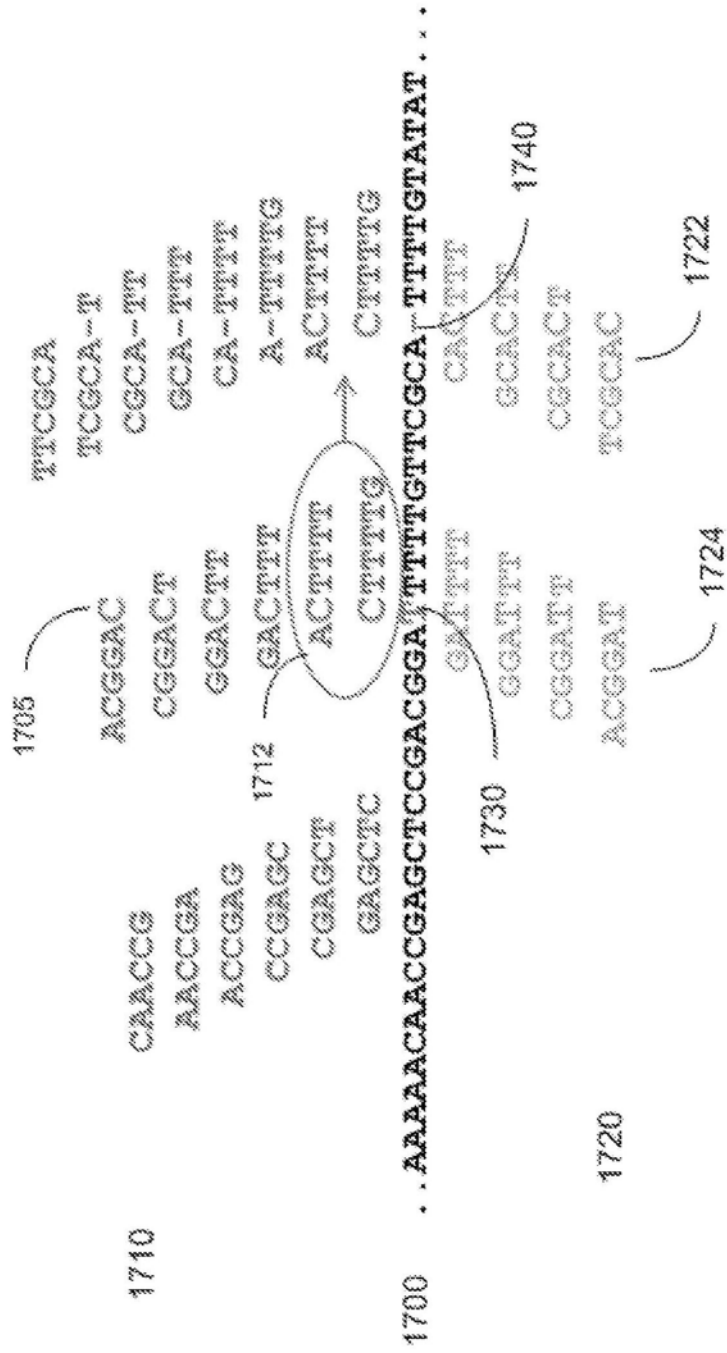


图17

