



(12) 发明专利申请

(10) 申请公布号 CN 115044645 A

(43) 申请公布日 2022. 09. 13

(21) 申请号 202210236083.9

C12Q 1/6855 (2018.01)

(22) 申请日 2016.11.10

C40B 50/06 (2006.01)

(30) 优先权数据

C12Q 1/6886 (2018.01)

62/254,110 2015.11.11 US

(62) 分案原申请数据

201680076637.6 2016.11.10

(71) 申请人 分析生物科学有限公司

地址 美国华盛顿州

(72) 发明人 克里斯多弗·K·雷蒙德 林继力

(74) 专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

专利代理师 王达佐 洪欣

(51) Int. Cl.

C12Q 1/6806 (2018.01)

C12N 15/10 (2006.01)

权利要求书3页 说明书32页

序列表20页 附图5页

(54) 发明名称

DNA文库的高效率构建

(57) 摘要

本公开提供一种用于有效DNA文库构建以及所述文库的靶向遗传分析的方法。文库可用于预测、诊断或监测受试者的遗传疾病。

1. 一种提高衔接子连接至一个或多个DNA片段的效率的方法,所述方法包括:

(a) 除去一个或多个DNA片段的末端磷酸酯残基;

(b) 用一种或多种末端修复酶处理所述去磷酸化的DNA片段以产生末端修复的DNA;

(c) 将一个或多个双链DNA (dsDNA) 前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;

(d) 用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的所述非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;以及

(e) 用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;

其中与将去磷酸化的衔接分子连接至磷酸化的DNA片段的方法相比,衔接子连接的效率提高。

2. 一种构建DNA文库的方法,所述方法包括:

(a) 除去一个或多个DNA片段的末端磷酸酯残基;

(b) 用一种或多种末端修复酶处理所述去磷酸化的DNA片段以产生末端修复的DNA;

(c) 将一个或多个双链DNA (dsDNA) 前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;

(d) 用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的所述非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;以及

(e) 用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库。

3. 如权利要求1或权利要求2所述的方法,其中所述非连接配偶体链寡核苷酸在3'末端包含防止其连接至所述末端修复的DNA的5'末端和/或衔接子二聚体形成的修饰。

4. 如权利要求1至3中任一项所述的方法,其中所述一个或多个DNA片的来源是选自由以下各项组成的组的DNA:基因组DNA (gDNA)、互补DNA (cDNA) 和无细胞DNA (cfDNA)。

5. 如权利要求4所述的方法,其中所述DNA的来源是选自由以下各项组成的组的生物样品:血液、皮肤、毛发、毛囊、唾液、口腔粘液、阴道粘液、汗液、泪液、上皮组织、尿液、精液、精液、精浆、前列腺液、射精前液(考珀氏液)、排泄物、活组织检查物、腹水、脑脊液、淋巴以及组织提取物样品或活组织检查样品。

6. 如权利要求4所述的方法,其中所述DNA的来源是选自由以下各项组成的组的生物样品:羊水、血液、血浆、血清、精液、淋巴液、脑脊液、眼内液、尿液、唾液、粪便、粘液和汗液。

7. 如权利要求5或权利要求6所述的方法,其还包括从受试者的生物样品分离所述DNA。

8. 一种构建cfDNA文库的方法,所述方法包括:

(a) 从受试者的生物样品分离或获得cfDNA;

(a) 除去所述cfDNA的末端磷酸酯残基;

(b) 用一种或多种末端修复酶处理所述去磷酸化的cfDNA以产生末端修复的cfDNA并任

选地修复DNA损伤；

(c) 将一个或多个双链DNA (dsDNA) 前衔接子连接至所述末端修复的cfDNA的每条链的3'末端以形成前衔接子/末端修复的cfDNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的cfDNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;

(d) 用修复寡核苷酸置换来自所述前衔接子/末端修复的cfDNA复合物的所述非连接配偶体链寡核苷酸以形成衔接子/末端修复的cfDNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;

(e) 用一种或多种酶处理所述衔接子/末端修复的cfDNA复合物以形成连续的双链cfDNA文库;以及

(f) 扩增所述cfDNA文库以产生无细胞DNA克隆文库。

9. 一种预测、诊断或监测受试者的遗传疾病的方法,所述方法包括:

(a) 从受试者的生物样品分离或获得DNA;

(b) 除去所述DNA的末端磷酸酯残基;

(c) 用一种或多种末端修复酶处理所述去磷酸化的DNA以产生末端修复的DNA;

(d) 将一个或多个双链DNA (dsDNA) 前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;

(e) 用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的所述非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;

(f) 用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;

(g) 扩增所述DNA文库以产生DNA克隆文库;

(h) 测定所述DNA克隆文库中的基因组等效物的数量;以及

(i) 进行所述DNA克隆文库中与所述遗传疾病相关的一个或多个靶遗传基因座的定量遗传分析,其中所述一个或多个靶遗传基因座中一种或多种遗传损伤的鉴定或检测预后、诊断或监测所述遗传疾病的进展。

10. 一种伴随诊断遗传疾病的方法,所述方法包括:

(a) 从受试者的生物样品分离或获得DNA;

(b) 除去所述DNA的末端磷酸酯残基;

(c) 用一种或多种末端修复酶处理所述去磷酸化的DNA以产生末端修复的DNA;

(d) 将一个或多个双链DNA (dsDNA) 前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;

(e) 用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的所述非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;

(f) 用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;

- (g) 扩增所述DNA文库以产生DNA克隆文库;
- (h) 测定所述DNA克隆文库中的基因组等效物的数量;以及
- (i) 进行所述DNA克隆文库中与所述遗传疾病相关的一种或多种生物标志物的定量遗传分析,其中检测到或未能检测到所述一种或多种生物标志物中的至少一种指示所述受试者是否应针对所述遗传疾病进行治疗。

## DNA文库的高效率构建

[0001] 背景

[0002] 相关申请的交叉引用

[0003] 本申请是申请号为2016800766376的中国专利申请的分案申请,并且要求2015年11月11日提交的美国临时申请序列号62/254,110的优先权,所述申请的公开内容特此以引用的方式整体并入。

[0004] 以电子方式提交的文本文件的描述

[0005] 同此以电子方式提交的文本文件的内容以引用的方式整体并入本文:序列表的计算机可读格式副本(文件名称:CLFK\_003\_01W0\_ST25,记录日期:2016年11月10日;文件大小:23kB)。

### 技术领域

[0006] 本发明总体上涉及用于构建DNA文库的改进的组合物和方法。具体地说,本发明涉及有效构建用于定量遗传分析的DNA克隆文库。

[0007] 相关技术的描述

[0008] 对于下游分析有价值的各种DNA样本以微量收集。作为举例,从全血的血浆级分收集的无细胞DNA(cfDNA)通常以纳克量/mL血浆存在。鉴于一个二倍体人基因组的重量为6皮克,这意味着存在可从单次抽血中分离出几百到几千个总信息基因组。

[0009] 在癌症患者中,肿瘤DNA以在 $\leq 0.1\%$ 至 $\geq 10\%$ 的总循环DNA范围内的高度可变的量流入血流中。抽血仅含有几纳克的DNA,并且如果肿瘤基因组以总循环DNA的 $0.1\%$ 存在,则仅存在肿瘤基因组的1至10个总拷贝。为了通过序列分析明确鉴定肿瘤DNA,有必要观察肿瘤特异性遗传损伤的两个或更多个拷贝。然而,需要使DNA的检测灵敏度最大化,从而意味着准确检测在 $0.1\%$ 范围内的肿瘤DNA尚未实现。

[0010] 这些考虑阐述了以下根本性问题:使用血液进行实体瘤的可靠遗传分析部分地由分离和分析稀有基因组片段的能力决定。此外,许多治疗可行的肿瘤损伤涉及基因融合、DNA序列的显著插入或缺失和/或基因拷贝数的变化。此类改变难以通过PCR分析,其中两个相邻的引物结合位点必须是已知的,并且其中拷贝变异被许多轮目标扩增掩盖。

[0011] 目前,靶标检索方法被用于循环肿瘤DNA中的潜在损伤的全面分析。此类检索方法依赖于DNA克隆文库的创建。不幸的是,目前用于创建这些DNA文库的方法是无效的,只有很小比例的DNA片段被成功转化为有用的文库克隆。

[0012] 简述

[0013] 本发明总体上涉及用于将DNA衔接子高效率连接至DNA片段以产生用于定量遗传分析的DNA文库的组合物和方法。

[0014] 在各种实施方案中,提供了一种提高衔接子连接至一个或多个DNA片段的效率的方法,所述方法包括:除去一个或多个DNA片段的末端磷酸酯残基;用一种或多种末端修复酶处理去磷酸化的DNA片段以产生末端修复的DNA;将一个或多个双链DNA(dsDNA)前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其

中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;用修复寡核苷酸置换来自前衔接子/末端修复的DNA复合物的非连接配偶体链寡核苷酸以形成所述衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;以及用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;其中与将去磷酸化的衔接分子连接至磷酸化的DNA片段的方法相比,衔接子连接的效率被提高。

[0015] 在各种实施方案中,提供了一种构建DNA文库的方法,所述方法包括:除去一个或多个DNA片段的末端磷酸酯残基;用一种或多种末端修复酶处理去磷酸化的DNA片段以产生末端修复的DNA;将一个或多个双链DNA(dsDNA)前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;以及用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库。

[0016] 在具体实施方案中,所述非连接配偶体链寡核苷酸在所述3'末端包含防止其连接至末端修复的DNA的5'末端和/或衔接子二聚体形成的修饰。

[0017] 在某些实施方案中,所述一种或多种DNA片的来源是选自由以下各项组成的组的DNA:基因组DNA(gDNA)、互补DNA(cDNA)和无细胞DNA(cfDNA)。

[0018] 在其他实施方案中,所述DNA的来源是选自由以下各项组成的组的生物样品:血液、皮肤、毛发、毛囊、唾液、口腔粘液、阴道粘液、汗液、泪液、上皮组织、尿液、精液(semen)、精液(seminal fluid)、精浆、前列腺液、射精前液(考珀氏液)、排泄物、活组织检查物、腹水、脑脊液、淋巴以及组织提取物样品或活组织检查样品。

[0019] 在具体实施方案中,所述DNA的来源是选自由以下各项组成的组的生物样品:羊水、血液、血浆、血清、精液、淋巴液、脑脊液、眼内液、尿液、唾液、粪便、粘液和汗液。

[0020] 在其他实施方案中,所述方法还包括从受试者的生物样品分离所述DNA。

[0021] 在一些实施方案中,所述方法还包括将来自受试者的生物样品的DNA片段化。

[0022] 在某些实施方案中,所述方法还包括在连接之前修复所述一个或多个DNA片段的损伤。

[0023] 在具体实施方案中,所述损伤是脱氨基化的胞嘧啶(尿嘧啶)、无碱基位点、鸟嘌呤甲基化为O6MeG、DNA切口、缺口或胸腺嘧啶二聚体。

[0024] 在各种实施方案中,提供了一种构建cfDNA文库的方法,所述方法包括:从受试者的生物样品分离或获得cfDNA;除去所述cfDNA的末端磷酸酯残基;用一种或多种末端修复酶处理去磷酸化的cfDNA以产生末端修复的cfDNA并任选地修复DNA损伤;将一个或多个双链DNA(dsDNA)前衔接子连接至所述末端修复的cfDNA的每条链的3'末端以形成前衔接子/末端修复的cfDNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的cfDNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;用修复寡核苷酸置换来自所述前衔接子/末端修复的cfDNA复合物的非连接配偶体链寡核苷酸以形成衔接子/末端修复的cfDNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;用一种

或多种酶处理所述衔接子/末端修复的cfDNA复合物以形成连续的双链cfDNA文库;以及扩增所述cfDNA文库以产生无细胞DNA克隆文库。

[0025] 在具体实施方案中,所述连接链寡核苷酸包含一种或多种修饰以防止衔接子二聚体形成,任选地其中所述非连接配偶体链寡核苷酸的3'末端的修饰防止衔接子二聚体形成。

[0026] 在某些实施方案中,所述连接链寡核苷酸包含锚定序列、读取代码或PCR引物结合位点。

[0027] 在其他实施方案中,所述连接链寡核苷酸包含锚定序列、读取代码和PCR引物结合位点。

[0028] 在一些实施方案中,所述连接链寡核苷酸包含用于PCR扩增所述一个或多个连续双链DNA文库分子的一个或多个PCR引物结合位点。

[0029] 在具体实施方案中,所述连接链寡核苷酸包含一个或多个独特的读取代码。

[0030] 在具体实施方案中,所述连接链寡核苷酸包含用于样品复用的一个或多个样品代码。

[0031] 在某些实施方案中,所述连接链寡核苷酸包含用于DNA测序的一个或多个序列。

[0032] 在其他实施方案中,所述连接链寡核苷酸包含锚定序列。

[0033] 在其他实施方案中,所述修复寡核苷酸包含锚定序列、读取代码或PCR引物结合位点。

[0034] 在某些实施方案中,所述修复寡核苷酸包含锚定序列、读取代码和PCR引物结合位点。

[0035] 在具体实施方案中,所述修复寡核苷酸包含用于PCR扩增所述一个或多个连续双链DNA文库分子的一个或多个引物结合位点。

[0036] 在一些实施方案中,所述修复寡核苷酸包含一个或多个独特的读取代码。

[0037] 在某些实施方案中,所述修复寡核苷酸包含用于样品复用的一个或多个样品代码。

[0038] 在具体实施方案中,所述修复寡核苷酸包含用于DNA测序的一个或多个序列。

[0039] 在其他实施方案中,所述连接链寡核苷酸与所述修复寡核苷酸互补。

[0040] 在具体实施方案中,所述连接链寡核苷酸的锚定序列与所述修复寡核苷酸的锚定序列互补。

[0041] 在其他实施方案中,所述连接链寡核苷酸的PCR引物结合位点与所述修复寡核苷酸的PCR引物结合位点互补。

[0042] 在具体实施方案中,所述一个或多个衔接子包含多个连接链寡核苷酸种类。

[0043] 在一些实施方案中,所述一个或多个衔接子包含多个修复寡核苷酸种类。

[0044] 在具体实施方案中,所述连接链寡核苷酸的引物结合位点不与所述修复寡核苷酸的引物结合位点互补。

[0045] 在某些实施方案中,所述连接链寡核苷酸的引物结合位点与所述修复寡核苷酸的引物结合位点显著不同。

[0046] 在某些实施方案中,结合所述连接链寡核苷酸的引物结合位点的引物基本上不结合所述修复寡核苷酸的引物结合位点。

- [0047] 在具体实施方案中,扩增所述DNA文库以产生DNA克隆文库。
- [0048] 在其他实施方案中,对所述DNA克隆文库进行qPCR并将qPCR测量值与已知基因组等效物的标准进行比较以确定所述DNA克隆文库的基因组等效物。
- [0049] 在具体实施方案中,用结合至Alu序列的引物和结合至衔接子中的序列的引物进行所述qPCR。
- [0050] 在一些实施方案中,对所述DNA克隆文库中的多个遗传基因座进行定量遗传分析。
- [0051] 在具体实施方案中,对多个DNA克隆文库中的多个遗传基因座进行定量遗传分析。
- [0052] 在具体实施方案中,定量遗传分析包括将一个或多个捕获探针与靶遗传基因座杂交以形成捕获探针模块-DNA克隆复合物。
- [0053] 在某些实施方案中,定量遗传分析包括分离所述捕获探针-DNA克隆复合物。
- [0054] 在其他实施方案中,所述定量遗传分析包括扩增所述分离的捕获探针-DNA克隆复合物中的DNA克隆序列。
- [0055] 在具体实施方案中,定量遗传分析包括DNA测序以产生多个测序读数。
- [0056] 在其他实施方案中,所述方法还包括所述多个测序读数的生物信息学分析。
- [0057] 在具体实施方案中,对所述DNA克隆文库中的多个遗传基因座进行定量遗传分析,并且其中生物信息学分析用于:定量所述DNA克隆文库中分析的基因组等效物的数量;检测靶遗传基因座中的遗传变体;检测靶遗传基因座内的突变;检测靶遗传基因座内的遗传融合;和/或测量靶遗传基因座内的拷贝数波动。
- [0058] 在某些实施方案中,所述定量遗传分析用于鉴定或检测导致遗传疾病或与遗传疾病相关的一种或多种遗传损伤。
- [0059] 在具体实施方案中,所述遗传损伤包括核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。
- [0060] 在某些实施方案中,所述遗传疾病是癌症。
- [0061] 在其他实施方案中,所述定量遗传分析用于鉴定或检测胎儿cfDNA中的一个或多个靶遗传基因座的一种或多种遗传变体或遗传损伤。
- [0062] 在一些实施方案中,所述捕获探针是捕获探针模块的组分,所述捕获探针模块任选地与半抗原标记的配偶体寡核苷酸复合,所述寡核苷酸与所述捕获探针模块中的尾部序列杂交。
- [0063] 在各种实施方案中,提供了一种预测、诊断或监测受试者的遗传疾病的方法,所述方法包括:从受试者的生物样品分离或获得DNA;除去所述DNA的末端磷酸酯残基;用一种或多种末端修复酶处理去磷酸化的DNA以产生末端修复的DNA;将一个或多个双链DNA(dsDNA)前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;扩增所述DNA文库以产生DNA克隆文库;测定所述DNA克隆文库中的基因组等效物的数量;以及进行所述DNA克隆文库中与所述遗传疾病相关的一个或多个靶遗传基因座的定量遗传分析,其中所述一个或多个靶遗传基因座



中一种或多种遗传损伤的鉴定或检测预后、诊断或监测所述遗传疾病的进展。

[0064] 在某些实施方案中,所述DNA是基因组DNA、来自福尔马林固定的石蜡包埋的(FFPE)样品的DNA、cDNA或cfDNA。

[0065] 在具体实施方案中,所述cfDNA是从选自以下各项的组的生物样品分离:羊水、血液、血浆、血清、精液、淋巴液、脑脊液、眼内液、尿液、唾液、粪便、粘液和汗液。

[0066] 在其他实施方案中,所述遗传损伤包括核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。

[0067] 在具体实施方案中,所述遗传疾病是癌症。

[0068] 在各种实施方案中,提供了一种伴随诊断遗传疾病的方法,所述方法包括:从受试者的生物样品分离或获得DNA;除去所述DNA的末端磷酸酯残基;用一种或多种末端修复酶处理去磷酸化的DNA以产生末端修复的DNA;将一个或多个双链DNA(dsDNA)前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;扩增所述DNA文库以产生DNA克隆文库;测定所述DNA克隆文库中的基因组等效物的数量;以及进行所述DNA克隆文库中与所述遗传疾病相关的一种或多种生物标志物的定量遗传分析,其中检测到或未能检测到所述一种或多种生物标志物中的至少一种指示所述受试者是否应针对所述遗传疾病进行治疗。

