



(12) 发明专利

(10) 授权公告号 CN 107109485 B

(45) 授权公告日 2020.12.08

(21) 申请号 201580066291.7

(22) 申请日 2015.10.08

(65) 同一申请的已公布的文献号
申请公布号 CN 107109485 A

(43) 申请公布日 2017.08.29

(30) 优先权数据
62/062,612 2014.10.10 US
62/062,616 2014.10.10 US

(85) PCT国际申请进入国家阶段日
2017.06.05

(86) PCT国际申请的申请数据
PCT/IB2015/057679 2015.10.08

(87) PCT国际申请的公布数据
W02016/055956 EN 2016.04.14

(73) 专利权人 因维蒂公司
地址 美国加利福尼亚州

(72) 发明人 E·奥利瓦瑞斯

(74) 专利代理机构 北京市金杜律师事务所
11256

代理人 陈文平

(51) Int.Cl.
C12Q 1/6816 (2018.01)

(56) 对比文件
US 2014243232 A1,2014.08.28
WO 2014008447 A1,2014.01.09
Nadin Rohland等.Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture.《Genome Research》.2012,
Maren F. Hansen等.A massive parallel sequencing workflow for diagnostic genetic testing of mismatch repair genes.《Molecular Genetics & Genomic Medicine》.2014,

审查员 魏应亮

权利要求书2页 说明书35页 附图1页

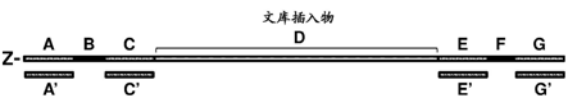
(54) 发明名称

用于多重捕获反应的通用阻断寡聚物系统和改进的杂交捕获的方法

(57) 摘要

在一些实施方式中,本文提供了用于核酸操作和分析的新型组合物和改进方法,其可用于多重核酸测序。在某些实施方式中,本文提出的新型组合物和方法更具成本效益,更有利于自动化,并且比传统方法更快。本文还提供了新型阻断核酸。

四寡聚物
阻断策略



1. 一种用于大规模并行核酸测序的组合物,其包含:

a) 包含多个文库插入物的单链核酸的文库,其中所述文库的每个单链核酸包含 (i) 从四个或更多个样品之一获得的至少一个文库插入物, (ii) 第一非天然核酸,和 (iii) 第二非天然核酸,其中所述第一非天然核酸和所述第二非天然核酸位于所述至少一个文库插入物的相对侧上,并且所述第一非天然核酸包含第一可区分核酸条形码,所述第二非天然核酸包含第二可区分核酸条形码,其中所述第一和第二可区分核酸条形码对于所述四个或更多个样品之一是独特的;和

b) 四个单链通用阻断 (U阻断) 核酸,其中 (i) 第一和第二U阻断核酸配置为与所述第一可区分核酸条形码的相对侧上的所述第一非天然核酸杂交, (ii) 第三和第四U阻断核酸配置为与所述第二可区分核酸条形码的相对侧上的所述第二非天然核酸杂交, (iii) 所述单链U阻断核酸各自基本上不与所述第一或第二可区分核酸条形码的部分杂交,和 (iv) 所述四个单链U阻断核酸各自包含10至40个核苷酸的长度和至少65°C的熔融温度。

2. 权利要求1所述的组合物,其中所述核酸的文库包含至少八个可区分核酸条形码,并且任选地,其中所述至少八个可区分核酸条形码各自存在于所述文库的不同核酸上。

3. 权利要求1或2所述的组合物,其中所述第一和第二U阻断核酸与所述第一非天然核酸的部分基本上互补,并且所述第三和第四U阻断核酸与所述第二非天然核酸的部分基本上互补。

4. 权利要求1至3中任一项所述的组合物,其包含不超过四个U阻断核酸。

5. 权利要求1至4中任一项所述的组合物,其包含一个或多个捕获核酸,其中

(i) 所述捕获核酸包含结合对的成员;和

(ii) 所述捕获核酸各自配置为与所述文库的核酸的亚组特异性杂交。

6. 权利要求1至5中任一项所述的组合物,其中所述核酸的文库包含十个或更多个可区分核酸条形码。

7. 权利要求1至6中任一项所述的组合物,其中所述第一和第二非天然核酸包含衔接子核酸。

8. 权利要求1至7中任一项所述的组合物,其中所述四个U阻断核酸各自包含10至30个核苷酸的长度。

9. 权利要求1至8中任一项所述的组合物,其中所述四个U阻断核酸各自包含锁核酸。

10. 权利要求1至9中任一项所述的组合物,其中所述四个U阻断核酸各自包含桥连核酸。

11. 权利要求1至10中任一项所述的组合物,其中所述四个U阻断核酸各自包含65°C至90°C的熔融温度。

12. 权利要求1至11中任一项所述的组合物,其中所述四个U阻断核酸基本上不与可区分核酸条形码杂交。

13. 权利要求1至12中任一项所述的组合物,其中所述四个U阻断核酸包含链终止子。

14. 一种分析核酸文库的方法,其包括:

a) 获得包含包含多个文库插入物的单链核酸的文库,其中所述文库的每个单链核酸包含 (i) 从四个或更多个样品之一获得的至少一个文库插入物, (ii) 第一非天然核酸,和 (iii) 第二非天然核酸,其中所述第一非天然核酸和所述第二非天然核酸位于所述至少一

个文库插入物的相对侧上,并且所述第一非天然核酸包含第一可区分核酸条形码,所述第二非天然核酸包含第二可区分核酸条形码,其中所述第一和第二可区分核酸条形码对于所述四个或更多个样品之一是独特的;

b) 使所述核酸的文库与四个单链通用阻断(U阻断)核酸接触,其中(i)第一和第二U阻断核酸配置为与所述第一可区分核酸条形码的相对侧上的所述第一非天然核酸杂交,(ii)第三和第四U阻断核酸配置为与所述第二可区分核酸条形码的相对侧上的所述第二非天然核酸杂交,(iii)所述单链U阻断核酸各自基本上不与所述第一或第二可区分核酸条形码的部分杂交,和(iv)所述四个单链U阻断核酸各自包含10至40个核苷酸的长度和至少65℃的熔融温度;和

c) 使所述核酸的文库与一个或多个捕获核酸接触,所述一个或多个捕获核酸各自包含结合对的第一成员,其中所述一个或多个捕获核酸配置为与所述文库的核酸的亚组特异性杂交;

d) 捕获所述捕获核酸,从而提供包含所述文库的核酸的亚组的捕获的核酸;

e) 使所述捕获的核酸在扩增条件下与引物组接触,从而提供扩增子;和

f) 分析所述扩增子。

15. 权利要求14所述的方法,其中所述核酸的文库包含十个或更多个可区分核酸条形码,和所述四个单链U阻断核酸各自包含锁核酸或桥连核酸。

用于多重捕获反应的通用阻断寡聚物系统和改进的杂交捕获的方法

[0001] 相关专利申请

[0002] 本专利申请要求于2015年10月10日提交的题为“UNIVERSAL BLOCKING OLIGO SYSTEM FOR MULTIPLEXED CAPTURE REACTIONS”,发明人名为Eric Olivares,并标示为代理人案卷No.055911-0432232的美国临时专利申请No.62/062612,以及于2014年10月10日提交的题为“METHODS OF HYBRIDIZATION CAPTURE USING NUCLEIC ACID BAITS FROM PAIRED-END SEQUENCING”,发明人名为Eric Olivares,并标示为代理人案卷No.055911-0432231的美国临时专利申请No.62/062616的权益。前述专利申请的全部内容通过引用并入本文,包括所有文本、表格和附图。

技术领域

[0003] 本技术部分涉及核酸操作、分析和高通量测序的组合物和方法。

背景技术

[0004] 活生物体(例如动物、植物、微生物、病毒)的遗传信息编码在脱氧核糖核酸(DNA)或核糖核酸(RNA)中。遗传信息是代表核酸的一级结构的一连串核苷酸或修饰核苷酸。生物体的核酸内容(例如DNA)通常称为基因组。在人类中,完整的基因组通常含有位于二十四(24)条染色体上的约30,000个基因。大多数基因编码特定蛋白,其在通过转录和翻译表达后在活细胞内实现特定生物化学功能。

[0005] 许多医学病症是由基因组内的一个或多个遗传变异引起。一些遗传变异可以使个体易患或产生大量疾病例如糖尿病、动脉硬化、肥胖症、各种自身免疫性疾病和癌症(例如结直肠癌、乳腺癌、卵巢癌、肺癌)中的任一种。这样的遗传疾病可以由基因组内的一个或多个核苷酸的添加、置换、插入或缺失引起。

[0006] 遗传变异可以通过对通常从多个来源获得的核酸混合物的多重分析识别,例如通过使用下一代测序技术。这样的多重分析通常涉及在分析之前的大量核酸操作,涉及不利于高通量处理的许多不同步骤。此外,现有核酸操作方法通常昂贵、耗时,并且经常存在可以导致样品污染的重大缺陷。本文的组合物和方法提供了相对于现有核酸操作和分析技术的明显改进,其更有利于高通量自动化、更具成本效益、更少耗时和/或提供更少污染风险。

发明内容

[0007] 在一些方面,本文提出了一种用于大规模并行核酸测序的组合物,其包含a)包含多个文库插入物的核酸的文库,其中所述文库的每个核酸包含(i)从四个或更多个样品之一获得的至少一个文库插入物,(ii)第一非天然核酸,和(iii)第二非天然核酸,其中所述第一非天然核酸和所述第二非天然核酸位于所述至少一个文库插入物的相对侧上,并且所述第一非天然核酸包含第一可区分核酸条形码,所述第二非天然核酸包含第二可区分核酸条形码,其中所述第一和第二可区分核酸条形码对于所述四个或更多个样品之一是独特

的;和b)四个U阻断核酸,其中(i)第一和第二U阻断核酸配置为与所述第一可区分核酸条形码的相对侧上的所述第一非天然核酸杂交,和(ii)第三和第四U阻断核酸配置为与所述第二可区分核酸条形码的相对侧上的所述第二非天然核酸杂交,和(iii)所述U阻断核酸各自基本上不与所述第一或第二可区分核酸条形码的部分杂交。在某些方面,核酸的文库包含至少八个可区分核酸条形码。

[0008] 在一些方面,组合物还包含一种或多种捕获核酸,其中(i)捕获核酸包含结合对的成员,和(ii)捕获核酸各自配置为与一个或多个文库插入物的亚组特异性杂交。

[0009] 在某些实施方式中,本文还提出了一种分析核酸文库的方法,其包括a)获得包含多个文库插入物的核酸的文库,其中所述文库的每个核酸包含(i)从四个或更多个样品之一获得的至少一个文库插入物,(ii)第一非天然核酸,和(iii)第二非天然核酸,其中所述第一非天然核酸和所述第二非天然核酸位于所述至少一个文库插入物的相对侧上,并且所述第一非天然核酸包含第一可区分核酸条形码,所述第二非天然核酸包含第二可区分核酸条形码,其中所述第一和第二可区分核酸条形码对于所述四个或更多个样品之一是独特的;b)使所述核酸的文库与四个U阻断核酸接触,其中(i)第一和第二U阻断核酸配置为与所述第一可区分核酸条形码的相对侧上的所述第一非天然核酸杂交,和(ii)第三和第四U阻断核酸配置为与所述第二可区分核酸条形码的相对侧上的所述第二非天然核酸杂交,和(iii)所述U阻断核酸各自基本上不与所述第一或第二可区分核酸条形码的部分杂交;和c)使所述核酸的文库与一个或多个捕获核酸接触,所述一个或多个捕获核酸各自包含结合对的第一成员,其中所述一个或多个捕获核酸配置为与所述文库的核酸的亚组特异性杂交;d)捕获所述捕获核酸,从而提供包含所述文库的核酸的亚组的捕获的核酸;e)使所述捕获的核酸在扩增条件下与引物组接触,从而提供扩增子;和f)分析所述扩增子。

[0010] 在某些方面,分析包括提供序列读数。在一些方面,序列读数可以通过包括大规模并行测序和/或双端测序(pair-end sequencing)的方法获得。