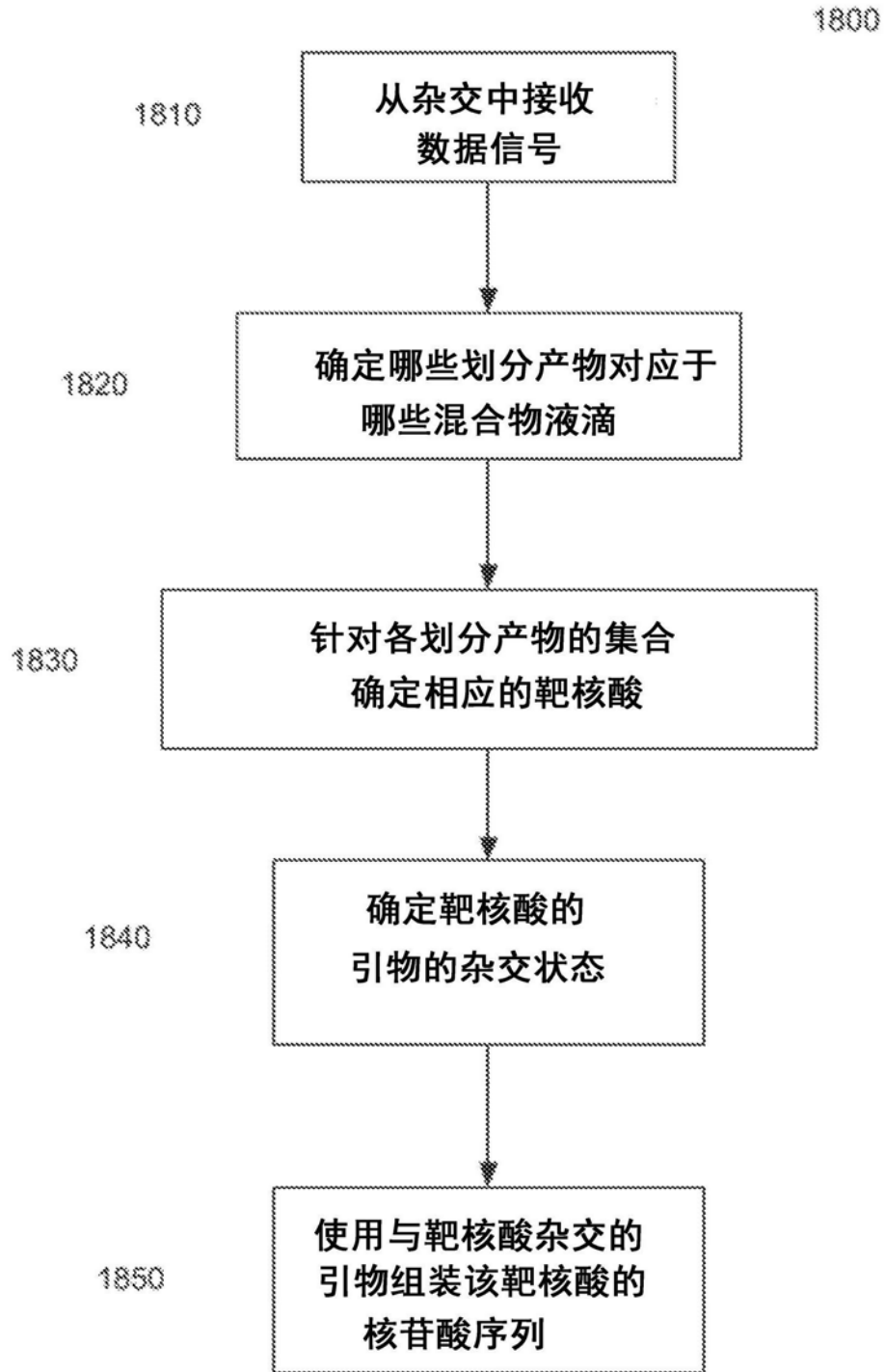


图18