[0069] 在其他实施方案中,所述DNA是基因组DNA、来自福尔马林固定的石蜡包埋的(FFPE)样品的DNA、cDNA或cfDNA。

[0070] 在具体实施方案中,所述cfDNA是从选自由以下各项组成的组的生物样品分离:羊水、血液、血浆、血清、精液、淋巴液、脑脊液、眼内液、尿液、唾液、粪便、粘液和汗液。

[0071] 在具体实施方案中,所述生物标志物是遗传损伤。

[0072] 在其他实施方案中,所述遗传损伤包括核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。

[0073] 在某些实施方案中,所述遗传疾病是癌症。

[0074] 附图的若干视图简述

[0075] 图1示出常规对比高效率(HE)连接技术。(A)在DNA与未磷酸化的双链体衔接子连接之前,将靶DNA片段进行5'磷酸化。(B)常规连接方法的常见低效率是缺少磷酸酯基团的5'靶DNA片段末端未能与未磷酸化的双链体衔接子连接。(C)将靶DNA片段彼此连接。(D)包含所需的5'磷酸酯基团的双链体衔接子与靶DNA片段的3'末端连接。衔接子的配偶体链寡核苷酸上的阴影圆圈表示3'封闭基团。(E)3'封闭基团还阻止衔接子彼此连接。(F)缺少5'磷酸酯的偶然衔接子双链体不能连接至靶片段。

[0076] 图2示出RsaI消化的、去磷酸化的pUC19质粒与一系列九种不同的HE衔接子的完全连接的琼脂糖凝胶的代表性图像。未连接的片段在对照泳道中显示,所述对照泳道与740、500、300和150bp(从上到下)的分子量(MW)标记物相邻。三个载体片段(箭头)的迁移率的完全变动表明衔接子与所有九个衔接头两端的完全连接。这些结果表明HE连接技术是可泛化

的。

[0077] 图3示出完成HE连接的构建体的示例性方法。(A) HE连接产物是连接至片段的3'末端的3'延伸衔接子寡核苷酸。(B) 用于“修复”初始连接产物的一种策略是添加修复寡核苷酸(顶链;绿色);T4多核苷酸激酶可将磷酸酯(P)添加至靶片段的5'末端,并且可使用切口封闭连接酶如Taq DNA连接酶来将修复寡核苷酸连接至靶片段。(C) 另一种策略是组合互补衔接子寡核苷酸即具有5'至3'外切核酸酶活性的DNA聚合酶(例如,BstI DNA聚合酶)和Taq DNA连接酶。BstI DNA聚合酶通过从靶片段除去5'碱基以暴露通过Taq DNA连接酶实现切口封闭连接的5'磷酸酯而延伸修复寡核苷酸。(D) 互补衔接子寡核苷酸也可被设计为将另外的序列特征引入具有BstI DNA聚合酶的原始HE连接链中。

[0078] 图4示出使用HE连接技术制备DNA文库。(A) DNA片段可拥有可能具有或可能不具有磷酸酯基团(P)的异质“杂乱”末端。用磷酸酶处理DNA片段除去暴露的5'和3'磷酸酯。然后可用修复DNA损伤并将5'或3'突出端“精修”至钝端的酶如脱氨基化的胞嘧啶(U)、无碱基位点(↑)和胸腺嘧啶二聚体处理DNA。(B) 将衔接子分两步添加至DNA片段。首先,将包含5'磷酸化连接链和3'封闭配偶体链的双链体衔接子连接至靶片段。具有约30℃的解链温度的配偶体链在温度 $\geq 37^\circ\text{C}$ 发生的后续步骤中移除。其次,修复寡核苷酸与衔接子连接片段退火;使用激酶/连接酶策略或聚合酶/外切核酸酶/连接酶策略,修复寡核苷酸变得共价连接至靶片段的5'末端(图3)。初始连接链的引物延伸将修复寡核苷酸信息拷贝到适合下游分析的全长衔接子双链体中。

[0079] 图5示出用于产生双重PCR引物适应性DNA片段的HE连接技术策略。在此方案中,连接链携带充当独立引物结合位点的另外序列(引物2)。修复寡核苷酸虽然与连接链的一部分互补,但具有充当第二PCR引物结合位点(引物1)的其自身发散序列。完全完成的衔接子允许使用更常规的通用双重引物PCR方法扩增DNA样本片段。

[0080] 详述

[0081] A. 综述

[0082] 本发明部分考虑了用于解决定量遗传分析领域的急需的组合物和方法以获得用于克隆此类DNA片段以进行下游分析的改进的高效方法。

[0083] 用于DNA分析的当前方法包括将特化衔接子连接至DNA片段(图1)。在常规技术中,在DNA连接之前将靶DNA片段进行5'磷酸化,以实现与未磷酸化的双链体衔接子共价连接。靶DNA片段和衔接子可以是平末端的,或者它们可共有互补突出端(例如T/A)。(图1A)。这是严重的缺点,因为不可能确保所有靶DNA片段的两个末端均被磷酸化,并且未磷酸化末端不能连接并且这些靶片段从随后的文库中丢失。(图1B)。作为非限制性实例,如果70%的靶DNA片段末端具有5'磷酸酯,则最多仅49%的片段( $0.7 \times 0.7 \times 100\%$ )可连接在所述片段的两个末端上并且克隆需要连接至两个末端。此外,靶DNA片段上存在5'磷酸酯促进单独的不希望的假象,其中DNA片段可彼此连接(图1C)。这产生可混淆检测疾病特异性染色体重排的人工染色体序列融合事件。

[0084] 在各种实施方案中,本发明部分考虑了用于将衔接子序列有效连接至靶DNA片段的组合物和方法。在具体实施方案中,从靶DNA片段的5'和3'末端除去磷酸酯。然后将这些去磷酸化的片段用产生平末端DNA末端的酶处理,并且任选地用修复可能已经对DNA造成的许多类型的DNA损伤的酶处理,所述DNA损伤例如脱氨基化的胞嘧啶(尿嘧啶)、无碱基位点、

鸟嘌呤甲基化为 $O^6$ MeG、切口、双链断裂或胸腺嘧啶二聚体。衔接子包含与非连接配偶体链寡核苷酸复合的连接链寡核苷酸。衔接子的连接链携带连接至靶DNA片段所需的5'磷酸酯基团,并且配偶体链包含3'封闭基团(图1D)。3'封闭基团防止衔接子:衔接子二聚体的形成(图1E)。与DNA片段一样,并非所有衔接子序列都将具有5'磷酸酯(溶剂暴露的末端磷酸酯键是固有化学不稳定的)。虽然此类未磷酸化的衔接子将存在,但它们将仅瞬时接合连接机器(图1F);此类衔接子与片段的非生产性配对迅速解离并被可提供有效共价连接的衔接子:靶DNA片段配对置换。最终约100%的靶DNA片段变得在两个末端上与衔接分子连接,这说明本文考虑的组合物和方法用于构建DNA文库的高效率。

[0085] 在各种实施方案中,本文考虑的用于DNA文库的高效率构建的组合物和方法提供使用可从各种生物来源获得的DNA解决分子遗传分析的新颖综合框架。克隆纯化的DNA引入标记的DNA序列,所述标记的DNA序列提供关于下游分析的信息并使所得克隆文库能够扩增。使用靶特异性寡核苷酸的杂合捕获用于检索用于随后分析的特定序列。对文库中存在的基因组数量的独立测量被应用于每种样品,并且这些测定提供用于估计测定的灵敏度的方式。本文考虑的测定提供用于分析、检测、诊断或监测遗传状态、病状或疾病的可靠的、可重复的和稳健的方法。

[0086] 除非相反地指明,否则本发明的具体实施方案的实践将采用在化学、生物化学、有机化学、分子生物学、微生物学、重组DNA技术、遗传学、免疫学和细胞生物学的在本领域技术范围内的常规方法,出于说明的目的,下文描述了所述方法中的许多。此类技术在文献中充分地解释。参见例如Sambrook,等人,Molecular Cloning:A Laboratory Manual(第3版,2001);Sambrook,等人,Molecular Cloning:A Laboratory Manual(第2版,1989);Maniatis等人,Molecular Cloning:A Laboratory Manual(1982);Ausubel等人,Current Protocols in Molecular Biology(John Wiley和Sons,最新版本2008年7月);Short Protocols in Molecular Biology:A Compendium of Methods from Current Protocols in Molecular Biology,Greenlee Pub.Associates and Wiley-Interscience;Glover,DNA Cloning:A Practical Approach,第I&II卷(IRL Press,Oxford,1985);Anand,Techniques for the Analysis of Complex Genomes,(Academic Press,New York,1992);Transcription and Translation(B.Hames和S.Higgins,编著,1984);Perbal,A Practical Guide to Molecular Cloning(1984);以及Harlow和Lane,Antibodies,(Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,1998)。

[0087] B. 定义

[0088] 除非另外定义,否则本文使用的所有技术和科学术语都具有与由本发明所属领域普通技术人员通常理解的含义相同的含义。尽管与本文所述的那些方法和材料类似或等效的任何方法和材料可在本发明的实践或测试中使用,但本文描述组合物、方法和材料的优选实施方案。出于本发明的目的,下文定义以下术语。

[0089] 本文所使用的冠词“一(个/种)”和“所述”是指一个(种)或多于一个(种)(即,至少一个(种))所述冠词的修饰对象。作为举例,“一个/种要素”是指一个/种要素或多于一个/种要素。

[0090] 使用替代(例如,“或”)应理解为是指多个替代中的任一、两者或其任何组合。

[0091] 术语“和/或”应被理解为是指替代中的一个或两个。

[0092] 如本文所用,术语“约”或“大约”是指量、水平、价值、数目、频率、百分比、尺寸、大小、量、重量或长度与参考量、水平、价值、数目、频率、百分比、尺寸、大小、量、重量或长度变化多达15%、10%、9%、8%、7%、6%、5%、4%、3%、2%或1%。在一个实施方案中,术语“约”或“大约”是指在约参考量、水平、价值、数目、频率、百分比、尺寸、大小、量、重量或长度的 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 9\%$ 、 $\pm 8\%$ 、 $\pm 7\%$ 、 $\pm 6\%$ 、 $\pm 5\%$ 、 $\pm 4\%$ 、 $\pm 3\%$ 、 $\pm 2\%$ 或 $\pm 1\%$ 的量、水平、价值、数目、频率、百分比、尺寸、大小、量、重量或长度的范围。

[0093] 在整篇本说明书中,除非上下文另外需要,否则词语“包含(comprise)”、“包含(comprises)”和“包含(comprising)”将被理解成暗示包括所述步骤或要素或一组步骤或要素,但不排除任何其它步骤或要素或任何其他组步骤或要素。在具体实施方案中,术语“包括”、“具有”、“含有”和“包含”同义地使用。

[0094] “由……组成”意味着包括但不限于,无论在短语“由……组成”后的是什么。因此,短语“由……组成”指示所列出的要素是必要的或强制性的,并且可不出现其他要素。

[0095] “实质上由……组成”意味着包括在所述短语后列出的任何要素,并且限于不干扰或影响针对所列出要素的在本公开中指定的活性或动作的其他要素。因此,短语“实质上由……组成”指示所列出要素是必需的或强制性的,但其他要素是任选的并且取决于它们是否影响所列出要素的活性或动作可或可不出现。

[0096] 贯穿本说明书所提到的“一个实施方案(one embodiment)”、“一个实施方案(an embodiment)”、“一个具体实施方案(a particular embodiment)”、“一个相关实施方案(a related embodiment)”、“某一实施方案”、“另一实施方案”或者“其他实施方案”表示结合所述实施方案描述的具体特征、结构或特性包括在本发明的至少一个实施方案中。因此,在本说明书全文的各种位置出现前述短语不一定都是指同一实施方案。此外,具体特征、结构或特性可以任何合适的方式组合在一个或多个实施方案中。

[0097] 如本文所用,术语“分离的”是指基本上或实质上不含天然状态中通常伴随的组分材料。在具体实施方案中,术语“获得的”或“衍生的”与分离的同义地使用。

[0098] 如本文所用,术语“DNA”是指脱氧核糖核酸。在各种实施方案中,术语DNA是指基因组DNA、重组DNA、合成DNA、互补DNA(cDNA)或无细胞DNA(cfDNA)。在一个实施方案中,DNA是指基因组DNA或cDNA。在一个实施方案中,DNA是指cfDNA。在具体实施方案中,DNA是包含“靶区”的DNA片段,所述DNA片段在某些实施方案中也被称为靶DNA片段。本文考虑的DNA文库包括基因组DNA文库、cfDNA文库以及从RNA构建的cDNA文库,例如RNA表达文库。在不同实施方案中,DNA文库包含一个或多个另外的DNA序列和/或标签。

[0099] “靶遗传基因座”或“DNA靶区”是指DNA序列内的目标区域。在各种实施方案中,对靶遗传基因座进行靶向遗传分析。在具体实施方案中,所述DNA靶区是基因的与特定遗传状态、遗传病状、遗传疾病、胎儿检测、遗传镶嵌性、亲子鉴定、预测对药物治疗的反应、诊断或监测医学病状、微生物组谱、病原体筛选或器官移植监测相关的区。

[0100] 如本文所用,术语“循环DNA”、“循环无细胞DNA”和“无细胞DNA”通常可互换使用并且是指为细胞外DNA、已经从细胞中挤出的DNA或已经从坏死或凋亡细胞释放的DNA的DNA。

[0101] 如本文所用,“受试者”、“个体”或“患者”包括表现出可用本文考虑的组合物检测或鉴定的病状的症状的任何动物。合适的受试者包括实验动物(如小鼠、大鼠、兔或豚鼠)、农场动物(如马、牛、羊、猪)和家畜或宠物(如猫或狗)。在具体实施方案中,所述受试者是哺

乳动物。在某些实施方案中,所述受试者是非人灵长类动物,并且在优选的实施方案中,所述受试者是人。

[0102] “反应容器”是指适合于进行本文考虑的反应之一的容器。适用于具体实施方案的反应容器的说明性实例包括但不限于试管、微量离心管(例如PCR管)、微量滴定板(例如,96孔板、384孔板、1536孔板)、载玻片、板、阵列和微阵列。

[0103] C.DNA文库的高效率构建

[0104] 在具体实施方案中,构建本文考虑的DNA文库的方法包括衔接子与靶DNA片段的高效率连接。

[0105] (a) DNA来源

[0106] 本文考虑的方法和组合物被设计为使用DNA作为分析物有效地分析、检测、诊断和/或监测遗传状态、遗传病状、遗传疾病、遗传镶嵌性、胎儿诊断、亲子鉴定、微生物组谱、病原体筛选和器官移植监测。适用于本文考虑的组合物和方法的DNA可来自本领域技术人员已知的任何来源。在具体实施方案中,所述DNA是从任何来源分离的基因组DNA,从RNA合成的拷贝DNA(cDNA)或无细胞DNA(cfDNA)。

[0107] 在一些实施方案中,所述DNA是高分子量DNA(>1000bp)。在本文考虑的组合物和方法中使用高分子量DNA通常包括片段化步骤。高分子量DNA可被片段化为约25至约750个碱基对、约25至约500个碱基对、约25至约250个碱基对、约25至约200个碱基对、约25至约150个碱基对、约25至约100个碱基对、约25至约50个碱基对、约100至约200个碱基对、约150至约180个碱基对、约150个碱基对、约155个碱基对、约160个碱基对、约165个碱基对、约170个碱基对、约175个碱基对或约180个碱基对。

[0108] 适用于本文考虑的组合物和方法的具体实施方案的用于片段化DNA的说明性方法包括但不限于:剪切、超声处理、酶消化;包括限制酶切消化以及其他方法。在具体实施方案中,本领域中已知用于片段化DNA的任何方法都可用于本发明。

[0109] 适用于本文考虑的组合物和方法的具体实施方案的基因组DNA和RNA(以产生cDNA)的说明性来源包括但不限于选自由以下各项组成的组的生物样品:脑组织、骨组织、眼组织、嗅觉组织、肌肉组织、心脏组织、肺组织、肝组织、胰腺组织、肾组织、胃组织、肠组织、结肠组织、血液、皮肤、毛发、毛囊、唾液、口腔粘液、阴道粘液、汗液、泪液、上皮组织、尿液、精液、精液、精浆、前列腺液、射精前液(考珀氏液)、排泄物、活组织检查物、腹水、脑脊液、淋巴以及组织提取物样品或活组织检查样品。

[0110] 在具体实施方案中,所述DNA是cfDNA。cfDNA的大小分布在约150bp至约180bp片段的范围内。片段化可以是核酸内切和/或核酸外切活性的结果,并且对cfDNA的准确、可靠和稳健分析提出重大挑战。用于分析cfDNA的另一个挑战是其在血流中的短半衰期,约为15分钟。不希望受任何特定理论束缚,本发明部分考虑cfDNA的分析是像“液体活组织检查”并且是当前生物过程的实时快照。

[0111] 在一些实施方案中,从血浆部分分离的cfDNA可基本上被从在收集方案期间溶解的有核血细胞释放的长(>10千碱基对)高分子量基因组DNA污染。这种长时间污染的DNA(如果不经片段化)不能良好地克隆和扩增,且因此在下文库制备过程中丢失。然而,在具体实施方案中,在不存在DNA片段化的情况下,本文考虑的高效率DNA文库构建方法选择性地克隆来自DNA样本中存在的一些片段大小的更短(<1000bp)片段。不希望受任何特定理论

束缚,来自为长片段和短片段的共混物的DNA样本的短cfDNA片段的选择性克隆在构建液体活组织检查中是有利的。

[0112] 在具体实施方案中为从其分离cfDNA的合适来源的生物样品的说明性实例包括但不限于羊水、血液、血浆、血清、精液、淋巴液、脑脊液、眼内液、尿液、唾液、粘液和汗液。

[0113] 在具体实施方案中,所述生物样品是血液或血浆。

[0114] 在某些实施方案中,所述DNA样品可源自包埋组织,如FFPE或细针抽吸物;源自意图询问存在的微生物组序列的拭子;源自法医样品,如毛发、衣服、指纹等;或源自需要对于构建来自低输入DNA样品的文库特别有效的本文考虑的文库构建方法的任何其他DNA来源。

[0115] 在某些实施方案中,可使用可商购的试剂盒和本领域技术人员已知的其他方法直接从患者的生物样品或从先前获得的和任选稳定化的生物样品中分离cfDNA,例如通过冷冻和/或添加螯合剂,包括但不限于EDTA、EGTA或对二价阳离子具有特异性的其他螯合剂。

[0116] (b) 输入DNA的去磷酸化

[0117] 在具体实施方案中,首先用除去末端磷酸酯残基的不耐热磷酸酶处理输入DNA,例如靶DNA片段。参见例如图4A。

[0118] 适用于本文考虑的组合物和方法的具体实施方案的不耐热磷酸酶的说明性实例包括但不限于APex<sup>TM</sup>不耐热碱性磷酸酶(Epicentre Biotechnologies)、NTPhos<sup>TM</sup>不耐热磷酸酶(Epicentre Biotechnologies)、HK<sup>TM</sup>不耐热磷酸酶(Epicentre Biotechnologies)和虾碱性磷酸酶(SAP;NEB)。