[0011] 在关于本文的组合物和方法的某些方面,非天然核酸包含通用核酸。在一些方面,文库的核酸包含四个或更多个、或十个或更多个条形码核酸。在一些方面,每个文库插入物包含一个或两个条形码序列。在某些方面,U阻断核酸包含10至40个核苷酸的长度。在某些方面,U阻断核酸包含10至20个核苷酸的长度。在一些方面,U阻断核酸包含锁核酸和/或桥连核酸。在某些方面,U阻断包含约65℃至约90℃的熔融温度。在某些方面,U阻断核酸包含至少65℃或至少75℃的熔融温度。在一些方面,U阻断核酸不包含简并(degenerate)核苷酸碱基。在一些方面U阻断核酸不包含3-硝基吡咯、5-硝基吡咯、肌苷、2'-脱氧肌苷、类似物、衍生物或其组合。

[0012] 在一些方面,本文提供了一种分析核酸文库的方法,其包括a)获得包含第一组扩增子的核酸的文库,其中每个扩增子包含第一非天然核酸和第二非天然核酸、一个或多个可区分标识符和从一个或多个样品之一获得的文库插入物,其中文库插入物位于第一和第二非天然核酸之间;b)制备混合物,其包括使文库的核酸与一种或多种阻断核酸和捕获核酸接触,其中(i)一种或多种阻断核酸配置为与第一和第二非天然核酸特异性杂交,(ii)捕获核酸包含结合对的第一成员,和(iii)捕获核酸配置为与第一组扩增子的亚组特异性杂交,c)纯化混合物,从而提供纯化核酸,其中纯化核酸包含文库的核酸,一个或多个阻断核酸和捕获核酸,d)在杂交条件下杂交纯化核酸,e)捕获捕获核酸,从而提供捕获的核酸,f)

在扩增条件下将捕获的核酸与引物组接触,从而提供第二组扩增子,以及g)分析第二组扩增子。在一些方面,扩增条件包括热稳定聚合酶和/或聚合酶链式反应。在某些方面,(b)中的制备包括使文库的核酸与竞争者核酸接触。在一些实施方式中,捕获核酸配置为与文库插入物的部分特异性杂交。在某些实施方式中,一个或多个阻断核酸配置为与第一非天然核酸和/或第二非天然核酸的部分特异性杂交。在某些实施方式中,一个或多个阻断核酸包含锁核酸和/或桥连核酸。

[0013] 在一些方面,包含结合对的第一成员的捕获核酸配置为与外显子的部分、内含子、所选染色体的部分和/或包含遗传变异(例如重复、多态性)的DNA区域特异性杂交。在一些实施方式中,结合对的第一成员包含生物素、抗原、半抗原、抗体或其部分。在一些方面,在(e)中的捕获包括使混合物与结合对的第二成员接触。在某些实施方式中,结合对的第二成员包含抗生物素蛋白、蛋白A、蛋白G、抗体或其结合部分。在某些实施方式中,结合对的第二成员包含基质。在某些实施方式中,基质包含磁性化合物。在某些实施方式中,基质包含珠。在某些实施方式中,基质包含聚苯乙烯、聚碳酸酯、琼脂糖凝胶或琼脂糖。在某些实施方式中,基质包含金属。

[0014] 在某些实施方式中,杂交条件包括变性。在某些实施方式中,杂交条件包括将捕获的核酸与第一组扩增子中的一个或多个的部分杂交。在某些实施方式中,杂交条件包括在约25℃至约70℃的温度下孵育捕获的核酸。在某些实施方式中,杂交条件包括在约35℃至约60℃的温度下孵育捕获的核酸。在某些实施方式中,杂交条件包括孵育约1小时至24小时或约12小时至20小时的时间量。在某些实施方式中,杂交条件不包括聚合酶。在某些实施方式中,在(d)中的杂交包括使混合物与杂交缓冲液接触。在某些实施方式中,在(d)中的杂交包括(i)使混合物与杂交缓冲液接触,(ii)变性和(iii)杂交的顺序步骤。

[0015] 在一些方面,分析包括提供序列读数。有时序列读数通过包括下一代测序(例如大规模并行测序)的方法获得。有时测序读数通过包括双端测序的方法获得。

[0016] 在某些实施方式中,第一非天然核酸包含至少一个核酸条形码。在某些实施方式中,第二非天然核酸包含至少一个核酸条形码。

[0017] 在某些实施方式中,本文所要求保护的方法不包括干燥步骤。在某些实施方式中,所述方法不包括(c)之前的变性步骤。在某些实施方式中,所述方法不包括(d)之前的变性步骤。在某些实施方式中,所述方法不包括在(d)之前加热到高于80℃的温度。在某些实施方式中,所述方法不包括在(d)之前加热到高于90℃的温度。在某些实施方式中,在(e)之前,混合物不固定在流动池或阵列的基质上。在某些实施方式中,(c)中的纯化不包括加入配置为结合结合对的第一成员的结合对的第二成员。

[0018] 在一些方面,样品可以从一种或多种物种、一个或多个组织、一种或多种哺乳动物或一个或多个受试者获得。

[0019] 某些实施方式在以下说明书、实施例、权利要求和附图中进一步描述。

附图说明

[0020] 附图说明了本技术的实施方式,而不是限制性的。出于清楚和易于说明,附图不是按比例,并且在一些情况下,各个方面可以扩大或放大显示,以便于理解具体实施方式。

[0021] 图1显示包含四个U阻断核酸(A'、C'、E'和G')的阻断方法的实施方式。图1显示包

含文库插入物(D)和可区分核酸条形码(B和F)的文库(Z)的代表性核酸,其中多个不同的插入物和不同的可区分条形码存在于文库的许多核酸中。图1显示U阻断核酸(A'、C'、E'和G'),其各自配置为与所示非天然核酸(A、C、E和G)的部分特异性杂交,并且所述U阻断核酸邻近核酸条形码(B或F)杂交。

具体实施方式

[0022] 下一代测序(NGS)允许通过比传统测序方法更快且更便宜的方法在全基因组范围内对核酸进行测序和分析。本文的方法和组合物提供可用于定位和鉴定遗传变异和/或相关疾病和病症的先进测序技术的改进。在一些实施方式中,本文提供了部分包括用于NGS的核酸混合物的操作和制备的方法。

[0023] 用基因组核酸作为靶材料的测序应用通常需要从高度复杂的混合物中选择感兴趣的核酸靶。测序工作的质量通常取决于选择过程的效率,其进而依赖于可以相对非靶序列多大程度上富集核酸靶。核酸文库的选择和富集有时候包括通过杂交捕获法捕获衔接子连接的插入物(例如基因组DNA插入物)。

[0024] 大多数下一代测序文库分子含有非天然序列(例如衔接子核酸、条形码序列、引物结合位点和通用序列),使其能够进行随后的测序。在杂交捕获反应期间,非天然序列可以退火到彼此,导致富集的核酸池的污染。这些不需要的序列的大部分通常是由于连接到文库插入物的末端衔接子序列的部分之间的不期望的杂交事件。有时,多个文库插入物可以通过它们的末端衔接子非特异性退火到彼此,从而导致连接和分离在一起的在其他方面不需要的DNA片段的“菊花链(daisy chain)”。

[0025] 减少所谓的“菊花链”效应的一种方法是利用意在与衔接子序列的大部分杂交的封闭核酸。对于传统的方法,衔接子每一侧都需要封闭核酸,并且每个封闭核酸都含有与每个衔接子中含有的衔接子序列(包括条形码序列(例如索引序列))的完美互补匹配。对于高通量多重测序方法,通常混合多个文库,每个文库由不同衔接子序列和不同条形码序列组成。对于这样的多重方法,需要合成多组传统封闭核酸,各自对每个文库的衔接子都具有特异性。这种方法是麻烦和昂贵的,并且需要制造许多不同的、相对长的寡核苷酸,其阻碍文库制备和测序过程的有效且有成本效益的自动化。

[0026] 在一些方面,本文提供了用于减少不需要的捕获事件的新型和改进的组合物以及方法。在一些实施方式中,本文提供了新型U阻断(即,通用封闭)核酸。包含本文提供的新型U阻断核酸的组合物和利用本文提供的U阻断核酸的方法比传统方法成本更低,增加效率和工作流程,并且更有利于自动化。

[0027] 此外,杂交捕获法的传统应用通常涉及将衔接子连接的文库插入物或其扩增子的池与C0t-1DNA和封闭寡核苷酸组合,然后进行干燥步骤。干燥步骤通常在真空中进行,其是耗时的并且在开放系统中进行,其提供样品之间交叉污染的高风险。干燥后,样品变性,然后退火数天。然后加入生物素化的捕获寡核苷酸(例如“诱饵”),并且杂交的核酸通常用抗生物素蛋白涂布的珠拉下。然后将保留的核酸池从珠洗脱,并将其引入到自动测序过程中。上述步骤是无效率且耗时的,不利于自动化并且可以导致交叉污染。

[0028] 在一些方面,本文提供了用于操作和制备用于分析(例如用于高通量测序)的核酸文库的改进方法,该方法不需要延长的孵育时间和/或干燥步骤。

[0029] 受试者

[0030] 受试者可以是任何有生命或无生命的生物体,包括但不限于人类、非人类动物、植物、细菌、真菌、病毒或原生生物。受试者可以是任何年龄(例如胚胎、胎儿、婴儿、儿童、成人)。受试者可以是任何性别(例如雄性、雌性或其组合)。受试者可以是怀孕的。受试者可以是患者(例如人类患者)。

[0031] 样品

[0032] 本文提供了用于分析样品的方法和组合物。样品(例如包含核酸的样品)可以从适合的受试者获得。样品可以直接从受试者或其部分分离或获得。在一些实施方式中,样品间接从个体或医疗专业人员获得。样品可以从受试者或其部分分离或获得的任何标本。样品可以从多个受试者分离或获得的任何标本。标本的非限制性实例包括来自受试者的液体或组织,包括但不限于血液或血液制品(例如血清、血浆、血小板、血沉棕黄层等)、脐带血、绒毛膜绒毛、羊水、脑脊液、脊髓液、灌洗液(如肺、胃、腹膜、导管、耳、关节镜的)、活检样品、腹腔穿刺(centesis)样品、细胞(血细胞、淋巴细胞、胎盘细胞、干细胞、骨髓衍生细胞、胚胎或胎儿细胞)或其部分(例如线粒体、核、提取物等)、尿液、粪便、痰液、唾液、鼻粘液、前列腺液、灌洗液、精液、淋巴液、胆汁、眼泪、汗液、母乳、乳腺液体等或其组合。从其提取核酸的液体或组织样品可以是非细胞的(例如无细胞的)。组织的非限制性实例包括器官组织(例如肝、肾、肺、胸腺、肾上腺、皮肤、膀胱、生殖器官、肠、结肠、脾、脑或其部分)、上皮组织、毛发、毛囊、导管、血管、骨、眼、鼻、口、喉、耳、指甲等、其部分或其组合。样品可以包括正常、健康、患病(例如感染)和/或癌性(例如癌细胞)的细胞或组织。从受试者获得的样品可以包含多种生物体的细胞或细胞材料(例如核酸)(例如病毒核酸、胎儿核酸、细菌核酸、寄生虫核酸)。

[0033] 在一些实施方式中,样品包含核酸或其片段。样品可以包含从一个或多个受试者获得的核酸。在一些实施方式中,样品包含从单个受试者获得的核酸。在一个实施方式中,样品包含核酸混合物。核酸混合物可以包含具有不同核苷酸序列、不同片段长度、不同来源(例如基因组起源、细胞或组织来源、受试者来源等或其组合)或其组合的两种或更多种核酸物质。样品可以包含合成核酸。

[0034] 核酸

[0035] 术语“核酸”是指来自例如DNA(例如互补DNA(cDNA)、基因组DNA(gDNA)等)、RNA(例如信使RNA(mRNA)、短抑制RNA(siRNA)、核糖体RNA(rRNA)、tRNA、microRNA)、和/或DNA或RNA类似物(例如含有碱基类似物、糖类似物和/或非天然骨架等)、RNA/DNA杂交体和聚酰胺核酸(PNA)的任何组合物的一种或多种核酸(例如核酸组或核酸亚组),所有这些都可以是单链或双链形式,除非另有说明,可以包括天然核苷酸的已知类似物,其可以以与天然存在的核苷酸类似的方式起作用。除非特别限定,该术语包括包含脱氧核糖核苷酸、核糖核苷酸和天然核苷酸的已知类似物的核酸。核酸可以包括作为等同物、衍生物或其变体的从核苷酸类似物、单链(“正义”或“反义”、“正”链或“负”链、“前向”阅读框或“反向”阅读框)和双链多核苷酸合成的RNA或DNA的适合的类似物。核酸可以是单链或双链。核酸可以具有2个或更多、3个或更多、4个或更多或5个或更多连续核苷酸的任何长度。核酸可以包含特定5'至3'顺序的本领域已知核苷酸作为序列(例如核酸序列)。