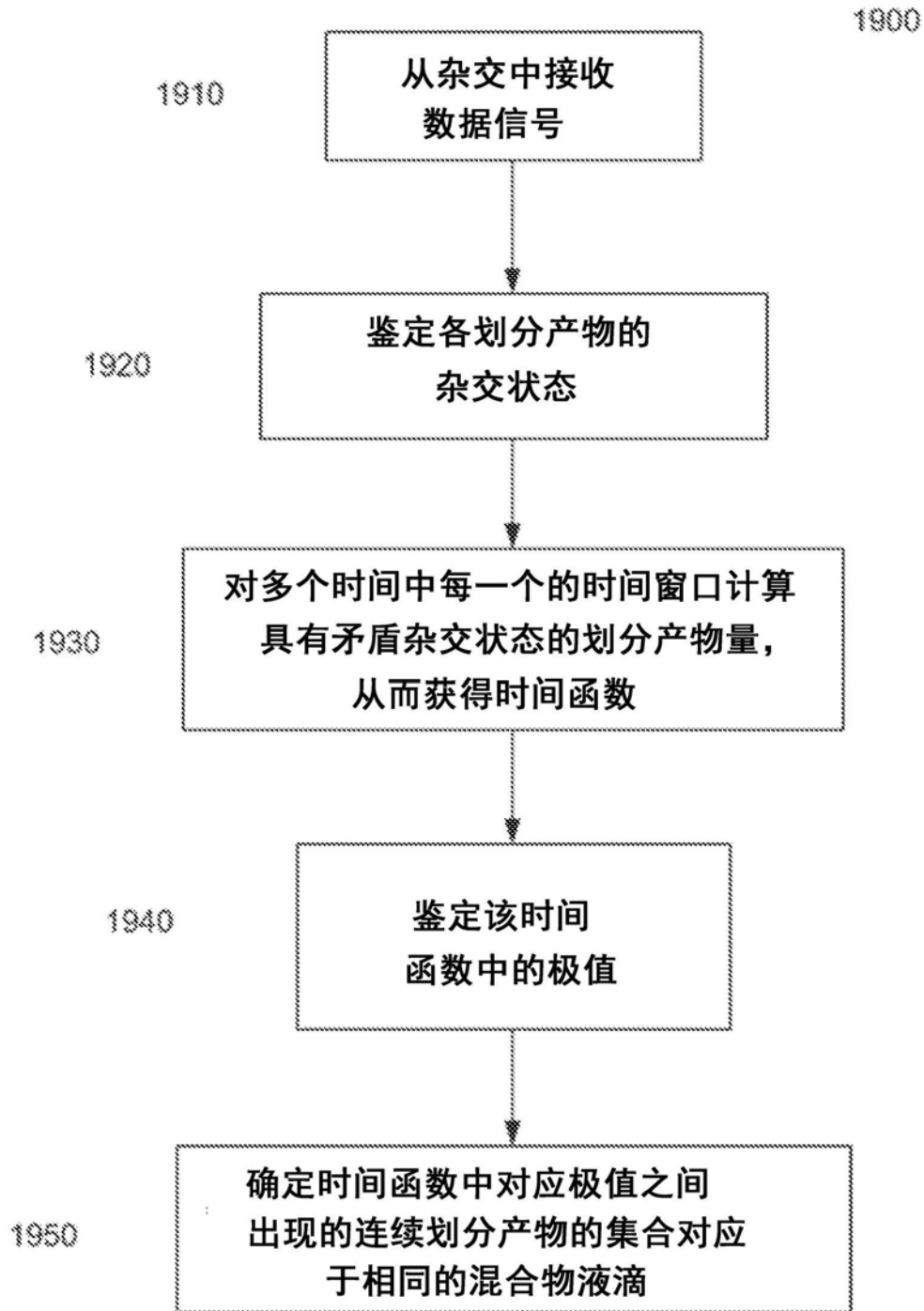


图19

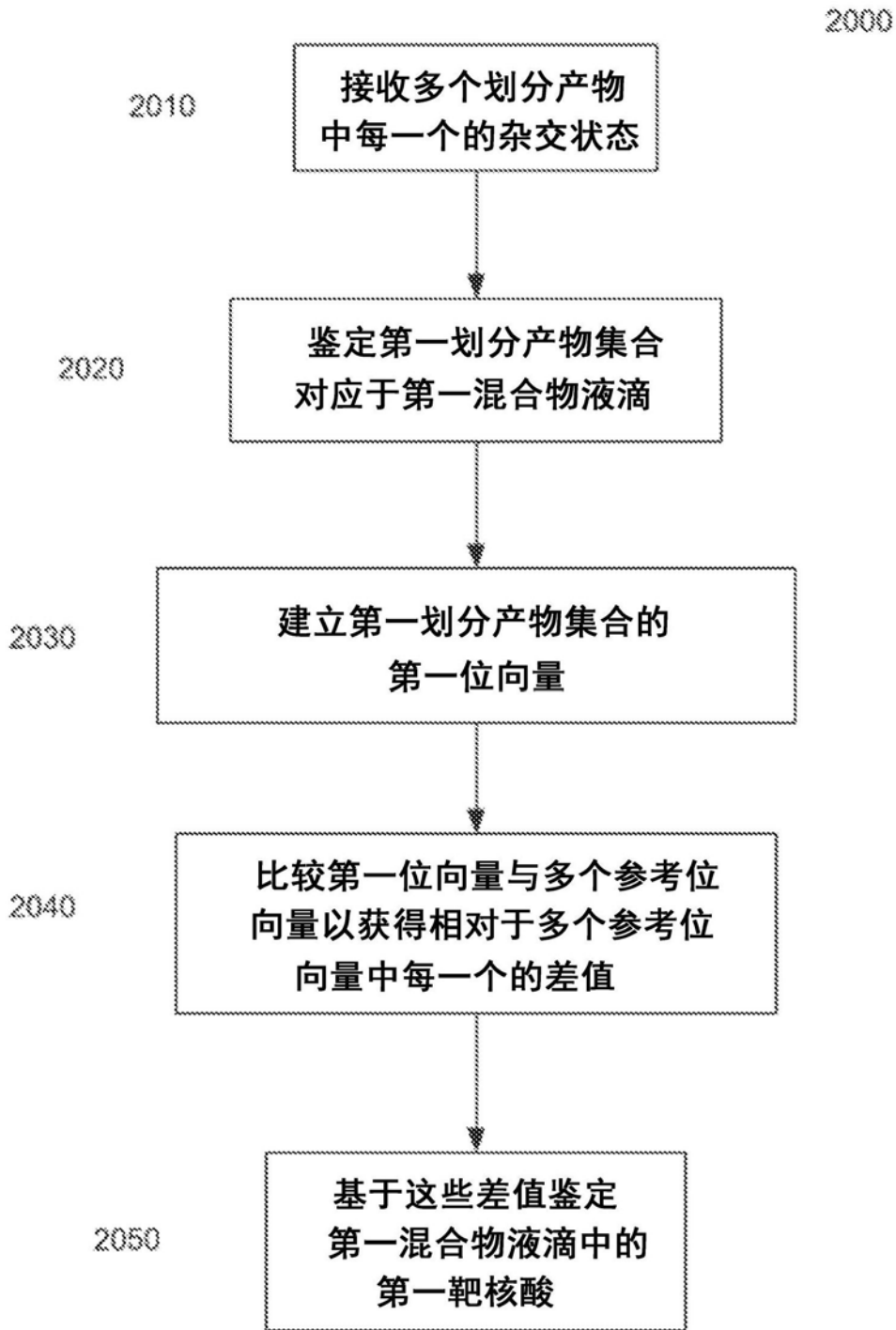