[0119] 在一个实施方案中,所述不耐热磷酸酶是SAP。

[0120] (c) 逆转靶DNA片段中的DNA损伤

[0121] 在具体实施方案中,还用一种或多种逆转常见来源的DNA损伤的酶处理输入DNA或去磷酸化的DNA,所述DNA损伤如胞嘧啶脱氨基化为尿嘧啶、氧化添加至鸟嘌呤、胸苷二聚体、导致无碱基位点的碱基缺失、双链体DNA的一条链上的切口或缺口等。参见例如图4A。

[0122] 在一个实施方案中,使用包含一种或多种以下酶的组合物逆转对DNA的内部损伤: Taq DNA连接酶、内切核酸酶IV、Bst DNA聚合酶、Fpg (8-氧鸟嘌呤DNA糖基化酶)、尿嘧啶-DNA糖基化酶(UDG)、T4 PDG (T4内切核酸酶V)、内切核酸酶VIII和T4 DNA聚合酶。

[0123] 在一个实施方案中,使用包含Taq DNA连接酶、内切核酸酶IV、Bst DNA聚合酶、Fpg、尿嘧啶-DNA糖基化酶(UDG)、T4 PDG (T4内切核酸酶V)、内切核酸酶VIII和T4DNA聚合酶的组合物逆转对DNA的内部损伤。

[0124] (d) 产生末端修复的DNA

[0125] 在具体实施方案中,本文考虑的组合物和方法包括产生末端修复的DNA片段。在某些实施方案中,对所述DNA片段进行末端修复以产生具有平末端、5'-突出端或3'-突出端的末端修复的DNA片段。参见例如图4A。在具体实施方案中,所述DNA是cfDNA。

[0126] 在一些实施方案中,所述末端修复的DNA含有平末端。在一些实施方案中,所述末端修复的DNA被加工为包含平末端。在优选实施方案中,所述DNA片段通过一种或多种末端修复酶进行末端修复以产生具有平末端的末端修复的DNA片段。

[0127] 在本文考虑的组合物和方法的具体实施方案中,适于产生平末端DNA片段的末端修复酶的说明性实例包括保留聚合活性和3'→5'外切核酸酶活性但缺乏5'→3'外切核酸

酶活性的DNA聚合酶(例如,T4 DNA聚合酶、DNA聚合酶I的克列诺片段等)。DNA聚合酶用于填充在5'突出端或“chew back”3'突出端中,从而留下具有平末端的DNA片段。

[0128] 在一些实施方案中,所述末端修复的DNA的平末端进行进一步修饰以含有单碱基对突出端。在一些实施方案中,含有平末端的末端修复的DNA可进行进一步加工以含有腺嘌呤(A)/胸腺嘧啶(T)突出端。在一些实施方案中,含有平末端的末端修复的DNA可进行进一步加工以含有腺嘌呤(A)/胸腺嘧啶(T)突出端作为单碱基对突出端。在一些实施方案中,末端修复的DNA具有非模板化的3'突出端。在一些实施方案中,末端修复的DNA进行加工以含有3'突出端。在一些实施方案中,末端修复的DNA用末端转移酶(TdT)进行加工以含有3'突出端。在一些实施方案中,G-尾可通过TdT添加。在一些实施方案中,使用任何已知的限制性内切酶(例如用Sau3A酶等)进行部分消化加工末端修复的DNA以含有突出末端。

[0129] (e) 将前衔接子连接至末端修复的DNA

[0130] 在具体实施方案中,本文考虑的组合物和方法包括将双链DNA前衔接子连接至末端修复的DNA的每个末端。

[0131] 如本文所用,术语“前衔接子”是指包含连接链寡核苷酸和配偶体链寡核苷酸的双链DNA分子或DNA双链体。可使用任何合适的连接酶将前衔接子连接至末端修复的DNA片段。在一个实施方案中,连接酶是T4 DNA连接酶。参见例如图4B和图5。

[0132] “连接链寡核苷酸”是包含5'磷酸酯并且能够连接至末端修复的DNA片段的每个3'末端的多核苷酸。

[0133] “配偶体链寡核苷酸”与连接链寡核苷酸的一部分或全部核苷酸互补并退火。配偶体链寡核苷酸在其3'端包含防止或基本上抑制配偶体链寡核苷酸连接至另一个衔接子或靶DNA片段的磷酸化5'末端的修饰。可阻断连接的配偶体链的3'末端的化学修饰包括但不限于双脱氧核糖核苷酸类似物、2-羟基脱氧核糖核苷酸类似物和对核糖的多种化学修饰。

[0134] 若干考虑因素涉及用于前衔接子中的连接链寡核苷酸的序列设计和内容物。连接链寡核苷酸的长度可从在DNA连接酶有活性的温度下形成稳定的DNA双链体所需的最小长度(约5nt)至推动当前合成能力极限的寡核苷酸(>200nt)变化。在具体实施方案中,连接链寡核苷酸是约8至约60个核苷酸或约8至约15个核苷酸。

[0135] 作为与DNA片段的NGS分析有关的额外考虑因素,通过连接链并入的DNA碱基由测序仪器用于在整个DNA测序轮次中校准DNA碱基识别。这些仪器的仪器和软件要求所有四个DNA碱基都存在于所测序的初始8-15个核苷酸的整个长度的每一碱基位置,并且这通常包括嵌入连接衔接子链中的碱基。出于这个原因,经常使用在连接链序列长度上互相具有全部四个碱基的四条连接链的组。此类连接链寡核苷酸的非限制性实例显示在表1和表2中。

[0136] 在各种其他实施方案中,连接链寡核苷酸包含以下元件:(i)用于单引物文库扩增的PCR引物结合位点;(ii)用于唯一地鉴定每个测序读数的5核苷酸的读取代码;(iii)8至15个核苷酸的锚定序列,其充当样品鉴定序列,实现测序轮次内的样品复用,实现测序读数中校准正确的碱基识别并充当与配偶体链寡核苷酸杂交的锚。

[0137] 在各种其他实施方案中,连接链寡核苷酸包含8至15个核苷酸的锚定序列,所述锚定序列充当样品鉴定序列,实现测序轮次内的样品复用,实现测序读数中正确碱基识别的校准并充当与配偶体链寡核苷酸杂交的锚。

[0138] 在具体实施方案中,连接链寡核苷酸包含为有效连接底物的一个或多个PCR引物

序列、一个或多个读取代码、一个或多个样品代码、一个或多个锚定序列或两个或更多个3'核苷酸。在另外的实施方案中,连接链寡核苷酸还包含一个或多个测序引物结合位点。

[0139] 在具体实施方案中,连接链寡核苷酸包含用于扩增DNA文库的一个或多个PCR引物结合序列。在一个实施方案中,所述PCR引物结合序列是约12至约40个核苷酸、约18至约40个核苷酸、约20至约35个核苷酸或约20至约30个核苷酸。在另一个实施方案中,所述PCR引物结合序列是约12个核苷酸、约13个核苷酸、约14个核苷酸、约15个核苷酸、约16个核苷酸、约17个核苷酸、约18个核苷酸、约19个核苷酸、约20个核苷酸、约21个核苷酸、约22个核苷酸、约23个核苷酸、约24个核苷酸、约25个核苷酸、约26个核苷酸、约27个核苷酸、约28个核苷酸、约29个核苷酸、约30个核苷酸、约31个核苷酸、约32个核苷酸、约33个核苷酸、约34个核苷酸、约35个核苷酸、约36个核苷酸、约37个核苷酸、约38个核苷酸、约39个核苷酸或约40个核苷酸或更多。

[0140] 在一个实施方案中,所述PCR引物结合序列是约25个核苷酸。

[0141] 在具体实施方案中,连接链寡核苷酸包含一个或多个读取代码序列。如本文所用,术语“读取代码”是指用于鉴定独特测序读数的多核苷酸。在一个实施方案中,读取代码是核苷酸的随机序列。在一个实施方案中,读取代码是约1个核苷酸、约2个核苷酸、约3个核苷酸、约4个核苷酸、约5个核苷酸、约6个核苷酸、约7个核苷酸、约8个核苷酸、约9个核苷酸、约10个核苷酸或更多。

[0142] 作为非限制性实例,5核苷酸读取代码由256个可能的独特序列组成,其中每个选择的代码是与所述组中每个其他代码不同的2个核苷酸。这种特征使独特且不同的读数能够区别于由于代码区域中的测序误差而显得独特的读数。在具体实施方案中,由于特定序列组合,已经凭经验确定为干扰衔接子功能的代码可被排除使用,例如,256个中的7个代码具有G核苷酸的过代表并被排除。

[0143] 在其他实施方案中,5、6、7、8、9、10个或更多个核苷酸的每个读取代码可与每个其他读取代码相差2、3、4或5个核苷酸。

[0144] 在一个实施方案中,所述读取代码是约5个核苷酸并任选地与每个其他读取代码相差2个核苷酸。

[0145] 在具体实施方案中,连接链寡核苷酸包含一个或多个样品代码序列。如本文所用,术语“样品代码”是指用于鉴定样品的多核苷酸。样品代码还适用于建立多重测序反应,因为每个样品代码对样品都是唯一的,且因此可于鉴定来自多重测序反应内的特定样品的读数。

[0146] 在一个实施方案中,所述样品代码包含约1个核苷酸、约2个核苷酸、约3个核苷酸、约4个核苷酸或约5个核苷酸或更多个核苷酸的序列。在另一个实施方案中,2、3、4、5或更多个核苷酸的每个样品代码可与每个其他样品代码相差2、3、4或5个核苷酸。

[0147] 在一个实施方案中,所述样品代码是约三个核苷酸,并且与其他样品中使用的每个其他样品代码相差两个核苷酸。

[0148] 在具体实施方案中,连接链寡核苷酸包含一个或多个锚定序列。如本文所用,“锚定序列”是指与配偶体链寡核苷酸杂交并包含以下性质的至少8个核苷酸、至少10个核苷酸、至少12个核苷酸、至少14个核苷酸或至少16个核苷酸的核苷酸序列:(1) 每个锚定序列是四个锚定序列家族的一部分,所述四个锚定序列共同代表延伸内每个位点处的四个可能



的DNA碱基中的每一个;在具体实施方案中,这一特征,平衡碱基表示适用于校准测序读数中的正确碱基识别;以及(2)每个锚定序列由相同数目的A+C和G+T组成,且因此每个锚定序列与一组四个中每个其他锚定序列共有大致相同的解链温度和双链体稳定性。在一个实施方案中,所述锚定序列或其一部分还用于鉴定样品,实现测序轮次内的样品复用,实现测序读数中正确碱基识别的校准并充当与配偶体链寡核苷酸杂交的锚。

[0149] 此外,在设计非连接配偶体链寡核苷酸涉及若干考虑因素。配偶体链寡核苷酸与形成磷酸化平末端的区域中的连接链寡核苷酸至少部分互补(>5nt)。其次,对配偶体链寡核苷酸的3'末端进行修饰以阻断或基本上抑制寡核苷酸变成连接底物,特别是形成自连接的衔接子二聚体。配偶体链寡核苷酸被设计为在进行连接的温度( $\leq 22^{\circ}\text{C}$ )下与连接链形成稳定的双链体,但也被设计为在修复寡核苷酸并入衔接子的温度下( $\geq 37^{\circ}\text{C}$ )与连接链寡核苷酸解离。这种设计考虑被描绘为图4B和图5中所示的产生衔接子/末端修复的DNA复合物中解离的配偶体链寡核苷酸,因为所述反应从连接转换为由修复寡核苷酸介导的衔接子完成步骤。

[0150] 在具体实施方案中,本文考虑的组合物和方法包含连接步骤,其中前衔接子被连接至末端修复的DNA以产生“标记的”DNA文库。在一些实施方案中,采用单一物种的前衔接子。在一些实施方案中,采用2、3、4或5种物种的前衔接子。在一些实施方案中,相同序列的前衔接子被连接至片段化末端修复的DNA的每个末端。

[0151] 在一个实施方案中,将多种预衔接子物种连接至末端修复的DNA文库。多个预衔接子中的每个可包含用于扩增DNA文库的一个或多个引物结合位点、一个或多个读取代码序列、一个或多个用于样品复用的序列、一个或多个锚定序列或一个或多个用于DNA测序的序列。

[0152] (f) 衔接子/末端修复的DNA复合物的形成

[0153] 在具体实施方案中,本文考虑的组合物和方法包括将配偶体链寡核苷酸从前衔接子/末端修复的DNA复合体移位并用修复寡核苷酸置换移位的配偶体链寡核苷酸以产生衔接子/末端修复的DNA复合物。参见例如图3。在具体实施方案中,可操纵衔接子的设计以实现单引物或双引物扩增策略。参见例如图4A和图5。

[0154] 在具体实施方案中,本文考虑的组合物和方法包含连接步骤,其中包含连接链寡核苷酸和修复寡核苷酸的衔接子被连接至末端修复的DNA以产生“标记的”DNA文库。在一些实施方案中,采用单一物种的衔接子。在一些实施方案中,采用2、3、4或5种物种的衔接子。在一些实施方案中,相同序列的衔接子被连接至片段化末端修复的DNA的每个末端。

[0155] 配偶体链寡核苷酸的设计考虑使其从前衔接子/末端修复的DNA复合物移位,因为其在修复寡核苷酸与连接链寡核苷酸退火的温度(例如 $>37^{\circ}\text{C}$ )下并在进行酶步骤以将修复寡核苷酸并入衔接子/末端修复的DNA复合物以产生连续的双链DNA文库分子的温度(例如 $>37^{\circ}\text{C}$ )下从连接链寡核苷酸解离。

[0156] 如本文所用,术语“修复寡核苷酸”是指与连接链寡核苷酸的一部分或全部核苷酸互补并退火的多核苷酸序列。修复寡核苷酸的长度可从在DNA连接酶有活性的温度下形成稳定的DNA双链体所需的最小长度(约8nt)至推动当前合成能力极限的寡核苷酸( $>200\text{nt}$ )变化。在具体实施方案中,“修复寡核苷酸”包括不一定存在于连接链寡核苷酸中的另外的功能性DNA序列。

[0157] 在具体实施方案中,连接链寡核苷酸是约8至约15个核苷酸,并且修复寡核苷酸是35至60个核苷酸。在这种设计中,配体链寡核苷酸的序列通过引物延伸而延伸并产生与修复寡核苷酸互补的核苷酸序列。这种设计将产生相同的PCR引物结合位点。相同的PCR引物结合位点允许单引物文库扩增策略。参见例如图3D和图4A。

[0158] 在具体实施方案中,连接链寡核苷酸是约35至约60个核苷酸,并且修复寡核苷酸与连接链寡核苷酸完全互补。相同的PCR引物结合位点允许单引物文库扩增策略。参见例如图4A。

[0159] 在具体实施方案中,连接链寡核苷酸是约35至约60个核苷酸,并且修复寡核苷酸是约35至约60个核苷酸,并且所述两种寡核苷酸是互补的,但是对于PCR引物结合位点是互补的。不同的PCR引物结合位点允许双引物文库扩增策略。参见例如图5。

[0160] 在优选的实施方案中,连接链寡核苷酸包含以下元件:(i)用于单引物文库扩增的PCR引物结合位点;(ii)5个核苷酸读取代码,其用于唯一地鉴定每个测序读取;(iii)8至15个核苷酸的锚定序列,其与连接链寡核苷酸的锚定序列部分或完全互补。

[0161] 在其他实施方案中,所述连接链寡核苷酸包含与连接链寡核苷酸的锚定序列部分或完全互补的8至15个核苷酸的锚定序列。

[0162] 在具体实施方案中,修复寡核苷酸包含为有效连接底物的一个或多个PCR引物序列、一个或多个读取代码、一个或多个样品代码、一个或多个锚定序列或两个或更多个3'核苷酸。在另外的实施方案中,修复寡核苷酸还包含一个或多个测序引物结合位点。

[0163] 在具体实施方案中,修复寡核苷酸包含(i)与连接链寡核苷酸中的PCR引物结合位点互补的一个或多个PCR引物结合序列(实现单引物DNA文库扩增)或(ii)不与连接链寡核苷酸中的PCR引物结合位点互补的一个或多个PCR引物结合序列(实现双引物DNA文库扩增)。在一个实施方案中,所述PCR引物结合序列是约12至约40个核苷酸、约18至约40个核苷酸、约20至约35个核苷酸或约20至约30个核苷酸。在另一个实施方案中,所述PCR引物结合序列是约12个核苷酸、约13个核苷酸、约14个核苷酸、约15个核苷酸、约16个核苷酸、约17个核苷酸、约18个核苷酸、约19个核苷酸、约20个核苷酸、约21个核苷酸、约22个核苷酸、约23个核苷酸、约24个核苷酸、约25个核苷酸、约26个核苷酸、约27个核苷酸、约28个核苷酸、约29个核苷酸、约30个核苷酸、约31个核苷酸、约32个核苷酸、约33个核苷酸、约34个核苷酸、约35个核苷酸、约36个核苷酸、约37个核苷酸、约38个核苷酸、约39个核苷酸或约40个核苷酸或更多。

[0164] 在一个实施方案中,所述PCR引物结合序列是约25个核苷酸。

[0165] 在具体实施方案中,修复寡核苷酸包含一个或多个读取代码序列。在一个实施方案中,读取代码是核苷酸的随机序列。在一个实施方案中,读取代码是约1个核苷酸、约2个核苷酸、约3个核苷酸、约4个核苷酸、约5个核苷酸、约6个核苷酸、约7个核苷酸、约8个核苷酸、约9个核苷酸、约10个核苷酸或更多。

[0166] 作为非限制性实例,5核苷酸读取代码由256个可能的独特序列组成,其中每个选择的代码是与所述组中每个其他代码不同的2个核苷酸。这种特征使独特且不同的读数能够区别于由于代码区域中的测序误差而显得独特的读数。在具体实施方案中,由于特定序列组合,已经凭经验确定为干扰衔接子功能的代码可被排除使用,例如,256个中的7个代码具有G核苷酸的过代表并被排除。

[0167] 在其他实施方案中,5、6、7、8、9、10个或更多个核苷酸的每个读取代码可与每个其他读取代码相差2、3、4或5个核苷酸。

[0168] 在一个实施方案中,所述读取代码是约5个核苷酸并任选地与每个其他读取代码相差2个核苷酸。

[0169] 在具体实施方案中,修复寡核苷酸包含一个或多个样品代码序列。在一个实施方案中,所述样品代码包含约1个核苷酸、约2个核苷酸、约3个核苷酸、约4个核苷酸或约5个核苷酸或更多个核苷酸的序列。在另一个实施方案中,2、3、4、5或更多个核苷酸的每个样品代码可与每个其他样品代码相差2、3、4或5个核苷酸。