[0036] 核酸可以是天然存在的和/或可以通过人手合成、复制或改变。例如,核酸可以是

扩增子。核酸可以来自核酸文库,例如gDNA、cDNA或RNA文库。核酸可以是合成的(例如化学合成)或产生(例如通过体外聚合酶延伸,例如通过扩增,例如通过PCR)。核酸可以是或可以来自质粒、噬菌体、病毒、自主复制序列(ARS)、着丝粒、人工染色体、染色体或者在某些实施方式中能够在体外或宿主细胞、细胞、细胞的细胞核或细胞质中复制或被复制的其他核酸。核酸(例如核酸的文库)可以包含来自一个样品或来自两个或更多个样品(例如,来自1个或多个、2个或多个、3个或多个、4个或多个、5个或多个、6个或多个、7个或多个、8个或多个、9个或多个、10个或多个、11个或多个、12个或多个、13个或多个、14个或多个、15个或多个、16个或多个、17个或多个、18个或多个、19个或多个或20个或多个样品)的核酸。为本文所述的过程或方法提供的核酸可包含来自1至1000、1至500、1至200、1至100、1至50、1至20或1至10个样品的核酸。

[0037] 术语“基因”指涉及产生多肽链的DNA区段,并且可以包括涉及基因产物的转录/翻译和转录/翻译的调控的编码区之前和之后的区域(前导区和尾随区),以及各个编码区段(外显子)之间的插入序列(内含子)。由于基因序列中的遗传变异(例如基因的编码和非编码部分中的突变),基因可能不一定产生肽或可能产生截短或非功能性蛋白。基因无论是功能性还是非功能性的,通常可以通过与参考基因组中的基因的同源性而鉴定。

[0038] 寡核苷酸是相对短的核酸。寡核苷酸的长度可以为约2至150、2至100、2至50或2至约35个核酸。在一些实施方式中,寡核苷酸是单链。在某些实施方式中,寡核苷酸是引物。引物通常配置为与所选互补核酸杂交,并且配置为在杂交后由聚合酶延伸。

[0039] 核酸分离和纯化

[0040] 可以使用本领域已知的适合方法从一个或多个受试者、一个或多个样品或一种或多种来源衍生、分离、提取、纯化或部分纯化核酸。可以使用任何适合的方法分离、提取和/或纯化核酸。

[0041] 本文所用的术语“分离”是指从其原始环境(例如天然环境,如果其是天然存在的;或宿主细胞,如果是外源性表达的)中移除,并因此通过人为干预(例如“通过人手”)从其原始环境改变的核酸。本文所用的术语“分离的核酸”可以指从受试者(例如人类受试者)移除的核酸。分离的核酸可以具有比源样品中存在的成分的量更少的非核酸成分(例如蛋白质、脂质)。包含分离的核酸的组合物可以不含非核酸组分约50%至大于99%。包含分离的核酸的组合物可以不含非核酸组分的约90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或大于99%。本文所用的术语“纯化”可以指核酸,条件是含有比在使核酸进行纯化程序之前存在的非核酸组分的量更少的非核酸组分(例如蛋白质、脂质、碳水化合物、盐、缓冲液、洗涤剂,或其组合)。包含纯化的核酸的组合物可以不含其他非核酸组分的至少约60%、70%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或大于99%。包含纯化的核酸的组合物可以包含在施用纯化方法之前存在于样品中的总核酸的至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或大于99%。

[0042] 在一些实施方式中,纯化混合物(例如纯化混合物中的核酸)提供纯化的核酸。在某些实施方式中,纯化包含文库核酸、阻断核酸、捕获核酸、竞争者核酸和/或其组合的混合物,从而提供纯化的核酸。核酸纯化有时候包含DNA净化柱或DNA净化珠。各种核酸净化柱、树脂、基质和试剂盒是本领域已知的。任何适合的核酸纯化方法、树脂、珠、基质或试剂可以

用本文的方法使用。例如,核酸纯化方法可以包括将核酸与包含金属的适合的阳离子交换树脂(例如阳离子珠)结合(例如非共价结合),通过使用磁体将结合的核酸复合物拉下,然后通过加入低盐缓冲液洗脱结合的核酸。例如,在某些实施方式中,核酸纯化包括AMPureXP磁珠(Beckman Coulter, Inc., Indianapolis IN, (USA))等的使用。本文使用的核酸纯化方法通常被改变以最佳地回收短核酸(例如阻断核酸)以及文库插入物(例如衔接子连接的插入物及其扩增子)。在某些实施方式中,本文的核酸纯化方法被改变以最佳地回收具有约5至约1000个核苷酸、5至约800个核苷酸或5至约500个核苷酸的平均或绝对长度的核酸。在某些实施方式中,本文使用的核酸纯化方法使用核酸结合树脂(例如核酸结合珠、核酸结合基质、例如100%浆液)与含核酸混合物的比例为1.8:1、1.9:1、2:1、2.1:1、2.2:1、2.3:1、2.4:1、2.5:1、2.6:1、2.7:1、2.8:1、2.9:1或3:1(体积:体积)。

[0043] 在一些实施方式中,纯化过程包含洗涤步骤。在一些实施方式中,纯化过程包含洗脱步骤。

[0044] 在一些实施方式中,本文所述的纯化过程不包含干燥步骤、真空(例如快速真空)和/或冻干的使用。这样的方法导致交叉污染的高风险。在一些实施方式中,尽管混合物的捕获核酸通常包含结合对的成员,但是本文所述的纯化过程不包含结合对的第二成员的使用。在一些实施方式中,在不存在杂交缓冲液的情况下进行纯化方法。在一些实施方式中,在不存在加入的钙或镁盐的情况下进行纯化方法。在一些实施方式中,在不存在洗涤剂(例如SDS)、Ficoll、BAS和/或聚乙烯吡咯烷酮的情况下进行纯化方法。

[0045] 杂交

[0046] 基本上互补的单链核酸可以在杂交条件下彼此杂交,从而形成部分或完全双链的核酸。在一些实施方式中,核酸序列的全部或部分可以与另一个核酸序列基本上互补。如本文所述,“基本上互补”指在适合的杂交条件下可以彼此杂交的核苷酸序列。可以改变杂交条件以忍受基本上互补的互补核酸内的不同量的序列错配。可以彼此杂交的核酸的基本上互补的部分可以彼此75%或更高、76%或更高、77%或更高、78%或更高、79%或更高、80%或更高、81%或更高、82%或更高、83%或更高、84%或更高、85%或更高、86%或更高、87%或更高、88%或更高、89%或更高、90%或更高、91%或更高、92%或更高、93%或更高、94%或更高、95%或更高、96%或更高、97%或更高、98%或更高或99%或更高互补。在一些实施方式中,可以彼此杂交的核酸的基本上互补的部分是100%互补。配置为彼此杂交的核酸或其部分通常包含彼此基本上互补的核酸序列。

[0047] 如本文所用,“特异性杂交”指在杂交条件下的优先杂交,其中基本上互补的两个核酸或其部分彼此杂交,而不与与两个核酸中的任一个基本上不互补的其他核酸杂交。例如,特异性杂交包括捕获核酸与与捕获核酸基本上互补的靶扩增子的部分杂交。在一些实施方式中,配置为特异性杂交的核酸或其部分通常在核酸序列的连续部分上彼此约80%或更高、81%或更高、82%或更高、83%或更高、84%或更高、85%或更高、86%或更高、87%或更高、88%或更高、89%或更高、90%或更高、91%或更高、92%或更高、93%或更高、94%或更高、95%或更高、96%或更高、97%或更高、98%或更高、99%或更高、或100%互补。特异性杂交与非特异性杂交相互作用(例如未配置为特异性杂交的两个核酸、例如80%或更少、70%或更少、60%或更少或50%或更少的互补性的两个核酸)有约2倍或更高、通常约10倍或更高,有时约100倍或更高、1000倍或更高、10,000倍或更高、100,000倍或更高、或1,000,

000倍或更高的差别。

[0048] 在一些实施方式中,本文所述的方法包括在杂交条件下杂交核酸。有利于基本上互补的核酸的杂交的条件在本文中称为杂交条件。改变杂交条件严格性的方法是本领域熟知的。

[0049] 杂交条件可以根据试验中所用核酸特性确定和/或调整。用于优化杂交条件的方法是本领域熟知的,并且可见于Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989)。核酸序列含量(例如GC含量、错配度)和/或长度有时可以影响基本上互补的核酸的杂交。杂交条件通常包括参数,其可以被调节以使感兴趣的两个或更多个基本上互补的核酸最佳地退火。这样的可调节参数的非限制性实例包括温度、一价或二价离子和/或阳离子浓度(例如Mg浓度)、缓冲液浓度、磷酸盐浓度、甘油浓度、DMSO浓度、核酸浓度等或其组合。根据基本上互补的核酸之间的错配程度,可以调整杂交条件以影响退火和/或以选择所选核酸的特异性杂交(例如具有已知或预测的熔融温度的寡核苷酸或引物)。

[0050] 杂交条件通常包括将包含核酸的样品加热或冷却至适合的温度。杂交的适合温度有时为约0℃至80℃、约25℃至80℃、约25℃至70℃、约30℃至70℃、约35℃至70℃、约40℃至70℃、约35℃至65℃、约35℃至60℃、约35℃至55℃、或约40℃至50℃。在一些实施方式中,杂交条件包括将样品冷却至约40℃、41℃、42℃、43℃、44℃、45℃、46℃、47℃、48℃、49℃、50℃、51℃、52℃、53℃、54℃或约55℃的温度。在一些实施方式中,在杂交条件下杂交纯化的核酸混合物包括变性步骤,随后是杂交步骤(例如,在适合杂交的温度下孵育一段时间)。杂交条件有时包括使核酸混合物变性,然后立即将混合物冷却(例如温度迅速降低)至适合的杂交温度。

[0051] 在某些实施方式中,杂交条件包括变性过程。变性通常包括将包含核酸的样品的温度升高(例如通过加热)至等于或高于样品中的一个或多个双链核酸的熔点的温度。在一些实施方式中,变性包括将样品的温度从约70℃升至约120℃、约85℃至约105℃、约90℃至约105℃或约95℃至约105℃。在一些实施方式中,变性包括将样品的温度升高至约70℃或更高、约75℃或更高、约80℃或更高、约85℃或更高、或至约90℃、约91℃、约92℃、约93℃、约94℃、约95℃、约96℃、约97℃、约98℃、约99℃、约100℃、约101℃、约102℃、约103℃、约104℃或至约105℃。在一些实施方式中,核酸在期望变性温度下变性约1秒至约30分钟或更长、约15秒至约30分钟、约30秒至约30分钟、约1分钟至约30分钟、约1分钟至约20分钟、约1分钟至约15分钟、或约5分钟至约10分钟。

[0052] 杂交条件通常包含孵育期,其中将样品在期望杂交温度保持一段时间。杂交文库核酸、阻断和/或捕获核酸的传统方法需要超过48小时的杂交时间。虽然可以使用任何适合的条件进行杂交,但是本文提供的某些方法提供显著减少的杂交时间。在某些实施方式中,杂交条件包括在期望杂交温度下孵育混合物约1分钟至约24小时、约5分钟至约24小时、约10分钟至约24小时、约15分钟至约24小时、约30分钟至约24小时、约1小时至约24小时、约2小时至约24小时、约8小时至约24小时、约10小时至约24小时、约10小时至约20小时、或约12小时至约20小时。在一些实施方式中,杂交条件包括在期望杂交温度下孵育混合物约10、11、12、13、14、15、16、17、18、19或20小时。在一些实施方式中,杂交条件包括在期望杂交温度下孵育样品不长于约48小时、不长于约24小时或不长于约18小时。