图20

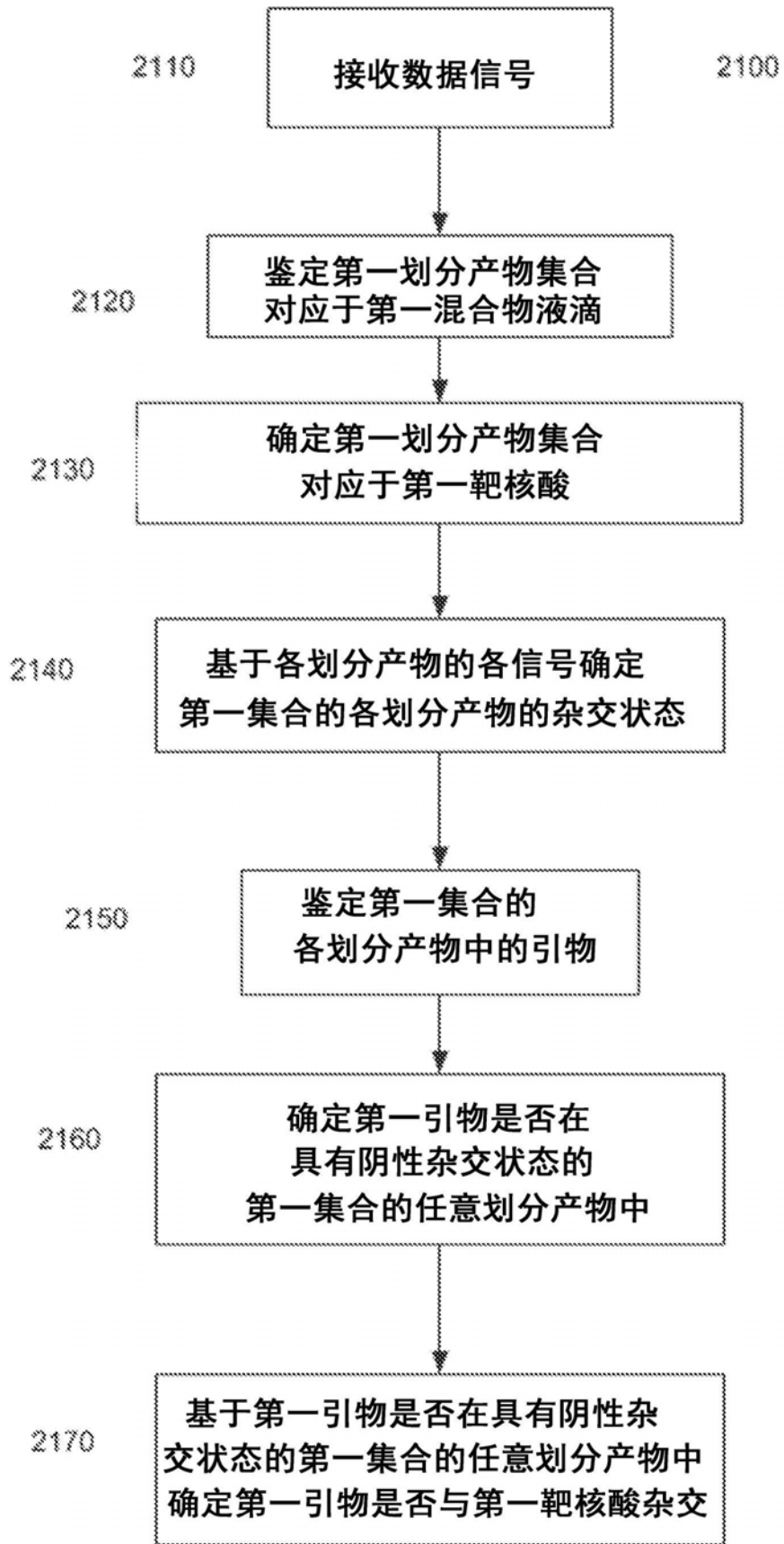