[0170] 在一个实施方案中,所述样品代码是约三个核苷酸,并且与其他样品中使用的每个其他样品代码相差两个核苷酸。

[0171] 在具体实施方案中,修复寡核苷酸包含与连接链寡核苷酸的一个或多个锚定序列互补的一个或多个锚定序列。

[0172] 不希望受任何特定理论束缚,考虑将修复寡核苷酸并入衔接子/末端修复DNA复合物中的至少两种示例性策略。

[0173] 在一个实施方案中,使配偶体链寡核苷酸从前衔接子/末端修复的DNA复合物移位;添加修复寡核苷酸并使其与连接链退火;使用多核苷酸激酶,例如T4多核苷酸激酶来向所述末端修复的DNA片段的5'末端添加磷酸酯基团;并使用DNA连接酶来修复存在于所述修复寡核苷酸的5'末端与所述末端修复的DNA片段的3'末端之间的切口。在具体实施方案中,所述DNA连接酶是在广泛温度范围内具有活性的热稳定的切口特异性连接酶,包括但不限于Taq DNA连接酶、大肠杆菌DNA连接酶、9°North连接酶(NEB)和任何其他可封闭磷酸化切口的连接酶。参见例如图3A和图3B。

[0174] 在另一个实施方案中,配偶体链寡核苷酸从前衔接子/末端修复的DNA复合物移位;添加修复寡核苷酸并使其与连接链退火;具有5'→3'外切核酸酶活性(并且没有内在3'→5'外切核酸酶活性)的低持续合成能力DNA聚合酶延伸连接链寡核苷酸的3'末端,并且此外除去具有5'→3'外切核酸酶活性的去磷酸化的5'末端核苷酸和邻近的核苷酸,从而暴露可连接的5'磷酸酯基团并用并入的碱基置换它们,所述碱基在酶解离时留下切口;并且DNA连接酶(例如Taq DNA连接酶)用于修复切口。

[0175] 适用于本文考虑的组合物和方法的具体实施方案的低持续合成能力DNA聚合酶的说明性实例包括但不限于Taq DNA聚合酶和BstI DNA聚合酶。

[0176] D. DNA文库扩增

[0177] 在具体实施方案中,本文考虑的方法包括扩增DNA文库以产生DNA克隆文库或DNA克隆的文库。在具体实施方案中,所述DNA是cfDNA。DNA文库的每个分子包含与末端修复的DNA的每个末端连接的衔接子,并且每个衔接子包含一个或多个PCR引物结合位点。在一个实施方案中,不同的衔接子被连接至末端修复的DNA的不同末端。

[0178] 在一个实施方案中,同一衔接子被连接至DNA的两个末端。将同一衔接子连接至末端修复的DNA的两个末端允许用单引物序列进行PCR扩增。在具体实施方案中,将使用标准PCR技术用驱动扩增的单引物序列来扩增衔接子连接的DNA文库的一部分。在一个实施方案中,单引物序列是约25个核苷酸,任选地在标准离子强度条件下具有 $\geq 55^{\circ}\text{C}$ 的预测 $T_m$ 。

[0179] 在一个实施方案中,与末端修复的DNA片段的3'末端连接的衔接子包含与连接至

末端修复的DNA片段的5'末端的衔接子不同的PCR引物结合位点。在具体实施方案中,将使用标准PCR技术用驱动扩增的两种引物来扩增衔接子连接的DNA文库的一部分。

[0180] 在具体实施方案中,皮克的初始DNA文库被扩增成微克的DNA克隆,从而意味着10,000倍扩增。扩增产物的量可使用本领域中已知的方法测量,例如在Qubit 2.0或Nanodrop仪器上定量。

[0181] E.DNA的遗传分析的方法

[0182] 在各种实施方案中,提供一种用于DNA的遗传分析的方法。在具体实施方案中,所述DNA是cfDNA。cfDNA是在血浆或其他体液中发现的无细胞DNA。

[0183] 在具体实施方案中,用于DNA的遗传分析的方法包括:产生并扩增DNA文库,测定所述DNA文库中基因组等效物的数量;以及进行一个或多个基因组靶基因座的定量遗传分析。

[0184] 1.测定基因组等效物的数量

[0185] 在各种实施方案中,用于DNA的遗传分析的方法包括测定所述DNA克隆文库中基因组等效物的数量。如本文所用,术语“基因组等效物”是指每个文库中基因组拷贝的数量。本文考虑的组合物和方法所遇到的重要挑战是实现足够的测定灵敏度以检测并分析罕见遗传突变或遗传序列的差异。为了逐个样品地确定测定灵敏度值,通过测量存在于测序文库中的基因组等效物的数量来测量每个样品中存在的不同且独特的序列的数量。为了建立灵敏度,必须针对每个样品文库测量基因组等效物的数量。

[0186] 基因组等效物的数量可通过qPCR测定或通过在进行测序后使用基于生物信息学的计数来确定。在临床样品的工艺流程中,基因组等效物的qPCR测量被用作DNA文库的QC步骤。它在序列分析之前建立了测定灵敏度的预期,并且如果其相应的DNA克隆文库缺乏所需的基因组等效物深度,则可将样品从分析中排除。最终,基因组等效物的基于生物信息学的计数也可用于鉴定每个给定DNA克隆文库的基因组等效物,且因此鉴定测定灵敏度和假阴性估计值。

[0187] 经验性qPCR测定和统计计数分析应良好地相关。在测序未能揭示DNA克隆文库中的序列深度的情况下,可能需要对DNA克隆文库进行再处理和/或另外的测序。

[0188] 在一个实施方案中,使用定量PCR (qPCR) 测定来确定DNA克隆文库中的基因组等效物。在一个具体实施方案中,使用已知浓度的标准文库来构建标准曲线,并将来自qPCR测定的测量值拟合至所得标准曲线,并从拟合中导出基因组等效物的值。通过基于重复序列的测定测量的基因组等效物的数量提供更一致的文库至文库性能以及在测序轮次中的基因组等效物的qPCR估计值与生物信息学计数的标签等效物之间的更好比对。

[0189] 适用于本文考虑的基于重复序列的基因组等效物测定的重复的说明性实例包括但不限于:短散布核元件(SINE),例如Alu重复序列;长散布核元件(LINE),例如LINE1、LINE2、LINE3;微卫星重复元件,例如短串联重复序列(STR)、简单序列重复序列(SSR);以及哺乳动物全散布重复序列(MIR)。

[0190] 在一个实施方案中,重复序列是Alu重复序列。

[0191] 2.定量遗传分析

[0192] 在各种实施方案中,用于DNA的遗传分析的方法包括DNA文库克隆的一个或多个靶遗传基因座的定量遗传分析。定量遗传分析包括以下步骤中的一个或多个或全部:捕获包含靶遗传基因座的DNA克隆;扩增所捕获的靶向遗传基因座;对所扩增的捕获的靶向遗传基

因座进行测序;以及对所得序列读数进行生物信息学分析。

[0193] (a) 靶遗传基因座的捕获

[0194] 本发明部分地考虑捕获探针模块,所述捕获探针模块是多功能的并被设计为保留较大探针的效率和可靠性,但是使DNA克隆文库中的无信息序列生成最小化。“捕获探针模块”是指包含捕获探针序列和尾部序列的多核苷酸。在具体实施方案中,捕获探针模块序列或其部分用作一种或多种测序引物的引物结合位点。

[0195] 在具体实施方案中,捕获探针模块包括捕获探针。如本文所用,“捕获探针”是指包含能够与特定DNA靶区杂交的区域的核苷酸。由于DNA的平均大小相对较小并且高度片段化,因此本文考虑的组合物和方法包括使用高密度和相对较短的捕获探针来查询目标DNA靶区。

[0196] 在具体实施方案中,捕获探针模块与任选地包含半抗原并与尾部序列杂交以产生捕获探针模块双链体的配偶体寡核苷酸组合。

[0197] 使用高密度捕获探针的一个特别的问题是,通常捕获探针是使用特定“序列规则”来设计的。例如,在设计捕获探针时通常不包括冗余序列的区域或表现出极端碱基组成偏差的区域。然而,本发明人已经发现,捕获探针设计规则缺乏灵活性不会显著影响探针性能。相比之下,通过位置约束严格选择的捕获探针提供中靶序列信息;表现出非常少的脱靶和不可映射的读取捕获;并产生统一的、有用的中靶读数,只有少数例外。此外,近距探针间距处的高冗余度足以补偿偶然性能差的捕获探针。

[0198] 在具体实施方案中,靶区被多个捕获探针靶向,其中任何两个或更多个捕获探针被设计为结合至在彼此的10个核苷酸内、彼此的15个核苷酸内、彼此的20个核苷酸内、彼此的25个核苷酸内、彼此的30个核苷酸内、彼此的35个核苷酸内、彼此的40个核苷酸内、彼此的45个核苷酸内或彼此的50个核苷酸或更多内以及所有居间核苷酸长度的靶区。

[0199] 在一个实施方案中,捕获探针是约25个核苷酸、约26个核苷酸、约27个核苷酸、约28个核苷酸、约29个核苷酸、约30个核苷酸、约31个核苷酸、约32个核苷酸、约33个核苷酸、约34个核苷酸、约35个核苷酸、约36个核苷酸、约37个核苷酸、约38个核苷酸、约39个核苷酸、约40个核苷酸、约41个核苷酸、约42个核苷酸、约43个核苷酸、约44个核苷酸或约45个核苷酸。

[0200] 在一个实施方案中,捕获探针是约100个核苷酸、约200个核苷酸、约300个核苷酸、约400个核苷酸或约100个核苷酸。在另一个实施方案中,捕获探针是约100个核苷酸至约500个核苷酸、约200个核苷酸至约500个核苷酸、约300个核苷酸至约500个核苷酸或约400个核苷酸至约500个核苷酸,或其任何居间范围。

[0201] 在具体实施方案中,捕获探针是60个核苷酸。

[0202] 在具体实施方案中,捕获探针不是60个核苷酸。

[0203] 在另一个实施方案中,捕获探针基本上小于60个核苷酸,但比靶向同一DNA靶区的60个核苷酸捕获探针可比较地以及更好地杂交。

[0204] 在某一实施方案中,捕获探针是40个核苷酸。

[0205] 在某些实施方案中,捕获探针模块包含尾部序列。如本文所用,术语“尾部序列”是指在捕获探针模块的5'末端处的多核苷酸,其在特定实施方案中可充当引物结合位点。在具体实施方案中,测序引物与尾部区域中的引物结合位点结合。

[0206] 在具体实施方案中,尾部序列是约5至约100个核苷酸、约10至约100个核苷酸、约5至约75个核苷酸、约5至约50个核苷酸、约5至约25个核苷酸或约5至约20个核苷酸。在某些实施方案中,第三区域是约10至约50个核苷酸、约15至约40个核苷酸、约20至约30个核苷酸或约20个核苷酸,或任何居间数量的核苷酸。

[0207] 在具体实施方案中,尾部序列是约30个核苷酸、约31个核苷酸、约32个核苷酸、约33个核苷酸、约34个核苷酸、约35个核苷酸、约36个核苷酸、约37个核苷酸、约38个核苷酸、约39个核苷酸或约40个核苷酸。

[0208] 在各种实施方案中,捕获探针模块包含结合对的特定成员,以使得能够分离和/或纯化与所述捕获探针杂交的标记的和/或扩增的DNA文库的一个或多个捕获的片段。在具体实施方案中,捕获探针模块与生物素或另一种合适的半抗原,例如二硝基苯酚、地高辛缀合。

[0209] 在各种实施方案中,捕获探针模块与标记的和任选扩增的DNA文库杂交以形成复合物。在一些实施方案中,多功能捕获探针模块基本上与DNA文库中的特定基因组靶区杂交。

[0210] 杂交或杂交条件可包括其中两个核苷酸序列形成稳定复合物的任何反应条件;例如标记的DNA文库和捕获探针模块形成稳定的标记DNA文库-捕获探针模块复合物。此类反应条件是本领域众所周知的,并且本领域技术人员将理解,可适当地修改此类条件,例如用更短长度的捕获探针降低退火温度,并且在本发明的范围内。当捕获探针复合物的第二区域表现出与标记的DNA文库的区域的100%、99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、89%、88%、85%、80%、75%或70%序列同一性、同源性或互补性时,可发生实质性杂交。

[0211] 在具体实施方案中,捕获探针是约40个核苷酸并具有约44℃至约47℃的最佳退火温度。

[0212] 在某些实施方案中,本文考虑的方法包括分离标记的DNA文库-捕获探针模块复合物。在具体实施方案中,用于分离DNA复合物的方法是本领域技术人员熟知的,并且本领域技术人员认为适当的任何方法都可用于本发明的方法(Ausubel等人,Current Protocols in Molecular Biology,2007-2012)。在具体实施方案中,使用生物素-链霉抗生物素蛋白分离技术分离复合物。在一些实施方案中,对能够与多功能捕获探针模块的尾部序列杂交的捕获配偶体寡核苷酸进行修饰以在5'末端或3'末端含有生物素,其能够与连接至柱、珠或用于DNA复合物分离方法的其他底物相互作用。

[0213] 在一个实施方案中,对能够与多功能捕获探针模块的尾部序列杂交的捕获配偶体寡核苷酸进行修饰以在3'末端含有生物素,其能够与连接至柱、珠或用于DNA复合物分离方法的其他底物相互作用。

[0214] 在具体实施方案中,多功能捕获探针模块的尾部序列结合至捕获配偶体寡核苷酸。在一些实施方案中,多功能捕获探针模块在形成标记的DNA文库-多功能捕获探针模块复合物之前结合至捕获配偶体寡核苷酸。在一些实施方案中,多功能捕获探针模块在形成标记的DNA文库-多功能捕获探针模块复合物之后结合至捕获配偶体寡核苷酸。在一些实施方案中,多功能捕获探针模块与形成标记的DNA文库-多功能捕获探针模块复合物同时结合至捕获配偶体寡核苷酸。在一些实施方案中,对捕获配偶体寡核苷酸进行化学修饰。在一个

实施方案中,捕获配偶体寡核苷酸通过向5'或3'末端添加半抗原而进行修饰。在一个实施方案中,半抗原是生物素。

[0215] 在具体实施方案中,考虑从分离的标记的DNA文库-捕获探针模块复合物中除去单链3'-末端。在某些实施方案中,所述方法包括分离的标记的DNA文库-多功能捕获探针模块复合物的3'-5'外切核酸酶加工以除去单链3'末端。

[0216] 在某些其他实施方案中,所述方法包括利用分离的标记的DNA文库片段作为模板进行多功能捕获探针的5'-3'DNA聚合酶延伸。

[0217] 在某些其他实施方案中,所述方法包括通过5'FLAP内切核酸酶的协同作用、DNA聚合和通过DNA连接酶的切口闭合产生杂合捕获探针分离的标记的DNA靶分子。

[0218] 多种酶可用于分离的标记的DNA文库-多功能捕获探针模块复合物的3'-5'外切核酸酶加工。可用于具体实施方案中的表现出3'-5'外切核酸酶活性的合适酶的说明性实例包括但不限于:T4或外切核酸酶I、III、V(还参见,Shevelev IV,Hübscher U.,“The 3'5'exonucleases,”Nat Rev Mol Cell Biol.3(5):364-76(2002))。在具体实施方案中,包含3'-5'外切核酸酶活性的酶是T4。在具体实施方案中,可使用表现出3'-5'外切核酸酶活性并能够引物模板延伸的酶,包括例如T4或外切核酸酶I、III、V。同上。

[0219] 在一些实施方案中,本文考虑的方法包括对上文和本文其他地方论述的3'-5'外切核酸酶加工的复合物进行测序和/或PCR。在具体实施方案中,对捕获探针分子的尾部部分进行拷贝以产生杂合核酸分子。在一个实施方案中,所产生的杂合核酸分子包含能够与捕获探针模块和捕获探针模块尾部序列的补体杂交的靶区。

[0220] 在一个具体的实施方案中,遗传分析包括a) 将一个或多个捕获探针模块与多个DNA文库克隆中的一个或多个靶遗传基因座杂交以形成一种或多种捕获探针模块-DNA文库克隆复合物;b) 从a) 中分离所述一种或多种捕获探针模块-DNA文库克隆复合物;c) 对来自步骤b) 的一种或多种分离的捕获探针模块-DNA文库克隆复合物进行酶加工;d) 对来自c) 的酶加工的复合物进行PCR,其中拷贝捕获探针分子的尾部部分以便产生扩增的杂合核酸分子,其中所述扩增的杂合核酸分子在靶基因组基因座中包含能够与捕获探针和捕获探针模块尾部序列的补体杂交的靶序列;以及e) 对来自d) 的扩增的杂合核酸分子进行定量遗传分析。

[0221] 在具体实施方案中,考虑了用于确定特定靶遗传基因座的拷贝数的方法,所述包括:a) 将一个或多个捕获探针模块与多个DNA文库克隆中的一个或多个靶遗传基因座杂交以形成一种或多种捕获探针模块-DNA文库克隆复合物;b) 从a) 中分离所述一种或多种捕获探针模块-DNA文库克隆复合物;c) 对来自步骤b) 的一种或多种分离的捕获探针模块-DNA文库克隆复合物进行酶加工;d) 对来自c) 的酶加工的复合物进行PCR,其中拷贝捕获探针分子的尾部部分以便产生扩增的杂合核酸分子,其中所述扩增的杂合核酸分子在靶遗传基因座中包含能够与捕获探针和捕获探针模块尾部序列的补体杂交的靶序列;e) 进行d) 中的扩增的杂合核酸分子的PCR扩增;以及f) 定量e) 中的PCR反应,其中所述定量允许确定特定靶区的拷贝数。

[0222] 在一个实施方案中,步骤c) 的酶加工包括使用具有3'-5'外切核酸酶活性的酶对来自b) 的一种或多种捕获探针模块-DNA文库克隆复合物进行3'-5'外切核酸酶加工以除去单链3'末端;通过5'FLAP内切核酸酶的协同作用、DNA聚合和通过DNA连接酶进行的切口

闭合产生一个或多个杂合捕获探针模块-DNA文库克隆分子;或使用所述复合物中的分离的DNA克隆作为模板进行捕获探针的5'-3'DNA聚合酶延伸。

[0223] 在一个实施方案中,步骤c)的酶加工包括使用所述复合物中的分离的DNA克隆作为模板进行捕获探针的5'-3'DNA聚合酶延伸。

[0224] 在具体实施方案中,PCR可使用本领域技术人员熟知的任何标准PCR反应条件进行。在某些实施方案中,e)中的PCR反应使用两种PCR引物。在一个实施方案中,e)中的PCR反应使用第一PCR引物,其与靶遗传基因座内的重复序列杂交。在一个具体实施方案中,e)中的PCR反应使用第二PCR引物,其与靶遗传基因座/尾部接点处的杂合核酸分子杂交。在某些实施方案中,e)中的PCR反应使用与靶遗传基因座杂交的第一PCR引物,并且第二PCR引物与靶遗传基因座/尾部接点处的扩增的杂合核酸分子杂交。在具体实施方案中,所述第二引物与靶遗传基因座/尾部接点杂交,以使得引物的至少一个或多个核苷酸与靶遗传基因座杂交,并且引物的至少一个或多个核苷酸与尾部序列杂交。