[0053] 在一些实施方式中,杂交包括将核酸与杂交缓冲液接触。在一些实施方式中,杂交条件包括核酸与适合的杂交缓冲液的混合物。杂交缓冲液是本领域已知的并且可以商购。杂交缓冲液可以包含洗涤剂(例如SDS)、Ficoll、甘油、BAS、聚乙烯吡咯烷酮、葡聚糖甘油、二价阳离子(例如钙和/或镁)、一价阳离子、磷酸盐等或其组合。

[0054] 在一些实施方式中,本文所述的方法不包括在杂交条件下杂交纯化的核酸的步骤之前的变性步骤。

[0055] 在某些实施方式中,杂交条件不包括聚合酶。在一些实施方式中,杂交条件不包括聚合酶链式反应。在某些实施方式中,在捕获混合物的核酸之后,混合物不包含聚合酶。

[0056] 扩增

[0057] 可以通过适合的方法扩增核酸。本文所用的术语“扩增”指使样品中的靶核酸进行线性或指数地产生具有与靶核酸或其片段相同或基本上相同(例如基本上同一)的核苷酸序列的扩增子核酸的方法。在一些实施方式中,扩增反应包括适合的热稳定聚合酶。热稳定聚合物是本领域已知的,并且当与大多数哺乳动物中发现的普通聚合酶相比时,在高于80℃的温度下在延长的时间段内是稳定的。在某些实施方式中,术语“扩增”指包括聚合酶链式反应(PCR)的方法。有利于扩增的条件(即扩增条件)是众所周知的,并且通常包括至少适合的聚合酶、适合的模板、适合的引物或引物组、适合的核苷酸(例如dNTP)、适合的缓冲液以及应用适合的退火、杂交和/或延伸时间和温度。在某些实施方式中,扩增产物(例如扩增子)可以含有与从其产生扩增子的模板序列或其部分相比一个或多个另外的和/或不同的核苷酸(例如引物可以含有“额外的”核苷酸)。

[0058] 核酸可以通过热循环法或通过等温扩增法扩增。在一些实施方式中,使用滚环扩增法。在一些实施方式中,扩增发生在固体载体上(例如在流动池内),其中核酸、核酸文库或其部分被固定。在某些测序方法中,将核酸文库加入到流动池中,并通过在适合条件下与锚杂交而固定。这种类型的核酸扩增通常被称为固相扩增。在固相扩增的一些实施方式中,扩增产物的全部或部分通过从固定化引物起始的延伸合成。固相扩增反应类似于标准溶液相扩增,除了将扩增寡核苷酸(例如引物)中的至少一个固定在固体载体上。

[0059] 在一些实施方式中,固相扩增包括核酸扩增反应,其包含仅一种固定到表面的寡核苷酸引物。在某些实施方式中,固相扩增包括多种不同的固定化寡核苷酸引物种类。在一些实施方式中,固相扩增可以包括核酸扩增反应,其包含固定在固体表面上的一种寡核苷酸引物和溶液中的第二种不同的寡核苷酸引物。可以使用多种不同种类的固定的或基于溶液的引物。固相核酸扩增反应的非限制性实例包括界面扩增、桥扩增、乳液PCR、WildFire扩增(例如美国专利公开US20130012399)等或其组合。

[0060] 核酸文库

[0061] 在某些实施方式中,核酸文库(例如核酸的文库)是从一个或多个受试者获得的总gRNA、RNA或cRNA的集合或亚组。核酸文库可以包含单链和/或双链核酸。核酸文库通常从一个或多个样品产生,并且包含从其获得样品的一个或多个受试者或生物体的内源或天然的核酸。核酸文库通常包含从其获得样品的一个或多个生物体的内源或天然的多个核酸或核酸片段。这样的内源或天然的核酸有时在本文中称为文库插入物。因此,多个核酸可以指 10^3 至 10^{20} 个核酸。在一些实施方式中,多个核酸指 10^3 或更多、 10^4 或更多、 10^5 或更多、 10^6 或更多、 10^7 或更多、 10^8 或更多、 10^9 或更多、 10^{10} 或更多、或 10^{12} 或更多个核酸(或插入物)。文库

插入物可以是基因组DNA (例如基因组DNA文库)、RNA (例如RNA文库) 或cDNA (例如cDNA文库) 的片段。文库插入物可以包含全长基因、cDNA、内含子、外显子、非翻译区 (例如启动子、增强子、调控序列等) 基因、其部分或其组合。从任一来源或受试者获得的核酸的文库通常包含1000个或更多、10,000个或更多或100,000个或更多的文库插入物,其是不同的且经常可以彼此区分。在一些实施方式中,核酸的文库包含多个文库插入物,其中针对特定过程制备、组装和/或改变核酸,所述过程的非限制性实例包括固定在固相 (例如固体载体,如流动池、珠) 上、富集、扩增、克隆、检测和/或用于核酸测序。在一些实施方式中,核酸的文库包含从一个或多个样品 (例如一个或多个受试者、一个或多个组织、一种或多种物种或其组合) 获得的一个或多个文库插入物。在一些实施方式中,文库的每个核酸包含至少一个文库插入物 (例如一个、两个、三个或更多个插入物)、包含一个或多个核酸条形码的一个或多个非天然核酸。在某些实施方式中,在测序过程之前或期间制备核酸文库。核酸文库 (例如测序文库) 可以通过本领域已知的适合方法制备。核酸文库可以通过靶向或非靶向制备过程制备。

[0062] 在一些实施方式中,核酸的文库被改变以包含一种或多种非天然核酸,其通常具有已知组成 (例如合成核酸、异源核酸)。在一些实施方式中,文库的每个核酸包含1、2、3、4、5、6、7、8、9或10个或更多个非天然核酸。在一些实施方式中,文库的每个核酸包含一个或两个非天然核酸。非天然核酸可以在适合的位置加入,例如加入到文库插入物的一端、另一端或两端上。在一些实施方式中,文库的核酸包含在文库插入物的相对端上的非天然核酸。例如,文库的核酸可以包含位于文库插入物的5' 端和/或3' 端处的非天然核酸。在一些实施方式中,非天然核酸共价地结合到文库插入物,例如通过适合的磷酸二酯键。在一些实施方式中,文库的核酸包含文库插入物、第一非天然核酸和第二非天然核酸,其中第一和第二非天然核酸位于文库插入物的相对侧 (例如5' 侧和3' 侧) 上。

[0063] 在某些实施方式中,非天然核酸对于从其获得核酸文库的一个或多个受试者或生物体不是天然的和/或不是内源的。对于从其获得核酸文库插入物的受试者或生物体不是天然的和/或不是内源的非天然核酸通常不包含衍生自所述受试者或生物体的基因组DNA或RNA (例如cDNA)。

[0064] 在某些实施方式中,非天然核酸包含适合的外源核酸和/或适合的合成核酸。合成核酸的非限制性实例包括可区分标识符、核酸条形码 (例如可区分核酸条形码)、捕获核酸、序列标签、随机核酸序列、衔接子核酸、限制性酶位点、突出端、启动子、增强子、复制起点、茎环、引物结合位点、寡核苷酸退火位点、适合的整合位点 (例如转座子、病毒整合位点)、一个或多个修饰的核苷酸等、其部分或其组合。非天然核酸可具有相同或不同核酸序列。在一些实施方式中,非天然核酸配置为与一个或多个捕获核酸、阻断核酸 (例如U阻断核酸) 或引物杂交。

[0065] 在某些实施方式中,核酸文库通过基于连接的文库制备方法制备。在一些实施方式中,通过基于连接的文库制备方法将非天然核酸加入到文库的核酸中。基于连接的文库制备方法通常使用一个或多个衔接子。衔接子通常是合成核酸 (例如通过人手制成),其包含核酸序列,其不是从其衍生文库的生物体的内源的或者不存在于从其衍生文库的生物体中。在一些实施方式中,非天然核酸包含衔接子。在某些实施方式中,衔接子可用于制备用于分析 (例如单读数测序、双端测序和多重测序) 的文库插入物。在一些实施方式中,核酸文库制备包括将一个或多个衔接子连接到多个文库插入物。衔接子可以是相对短的双链或单

链寡核苷酸(例如约2至约10、2至约30、2至约50或2至约100个核酸或更多),其可以包括,例如,可区分标识符、可区分核酸条形码和/或结合对的一个或多个成员。衔接子通常位于文库插入物的5' 和/或3'。衔接子通常位于文库插入物的侧面。在一些实施方式中,双链衔接子的仅一条链被引入文库的核酸中(例如连接到文库插入物)。有时双链衔接子的两条链都被引入到文库中。在一些实施方式中,文库的单链核酸包含5' 衔接子和3' 衔接子,其中5' 和3' 衔接子的序列基本上不同和/或基本上不互补。

[0066] 在某些实施方式中,衔接子包含与流动池锚基本上互补的部分,该部分有时用于将核酸文库固定到固体载体,例如流动池的内表面。在某些实施方式中,衔接子包含与扩增引物或测序引物基本上互补的部分,其可以是衔接子的相同、不同或重叠部分。在一些实施方式中,文库的核酸的至少一个衔接子(例如5' 或3' 衔接子)包含可区分标识符(例如可区分条形码序列)。在一些实施方式中,双链衔接子的两条链都包含核酸条形码,其中衔接子的第一链的核酸条形码与衔接子的第二链的核酸条形码基本上互补。

[0067] 在某些实施方式中,两个或更多个非天然核酸包含基本上相同的部分。基本上相同的非天然核酸的部分有时称为通用核酸(例如通用核酸序列)。在一些实施方式中,非天然核酸包含一个或多个通用核酸。在某些实施方式中,非天然核酸包含第一通用核酸、第二通用核酸和核酸条形码,其中条形码位于第一和第二通用核酸之间。这样的非天然核酸可以位于文库插入物的5' 和/或3'。例如,在一些实施方式中,文库的核酸可以包含第一通用核酸、第一核酸条形码和位于文库插入物的5' 的第二通用核酸,和第三通用核酸、第二条形码和位于文库插入物的3' 的第四通用核酸。在一些实施方式中,文库的核酸可以包含第一非天然核酸和位于文库插入物的5' 的第一核酸条形码以及第二非天然核酸和位于文库插入物的3' 的第二条形码。在某些实施方式中,非天然核酸被设计和/或配置为包含在可区分条形码侧面的通用序列。在一些实施方式中,本文的U阻断核酸被设计和/或配置为与通用核酸序列特异性杂交。

[0068] 在一些实施方式中,核酸文库或其部分被扩增(例如通过基于PCR的方法扩增)。在一些实施方式中,测序方法包含核酸文库的扩增。核酸文库可以在固定在固体载体(例如流动池中的固体载体)上之前或之后扩增。在一些实施方式中,核酸文库包含扩增子。核酸文库的扩增子可以是单链或双链。在一些实施方式中,核酸文库的扩增子包含文库插入物和衔接子序列(例如侧面是衔接子序列的文库插入物)。在一些实施方式中,核酸文库包含多个扩增子,其有时在本文中称为扩增子的文库。

[0069] 在某些实施方式中,扩增子文库的每个扩增子包含文库插入物和一个或多个非天然核酸。例如,在一些实施方式中,扩增子包含文库插入物和1、2、3、4、10或更多个、或50或更多个非天然核酸。扩增子通常包含位于一个或多个5' 非天然核酸与一个或多个3' 非天然核酸之间的文库插入物。在某些实施方式中,扩增子包含位于文库插入物的5' 的1、2、3、4或5个非天然核酸和位于文库插入物的3' 的1、2、3、4或5个非天然核酸。在一些实施方式中,扩增子包含位于文库物的5' 的一个或多个可区分条形码和位于文库插入物的3' 的一个或多个可区分条形码。在某些实施方式中,扩增子包含位于文库插入物的5' 的1、2、3、4或5个可区分条形码和位于文库插入物的3' 的1、2、3、4或5个可区分条形码。

[0070] 在一些实施方式中,文库的扩增子包含1个或多个基本上相同的非天然核酸。基本上相同的核酸是在核酸序列中至少85%、至少86%、至少87%、至少88%、至少89%、至少

90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少100%相同。类似地,基本上相同的两个或更多个核酸是指包含基本上相同的核酸序列的两个或更多个核酸。

[0071] 基本上不同的核酸是指在核酸序列中小于50%、小于60%、小于70%、小于80%或小于85%相同的两个或更多个核酸序列。基本不同的两个或更多个核酸指包含基本不同的核酸序列的两个或更多个核酸。