图21

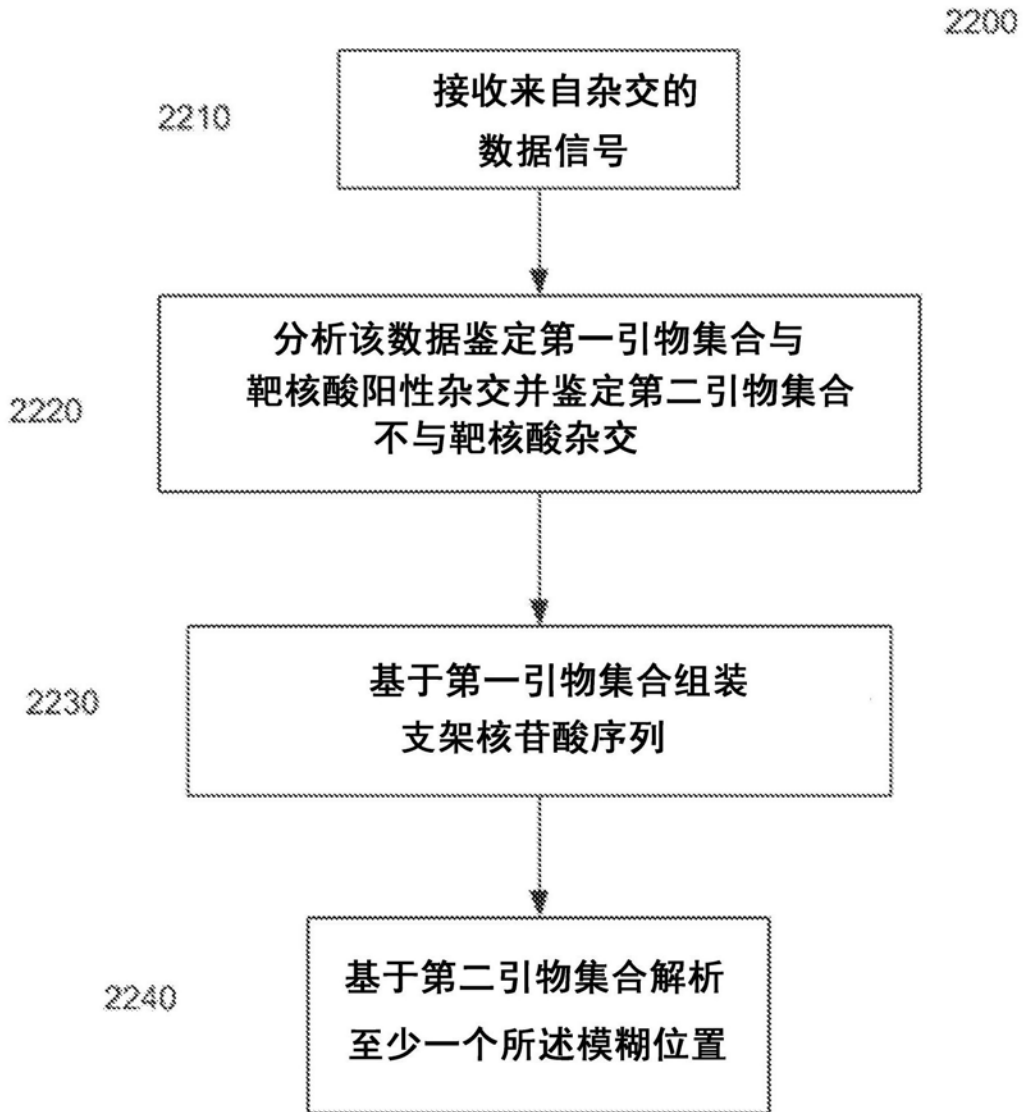