[0225] 在某些实施方案中,对从步骤e)获得的扩增的杂合核酸分子进行测序,并且将所述序列水平地比对,即彼此比对但不与参考序列比对。在具体实施方案中,步骤a)至e)用一个或多个捕获探针模块重复一次或多次。捕获探针模块可相同或不同,并被设计成靶向靶遗传基因座的任一DNA链。在一些实施方案中,当捕获探针不同时,它们在标记的DNA克隆文库中的靶遗传基因座内的重叠或相邻的靶序列处杂交。在一个实施方案中,使用高密度捕获探针策略,其中多个捕获探针与靶遗传基因座杂交,并且其中所述多个捕获探针中的每个与杂交于标记的DNA克隆文库中的靶遗传基因座的任何其他捕获探针的约5、10、15、20、25、30、35、40、45、50、100、200bp(包括所有居间距离)内的靶遗传基因座杂交。

[0226] 在一些实施方案中,所述方法可使用每个靶遗传基因座两个捕获探针模块进行,其中一个捕获探针模块与靶区上游的“沃森”链(非编码或模板链)杂交,并且一个捕获探针模块与靶区下游的“克里克”链杂交(编码或非模板链)。

[0227] 在具体实施方案中,本文考虑的方法可进一步用任何数量的捕获探针模块进行多次,例如每个靶遗传基因座2、3、4、5、6、7、8、9或10个或更多个捕获探针模块,任何数量的所述捕获探针模块以任何组合与沃森或克里克链杂交。在一些实施方案中,所获得的序列可彼此进行比对以便鉴定许多差异中的任何差异。

[0228] 在某些实施方案中,查询多个靶遗传基因座,例如使用一个或多个捕获探针模块在单一反应中查询100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、3500、4000、4500、5000、10000、50000、100000、500000或更多个靶遗传基因座。

[0229] (b) 测序

[0230] 在具体实施方案中,定量遗传分析包括对多个杂合核酸分子进行测序,如上文本其他地方所述,以产生足够的测序深度以获得多个独特的测序读数。独特的读数被定义为来自读数的全部在DNA内具有相同的读取代码和序列起始点的“家族”中读取的单个共有序列。每个捕获探针产生一组独特的读数,所述读数通过分组到家族中而从总读数中计算得到。然后将给定样品的独特读数计算为在逐探针基础上观察到的所有独特读数的平均值。从用于计算平均值的数据集中排除存在明显拷贝数变化的情况。独特的读数是重要的,因为每个独特的读数必须源自独特的DNA克隆。每个独特的读数代表基因组DNA的单倍体等效物的输入和分析。独特读数的总和是所分析的单倍体基因组的总和。所分析的基因组的



数量进而决定测序测定的灵敏度。作为非限制性实例,如果平均独特读数计数是100基因组等效物,则所述特定测定具有能够检测100中读取的一个突变体或1%的灵敏度。任何小于这个的观察结构都是不可抵抗的。

[0231] 在具体实施方案中,定量遗传分析包括源自多个样品的杂合核酸分子的多重测序。

[0232] 在各种实施方案中,定量遗传分析包括获得一个或多个或多个标记的DNA文库克隆,每个克隆包含第一DNA序列和第二DNA序列,其中所述第一DNA序列包含靶向遗传基因座中的序列,并且所述第二DNA序列包含捕获探针序列;对所述一个或多个克隆进行配对末端测序反应,并获得一个或多个测序读数,或对所述一个或多个克隆进行测序反应,其中获得大于约100、200、300、400、500个或更多个核苷酸的单个长测序读数,其中所述读数足以鉴定所述第一DNA序列和所述第二DNA序列两者;以及根据测序读数的探针序列对一个或多个克隆的测序读数进行排序或聚类。

[0233] (c) 生物信息学分析

[0234] 在各种实施方案中,定量遗传分析还包括对测序读数的生物信息学分析。生物信息学分析不包括在不在于用于测序的组合物或方法的情况下进行的任何纯心理分析。在某些实施方案中,生物信息学分析包括但不限于:序列比对;基因组等效物分析;单核苷酸变体(SNV)分析;基因拷贝数变异(CNV)分析;以及遗传损伤的检测。在具体实施方案中,生物信息学分析适用于定量在DNA克隆文库中分析的基因组等效物的数量;检测靶遗传基因座的遗传状态;检测靶遗传基因座中的遗传损伤;以及测量靶遗传基因座内的拷贝数波动。

[0235] 序列比对可在序列读数与一个或多个参考DNA序列之间进行。在具体实施方案中,测序比对可用于检测靶遗传基因座中的遗传损伤,包括但不限于检测核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。检测作为因果关系或预后指标的遗传损伤可有助于特定遗传病状或疾病的诊断、预后、治疗和/或监测。

[0236] 本文还考虑了用于序列比对分析的方法,所述方法可在不需要与参考序列比对的情况下进行,在本文中称为水平序列分析。这种分析可对通过本文考虑的方法或任何其他方法产生的任何序列进行。在具体实施方案中,序列分析包括对通过本文考虑的方法获得的读数进行序列比对。

[0237] 在一个实施方案中,在进行测序后,使用基于生物信息学的计数来确定DNA克隆文库中的基因组等效物。每个测序读取与特定捕获探针相关,并且分配给每个捕获探针的读取集合被解析成组。在一组内,单独读数集在基因组序列内共有同一读取代码和同一DNA序列起始位置。这些单独读数被分组为“家族”,并且代表这一家族的单一共有序列作为“独特读数”结转。构成家族的所有单独读数源自单一连接事件,且因此它们是来自扩增的彼此的“同胞”。每个独特读数被认为是独特的连接事件,并且独特读数的和被认为等效于所分析的基因组等效物的数量。

[0238] 随着独特克隆的数量接近可能的序列组合的总数,概率指示同一代码和起始位点组合将由独立事件创建,并且这些独立事件将不适合分组在单一家族中。最终的结果将低估所分析的基因组等效物,并且罕见的突变体读数可作为测序误差而被丢弃,因为它们与携带相同标识符的野生型读数重叠。

[0239] 在具体实施方案中,为了提供DNA克隆文库的准确分析,所分析的基因组等效物的

数量是可能的独特克隆数量的约1/10、约1/12、约1/14、约1/16、约1/18、约1/20、约1/25或更少。应理解,以上概述的程序仅是说明性而非限制性的。

[0240] 在一些实施方案中,可能需要增加待分析的基因组等效物的数量。为了扩展基因组等效物的深度,可考虑至少两种解决方案。第一种解决方案是每个样品使用多于一个衔接子组。通过组合衔接子,有可能以倍增方式扩展可能克隆的总数,且因此扩展基因组输入的舒适限制。第二种解决方案是将读取代码扩展1、2、3、4或5个或更多个碱基。与每个其他阅读代码相差至少2个碱基的可能阅读代码的数量按照 $4^{(n-1)}$ 缩放,其中n是阅读代码内的碱基的数量。因此,在非限制性实例中,如果读读取代码是5个核苷酸且 $4^{(5-1)} = 256$ ;因此,包含额外的碱基将每个额外碱基的可用组库扩展四倍。

[0241] 在一个实施方案中,定量遗传分析包括测序读取的生物信息学分析以鉴定罕见单核苷酸变体(SNV)。

[0242] 新一代测序具有大约0.02%-0.02%的固有误差,这意味着从1/200至1/500碱基识别的任何地方都不正确。为了检测以低于这个频率的频率(例如在每1000个序列1个的频率下)发生的变体和其他突变,有必要调用分子注释策略。作为非限制性实例,使用靶向序列捕获技术对5000个独特分子的分析将产生(在>50,000个读数的足够测序深度下)5000个独特读数的集合,其中每个独特读数属于全部拥有同一阅读代码的读数的“家族”。在家族中出现的SNV是罕见变体的候选者。当在多于一个家族中观察到同一变体时,它成为在起始样品中存在的罕见变体的非常有力的候选者。相比之下,在家族内偶发出现的变体可能是测序误差,并且在一个且仅一个家族内出现的变体是罕见的,或是离体发生的碱基改变的结果(例如,DNA碱基的氧化或PCR-引入误差)。

[0243] 在一个实施方案中,检测SNV的方法包括引入10倍以上的基因组输入(基因组或基因组等效物)作为测定的所需靶标灵敏度。在一个非限制性实例中,如果所需的灵敏度是2%(100中2个),则实验靶标是2000个基因组的输入。

[0244] 在具体实施方案中,测序数据的生物信息学分析用于检测或鉴定与遗传状态、病状或疾病、遗传镶嵌性、胎儿检测、亲子鉴定、预测对药物治疗的反应、诊断或监测医学病状、微生物组谱、病原体筛选或监测器官移植相关的SNV。

[0245] 在各种实施方案中,提供了用于拷贝数测定分析的方法,所述方法包括获得一个或多个或多个克隆,每个克隆包含第一DNA序列和第二DNA序列,其中所述第一DNA序列包含靶向遗传基因座中的序列并且第二DNA序列包含捕获探针序列。在相关实施方案中,对一个或多个克隆进行配对末端测序反应,并且获得一个或多个测序读数。在另一个实施方案中,对一个或多个克隆进行测序反应,其中获得大于约100个核苷酸的单个长测序读数,其中所述读数足以鉴定第一DNA序列和第二DNA序列。可根据测序读数的探针序列对一个或多个克隆的测序读数进行排序或聚类。

[0246] 拷贝数分析包括但不限于检查给定基因组DNA样品中出现的特定基因或突变的拷贝数并可进一步包括给定样品中给定基因的拷贝数或序列差异的定量测定的分析。在具体实施方案中,使用拷贝数分析来检测或鉴定与遗传状态、病状或疾病、胎儿检测、遗传镶嵌性、亲子鉴定、预测对药物治疗的反应、诊断或监测医学病状、微生物组谱、病原体筛选以及监测器官移植相关的基因扩增。

[0247] 在具体实施方案中,测序数据的生物信息学分析用于检测或鉴定靶基因座中的一

个或多个序列或遗传损伤,包括但不限于检测核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。检测作为因果关系或预后指标的遗传损伤可有助于特定遗传病状或疾病的诊断、预后、治疗和/或监测。在一个实施方案中,遗传损伤与遗传状态、病状或疾病、胎儿检测、遗传镶嵌性、亲子鉴定、预测对药物治疗的反应、诊断或监测医学病状、微生物组谱、病原体筛选以及监测器官移植相关。

[0248] F. 定量遗传分析的临床应用

[0249] 在各种实施方案中,本发明考虑了一种检测、鉴定、预测、诊断或监测受试者的病状或疾病的方法。

[0250] 在具体实施方案中,检测、鉴定、预测、诊断或监测受试者的遗传状态、病状或疾病的方法包括对DNA克隆文库中的一个或多个靶遗传基因座进行定量遗传分析以检测或鉴定在一个或多个靶遗传基因座处的序列的变化。在一个实施方案中,DNA是cfDNA。

[0251] 在具体实施方案中,检测、鉴定、预测、诊断或监测选自由以下各项组成的组的遗传状态或遗传病状或疾病的方法:遗传疾病;遗传镶嵌性;胎儿检测;亲子鉴定;亲子鉴定;预测对药物治疗的反应;诊断或监测医学病状;微生物组谱;病原体筛选;以及器官移植监测包括进行DNA克隆文库中的一个或多个靶遗传基因座的定量遗传分析以检测或鉴定一个或多个靶遗传基因座处的序列中的核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。

[0252] 可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的遗传疾病的说明性实例包括但不限于癌症、阿尔茨海默氏病(APOE1)、腓骨肌萎缩症、Leber遗传性视神经病变(LHON)、安格尔曼综合征(UBE3A, 遍在蛋白-蛋白质连接酶E3A)、普拉德-威利综合征(染色体15中的区域)、 $\beta$ -地中海贫血(HBB, $\beta$ -球蛋白)、戈谢病(I型)(GBA, 葡糖脑苷脂酶)、囊性纤维化(CFTR上皮氯离子通道)、镰状细胞病(HBB, $\beta$ -球蛋白)、泰-萨克斯病(HEXA, 己糖胺酶A)、苯酮尿症(PAH, 苯丙氨酸水解酶)、家族性高胆固醇血症(LDLR, 低密度脂蛋白受体)、成人多囊肾病(PKD1, 多囊蛋白)、亨廷顿病(HDD, 亨廷顿蛋白)、I型神经纤维瘤病(NF1, NF1肿瘤抑制基因)、肌强直性营养不良(DM, 肌强直蛋白)、结节性硬化(TSC1, 马铃薯球蛋白)、软骨发育不良(FGFR3, 成纤维细胞生长因子受体)、脆性X综合征(FMR1, RNA结合蛋白)、杜兴肌营养不良症(DMD, 肌萎缩蛋白)、血友病A(F8C, 凝血因子VIII)、莱施-尼汉综合征(HPRT1, 次黄嘌呤鸟嘌呤核糖基转移酶1)以及肾上腺脑白质营养不良(ABCD1)。

[0253] 可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的癌症的说明性实例包括但不限于:B细胞癌,例如多发性骨髓瘤、黑色素瘤、乳腺癌、肺癌(如非小细胞肺癌或NSCLC)、支气管癌、结直肠癌、前列腺癌、胰腺癌、胃癌、卵巢癌、膀胱癌、脑或中枢神经系统癌症、外周神经系统癌症、食道癌、宫颈癌、子宫癌或子宫内膜癌、口腔癌或咽癌、肝癌、肾癌、睾丸癌、胆道癌、小肠或阑尾癌、唾液腺癌、甲状腺癌、肾上腺癌、骨肉瘤、软骨肉瘤、血液组织癌症、腺癌、炎性肌成纤维细胞瘤、胃肠道间质瘤(GIST)、结肠癌、多发性骨髓瘤(MM)、骨髓增生异常综合征(MDS)、骨髓增生性病症(MPD)、急性淋巴细胞性白血病(ALL)、急性髓细胞性白血病(AML)、慢性髓细胞性白血病(CML)、慢性淋巴细胞性白血病(CLL)、真性红细胞增多症、霍奇金淋巴瘤、非霍奇金淋巴瘤(NHL)、软组织肉瘤、纤维肉瘤、粘液肉瘤、脂肪肉瘤、成骨肉瘤、脊索瘤、血管肉瘤、内皮肉瘤、淋巴管肉瘤、淋巴管内皮肉瘤、滑膜瘤、间皮瘤、尤因氏肿瘤、平滑肌肉瘤、横纹肌肉瘤、鳞状细胞癌、基底细胞癌、腺癌、汗腺癌、皮脂腺癌、

乳头状癌、乳头状腺癌、髓样癌、支气管癌、肾细胞癌、肝细胞癌、胆管癌、绒毛膜癌、精原细胞瘤、胚胎性癌、维尔姆斯瘤、膀胱癌、上皮癌、神经胶质瘤、星形细胞瘤、成神经管细胞瘤、颅咽管瘤、室管膜瘤、松果体瘤、成血管细胞瘤、听神经瘤、少突神经胶质瘤、脑膜瘤、成神经细胞瘤、成视网膜细胞瘤、滤泡性淋巴瘤、弥漫性大B细胞淋巴瘤、套细胞淋巴瘤、肝细胞癌、甲状腺癌、胃癌、头颈癌、小细胞癌、原发性血小板增多症、特发性髓样化生、高嗜酸粒细胞综合征、系统性肥大细胞增多症、家族性嗜酸性粒细胞增多症、慢性嗜酸细胞性白血病、神经内分泌癌、类癌瘤等。

[0254] 在一个实施方案中,遗传损伤是在Cosmic数据库中注释的损伤(损伤和序列数据可从cancer.sanger.ac.uk/cosmic/census下载)或在癌症基因组图谱中注释的损伤(损伤和序列数据可从tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp下载)。

[0255] 带有与可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的癌症相关的一种或多种遗传损伤的基因的说明性实例包括但不限于ABCB1、ABCC2、ABCC4、ABCG2、ABL1、ABL2、AKT1、AKT2、AKT3、ALDH4A1、ALK、APC、AR、ARAF、ARFRP1、ARID1A、ATM、ATR、AURKA、AURKB、BCL2、BCL2A1、BCL2L1、BCL2L2、BCL6、BRAF、BRCA1、BRCA2、C1orf144、CARD11、CBL、CCND1、CCND2、CCND3、CCNE1、CDH1、CDH2、CDH20、CDH5、CDK4、CDK6、CDK8、CDKN2A、CDKN2B、CDKN2C、CEBPA、CHEK1、CHEK2、CRKL、CRLF2、CTNNB1、CYP1B1、CYP2C19、CYP2C8、CYP2D6、CYP3A4、CYP3A5、DNMT3A、DOT1L、DPYD、EGFR、EPHA3、EPHA5、EPHA6、EPHA7、EPHB1、EPHB4、EPHB6、EPHX1、ERBB2、ERBB3、ERBB4、ERCC2、ERG、ESR1、ESR2、ETV1、ETV4、ETV5、ETV6、EWSR1、EZH2、FANCA、FBXW7、FCGR3A、FGFR1、FGFR2、FGFR3、FGFR4、FLT1、FLT3、FLT4、FOXP4、GATA1、GNA11、GNAQ、GNAS、GPR124、GSTP1、GUCY1A2、HOXA3、HRAS、HSP90AA1、IDH1、IDH2、IGF1R、IGF2R、IKBKE、IKZF1、INHBA、IRS2、ITPA、JAK1、JAK2、JAK3、JUN、KDR、KIT、KRAS、LRP1B、LRP2、LTK、MAN1B1、MAP2K1、MAP2K2、MAP2K4、MCL1、MDM2、MDM4、MEN1、MET、MITF、MLH1、MLL、MPL、MRE11A、MSH2、MSH6、MTHFR、MTOR、MUTYH、MYC、MYCL1、MYCN、NF1、NF2、NKX2-1、NOTCH1、NPM1、NQO1、NRAS、NRP2、NTRK1、NTRK3、PAK3、PAX5、PDGFRA、PDGFRB、PIK3CA、PIK3R1、PKHD1、PLCG1、PRKDC、PTCH1、PTEN、PTPN11、PTPRD、RAF1、RARA、RB1、RET、RICTOR、RPTOR、RUNX1、SLC19A1、SLC22A2、SLC01B3、SMAD2、SMAD3、SMAD4、SMARCA4、SMARCB1、SMO、SOD2、SOX10、SOX2、SRC、STK11、SULT1A1、TBX22、TET2、TGFB2、TMPRSS2、TNFRSF14、TOP1、TP53、TPMT、TSC1、TSC2、TYMS、UGT1A1、UMPS、USP9X、VHL以及WT1。