[0072] 在一些实施方式中,本文的方法包括获得包含从一个或多个样品获得的一个或多个核酸插入物的核酸的文库。在一些实施方式中,核酸文库可以通过如本文所述地产生文库或通过本领域已知的方法获得。在某些实施方式中,核酸文库从第三方获得,其中第三方产生核酸文库。在一些实施方式中,核酸文库从供应商处购买。在一些实施方式中,本文的方法包括获得包含从一个或多个样品获得的一个或多个文库插入物的核酸的文库,其中文库的每个核酸包含至少一个核酸插入物、第一非天然核酸、第二非天然核酸和一个或多个核酸条形码,其中第一非天然核酸位于至少一个文库插入物的5' 和第二非天然核酸位于文库插入物的3'。

[0073] 可区分标识符

[0074] 在一些实施方式中,核酸包含一个或多个可区分标识符。可以将可区分标识符引入或连接(例如共价地、非共价地、不可逆地或可逆地连接)至核酸(例如多核苷酸),其允许包含标识符的核酸的检测和/或鉴定。在一些实施方式中,可区分标识符在测序方法之前或期间被引入或连接至核酸(例如通过聚合酶)。任何适合的可区分标识符和/或可检测标识符可用于本文所述的组合物或方法。在某些实施方式中,可区分标识符可直接地或间接地联系到(例如结合至)与核酸相关(例如结合)。例如,可区分标识符可以共价地或非共价地结合到核酸。在一些实施方式中,可区分标识符与结合剂或结合对的成员结合或相关,所述结合剂或结合对的成员与核酸共价地或非共价地结合。在一些实施方式中,可区分标识符与核酸可逆地相关。在某些实施方式中,可以使用适合的方法(例如通过增加盐浓度、变性、洗涤、加入适合的溶剂和/或通过加热)从核酸中除去与核酸可逆地相关的可区分标识符。

[0075] 在一些实施方式中,1个或更多个、2个或更多个、3个或更多个、4个或更多个、5个或更多个、6个或更多个、7个或更多个、8个或更多个、9个或更多个、10个或更多个、20个或更多个、30个或更多个或50个或更多个可区分标识符用于本文所述的方法中(例如核酸检测、分析和/或测序方法)。

[0076] 在一些实施方式中,可区分标识符是标签。在一些实施方式中,核酸包含可检测标签,其非限制性实例包括放射性标签(例如同位素)、金属标签、荧光标签、发色团、化学发光标签、电-化学发光标签(例如Origen™)、磷光标签、猝灭剂(例如荧光团猝灭剂)、荧光共振能量转移(FRET)对(例如供体和受体)、染料、蛋白质(例如酶(例如碱性磷酸酶和辣根过氧化物酶)、抗体、抗原或其部分、接头、结合对的成员)、酶底物、小分子(例如生物素、抗生物素蛋白)、质量标签、量子点、纳米颗粒等或其组合。适合的荧光团可用作标签。可以通过多种适合技术,例如流式细胞术、定量聚合酶链式反应(qPCR)、凝胶电泳、基因芯片分析、微阵列、质谱、细胞荧光分析、荧光显微镜、共聚焦激光扫描显微镜、激光扫描细胞计数、亲和层析、手动批处理模式分离、电场悬浮、测序等及其组合,检测和/或定量发光标签。

[0077] 在一些实施方式中,可区分标识符是核酸条形码。

[0078] 核酸条形码

[0079] 在一些实施方式中,非天然核酸包含一个或多个可区分核酸条形码(例如索引核苷酸、序列标签或“条形码”核苷酸)。核酸条形码通常是具有特定序列的核酸,其被引入或附加于(例如,相关于)样品的特定核酸或核酸亚组以跟踪和/或鉴定在核酸混合物中的特定核酸或核酸亚组。在某些实施方式中,可区分核酸条形码包含可用作标识符的核苷酸的可区分序列,以允许在样品、方法或试验中明确鉴定一个或多个核酸(例如核酸亚组)。可区分核酸条形码通常配置为允许明确鉴定与条形码相关的核酸的来源或身份。在一些实施方式中,可区分核酸条形码(例如条形码)可以允许鉴定从不同来源获得的核酸的混合物中的特定核酸的来源。在一些实施方式中,可区分核酸条形码配置为允许明确鉴定与条形码相关的核酸的来源或身份。例如,在某些实施方式中,可区分核酸条形码对于从相同受试者或组织、特定核酸属或亚组、特定核酸种、来自相同染色体的核酸等或其组合获得的某个样品、样品来源、核酸的文库具有特异性和/或独特性。在一些实施方式中,包含衍生自样品、受试者或组织的插入物的核酸包含对于所述样品、受试者或组织具有特异性和独特性的核酸条形码,从而允许明确鉴定衍生自不同样品、受试者或组织的核酸和/或来自核酸的插入物。因此,对于样品、受试者或组织具有独特性的可区分核酸条形码通常与核酸混合物中的其他核酸条形码可区分且不同。在一些实施方式中,具有独特性的可区分核酸条形码与混合物中的其他条形码不同和/或可区分,所述组合物包含衍生自一种或多种来源的一个或多个样品(例如衍生自不同样品或来源的核酸的文库)。在一些实施方式中,对于样品、受试者或组织具有独特性的可区分核酸条形码与衍生自相同样品、受试者、组织的核酸或其特定亚组相关(或包含在内)。因此,在一些实施方式中,衍生自相同样品、受试者或组织的核酸通常包含具有相同序列的至少一个可区分核酸条形码,其与相同样品、受试者或组织的每个核酸相关。

[0080] 在一些实施方式中,可区分条形码包含具有4至10、4至15、4至20、4-50或者20或更多个连续核苷酸的可区分和/或独特序列。可区分的两个核酸条形码在序列中可以有1、2、3、4、5或更多个核苷酸不同。因此,在某些实施方式中,不同和/或可区分的两个核酸条形码可以至多99%相同,并且包含有至少一个核苷酸不同的核酸序列。在一些实施方式中,可区分条形码包含具有4、5、6、7、8、9、10、11、12、13、14或15或更多个连续核苷酸的可区分和/或独特的序列。在一些实施方式中,可区分条形码包含不超过10、不超过15或不超过20个连续核苷酸的可区分和/或独特的序列。可区分核酸条形码通常包含第一端和第二端。可区分核酸条形码的末端可以被鉴定为可区分核酸条形码序列的最5' 或最3' 核苷酸碱基。例如,在一些实施方式中,可以将第一端鉴定为可区分核酸条形码序列的最5' 核苷酸碱基,并且可以将第二端鉴定为可区分核酸条形码序列的最3' 核苷酸碱基。第一端和第二端通常位于可区分核酸条形码的相对端。在一些实施方式中,可以通过核酸测序方法区分和/或鉴定文库中的任何两个或更多个可区分核酸条形码。在一些实施方式中,可以通过杂交方法区分和/或鉴定两个或更多个可区分核酸条形码。

[0081] 在一些实施方式中,从多个来源或样品获得的核酸的文库包含多个可区分核酸条形码。在一些实施方式中,文库的每个可区分核酸条形码可用于鉴定混合文库的每个核酸的来源。例如,从多个来源获得的核酸的文库可以包含从第一受试者获得的核酸的第一文库,其包含第一和/或第二可区分核酸条形码,从第二受试者获得的核酸的第二文库,其包

含第三和/或第四可区分核酸条形码,从第三受试者获得的核酸的第三文库,其包含第五和/或第六可区分核酸条形码,从第四受试者获得的核酸的第四文库,其包含第七和/或第八可区分核酸条形码,等等。在一些实施方式中,从单个来源获得的文库的每个核酸包含1、2、3或4个可区分核酸条形码,其中每个可区分核酸条形码包含不同的序列并且其中每个可区分核酸条形码冗余地鉴定相同的来源。在一些实施方式中,包含从多个样品获得的多个文库插入物的核酸文库包含至少8个、至少10个、至少15个或至少20个可区分核酸条形码。在一些实施方式中,核酸文库,例如包含从多个样品或来源获得的多个文库插入物的核酸文库包含10个或更多个、20个或更多个、50个或更多个或100个或更多个可区分核酸条形码。

[0082] 在一些实施方式中,非天然核酸包含一个或多个可区分核酸条形码。在某些实施方式中,非天然核酸包含一个或两个可区分核酸条形码。在某些实施方式中,非天然核酸不包含可区分核酸条形码。在一些实施方式中,文库的每个核酸包含(i)文库插入物,(ii)第一非天然核酸,(iii)第二非天然核酸,和(iv)可区分核酸条形码,其中第一非天然核酸和第二非天然核酸位于文库插入物的相对侧上,并且第一非天然核酸或第二非天然核酸包含可区分核酸条形码。在一些实施方式中,文库的每个核酸包含(i)文库插入物,(ii)第一非天然核酸,(iii)第二非天然核酸,(iv)第一可区分核酸条形码和(v)第二可区分核酸条形码,其中第一非天然核酸和第二非天然核酸位于文库插入物的相对侧上,第一非天然核酸包含第一可区分核酸条形码并且第二非天然核酸包含第二可区分核酸条形码。

[0083] 包含可区分核酸条形码的非天然核酸通常包含邻近可区分核酸条形码的一端或两端的一个或两个部分。在一些实施方式中,邻近可区分核酸条形码的末端的非天然核酸的部分不包含可区分核酸条形码序列的任何核苷酸。在一些实施方式中,邻近可区分核酸条形码的末端的非天然核酸的部分包含可区分核酸条形码序列的1、2或3个连续核苷酸,其中1、2或3个连续核苷酸位于可区分核酸条形码序列的末端。邻近可区分核酸条形码的末端的非天然核酸的部分通常包含5至75个核苷酸、5至50个核苷酸、5至45个核苷酸、5至40个核苷酸、5至35个核苷酸、5至30个核苷酸、5至25个核苷酸、5至20个核苷酸或5至15个核苷酸,其位于可区分核酸条形码的5'或3'至一端。在一些实施方式中,邻近可区分核酸条形码的末端的非天然核酸的部分位于与可区分核酸条形码的末端相距0、1、2、3、4、5、6、7、8、9或10个核苷酸。在一些实施方式中,邻近可区分核酸条形码的末端的非天然核酸的部分与可区分核酸条形码序列的末端重叠1、2或3个核苷酸。

[0084] 结合对

[0085] 在一些实施方式中,本文所述的组合物或方法包含一个或多个结合对。在某些实施方式中,核酸包含结合对的一个或多个成员。在一些实施方式中,结合对包含彼此非共价结合的至少两个成员(例如,分子)。结合对的成员通常彼此特异性结合。结合对的成员通常彼此可逆地结合,例如其中结合对的两个成员的联系可以通过适合的方法解联。任何适合的结合对或其成员可用于本文所述的组合物或方法。结合对的非限制性实例包括互补核酸、抗体/抗原、抗体/抗体、抗体/抗体片段、抗体/抗体受体、抗体/蛋白A或蛋白G、半抗原/抗半抗原、巯基/马来酰亚胺、巯基/卤代乙酰基衍生物、胺/异三氰酸酯(isotriocyanate)、胺/琥珀酰亚胺酯、胺/磺酰卤、生物素/抗生物素蛋白、生物素/链霉亲和素、叶酸/叶酸结合蛋白、受体/配体、维生素B12/内在因子、其类似物、其衍生物、其结合部分等或其组合。在一

些实施方式中,结合对包含金属或磁性材料和磁体。结合对的成员的非限制性实例包括抗体、抗体片段、还原型抗体、化学改性抗体、抗体受体、抗原、半抗原、抗半抗原、肽、蛋白质、核酸(例如双链DNA(dsDNA)、单链DNA(ssDNA)或RNA)、核苷酸、核苷酸类似物或衍生物(例如溴脱氧尿苷(BrdU)、烷基部分(例如甲基化DNA或甲基化组蛋白上的甲基部分)、烷酰基(alkanoyl)部分(例如乙酰化蛋白质(例如乙酰化组蛋白)的乙酰基)、烷酸或烷酸酯部分(例如脂肪酸)、甘油部分(例如脂质)、磷酰基部分、糖基部分、泛素部分、凝集素、适体、受体、配体、金属离子、抗生物素蛋白、中性抗生物素蛋白、生物素、B12、内在因子、其类似物、其衍生物、其结合部分等或其组合。例如,在某些实施方式中,结合对的一个成员包含生物素或其类似物或衍生物,并且该对的另一个成员包含抗生物素蛋白或其类似物或衍生物。在另一个实例中,在某些实施方式中,结合对的一个成员包含适合的金属(例如包含金属、金属纳米颗粒、铁的基质)并且另一成员包含磁体。