图22

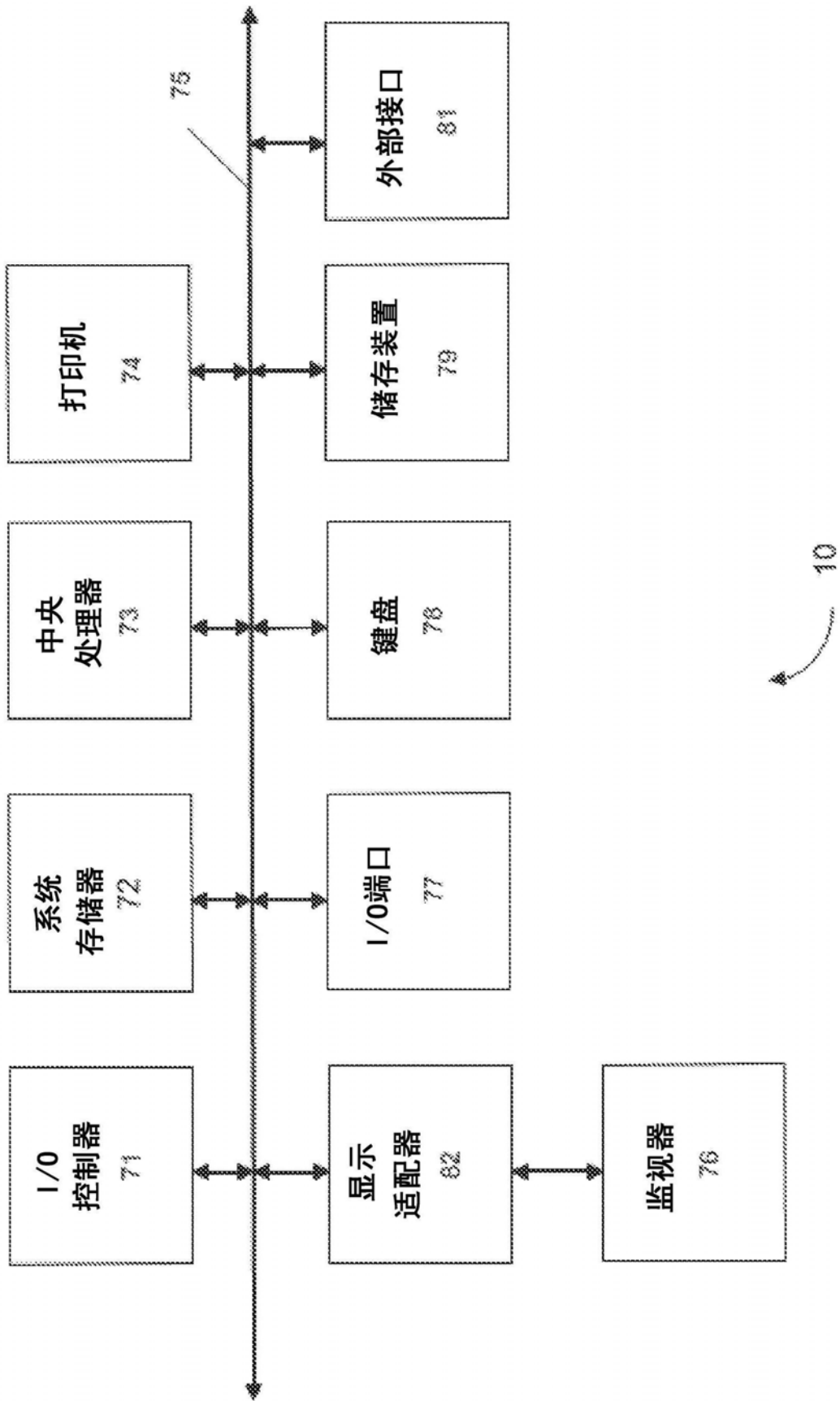


图23