[0256] 在具体实施方案中,所述遗传损伤包括核苷酸转换或颠换、核苷酸插入或缺失、基因组重排、拷贝数的变化或基因融合。

[0257] 在一个实施方案中,遗传损伤是将ALK基因的3'编码区融合至另一基因的基因融合。

[0258] 在一个实施方案中,遗传损伤是将ALK基因的3'编码区融合至EML4基因的基因融合。

[0259] 可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的适用于胎儿检测的病状的说明性实例包括但不限于:唐氏综合征(三体性21)、爱德华综合征(三体性18)、帕顿综合征(三体性13)、克莱恩费尔特综合征(XXY)、X三体综合征、XYY综合征、三体性8、三体性16、特纳综合征(XO)、罗伯逊易位、迪格奥尔格综合征和沃夫-贺许宏氏综合征。

[0260] 适用于可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的亲缘鉴定的

等位基因的说明性实例包括但不限于以下中的16种或更多种：D20S1082、D6S474、D12ATA63、D22S1045、D10S1248、D1S1677、D11S4463、D4S2364、D9S1122、D2S1776、D10S1425、D3S3053、D5S2500、D1S1627、D3S4529、D2S441、D17S974、D6S1017、D4S2408、D9S2157、牙釉蛋白、D17S1301、D1GATA113、D18S853、D20S482以及D14S1434。

[0261] 适用于可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的预测对药物治疗的反应的基因的说明性实例包括但不限于以下基因中的一种或多种：ABCB1 (ATP结合盒、亚家族B (MDR/TAP), 成员1)、ACE (血管紧张素I转化酶)、ADH1A (醇脱氢酶1A (I类),  $\alpha$ 多肽)、ADH1B (醇脱氢酶1B (I类),  $\beta$ 多肽)、ADH1C (醇脱氢酶1C (I类),  $\gamma$ 多肽)、ADRB1 (肾上腺素能 $\beta$ -1-受体)、ADRB2 (肾上腺素能 $\beta$ -2受体, 表面)、AHR (芳香烃受体)、ALDH1A1 (醛脱氢酶1家族, 成员A1)、ALOX5 (花生四烯酸5-脂氧合酶)、BRCA1 (乳腺癌1, 早发)、COMT (儿茶酚-O-甲基转移酶)、CYP2A6 (细胞色素P450, 家族2, 亚家族A, 多肽6)、CYP2B6 (细胞色素P450, 家族2, 亚家族B, 多肽6)、CYP2C9 (细胞色素P450, 家族2, 亚家族C, 多肽9)、CYP2C19 (细胞色素P450, 家族2, 亚家族C, 多肽19)、CYP2D6 (细胞色素P450, 家族2, 亚家族D, 多肽6)、CYP2J2 (细胞色素P450, 家族2, 亚家族J, 多肽2)、CYP3A4 (细胞色素P450, 家族3, 亚家族A, 多肽4)、CYP3A5 (细胞色素P450, 家族3, 亚家族A, 多肽5)、DPYD (二氢嘧啶脱氢酶)、DRD2 (多巴胺受体D2)、F5 (凝血因子V)、GSTP1 (谷胱甘肽S-转移酶pi)、HMGCR (3-羟基-3-甲基戊二酰基-辅酶A还原酶)、KCNH2 (钾电压门控通道, 亚家族H (eag相关的), 成员2)、KCNJ11 (钾内向整流通道, 亚家族J, 成员11)、MTHFR (5,10-亚甲基四氢叶酸还原酶 (NADPH))、NQO1 (NAD(P)H脱氢酶, 醌1)、P2RY1 (嘌呤能受体P2Y, G蛋白偶联, 1)、P2RY12 (嘌呤能受体P2Y, G-蛋白偶联, 12)、PTGIS (前列腺素I2 (前列环素) 合酶)、SCN5A (钠通道, 电压门控, V型,  $\alpha$  (长QT综合征3))、SLC19A1 (溶质载体家族19 (叶酸转运蛋白), 成员1)、SLC01B1 (溶质载体有机阴离子转运蛋白家族, 成员1B1)、SULT1A1 (磺基转移酶家族, 细胞溶质, 1A, 苯酚偏爱性, 成员1)、TPMT (硫嘌呤S-甲基转移酶)、TYMS (胸苷酸合成酶)、UGT1A1 (UDP葡萄糖醛酸基转移酶1家族, 多肽A1)、VDR (维生素D (1,25-二羟基维生素D3) 受体)、VKORC1 (维生素K环氧化物还原酶复合物, 亚基1)。

[0262] 可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的医学病状的说明性实例包括但不限于：中风、短暂性脑缺血发作、创伤性脑损伤、心脏病、心脏病发作、心绞痛、动脉粥样硬化和高血压。

[0263] 可用本文考虑的组合物和方法筛选的病原体的说明性实例包括但不限于：细菌真菌和病毒。

[0264] 可用本文考虑的组合物和方法筛选的细菌物种的说明性实例包括但不限于：分枝杆菌属、肺炎球菌属、埃希氏菌属、弯曲杆菌属、棒状杆菌属、梭菌属、链球菌属、葡萄球菌属、假单胞菌属、志贺菌属、密螺旋体属或沙门氏菌属。

[0265] 可用本文考虑的组合物和方法筛选的真菌物种的说明性实例包括但不限于：曲霉属、芽生菌属、假丝酵母属、球孢子菌属、隐球菌属、皮肤癣菌、癣属、毛癣菌属、小孢子菌属、镰刀菌属、组织胞浆菌属、毛霉菌属、肺孢子菌属、孢子丝菌属、突脐孢菌属或枝孢菌属。

[0266] 可用本文考虑的组合物和方法筛选的病毒的说明性实例包括但不限于：甲型流感 (如H1N1、H1N2、H3N2和H5N1 (禽流感))、乙型流感、丙型流感病毒、甲型肝炎病毒、乙型肝炎病毒、丙型肝炎病毒、丁型肝炎病毒、戊型肝炎病毒、轮状病毒、诺沃克病毒组的任何病毒、肠腺病毒、细小病毒、登革热病毒、猴痘、单链负义病毒、狂犬病病毒属 (如狂犬病病毒、兔头

蝙蝠病毒、莫科拉病毒、杜文海格病毒、欧洲蝙蝠病毒1和2和澳大利亚蝙蝠病毒)、暂时热病毒、水泡性病毒、水疱性口炎病毒(VSV)、疱疹病毒如单纯疱疹病毒1型和2型、水痘带状疱疹、巨细胞病毒、艾伯斯坦-巴尔病毒(EBV)、人疱疹病毒(HHV)、人疱疹病毒6型和8型、莫洛尼鼠白血病病毒(M-MuLV)、莫洛尼鼠肉瘤病毒(MoMSV)、哈维鼠肉瘤病毒(HaMuSV)、鼠乳腺肿瘤病毒(MuMTV)、长臂猿白血病病毒(GaLV)、猫白血病病毒(FLV)、泡沫病毒、弗罗德鼠白血病病毒、鼠干细胞病毒(MSCV)和劳氏肉瘤病毒(RSV)、HIV(人免疫缺陷病毒;包括HIV 1型和HIV 2型)、维斯纳-梅迪病毒(VMV)病毒、山羊关节炎-脑炎病毒(CAEV)、马传染性贫血病毒(EIAV)、猫免疫缺陷病毒(FIV)、牛免疫缺陷病毒(BIV)和猿免疫缺陷病毒(SIV)、乳头状瘤病毒、鼠 $\gamma$ 疱疹病毒、沙粒病毒(如阿根廷出血热病毒、玻利维亚出血热病毒、萨比亚相关出血热病毒、委内瑞拉出血热病毒、拉沙热病毒、马丘波病毒、淋巴细胞性脉络丛脑膜炎病毒(LCMV))、布尼亚病毒如克里米亚-刚果出血热病毒、汉坦病毒、致伴随肾综合征的出血热的病毒、裂谷热病毒、包括埃博拉出血热和马尔堡出血热的丝状病毒科(丝状病毒)、包括Kaysanur森林病病毒的黄病毒科、鄂木斯克出血热病毒、致蜱传脑炎病毒和副粘病毒科如亨德拉病毒和尼帕病毒、重型天花病毒和轻型天花病毒(天花)、甲病毒如委内瑞拉马脑炎病毒、东部马脑炎病毒、西部马脑炎病毒、SARS相关冠状病毒(SARS-CoV)、西尼罗河病毒和任何致脑炎病毒。

[0267] 适用于可用本文考虑的组合物和方法检测、鉴定、预测、诊断或监测的监测移植受体中的器官移植物的基因的说明性实例包括但不限于以下基因中的一种或多种:HLA-A、HLA-B、HLA-C、HLA-DR、HLA-DP以及HLA-DQ。

[0268] 在具体实施方案中,生物信息学分析用于定量所述DNA克隆文库中分析的基因组等效物的数量;检测靶遗传基因座中的遗传变体;检测靶遗传基因座内的突变;检测靶遗传基因座内的遗传融合;和/或测量靶遗传基因座内的拷贝数波动。

[0269] G. 伴随诊断

[0270] 在各种实施方案中,提供一种伴随诊断遗传疾病的方法,所述方法包括:从受试者的生物样品分离或获得DNA,例如cfDNA;除去所述DNA的末端磷酸酯残基;用一种或多种末端修复酶处理去磷酸化的DNA以产生末端修复的DNA;将一个或多个双链DNA(dsDNA)前衔接子连接至所述末端修复的DNA的每条链的3'末端以形成前衔接子/末端修复的DNA复合物,其中每个dsDNA前衔接子包含连接至所述末端修复的DNA的每条链的3'末端的连接链寡核苷酸,和非连接配偶体链寡核苷酸;用修复寡核苷酸置换来自所述前衔接子/末端修复的DNA复合物的非连接配偶体链寡核苷酸以形成衔接子/末端修复的DNA复合物,其中每个衔接子包含所述连接链寡核苷酸和所述修复寡核苷酸;以及用一种或多种酶处理所述衔接子/末端修复的DNA复合物以形成连续的双链DNA文库;扩增所述DNA文库以产生DNA克隆文库;测定所述DNA克隆文库中的基因组等效物的数量;以及进行所述DNA克隆文库中与所述遗传疾病相关的一种或多种生物标志物的定量遗传分析,其中检测到或未能检测到所述一种或多种生物标志物中的至少一种指示所述受试者是否应针对所述遗传疾病进行治疗。

[0271] 如本文所用,术语“伴侣诊断”是指与特定抗癌疗法有关的诊断测试。在一个具体实施方案中,诊断方法包括检测生物样品中与相关的生物标志物的遗传损伤,从而允许迅速鉴定患者应该或不应该用所述抗癌疗法治疗。

[0272] 抗癌疗法包括但不限于外科手术、放射、化学疗法、抗癌药物和免疫调节剂。

[0273] 抗癌药物的说明性实例包括但不限于：烷基化剂，如噻替派和环磷酰胺 (CYTOXAN<sup>TM</sup>)；烷基磺酸酯，如白消安、英丙舒凡和派泊舒凡；氮杂环丙院，如苯柔多巴、卡波醌、甲基多巴和普力多巴；乙烯亚胺和甲基三聚氰胺，包括六甲蜜胺、三亚乙基三聚氰胺、三亚乙基磷酰胺、三亚乙基硫代磷酰胺和三羟甲基三聚氰胺；氮芥，如苯丁酸氮芥、萘氮芥、氯磷酰胺、雌莫司汀、异环磷酰胺、二氯甲基二乙胺、盐酸甲氧氮芥、美法仑、新恩比兴、苯芥胆固醇、拨尼氮芥、曲磷胺、尿嘧啶芥末；亚硝基脲，如卡莫司汀、氯脲菌素、福莫司汀、洛莫司汀、嘧啶亚硝脲、雷莫司汀；抗生素，如阿克拉霉素、放线菌素、蒽霉素、重氮丝氨酸、博来霉素、放线菌素、卡奇霉素、卡拉比辛、洋红霉素、嗜癌菌素、色霉素、更生霉素、柔红霉素、地托比星、6-重氮基-5-氧代-L-正白氨酸、阿霉素及其聚乙二醇化制剂、表阿霉素、依索比星、伊达比星、麻西罗霉素、丝裂霉素、霉酚酸、诺加霉素、橄榄霉素、培洛霉素、博替罗霉素、嘌呤霉素、三铁阿霉素、罗多比星、链黑菌素、链脲菌素、杀结核菌素、乌苯美司、净司他丁、佐柔比星；抗代谢剂，如甲氨蝶呤和5-氟尿嘧啶 (5-FU)；叶酸类似物，如二甲叶酸、甲氨蝶呤、蝶罗呤、三甲喋呤；嘌呤类似物，如氟达拉滨、6-硫嘌呤、硫咪嘌呤、硫鸟嘌呤；嘧啶类似物，如安西他滨、氮杂胞苷、6-氮杂尿苷、卡莫氟、阿糖胞苷、二脱氧尿苷、去氧氟尿苷、依诺他滨、氟尿苷、5-FU；雄激素，如卡鲁睾酮、丙酸屈他雄酮、环硫雄醇、美雄烷、睾内酯；抗肾上腺，如氨鲁米特、米托坦、曲洛司坦；叶酸补充剂，如亚叶酸；醋葡醛内酯；醛磷酰胺葡萄糖苷；氨基乙酰丙酸；安吡啶；贝曲布索；比生群；依达曲沙；迪氟法迈；地美可辛；地吡酮；艾氟米辛；依利醋铵；依托格鲁；硝酸镓；羟基脲；香菇多糖；氯尼达明；米托胍脲；米托蒽醌；莫派达醇；二胺硝吡啶；喷司他丁；蛋氨酸芥；吡柔比星；鬼臼酸；2-乙基肼；甲基苄肼；**PSK®**；雷佐生；西佐喃；锗螺胺；细交链孢菌酮酸；三亚胺醌；2,2',2''-三氯三乙胺；聚氨酯；长春地辛；达卡巴嗪；甘露醇氮芥；二溴甘露醇；二溴卫矛醇；派泊溴烷；嘉胞苷；阿糖胞苷 ("Ara-C")；环磷酰胺；塞替派；紫杉烷类，例如紫杉醇 (**TAXOL®**, Bristol-Myers Squibb Oncology, Princeton, N.J.) 和多西他赛 (**TAXOTERE®**, Rhne-Poulenc Rorer, Antony, France)、苯丁酸氮芥；吉西他滨；6-硫鸟嘌呤；硫嘌呤；甲氨蝶呤；铂类似物，如顺铂和卡铂；长春花碱；铂；依托泊苷 (VP-16)；异环磷酰胺；丝裂霉素C；米托蒽醌；长春新碱；长春瑞滨；异长春花碱；诺消灵；替尼泊苷；氨基喋呤；希罗达；伊班膦酸盐；CPT-11；拓扑异构酶抑制剂RFS 2000；二氟甲基鸟氨酸 (DMFO)；维甲酸衍生物如Targretin<sup>TM</sup> (蓓萨罗丁)、Panretin<sup>TM</sup> (阿利维甲酸)；ONTAK<sup>TM</sup> (地尼白介素-白喉连接物)；埃斯波霉素；卡培他滨；以及上述任一种的药学上可接受的盐、酸或衍生物。此定义中还包括用于调控或抑制激素对癌症的作用的抗激素剂，如抗雌激素，包括例如他莫昔芬、雷洛昔芬、芳香酶抑制剂4(5)-咪唑、4-羟基他莫昔芬、曲沃昔芬、雷洛西芬、LY117018、奥那司酮和托瑞米芬 (Fareston)；和抗雄激素，如氟他胺、尼鲁米特、比卡鲁胺、醋酸亮丙瑞林和戈舍瑞林；以及上述任一种的药学上可接受的盐、酸或衍生物。

[0274] 免疫调节剂的说明性实例包括但不限于：环孢菌素、他克莫司、曲培莫司、吡美莫司、西罗莫司、维罗莫司、拉氟莫司、拉喹莫德和咪喹莫特，以及其类似物、衍生物、盐、离子和复合物。

[0275] 在本说明书中引用的所有公布、专利申请以及已授权专利以引用的方式并入本文，如同明确并个别地指示每个个别公布、专利申请或已授权专利以引用的方式并入一般。2013年12月10日提交的美国专利申请号14/102,285和2014年8月22日提交的美国专利申请



号14/466,741各自以引用的方式整体并入本文。

[0276] 尽管为了清楚理解的目的,已经通过举例说明和实例相当详细地描述了前述发明,但是根据本发明的教义,本领域的普通技术人员将显而易见的是,可另外对本发明进行某些改变和修改而不背离所附权利要求的精神和范围。以下实施例仅以说明方式提供,而并不起限制作用。本领域技术人员将容易地识别各种非关键性参数,所述参数可能发生改变或修改以产生基本上类似的结果。

## 实施例

### [0277] 实施例1

[0278] 高效率衔接子连接的原理的证据

[0279] 此实施例提供直接定量证据,即在不存在衔接子二聚体形成的情况下,本文考虑的高效率连接策略导致与DNA片段的两个末端连接(图2)。

[0280] 用限制酶RsaI消化来自克隆载体pUC19的质粒DNA以产生平末端并用南极碱性磷酸酶去磷酸化。将这些平末端的去磷酸化的DNA片段连接至九种不同的高效率衔接子的集合(表1)。在所有情况下,在连接反应(箭头)后观察到片段迁移率的定量变动,并且迁移率的变动等效于衔接子与每个DNA片段的每个末端的连接。此实施例提供了原理的证据,即本文考虑的组合物和方法导致衔接子与DNA片段的高效率连接,从而提高构建DNA文库的整体效率。

[0281] 表1用于创建测试衔接子的连接链和配偶体链寡核苷酸序列

SEQ ID NO:	衔接子ID	连接链(5'磷酸酯, 5'至3')	SEQ ID NO:	配偶体链(5'至3', 3' ddC)
1	1	GAGGGTCTACCTTCTTNNNNNTGTATTCGAATTCTCTGGTCCTGCA	10	AAGGTAGACCCT
2	2	GCTCTAGACGTCATCGNNNNNTGTATTCGAATTCTCTGGTCCTGCA	11	TGACGTCTAGAG
3	3	GGATACTCGTAGCGGNNNNNTGTATTCGAATTCTCTGGTCCTGCA	12	GCTACGAGTATC
4	4	GTCACGAGTAGAGAAANNNNNTGTATTCGAATTCTCTGGTCCTGCA	13	CTTACTCGTGA
5	5	GAGGGTCTACCTTAGTNNNNNTGTATTCGAATTCTCTGGTCCTGCA	14	AAGGTAGACCCT
6	6	GCTCTAGACGTCAGAGNNNNNTGTATTCGAATTCTCTGGTCCTGCA	15	TGACGTCTAGAG
7	7	GGATACTCGTAGCTTCNNNNNTGTATTCGAATTCTCTGGTCCTGCA	16	GCTACGAGTATC
8	8	GTCACGAGTAGAGCCANNNNNTGTATTCGAATTCTCTGGTCCTGCA	17	CTTACTCGTGA
9	9	GAGGGTCTACCTTGCTNNNNNTGTATTCGAATTCTCTGGTCCTGCA	18	AAGGTAGACCCT