[0086] 接头

[0087] 在一些实施方式中,可区分标识符和/或结合对的成员通过接头与核酸间接地相关或结合。在某些实施方式中,可区分标识符通过接头与结合对的成员间接地相关或结合。接头可以提供用于将可区分标识符和/或结合对的成员共价地连接到核酸或彼此的机制。任何适合的接头可用于本文所述的组合物或方法。适合的接头的非限制性实例包括:硅烷、硫醇、膦酸、聚乙二醇(PEG)。使用接头连接两个或更多个分子的方法是本领域公知的并且有时称为“交联”。交联的非限制性实例包括与N-羟基琥珀酰亚胺(NHS)酯、亚胺酯、五氟苯基(PFP)酯、羟甲基膦、环氧乙烷或任何其他羰基化合物反应的胺;与碳二亚胺反应的羧基;与马来酰亚胺、卤代乙酰、吡啶基二硫化物和/或乙烯基砜反应的巯基;与肼反应的醛;与双吡丙啶(diazirine)和/或芳基叠氮化物反应的任何非选择性基团;与异氰酸酯反应的羟基;与羰基化合物反应的羟胺等及其组合。

[0088] 核酸测序

[0089] 在某些实施方式中,核酸(例如扩增子;文库的核酸;分离的、纯化的和/或捕获的核酸)通过包括核酸测序的方法分析。在一些实施方式中,可以测序核酸。在一些实施方式中,获得完整或基本上完整的序列,并且有时获得部分序列。

[0090] 可以使用测序核酸的任何适合的方法,其非限制性实例包括Maxim&Gilbert、链终止法、合成测序、连接测序、质谱测序、基于显微镜的技术等或其组合。在一些实施方式中,第一代技术,例如包括自动化Sanger测序方法、包括微流体Sanger测序的Sanger测序方法,可用于本文提供的方法。在一些实施方式中,可以使用包括使用核酸成像技术(例如透射电子显微镜(TEM)和原子力显微镜(AFM))的测序技术。在一些实施方式中,使用高通量测序方法。高通量测序方法通常涉及以大规模平行方式测序的克隆扩增的DNA模板或单一DNA分子,有时是在流动池内。能够以大规模并行方式测序DNA的下一代(例如第二代和第三代)测序技术可用于本文所述的方法,并且在本文中统称为“大规模并行测序”(MPS)或“大规模并行核酸测序”。在一些实施方式中,MPS测序方法利用靶向方法,其中序列读数从感兴趣的特定染色体、基因或区域产生。感兴趣的特定染色体、基因或区域在本文中有时被称为靶向基因组区域。在某些实施方式中,使用非靶向方法,其中样品中的大部分或全部核酸片段被随机地测序、扩增和/或捕获。

[0091] MPS测序有时利用通过合成和某些成像过程进行测序。可用于本文所述方法的核

酸测序技术是合成测序和基于可逆终止子的测序(例如Illumina的Genome Analyzer; Genome Analyzer II; HISEQ 2000; HISEQ 2500 (Illumina, San Diego CA))。利用该技术, 数百万个核酸(例如DNA)片段可以并行测序。在这种类型的测序技术的一个实例中, 使用流动池, 其容纳光学透明的载玻片, 其表面上的8至16个单独泳道是结合的寡核苷酸锚(例如衔接子引物)。

[0092] 在一些实施方式中, 合成测序包括以模板指导的方式将核苷酸迭代加入(例如通过共价加入)至引物或预先存在的核酸链。检测核苷酸的每个迭代加入, 并重复该过程多次直到获得核酸链的序列。所获序列长度部分取决于所执行的加入和检测步骤的数量。在合成测序的一些实施方式中, 相同类型(例如A、G、C或T)的一个、两个、三个或更多个核苷酸在一轮核苷酸加入和检测。核苷酸可以通过任何适合的方法(例如酶促地或化学地)加入。例如, 在一些实施方式中, 聚合酶或连接酶以模板指导的方式将核苷酸加入至引物或预先存在的核酸链。在合成测序的一些实施方式中, 使用不同类型的核苷酸、核苷酸类似物和/或标识符。在一些实施方式中, 使用可逆终止子和/或可移除的(例如可切割的)标识符。在一些实施方式中, 使用荧光标记的核苷酸和/或核苷酸类似物。在某些实施方式中, 合成测序包括切割(例如标识符的切割和移除)和/或洗涤步骤。在一些实施方式中, 通过本文所述或本领域已知的适合方法检测一个或多个核苷酸的加入, 其非限制性实例包括任何适合的成像装置或机器、适合的相机、数码相机、基于CCD(电荷耦合器件)的成像装置(例如CDD照相机)、基于CMOS(互补金属氧化物硅)的成像装置(例如CMOS照相机)、光电二极管(例如光电倍增管)、电子显微镜、场效应晶体管(例如DNA场效应晶体管)、ISFET离子传感器(例如CHEMFET传感器)等或其组合。可用于进行本文方法的其他测序方法包括数字PCR和杂交测序。

[0093] 用于进行本文所述的方法的适合的MPS方法、系统或技术平台可用于获得核酸测序读数。MPS平台的非限制性实例包括Illumina/Solex/HiSeq(例如Illumina的Genome Analyzer; Genome Analyzer II; HISEQ 2000; HISEQ)、SOLiD、Roche/454、PACBIO和/或SMRT、Helicos True Single Molecule Sequencing、基于Ion Torrent和Ion semiconductor的测序(例如由Life Technologies开发)、WildFire、基于5500、5500x1 W和/或5500x1 W Genetic Analyzer的技术(例如由Life Technologies开发并销售, 美国专利公开no.US20130012399)、Polony测序、焦磷酸测序、大规模并行信号测序(MPSS)、RNA聚合酶(RNAP)测序、LaserGen系统和方法、基于纳米孔的平台、化学敏感场效应晶体管(CHEMFET)阵列、基于电子显微镜的测序(例如由ZS Genetics, Halcyon Molecular开发)、纳米球测序等。

[0094] 可用于进行本文方法的其他测序方法包括数字PCR和杂交测序。数字聚合酶链式反应(数字PCR或dPCR)可用于直接鉴定和定量样品中的核酸。在一些实施方式中, 可以在乳液中进行数字PCR。例如, 个体核酸被分离, 例如在微流体室装置中, 并且每种核酸通过PCR单独扩增。可以分离核酸使得每个孔有不超过一种核酸。在一些实施方式中, 可以使用不同的探针以区分各种等位基因(例如胎儿等位基因和母体等位基因)。可以列举等位基因以测定拷贝数。

[0095] 在某些实施方式中, 可以使用杂交测序。该方法包括使多个多核苷酸序列与多个多核苷酸探针接触, 其中多个多核苷酸探针中的每一个可以任选地束缚到基质。在一些实

施方式中,基质可以是具有已知核苷酸序列阵列的平坦表面。与阵列杂交的模式可用于测定样品中存在的多核苷酸序列。在一些实施方式中,每个探针被束缚到珠,例如磁珠等。可以鉴定与珠的杂交并用于鉴定样品中的多个多核苷酸序列。

[0096] 在一些实施方式中,进行染色体特异性测序。在一些实施方式中,利用DANSR(选定区域的数字分析)进行染色体特异性测序。选定区域的数字分析使得能够通过两个基因座特异性寡核苷酸经由中间“桥”寡核苷酸进行DNA依赖性连锁以形成PCR模板而同时定量数百个基因座。在一些实施方式中,通过产生富集染色体特异性序列的文库进行染色体特异性测序。在一些实施方式中,仅针对所选择的染色体组获得序列读数。

[0097] 竞争者核酸

[0098] 竞争者核酸通常在杂交过程之前加入以减少不想要的和非特异性的杂交事件。竞争者核酸可以包含重复核酸。重复的内源性核酸,例如Alu序列或LINE序列,通常存在于核酸文库中。有时内源性重复核酸可以彼此杂交,导致捕获的杂交混合物的污染。这种类型的污染可以部分地通过在杂交之前加入过量的外源竞争者核酸而减少。任何适合的竞争者核酸可以用于本文所述的组合物或方法。在一些实施方式中,竞争者核酸包含C0t-1DNA,其可以结合存在于核酸文库中的Alu、LINE和其他重复序列。C0t-1DNA可以从适合的生物体获得并且可以包含来自不同生物体的核酸的混合物。C0t-1DNA可以从生物体的适合组织获得。C0t-1DNA有时包含从胎盘分离的核酸。

[0099] 阻断核酸

[0100] 在一些方面,本文提供了改进的阻断方法和组合物。通常,通过杂交事件,不想要的核酸在杂交捕获方法结束后污染富集的核酸池。不想要的序列中的大部分有时是由于衔接子连接的文库插入物的终端衔接子序列(例如条形码、与流动池锚互补的部分和或引物等)的相同部分之间不期望的杂交事件。有时,不想要的文库插入物可以通过它们的终端衔接子退火到彼此,从而导致连接和分离在一起的在其他方面不想要的DNA片段的“菊花链”。以这种方式,单个期望片段的捕获可带来大量不期望的片段,其降低了富集方法的总体效率。

[0101] 在一些实施方式中,可以通过使用被引导与文库的非天然核酸的部分(例如衔接子序列)杂交的阻断核酸减少所谓的“菊花链”效应和其他不想要的杂交事件。阻断寡核苷酸是本领域已知的,并且经常配置为结合到并阻断文库的条形码序列之间的杂交。传统的阻断寡核苷酸通常是长度为50个或更多个核酸,并且配置为与条形码序列以及与位于核酸条形码的每侧的侧面的合成核酸部分杂交。因此传统的阻断寡核苷酸相对较长(例如>50个核苷酸),因此它们可以退火到条形码区域和侧面的非天然序列,并确保阻断寡核苷酸与它们的靶序列之间的高熔融温度。在一些实施方式中,本文的方法使用一种或多种阻断核酸(例如传统的阻断核酸)。

[0102] 在一些实施方式中,对于其中混合多个衔接子连接的文库的多重测序方法,必须合成多组阻断核酸,每组对每个文库的许多不同衔接子具有特异性。因此,高通量多重测序反应通常涉及8个、10个、15个或20个或更多个不同条形码序列,其需要合成8个、10个、15个或20个或更多个阻断寡核苷酸,每个配置为结合独特条形码序列中的每一个。该策略是昂贵且耗时的,因为必须设计并合成长度相对较长的许多不同阻断寡核苷酸用于混合的文库的多重测序。

[0103] 在一些实施方式中,阻断核酸是U阻断(U-block)核酸。在一些实施方式中,本文的组合物或方法包含U阻断(例如通用阻断)核酸。组合物的U阻断核酸可以是基本上相同或基本上不同的。U阻断核酸相对于传统阻断寡聚物具有多个优点。第一,U阻断核酸基本上不与核酸条形码序列杂交,U阻断核酸也不配置为与核酸条形码序列杂交。因此,含有8个、10个、15个或20个或更多个不同条形码序列的复杂核酸文库的操作、捕获和多重测序不需要对每个独特条形码序列具有特异性的U阻断核酸。同样,与传统的阻断寡核苷酸相比,U阻断核酸相对较短,使得U阻断核酸的核酸合成更经济。在某些实施方式中,本文提供的U阻断核酸具有45个核苷酸或更少、40个核苷酸或更少、35个核苷酸或更少、30个核苷酸或更少、25个核苷酸或更少、20个核苷酸或更少、15个核苷酸或更少、或10个核苷酸或更少的名义、平均、平均或绝对长度。在一些实施方式中,U阻断核酸具有8至约40个、8至约35个、8至约30个、8至约25个或8至约20个核苷酸的名义、平均、平均或绝对长度。U阻断核酸可以包含任何适合的核酸、核苷酸或核苷酸类似物。在一些实施方式中,U阻断核酸是合成的(例如通过人手合成)。U阻断核酸可以是寡核苷酸。

[0104] U阻断核酸通常配置为阻断不想要的杂交和/或核酸文库的非天然部分的后续扩增。在某些实施方式中,U阻断核酸配置为与在条形码核酸侧面的合成核酸区域(非天然核酸区域)杂交。因此,本文的组合物通常包含至少两个U阻断核酸,其配置为与可区分核酸条形码的相对侧杂交。在条形码核酸序列侧面的合成核酸区域通常相对较短(例如45个核苷酸或更少、40个核苷酸或更少、35个核苷酸或更少、30个核苷酸或更少、25个核苷酸或更少、20个核苷酸或更少、15个核苷酸或更少、或10个核苷酸或更少),并且因此提供供U阻断核酸杂交的相对短的核酸段。本文中由本发明人开发的U阻断核酸的早期原型排他地由标准核苷酸碱基组成,并且在阻断不想要的杂交事件时无效率。通过改变U阻断核酸的核苷酸碱基中的一些或全部,可以大量提高效率和阻断能力。因此,在一些实施方式中,部分地通过包含增加U阻断核酸的 T_m 的非标准或改变的核苷酸碱基,U阻断核酸配置为包含较高的熔融温度。在一些实施方式中,U阻断核酸包含 T_m ,其高于由选自鸟嘌呤、胞嘧啶、胸腺嘧啶、腺嘌呤和尿嘧啶的标准核苷酸组成的相似序列的未改性的U阻断核酸的 T_m 。可以使用任何适合的改变以增加U阻断核酸的 T_m 。在一些实施方式中,U阻断核酸包含改变的核苷酸、核苷酸类似物和/或改变的核苷酸键,其非限制性实例包括锁核酸(LNA、例如双环核酸)、桥核酸(BNA、例如约束核酸)、C5-改性的嘧啶碱基(例如5-甲基-dC、丙炔基嘧啶等)和交替的骨架化学物质,例如肽核酸(PNA)、吗啉基化合物等或其组合。在一些实施方式中,桥核酸是改变的RNA核苷酸。任何适合的BNA可用于本文所述的组合物或方法中。在某些实施方式中,BNA单体可以包含五元、六元或甚至七元桥连结构。新一代BNA单体的非限制性实例包括2',4'-BNANC[NH]、2',4'-BNANC[NMe]和2',4'-BNANC[NBn]。非碱基改性剂也可以引入U阻断核酸中以增加 T_m (或结合亲和力),其非限制性实例包括小沟结合剂(MGB)、精胺、G-夹、Uaq葱醌帽等或其组合。可以在U阻断核酸中使用多于一种类型的 T_m 增强改变,例如BNA核苷酸单体与末端MGB基团的组合。增加互补核酸的 T_m 的许多方法是本领域技术人员已知的,并且所有这样的改变的使用都被认为在本发明的范围内。在一些实施方式中,U阻断核酸包含至少40℃、至少45℃、至少50℃、至少55℃、至少60℃、至少65℃、至少70℃、至少75℃或至少80℃的熔融温度(T_m)。

[0105] 在某些实施方式中,U阻断核酸配置为与文库的一个或多个非天然核酸特异性杂

交。非天然核酸通常包含合成核酸序列。在一些实施方式中,非天然核酸不包含基因组DNA、基因、mRNA、cDNA或其部分。在一些实施方式中,U阻断核酸配置为与文库的一个或多个扩增子特异性杂交,其中一个或多个扩增子包含合成核酸(例如一个或多个衔接子序列、捕获序列或引物结合位点)。在一些实施方式中,U阻断核酸配置为与一个或多个衔接子或其部分特异性杂交。在某些实施方式中,U阻断核酸不配置为与文库插入物杂交。在某些实施方式中,U阻断核酸基本上不与核酸插入物杂交。在某些实施方式中,U阻断核酸不配置为与核酸条形码杂交并且不与条形码序列的大部分互补。在某些实施方式中,U阻断核酸基本上不与核酸条形码或核酸条形码的任何部分杂交。

[0106] U阻断核酸通常包含与非天然核酸基本上互补的核酸序列或由其组成。U阻断核酸有时配置为与核酸文库的一个或多个非天然核酸特异性杂交。在一些实施方式中,文库的每个核酸包含一个或多个(例如1、2、3、4或更多个)非天然核酸,其中非天然核酸对于文库的每个核酸是共同的(例如共有的)。例如,可以从两个或更多个样品产生核酸文库,其中衍生自每个样品的核酸包含引入到衔接子序列中的独特可区分条形码序列,并且其中衔接子包含一个或多个(例如1、2、3、4或更多个)相同的非天然核酸(例如合成核酸、通用核酸序列)。在文库的核酸之间基本上相同和共有的非天然核酸有时称为通用核酸。U阻断核酸通常与通用核酸序列(例如衔接子或其部分)基本上互补并配置为与其特异性杂交。

[0107] 在一些实施方式中,U阻断核酸配置为阻断通过聚合酶进行延伸。在一些实施方式中,U阻断核酸配置为阻断通过聚合酶延伸U阻断核酸。例如,U阻断核酸可以包含防止聚合酶延伸U阻断核酸的3'端(例如通过形成磷酸二酯键)的适合的3'链终止子(例如2',3'双脱氧核苷酸)或适合的官能团。因此,配置为阻断通过聚合酶进行延伸的U阻断核酸通常包含防止聚合酶延伸U阻断核酸的3'端的适合的3'链终止子(例如2',3'双脱氧核苷酸)或适合的官能团。因此,在某些实施方式中,U阻断核酸不是适合于通过聚合酶扩增(例如PCR)或延伸的核酸引物。

[0108] 在一些实施方法中,本文的组合物包含衍生自多个样品的一个或多个核酸文库(例如多个文库插入物)和2个或更多个、3个或更多个、4个或更多个、5个或更多个、6个或更多个、7个或更多个、8个或更多个、9个或更多个或10个或更多个独特可区分条形码。在某些实施方式中,这样的组合物包含不超过4个,有时不超过8个U阻断核酸,其各自配置为与文库的每个核酸的非天然核酸或其部分特异性杂交。在一些实施方式中,本文的组合物包含1、2、3、4、5、6、7或8个U阻断核酸和/或不超过1、2、3、4、5、6、7或8个U阻断核酸。在某些实施方式中,本文的组合物包含2至4个、2至6个、2至8个或4至8个U阻断核酸。在一些实施方式中,本文的组合物包含1、2、3、4、5、6、7或8个U阻断核酸,其中(i)U阻断核酸各自与第一和/或第二非天然核酸的部分基本上互补,(ii)U阻断核酸中的至少一个配置为邻近可区分核酸条形码的第一端杂交,(iii)U阻断核酸中的至少一个配置为邻近可区分核酸条形码的第二端杂交,和(iv)U阻断核酸各自基本上不与可区分核酸条形码的任何部分杂交。

[0109] 在一些实施方式中,U阻断核酸配置为邻近可区分核酸条形码的末端杂交。在一些实施方式中,配置为邻近可区分核酸条形码的末端杂交的U阻断核酸是指与非天然核酸的部分基本上互补的U阻断核酸,所述非天然核酸位于邻近可区分核酸条形码。U阻断核酸通常配置为与包含可区分核酸条形码的文库的非天然核酸杂交,其中U阻断核酸配置为与条形码序列的相对侧杂交。因此,当U阻断核酸与包含可区分核酸条形码的文库的非天然核酸

杂交时,杂交的U阻断核酸是在条形码两侧上的可区分核酸条形码的侧面。在某些实施方式中,U阻断核酸基本上不与条形码序列的任何部分杂交。在一些实施方式中,杂交的U阻断核酸可以与条形码序列重叠1、2、3、4、5或6个核苷酸。因此,在某些实施方式中,基本上不与条形码序列杂交的U阻断核酸可以与条形码序列的小部分(例如6个核苷酸或更少)杂交。

[0110] 在一些实施方式中,本文的组合物包含至多四个U阻断核酸。组合物的U阻断核酸可以是相同的。例如,在组合物包含四个U阻断核酸的情况下,U阻断核酸各自的核酸序列可以基本上同一(例如基本上相同)。在一些实施方式中,组合物包含四个U阻断核酸,其中U阻断核酸各自的核酸序列基本上不同。

[0111] 捕获核酸

[0112] 在一些实施方式中,本文的组合物或方法包含捕获核酸。捕获核酸通常包含核酸部分。在一些实施方式中,捕获核酸配置为与靶核酸特异性杂交,其中捕获核酸和其杂交的靶可以通过适合的技术(例如拉下(pull-down)方法)捕获。任何适合的捕获核酸或捕获核酸组可用于本文的方法或组合物。在一些实施方式中,捕获核酸是寡核苷酸。

[0113] 捕获核酸通常直接地或间接地结合(例如共价地或非共价地结合)至适合的结合对的成员。在某些实施方式中,捕获核酸包含适合的结合对的成员。在一些实施方式中,结合对的成员通过接头结合到捕获核酸。

[0114] 包含结合对成员的捕获核酸可以与其杂交的核酸靶一起通过使用适合的捕获方法捕获(例如通过捕获方法捕获)。捕获核酸的过程(例如通过捕获方法)通常提供捕获的核酸(例如捕获的核酸)。本文使用的术语“捕获的”和“富集的”可以指核酸或核酸亚组,条件是含有比捕获方法之前的样品中更少的核酸种类。包含捕获的核酸的组合物可以不含约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或大于99%的其它(例如不想要的)核酸种类。在一些实施方式中,捕获的核酸包含富集的核酸。富集的核酸可以包含与应用富集或捕获方法之前的一种或多种核酸种类的量相比富集至少1.2倍、1.3倍、1.4倍、1.5倍、2倍、5倍、10倍、100倍或1000倍的一定量的一种或多种核酸种类。捕获方法的非限制性实例包括拉下方法(例如重力拉下方法、离心拉下方法、磁性拉下方法)、免疫沉淀和各种柱纯化方法。捕获的核酸的分离和/或纯化通常涉及结合对的第一成员(例如结合到捕获核酸)与结合对的第二成员(例如结合到基质)的关联。捕获核酸有时与结合对的第一成员结合,其配置为与第二成员牢固地结合和/或关联,其中第二成员有时与适合的基质结合。在一些实施方式中,捕获核酸与包含基质的结合对的第一成员结合。在一些实施方式中,捕获核酸与基质间接地相关(例如通过接头或中间结合分子,例如抗体)。在一些实施方式中,捕获核酸结合到包含磁性基质(例如磁珠、含铁珠)的结合对的第一成员,并且结合对的第二成员是磁体。磁体和磁性材料可用于捕获(例如拉下)捕获核酸及其结合的靶。

[0115] 在一些实施方式中,核酸和/或扩增子(例如文库的核酸和/或扩增子)与捕获核酸接触。在某些实施方式中,方法包括将包含文库的核酸、阻断核酸和任选地一种或多种竞争者核酸的混合物与捕获核酸接触。捕获核酸和制造捕获核酸的方法是本领域已知的。在某些实施方式中,捕获核酸用于从包含从2个或更多个、10个或更多个、50个或更多个或100个或更多个受试者、样品或来源获得的多个文库插入物的核酸的文库中捕获感兴趣的核酸亚组。

[0116] 捕获核酸可以是配置为与文库的一个或多个核酸(例如选择的核酸、靶核酸、感兴趣的核酸)的部分特异性杂交的核酸。在一些实施方式中,配置为与文库的一个或多个核酸的部分特异性杂交的捕获核酸与文库的核酸的适合部分基本上互补。在一些实施方式中,捕获核酸配置为与非天然核酸的部分或其部分特异性杂交。在一些实施方式中,捕获核酸配置为与衔接子的部分特异性杂交。在一些实施方式中,捕获核酸配置为与一个或多个文库插入物的部分特异性杂交。捕获核酸可以配置为与启动子、增强子、内含子、外显子、聚A区段、聚T区段、任何适合的翻译或转录控制序列等或其组合杂交。捕获核酸组可以配置为与基因亚组(例如染色体的基因的组、例如表达酶家族的基因的组)或文库的核酸的任何亚组特异性杂交。在一些实施方式中,捕获核酸组配置为在怀疑包含遗传变异(例如缺失、插入、SNP等)的一个或多个基因组区域处或附近杂交。配置为与第二核酸特异性杂交的核酸通常与第二核酸基本上互补。捕获核酸通常与一个或多个核酸靶的部分(例如文库的核酸的亚组)基本上互补。