### [0283] 实施例2

[0284] 高效率DNA文库构建

[0285] I. 片段末端修复

[0286] 将无细胞DNA(cfDNA)片段末端去磷酸化,对DNA双链体的内部损伤被修复并且DNA末端被“精修”成平末端。所得片段被称为“末端修复的片段”。

[0287] 通过将81μl的纯化的cfDNA、10μl的New England Biolabs (NEB) CutSmart缓冲液(B7204S)和5μl的NEB虾碱性磷酸酶(M0371)组合来使cfDNA片段末端去磷酸化。将反应混合物在37℃孵育15分钟,且然后在65℃孵育5分钟。

[0288] 通过加入4μl在冰上制备的以下混合物来修复对去磷酸化的cfDNA片段的内部损伤:1.1μl的10mM dNTP混合物(NEB N0447)、2.2μl的PreCR酶混合物(M0309)以及1.1μl的T4 DNA聚合酶(M0203)。将此反应混合物在20℃孵育15分钟,且然后在70℃孵育10分钟。以这种方式修复和精修的cfDNA片段可直接用于DNA连接反应。



## [0289] II. 前衔接子设计

[0290] 使用若干设计考虑来产生前衔接子,所述前衔接子由连接链寡核苷酸和具有封闭的3'末端的互补配偶体链寡核苷酸组成。

[0291] 此实施例中使用的前衔接子被设计为具有以下特征:长度为10nt的连接链寡核苷酸;A/T或G/C残基的平衡;在每组前衔接子内的每个碱基位置处表示的四个DNA碱基中的每一个;10碱基序列的在50mM Na<sup>+</sup>(或K<sup>+</sup>)、10mM MgCl<sub>2</sub>中约37℃的预测解链温度;A/T和G/C nt两者作为每个前衔接子序列的前两个碱基;长度为8nt的互补配偶体寡核苷酸序列,所述寡核苷酸序列通过使用2-羟基核糖修饰的DNA碱基(MWG Eurofins)化学封闭并在50mM Na<sup>+</sup>或K<sup>+</sup>、10mM MgCl<sub>2</sub>中具有约25℃的解链温度。

[0292] 即使存在设计限制,也要进行衔接子组的经验性能筛选。在当前实验中,鉴定了五组具有可接受性能的衔接子(参见例如,表2)。标记为“分数”的栏显示每个衔接子相对于性能最佳的衔接子(组6-2)的克隆效率百分比。

[0293] 表2.经验验证的衔接子组。

SEQ ID NO.	连接链寡核苷酸		配偶体链寡核苷酸		分数
	名称	序列	名称	序列	
19	he lig 5-1	/5Phos/CTGAGCTAGT	he part 5-1	TAGCTCA[3-dG-Q]	60
20	he lig 5-2	/5Phos/GACTCGATAG	he part 5-2	ATCGAGT[3-dC-Q]	97
21	he lig 5-3	/5Phos/TCAGATCGTC	he part 5-3	CGATCTG[3-dA-Q]	27
22	he lig 5-4	/5Phos/AGTCTAGCCA	he part 5-4	GCTAGAC[3-dT-Q]	84
23	he lig 6-1	/5Phos/GGATTACCCT	he part 6-1	GGTAATC[3-dC-Q]	70
24	he lig 6-2	/5Phos/CTTACGGATG	he part 6-2	TCCGTAA[3-dG-Q]	100
25	he lig 6-3	/5Phos/ACCGATTGAC	he part 6-3	CAATCGG[3-dT-Q]	74
26	he lig 6-4	/5Phos/TAGCGCATGA	he part 6-4	ATGCGCT[3-dA-Q]	48
[0294] 27	he lig 8-1	/5Phos/ATGTCCAGCT	he part 8-1	CTGGACA[3-dT-Q]	24
28	he lig 8-2	/5Phos/CACAGGTTAG	he part 8-2	AACCTGT[3-dG-Q]	46
29	he lig 8-3	/5Phos/TGACATGCTC	he part 8-3	GCATGTC[3-dA-Q]	53
30	he lig 8-4	/5Phos/GCTGTACAGA	he part 8-4	TGTACAG[3-dC-Q]	48
31	he lig 11-1	/5Phos/TCAAGTCGGT	he part 11-1	CGACTTG[3-dA-Q]	83
32	he lig 11-2	/5Phos/GTTCAGACTG	he part 11-2	GTCTGAA[3-dC-Q]	95
33	he lig 11-3	/5Phos/CAGGTCTAAC	he part 11-3	TAGACCT[3-dG-Q]	33
34	he lig 11-4	/5Phos/AGCTCAGTCA	he part 11-4	ACTGAGC[3-dT-Q]	98
35	he lig 12-1	/5Phos/GATCCGTA	he part 12-1	TACGGAT[3-dC-Q]	91
36	he lig 12-2	/5Phos/ACAGTCGTAG	he part 12-2	ACGACTG[3-dT-Q]	98
37	he lig 12-3	/5Phos/TGGTAACCTC	he part 12-3	GGTTACC[3-dA-Q]	85
38	he lig 12-4	/5Phos/CTCAGTAGGA	he part 12-4	CTACTGA[3-dG-Q]	93

## [0295] III. 前衔接子的连接

[0296] 将前衔接子连接至此实施例的步骤I中产生的末端修复的片段。将25μl的末端修复的片段与10μl的10μM衔接子组合。通常,根据加入到反应中的单独衔接子的数量进行1至4个连接反应。将15μl连接混合物(5μl的10X T4-DNA连接缓冲液、7.5μl的50%PEG8000以及2.5μl的HC T4 DNA连接酶(NEB;M0202))以50μl的最终体积加入每个连接反应物中。将反应混合并在20℃孵育60分钟,然后在65℃孵育10分钟,然后冷却至室温。

[0297] 在连接反应后,加入50μl的TEzero(10mM Tris pH 8.0,0.1mM EDTA,0.05%Tween 20),并向每个反应中加入120μl的DNA纯化珠粒并充分混合。将反应在室温下孵育10分钟,

然后用200 $\mu$ l的70%乙醇/水(v/v)将珠粒洗涤两次,短暂风干(约5分钟)并用20 $\mu$ l的TEzero洗脱。

#### [0298] IV. 修复寡核苷酸

[0299] 此实施例中使用的修复寡核苷酸的完整列表在表3中示出。每个修复寡核苷酸是249个单独寡核苷酸的库。各修复寡核苷酸中的不变序列代表PCR引物结合位点并在表3的左侧部分中示出。

[0300] 249个寡核苷酸中的每一个均包含5个核苷酸的样品代码,在修复寡核苷酸序列中显示为“XXXXX”。5个核苷酸的样品在表3的右侧部分中示出。5核苷酸代码由256个可能的独特序列组成,所述独特序列被选择为与所述组中的每个其他代码不同的2个碱基变化。这种特征使独特且不同的读数能够区别于由于代码区域中的测序误差而显得独特的读数。其中G残基被过度代表并经验地显示干扰衔接子功能的七个代码被除去,从而留下249个随机代码:

[0301] 表3. 修复寡核苷酸及其相关的读取代码

SEQ ID NO:	与连接寡核苷酸相容的全长修复寡核苷酸(RO)	
	名称	序列
39	RO_5-1	TGCAGGACCAGAGAATTCGAATACAXXXXXACTAGCTCAG
40	RO_5-2	TGCAGGACCAGAGAATTCGAATACAXXXXXCTATCGAGTC
41	RO_5-3	TGCAGGACCAGAGAATTCGAATACAXXXXXGACGATCTGA
42	RO_5-4	TGCAGGACCAGAGAATTCGAATACAXXXXXTGGCTAGACT
43	RO_6-1	TGCAGGACCAGAGAATTCGAATACAXXXXXAGGGTAATCC
44	RO_6-2	TGCAGGACCAGAGAATTCGAATACAXXXXXCATCCGTAAG
45	RO_6-3	TGCAGGACCAGAGAATTCGAATACAXXXXXGTCAATCGGT

[0302]

46	RO 6-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTCATGCGCTA					
47	RO 7-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXACTGCTAGCA					
48	RO 7-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCAGCGATCAT					
49	RO 7-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGTCTATCGATG					
50	RO 7-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTGATAGCTGC					
51	RO 8-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXAGCTGGACAT					
52	RO 8-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCTAACCTGTG					
53	RO 8-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGAGCATGTCA					
54	RO 8-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTCTGTACAGC					
55	RO 11-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXACCGACTTGA					
56	RO 11-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCAGTCTGAAC					
57	RO 11-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGTTAGACCTG					
58	RO 11-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTGACTGAGCT					
59	RO 12-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXAGTACGGATC					
60	RO 12-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCTACGACTGT					
61	RO 12-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGAGGTTACCA					
62	RO 12-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTCTACTGAG					
XXXXX - 差异2个序列代码							
	Seq	Seq	Seq	Seq	Seq	Seq	Seq
	CGGGT	CACGG	GTGAT	ATCAG	GATAC	TGTCT	CCAAA
	CGGTG	CAGCG	GTGTA	ATCGA	GATCA	TGTTC	CCACC
	CGTGG	CAGGC	GTTAG	ATGAC	GCAAT	TTCGT	CCATT
	GCGGT	CCAGG	GTTGA	ATGCA	GCATA	TTCTG	CCCAC
	GCGTG	CCGAG	TAGGT	CAAGT	GCCCT	TTGCT	CCCCA
	GCTGG	CCGGA	TAGTG	CAATG	GCCTC	TTGTC	CCTAT
	GGCGT	CGACG	TATGG	CAGAT	GCTAA	TTTCG	CCTTA
	GGCTG	CGAGC	TGAGT	CAGTA	GCTCC	TTTGC	CTACT
	GGGCT	CGCAG	TGATG	CATAG	GCTTT	AAAAA	CTATC
	GGGTC	CGCGA	TGGAT	CATGA	GTAAC	AAACC	CTCAT
	GGTCG	CGGAC	TGGTA	CCCGT	GTACA	AAATT	CTCTA
	GGTGC	CGGCA	TGTAG	CCCTG	GTCAA	AACAC	CTTAC
	GTCGG	GAAAG	TGTGA	CCGCT	GTCCC	AACCA	CTTCA
	GTGCG	GAAGA	TTAGG	CCGTC	GTCTT	AATAT	TAAAT
	GTGGC	GACCG	TTGAG	CCTCG	GTTCT	AATTA	TAATA
	TGCGG	GACGC	TTGGA	CCTGC	GTTTC	ACAAC	TACCT
	TGGCG	GAGAA	AACGT	CGAAT	TAACG	ACACA	TACTC
	TGGGC	GAGCC	AACTG	CGATA	TAAGC	ACCAA	TATAA
	AAAGG	GAGTT	AAGCT	CGCCT	TACAG	ACCCC	TATCC
	AAGAG	GATGT	AAGTC	CGCTC	TACGA	ACCTT	TATTT
	AAGGA	GATTG	AATCG	CGTAA	TAGAC	ACTCT	TCACT

[0303]

[0304]

	ACCGG	GCACG	AATGC	CGTCC	TAGCA	ACTTC	TCATC
	ACGCG	GCAGC	ACAGT	CGTTT	TCAAG	ATAAT	TCCAT
	ACGGC	GCCAG	ACATG	CTAAG	TCAGA	ATATA	TCCTA
	AGAAG	GCCGA	ACGAT	CTAGA	TCCCG	ATCCT	TCTAC
	AGAGA	GCGAC	ACGTA	CTCCG	TCCGC	ATCTC	TCTCA
	AGCCG	GCGCA	ACTAG	CTCGC	TCGAA	ATTAA	TTAAA
	AGCGC	GGAAA	ACTGA	CTGAA	TCGCC	ATTCC	TTACC
	AGGAA	GGACC	AGACT	CTGCC	TCGTT	ATTTT	TTATT
	AGGCC	GGATT	AGATC	CTGTT	TCTGT	CAAAC	TTCAC
	AGGTT	GGCAC	AGCAT	CTTGT	TCTTG	CAACA	TTCCA
	AGTGT	GGCCA	AGCTA	CTTTG	TGAAC	CACAA	TTTAT
	AGTTG	GGTAT	AGTAC	GAACT	TGACA	CACCC	TTTTA
	ATGGT	GGTTA	AGTCA	GAATC	TGCAA	CACTT	
	ATGTG	GTAGT	ATACG	GACAT	TGCCC	CATCT	
	ATTGG	GTATG	ATAGC	GACTA	TGCTT	CATTC	

[0305] V. 修复寡核苷酸添加至前衔接子文库

[0306] 通过向衔接子添加修复寡核苷酸而完成文库构建。此实施例中说明的修复寡核苷酸含有PCR引物结合位点；样品代码；和锚定序列，所述锚定序列是充当鉴定序列的一种方式、实现测序读数中正确碱基识别的校准并充当与连接链寡核苷酸杂交的锚的随机序列标签。

[0307] 将4 $\mu$ l的1 $\mu$ M修复寡核苷酸库（参见例如，表3）加入20 $\mu$ l的来自此实施例的步骤III的纯化连接混合物中。

[0308] 接下来，通过在冰上将24 $\mu$ l修复寡核苷酸/连接酶混合物与16 $\mu$ l的以下混合物组合来制备40 $\mu$ l修复寡核苷酸反应物：11 $\mu$ l水、4.4 $\mu$ l缓冲液“B”（190 $\mu$ l的CutSmart缓冲液（NEB；B7204）和10 $\mu$ l的1M二硫苏糖醇（DTT；Sigma-Aldrich 646563）、1.32 $\mu$ l核苷酸混合物“N”（将50 $\mu$ l的10mM dNTP混合物（NEB；N0447）与25 $\mu$ l的100X NAD<sup>+</sup>（NEB；B9007）和0.88 $\mu$ l酶混合物“E”（将20 $\mu$ l的T4多核苷酸激酶（NEB；M0201）、10 $\mu$ l全长BstI聚合酶（NEB；M0328）和10 $\mu$ l的Taq DNA连接酶（NEB；M0208）组合）组合。将反应混合并在37℃孵育15分钟，然后在60℃孵育15分钟。

[0309] 将反应从热循环仪中移出并将48 $\mu$ l珠粒再悬浮溶液（19%PEG8000, 2M NaCl、10mM Tris pH 8.0、10mM EDTA、0.1%Tween 20）加入反应中并在室温下孵育10分钟。将珠粒用200 $\mu$ l的70%乙醇洗涤两次，短暂风干（约5分钟）并重新悬浮于25 $\mu$ l的TEzero中。使用磁体来定位珠粒，并将澄清的DNA文库转移至新鲜反应容器。

[0310] VI. 概述

[0311] 使用本文全文考虑和实施例1中描述的方法构建的所得DNA文库是准备扩增的并且适用于一个或多个靶遗传基因座的下一代测序、qPCR分析和其他定量遗传分析。

[0312] 通常，在以下权利要求书中，所使用的术语不应该解释为将权利要求限制于本说明书和权利要求书中公开的具体实施方案，而应解释为包括连同与所述权利要求所规定的全部范围等效物一起的所有可能的实施方案。因此，权利要求并不受公开内容所限制。

## 序列表

<110> 分析生物科学有限公司 (Resolution Bioscience, Inc.)

克里斯多弗·K·雷蒙德 (Raymond, Christopher K)

林继力 (Lim, Lee P)

克里斯多弗·D·阿莫尔 (Armour, Christopher D)

<120> DNA文库的高效率构建

<130> CLFK-003/01W0

<150> US 62/254,110

<151> 2015-11-11

<160> 62

<170> PatentIn 3.5版

<210> 1

<211> 46

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)

<223> n是A、C、T或G

<400> 1

gagggtctac cttcttnnnn ntgtattcga attctctggc cctgca 46

<210> 2

<211> 46

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)

<223> n是A、C、T或G

<400> 2

gctctagacg tcatcgnnnn ntgtattcga attctctggt cctgca 46

<210> 3

<211> 46

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)

<223> n是A、C、T或G

<400> 3

ggatactcgt agcggcnnnn ntgtattcga attctctggt cctgca 46

<210> 4

<211> 46

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)

<223> n是A、C、T或G

<400> 4

gtcacgagta gagaaannnn ntgtattcga attctctggt cctgca 46

<210> 5

<211> 46

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)  
<223> n是A、C、T或G  
<400> 5  
gagggtctac cttagtnnnn ntgtattcga attctctggt cctgca 46  
<210> 6  
<211> 46  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<220>  
<221> misc\_feature  
<222> (17) .. (21)  
<223> n是A、C、T或G  
<400> 6  
gctctagacg tcagagnnnn ntgtattcga attctctggt cctgca 46  
<210> 7  
<211> 46  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<220>  
<221> misc\_feature  
<222> (17) .. (21)  
<223> n是A、C、T或G  
<400> 7  
ggatactcgt agcttcnnnn ntgtattcga attctctggt cctgca 46  
<210> 8  
<211> 46  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>



<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)

<223> n是A、C、T或G

<400> 8

gtcacgagta gagccannnn ntgtattcga attctctggc cctgca 46

<210> 9

<211> 46

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的连接链寡核苷酸

<220>

<221> misc\_feature

<222> (17) .. (21)

<223> n是A、C、T或G

<400> 9

gagggtctac cttgctnnnn ntgtattcga attctctggc cctgca 46

<210> 10

<211> 12

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的配偶体链寡核苷酸

<400> 10

aaggtagacc ct 12

<210> 11

<211> 12

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<221>

<222>

<223> 合成的配偶体链寡核苷酸



<400> 11  
tgacgtctag ag 12  
<210> 12  
<211> 12  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的配偶体链寡核苷酸  
<400> 12  
gctacgagta tc 12  
<210> 13  
<211> 12  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的配偶体链寡核苷酸  
<400> 13  
ctctactcgt ga 12  
<210> 14  
<211> 12  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的配偶体链寡核苷酸  
<400> 14  
aaggtagacc ct 12  
<210> 15  
<211> 12  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>

<223> 合成的配偶体链寡核苷酸  
<400> 15  
tgacgtctag ag 12  
<210> 16  
<211> 12  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的配偶体链寡核苷酸  
<400> 16  
gctacgagta tc 12  
<210> 17  
<211> 12  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的配偶体链寡核苷酸  
<400> 17  
ctctactcgt ga 12  
<210> 18  
<211> 12  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的配偶体链寡核苷酸  
<400> 18  
aaggtagacc ct 12  
<210> 19  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>

<222>  
<223> 合成的连接链寡核苷酸  
<400> 19  
ctgagctagt 10  
<210> 20  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 20  
gactcgatag 10  
<210> 21  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 21  
tcagatcgtc 10  
<210> 22  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 22  
agtctagcca 10  
<210> 23  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>

<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 23  
ggattaccct 10  
<210> 24  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 24  
cttacggatg 10  
<210> 25  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 25  
accgattgac 10  
<210> 26  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 26  
tagcgcata 10  
<210> 27  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)

<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 27  
atgtccagct 10  
<210> 28  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 28  
cacaggttag 10  
<210> 29  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 29  
tgacatgctc 10  
<210> 30  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 30  
gctgtacaga 10  
<210> 31  
<211> 10  
<212> DNA

<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 31  
tcaagtcggt 10  
<210> 32  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 32  
gttcagactg 10  
<210> 33  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 33  
caggtctaac 10  
<210> 34  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 34  
agctcagtca 10  
<210> 35  
<211> 10

<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 35  
gatccgtact 10  
<210> 36  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 36  
acagtcgtag 10  
<210> 37  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 37  
tggtaacctc 10  
<210> 38  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的连接链寡核苷酸  
<400> 38  
ctcagtagga 10  
<210> 39

<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 39  
tgcaggacca gagaattcga atacannnnn actagctcag 40  
<210> 40  
<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 40  
tgcaggacca gagaattcga atacannnnn ctatcgagtc 40  
<210> 41  
<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中



<400> 41  
tgcaggacca gagaattcga atacannnnn gacgatctga 40  
<210> 42  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 42  
tgcaggacca gagaattcga atacannnnn tggctagact 40  
<210> 43  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 43  
tgcaggacca gagaattcga atacannnnn agggtaatcc 40  
<210> 44  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>

<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 44  
tgcaggacca gagaattcga atacannnnn catccgtaag 40  
<210> 45  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 45  
tgcaggacca gagaattcga atacannnnn gtcaatcggt 40  
<210> 46  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 46  
tgcaggacca gagaattcga atacannnnn tcatgcgcta 40  
<210> 47  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>

<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 47  
tgcaggacca gagaattcga atacannnnn actgctagca 40  
<210> 48  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 48  
tgcaggacca gagaattcga atacannnnn cagcgatcat 40  
<210> 49  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 49  
tgcaggacca gagaattcga atacannnnn gtcatcgatg 40  
<210> 50  
<211> 40  
<212> DNA

<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 50  
tgcaggacca gagaattcga atacannnnn tgatagctgc 40  
<210> 51  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 51  
tgcaggacca gagaattcga atacannnnn agctggacat 40  
<210> 52  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 52  
tgcaggacca gagaattcga atacannnnn ctaacctgtg 40

<210> 53  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 53  
tgcaggacca gagaattcga atacannnnn gagcatgtca 40  
<210> 54  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 54  
tgcaggacca gagaattcga atacannnnn tctgtacagc 40  
<210> 55  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)

<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 55  
tgcaggacca gagaattcga atacannnnn accgacttga 40  
<210> 56  
<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 56  
tgcaggacca gagaattcga atacannnnn cagtctgaac 40  
<210> 57  
<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 57  
tgcaggacca gagaattcga atacannnnn gttagacctg 40  
<210> 58  
<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸

<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 58  
tgcaggacca gagaattcga atacannnnn tgactgagct 40  
<210> 59  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 59  
tgcaggacca gagaattcga atacannnnn agtacggatc 40  
<210> 60  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 60  
tgcaggacca gagaattcga atacannnnn ctacgactgt 40  
<210> 61  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial Sequence)  
<220>

<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 61  
tgcaggacca gagaattcga atacannnnn gaggttacca 40  
<210> 62  
<211> 40  
<212> DNA  
<213> 人工序列(Artificial Sequence)  
<220>  
<221>  
<222>  
<223> 合成的修复寡核苷酸  
<220>  
<221> misc\_feature  
<222> (26) .. (30)  
<223> n是A、C、T或G,其中从位置26至30的特定碱基组列于表3中  
<400> 62  
tgcaggacca gagaattcga atacannnnn tcctactgag 40



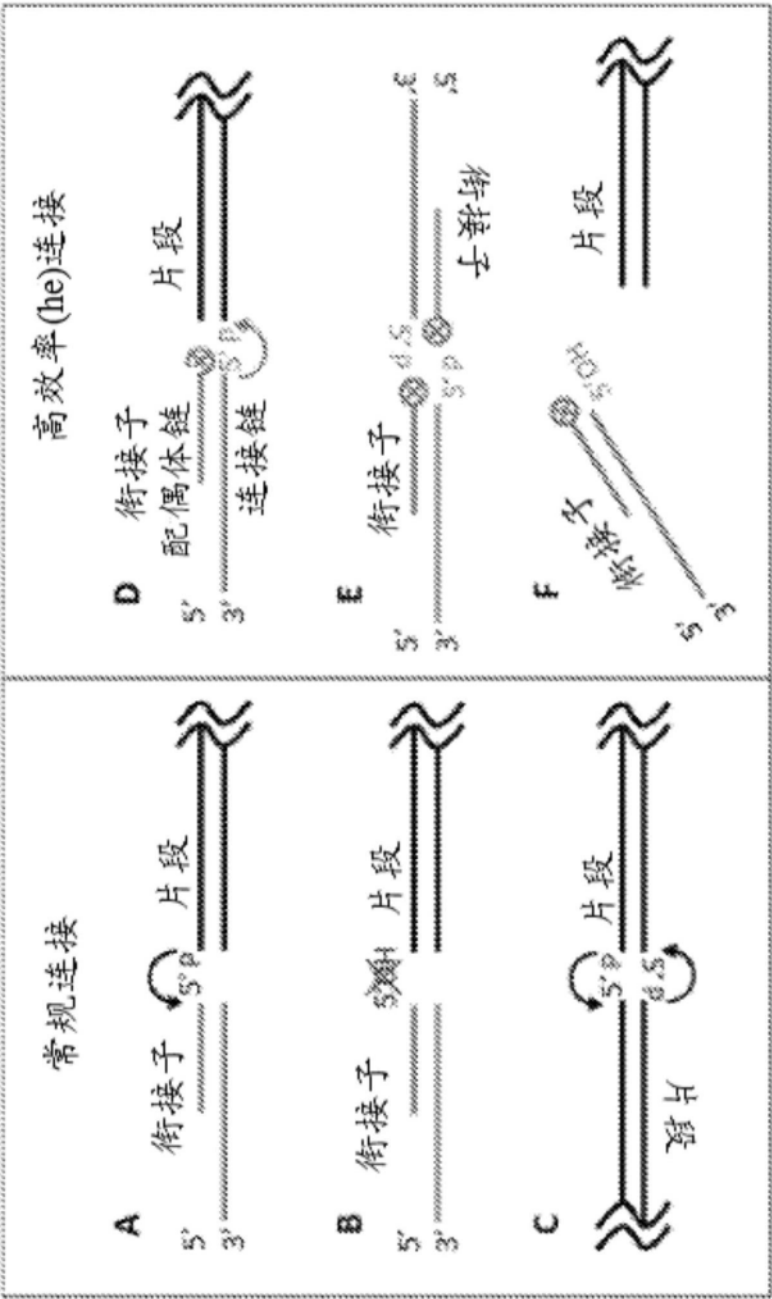


图1

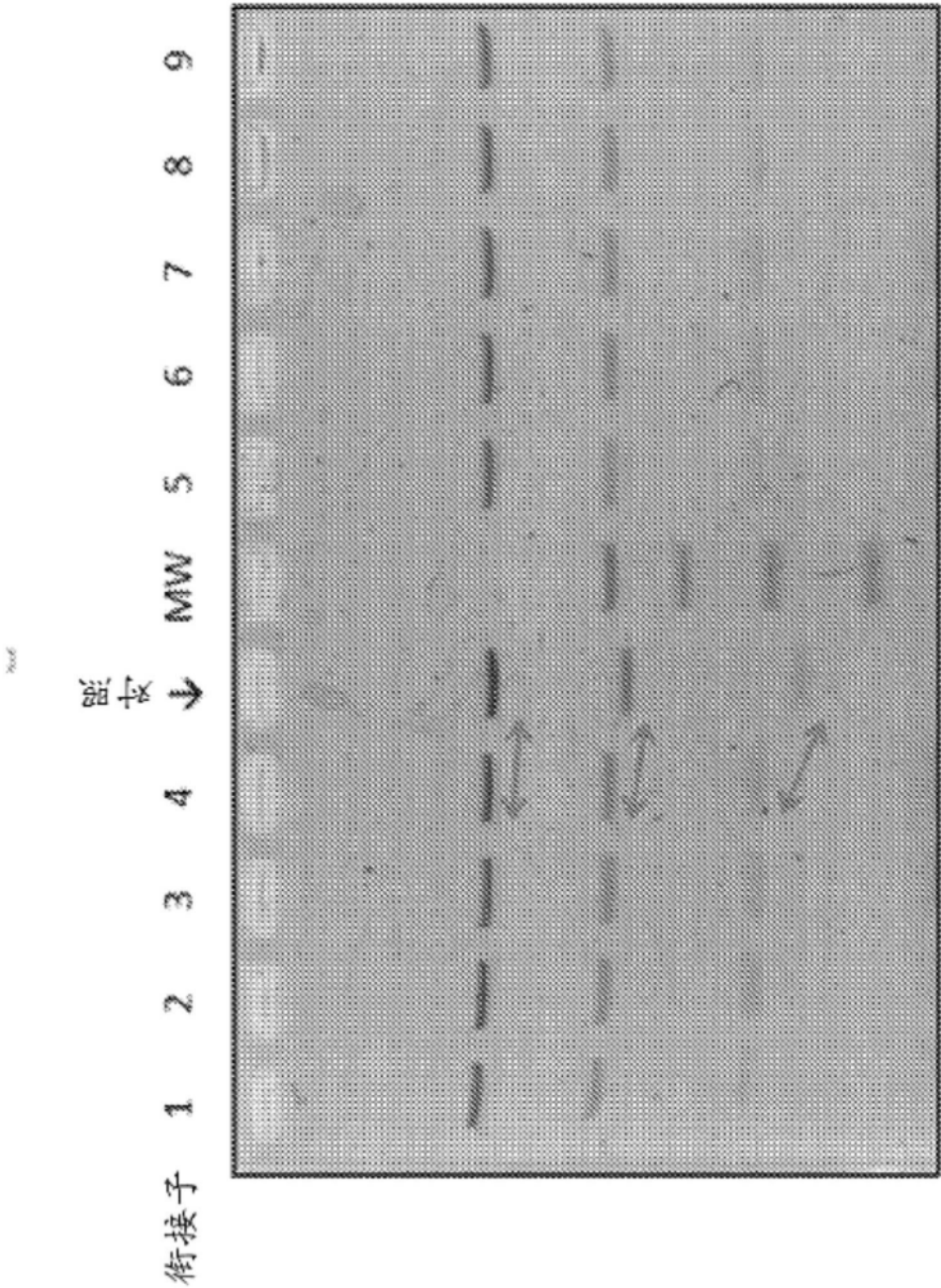


图2

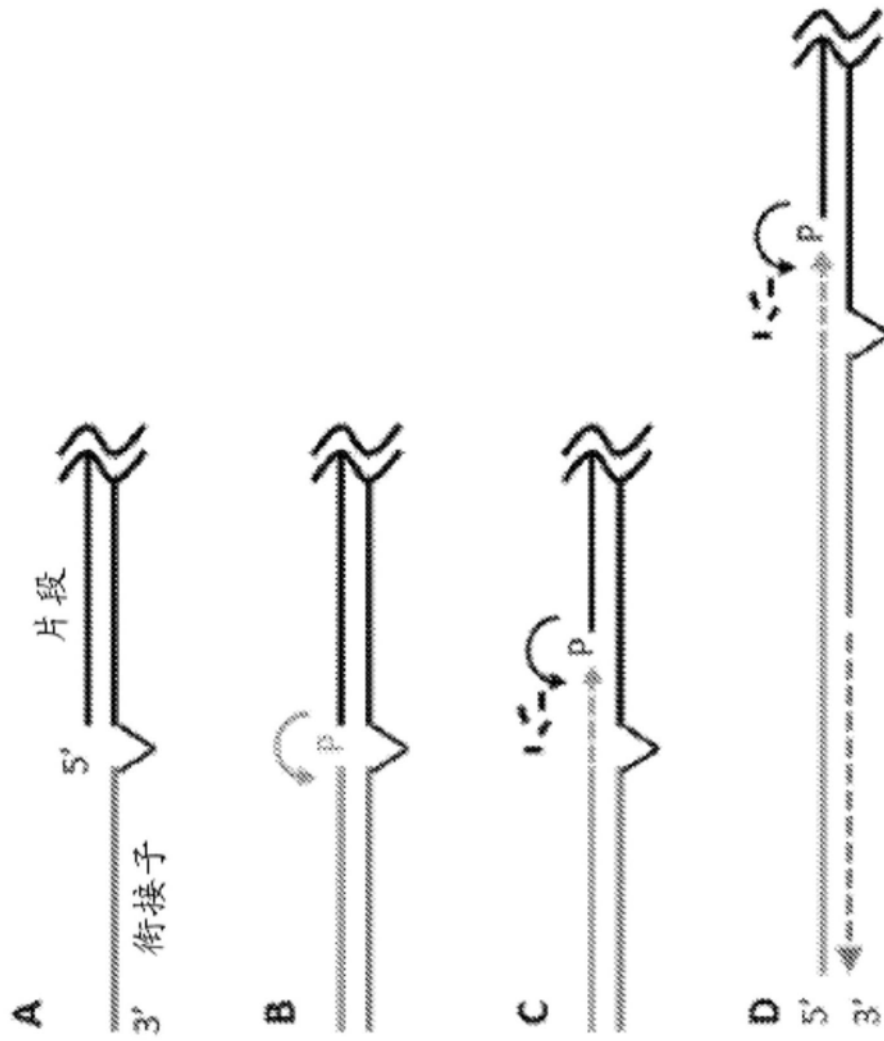


图3

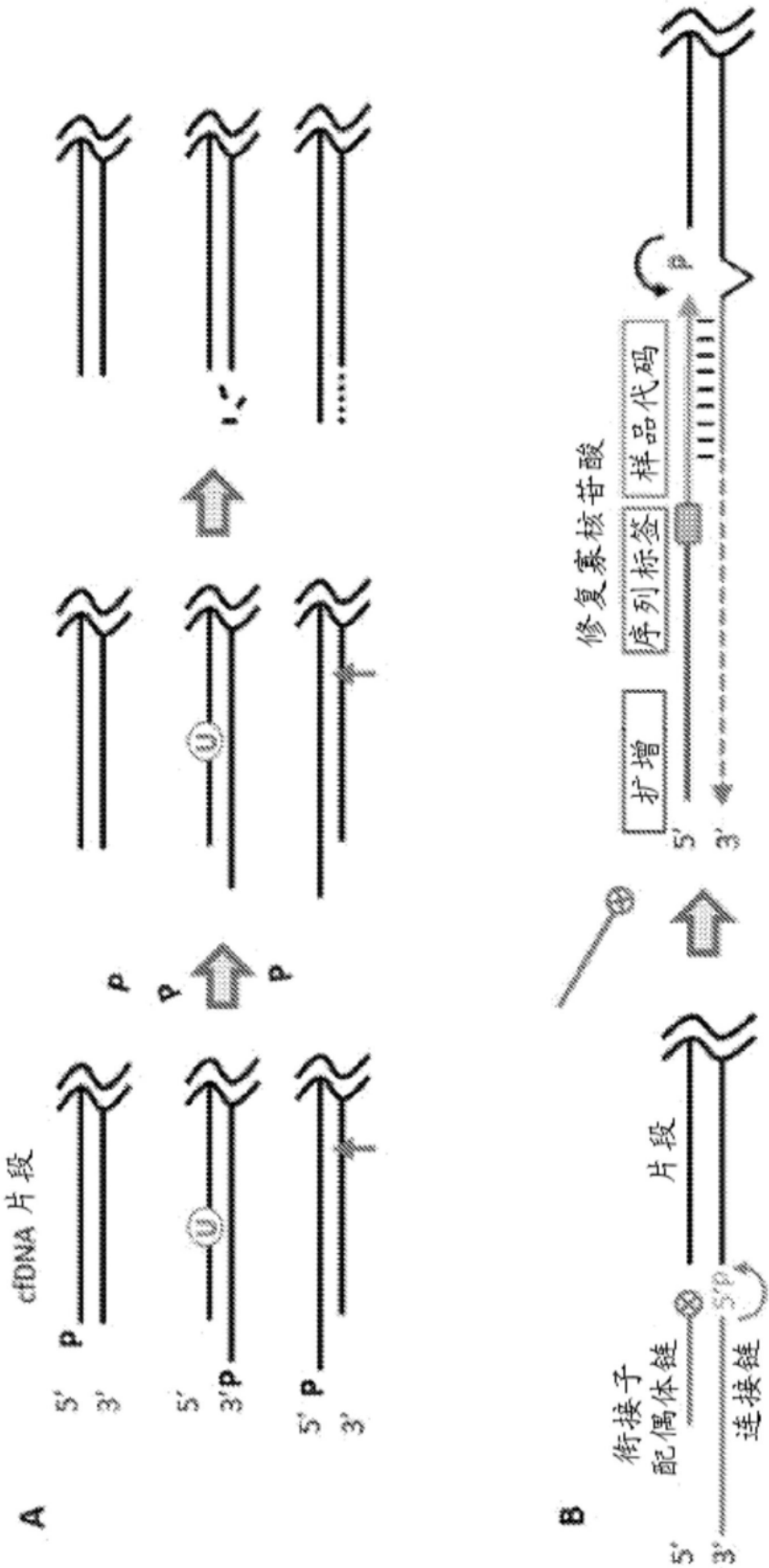


图4

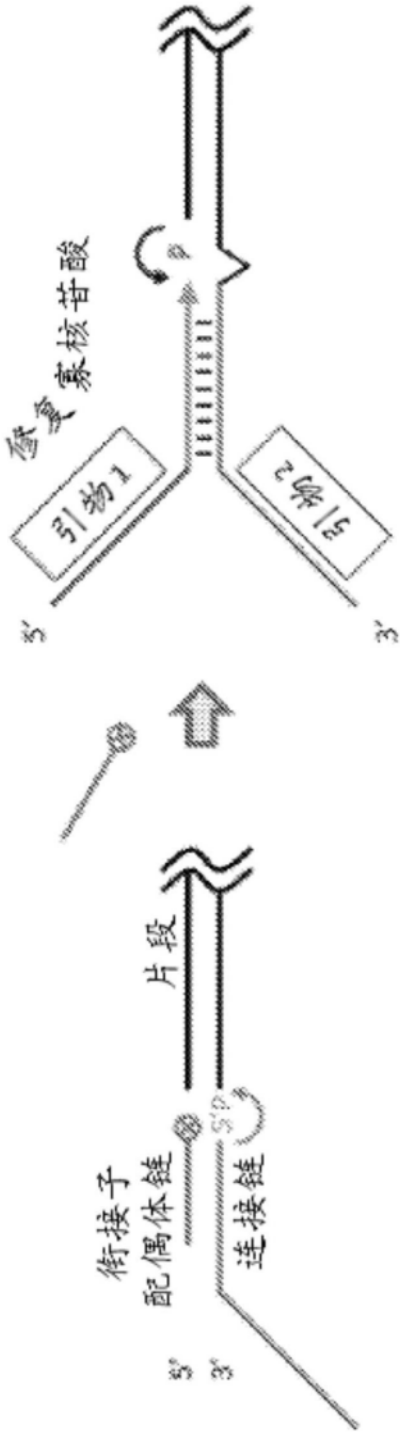


图5