[0117] 在一些实施方式中,方法包括制备包含核酸的文库、阻断核酸、捕获核酸和任选地竞争者核酸或基本上由其组成的混合物,其中使混合物经受捕获方法。例如,在一些实施方式中,加入链霉抗生物素蛋白珠,并回收与捕获核酸关联的(例如与捕获核酸杂交的)混合物的核酸(例如通过离心、过滤、免疫沉淀和/或通过磁性沉淀)。在一些实施方式中,使用洗涤步骤(例如通过用乙醇洗涤)。捕获的核酸(例如杂交的靶)可以使用适合的方法从捕获核酸洗脱。

[0118] 在一些实施方式中,在不存在结合对(例如结合对的两个成员)的情况下进行某些方法或方法的步骤。例如,混合物通常包含捕获核酸和结合对的第一成员(例如结合至捕获核酸),并且混合物不包含结合对的第二成员(例如,第二成员配置为与第一成员特异性结合)。在一些实施方式中,在不存在结合对的第二成员的情况下,文库的核酸与阻断核酸、包含结合对的第一成员的捕获核酸和/或竞争者核酸接触。在一些实施方式中,在不存在结合对的第二成员的情况下,纯化包含文库的核酸、阻断核酸、包含结合对的第一成员的捕获核酸和/或竞争者核酸的混合物。在一些实施方式中,在不存在结合对的第二成员的情况下,包含文库的核酸、阻断核酸、包含结合对的第一成员的捕获核酸和/或竞争者核酸的混合物在适合的杂交条件下杂交。

[0119] 在一些实施方式中,捕获核酸直接地或间接地结合(例如共价地或非共价地结合)至适合的基质。在某些实施方式中,结合对的成员直接地或间接地结合(例如共价地或非共价地结合)至适合的基质。可以使用任何适合的基质。在某些实施方式中,基质包含表面(例如,流动池的表面、管的表面、芯片的表面),例如金属表面(例如钢、金、银、铝、硅和铜)。在一些实施方式中,基质(例如基质表面)被涂覆和/或包含官能团和/或惰性材料。在某些实施方式中,基质包含例如珠、芯片、毛细管、板、膜、晶片(例如硅晶片)、梳子或针。在一些实施方式中,基质包含珠和/或纳米颗粒。基质可以由适合的材料制成,其非限制性实例包括塑料或适合的聚合物(例如,聚碳酸酯、聚(乙烯醇)、聚(二乙烯基苯)、聚苯乙烯、聚酰胺、聚酯、聚偏二氟乙烯(PVDF)、聚乙烯、聚氨酯、聚丙烯等)、硼硅酸盐、玻璃、尼龙、Wang树脂、Merrifield树脂、金属(例如,铁、金属合金)、琼脂糖凝胶(sepharose)、琼脂糖、聚丙烯酰胺、葡聚糖、纤维素等或其组合。在一些实施方式中,基质包含磁性材料(例如,铁、镍、钴、铂、铝等)。在某些实施方式中,基质包含磁珠(例如,DYNABEADS[®]、赤铁矿、AMPure

XP。磁体可用于纯化和/或捕获与某些基质结合的核酸(例如,包含金属或磁性材料的基质)。

[0120] 在某些实施方式中,通过适合的捕获方法捕获捕获核酸和与捕获核酸特异性杂交的核酸(例如靶核酸),从而提供捕获的核酸。包括捕获的过程通常提供富集的核酸(例如富集靶核酸的核酸)。捕获的核酸通常富集核酸的一种或多种种类。捕获的核酸可以包含一种或多种捕获核酸、与捕获核酸特异性杂交的核酸(例如,文库的捕获的核酸、靶核酸、富集的核酸)、结合对的成员和/或基质。可以通过适合的方法回收与捕获核酸特异性杂交的核酸(靶核酸)。例如,在一些实施方式中,通过捕获方法富集的核酸可以通过变性(例如通过加热)或通过增加或减少包含捕获的核酸的混合物的盐浓度而分离。在一些实施方式中,捕获方法包括使用过滤器、膜或柱回收富集的核酸。在某些实施方式中,重力和/或离心用于捕获核酸和/或回收富集的核酸。在某些实施方式中,捕获方法不使用离心。在某些实施方式中,磁体用于捕获和/或回收富集的核酸。通常可以通过适合的方法洗涤与基质结合的核酸,以去除未结合的或不想要的材料。在某些实施方式中,可以通过适合的方法调节洗涤溶液的严格性。

[0121] 捕获的核酸、与捕获核酸特异性杂交的核酸、富集的核酸和/或其扩增子可以通过适合的方法进行分析,所述方法可包括例如包括核酸测序或质谱的过程。

[0122] 固定

[0123] 在某些实施方式中,核酸被固定(例如固定在基质上)。核酸可以通过任何适合的方法固定。核酸可以直接地或间接地固定到适合的基质或材料上。固定到基质的核酸可以共价地或非共价地与基质结合。在某些实施方式中,核酸可逆地固定到基质并且可以使用适合的方法从基质解联或移除。在一些实施方式中,包含结合对的第一成员的核酸通过结合到与基质结合或关联的结合对的第二成员而固定到基质上。核酸可以非特异性地固定到基质,例如其中基质包含阴离子(例如阴离子交换基团,例如带正电荷的官能团)。在某些实施方式中,核酸通过磁性吸引力固定到基质。例如,核酸可以包含磁性材料(例如金属),并且核酸可以通过磁体的使用而固定到基质(例如,管的部分、表面)。在某些实施方式中,磁体不在溶液中和/或不与包含磁性材料的核酸直接接触。在某些实施方式中,磁体处于溶液中并和与核酸关联的磁性材料直接接触。在一些实施方式中,通过共价键的形成将核酸固定,例如通过将核酸的官能团(例如核酸类似物的官能团)交联到包含反应性基团的基质。在一些实施方式中,核酸(例如扩增子、文库的核酸、靶核酸)通过与另一个结合到基质(例如共价地或非共价地)的核酸杂交(例如,特异性杂交、退火)而固定到基质上。

[0124] 核酸可以在本文所述方法的任何适合步骤下固定。在一些实施方式中,核酸被固定到流动池(例如,流动池表面)或阵列(例如,芯片)。在某些实施方式中,本文所述的方法不包括将核酸固定到流动池或芯片。在某些实施方式中,本文所述的方法不包括将核酸固定到流动池或芯片,直至在捕获方法之后。在某些实施方式中,本文所述的方法不包括在分析核酸(例如通过核酸测序)之前将核酸固定到流动池或芯片。在一些实施方式中,将文库的核酸(例如文库的扩增子)连接到衔接子、与阻断核酸接触、与捕获核酸接触、与竞争者核酸接触,纯化和/或杂交(例如,进行变性和退火过程),并且在任何或所有上述过程之前或期间都不固定到流动池、阵列或芯片。在某些实施方式中,在不存在流动池、阵列或芯片的情况下,将文库的核酸(例如文库的扩增子)连接到衔接子,与阻断核酸接触,与捕获核酸接

触,与竞争者核酸接触,纯化和/或杂交(例如,进行变性和退火过程)。

[0125] 其他核酸方法

[0126] 在一些实施方式中,本文的方法包括制备用于后续杂交和/或捕获的混合物。在某些实施方式中,制备混合物包括将一个或多个核酸文库的核酸(例如扩增子)与一个或多个阻断核酸、一个或多个捕获核酸和/或一个或多个竞争者核酸接触。在一些实施方式中,在变性或杂交之前制备混合物。通过本文所述的方法制备的混合物可以包含一个或多个核酸文库(例如扩增子)、一个或多个阻断核酸、一个或多个捕获核酸和/或一个或多个竞争者核酸。通过本文所述的方法制备的混合物可以包含核酸的文库、阻断核酸和捕获核酸或基本上由其组成。在某些实施方式中,制备的混合物包含核酸的文库、阻断核酸、捕获核酸和竞争者核酸或基本上由其组成。基本上由核酸组成的制备的混合物可以含有水、EDTA、PEG、NaCl和/或缓冲液(例如,Tris或HEPES)。在一些实施方式中,基本上由核酸组成的混合物不含有杂交缓冲液。在某些实施方式中,制备核酸的混合物的方法不包括加入杂交缓冲液。在某些实施方式中,基本上由核酸组成的制备的混合物不含有钙或镁盐。在一些实施方式中,制备的核酸的混合物不含有钙或镁盐、洗涤剂(例如SDS)、Ficoll、BSA或聚乙烯吡咯烷酮。在一些实施方式中,基本上由核酸组成的混合物不含有洗涤剂(例如SDS)、Ficoll、BSA或聚乙烯吡咯烷酮。

[0127] 在一些实施方式中,如本文所述制备混合物,并且混合物不被加热直至混合物的核酸被纯化之后。例如,在从混合物中纯化核酸之前和/或直至混合物的核酸被纯化之后,混合物不被加热至大于约70℃、75℃、80℃、85℃、90℃、或高于95℃的温度。

[0128] 遗传变异和医疗状况

[0129] 可以使用本文所述的组合物或方法确定遗传变异的存在或不存在。遗传变异通常是存在于某些个体中的特定遗传表型,并且通常遗传变异存在于个体的统计学显著的亚群中。在一些实施方式中,遗传变异是染色体异常(例如,非整倍体、一个或多个染色体的复制、一个或多个染色体的丢失)、部分染色体异常或镶嵌(例如,染色体的一个或多个区段的丢失或获得)、易位或反转。遗传变异的非限制性实例包括一个或多个缺失、复制、插入、突变、多态性(例如,单核苷酸多态性)、融合、重复(例如短串联重复)等及其组合。插入、重复、缺失、复制、突变或多态性可以具有任何长度的,并且在一些实施方式中,长度为约1碱基或碱基对(bp)至约250兆碱基(Mb)。在一些实施方式中,插入、重复、缺失、复制、突变或多态性长度为约1碱基或碱基对(bp)至约50,000千碱基(kb)(例如,长度为约10bp、50bp、100bp、500bp、1kb、5kb、10kb、50kb、100kb、500kb、1000kb、5000kb或10,000kb)。

[0130] 在某些实施方式中,针对受试者确定其存在或不存在的遗传变异有时与医疗状况相关。医疗状况的非限制性实例包括与智力障碍相关的那些(例如唐氏综合症)、异常细胞增殖(如癌症)、非霍奇金淋巴瘤、骨髓增生异常综合征、威廉综合征、Langer-Giedion综合征、阿尔菲综合征、Rethore综合征、雅各布综合征、视网膜母细胞瘤、Smith-Magenis、爱德华综合征、乳头状肾细胞癌、DiGeorge综合征、Angelman综合征、猫眼综合征、家族性腺瘤性息肉病、米勒-迪克综合征、微生物核酸的存在(例如病毒、细菌、真菌、酵母)和子痫前期。

[0131] 实施例

[0132] 下述实施例说明某些实施方式,且不限本技术。

[0133] 实施例1:对于效率、成本降低和数据质量改善,工作流程在杂交捕获工作流程方

面取得进展

[0134] 本文示例的新型和改进的方法提供了以下优点。

[0135] 1. 通过磁珠使用零体积浓缩, 避免反应混合物基于真空和热的浓缩。

[0136] 2. 用于增加的反应动力学的DNA诱饵浓度

[0137] 3. 用于工作流程改进的短核酸浓度

[0138] 4. 结束的杂交反应的变性

[0139] 5. 高效和可自动化的链霉抗生物素蛋白珠操作

[0140] 用于杂交捕获试验的传统方案(例如, 2014年8月20日在http://www.nimblegen.com/products/lit/06588786001_SeqCapEZLibrary_SR_UGuide_v4p2.pdf访问的网站pdf文件: Roche-NimbleGen SeqCap EZ Library SR User's Guide)需要与高通量遗传测试方法绝对不兼容的许多操作。本文提供的方案导致效率和数据质量显著提高, 同时降低成本。

[0141] 传统杂交反应的设置涉及组合三个组分(DNA文库、阻断寡核苷酸和人Cot-1DNA), 然后进行通过在真空和热下同时离心打开的管去除所有液体的期望蒸发步骤(本领域的人称为“速度-真空(Speed-vac)”)。蒸发步骤是需要的, 部分因为杂交缓冲液必须以高浓度加入, 并且最初三种组分的额外稀释是不可接受的。蒸发步骤特别缓慢(>1小时), 容易发生交叉污染并且通常不兼容高通量处理。

[0142] 关于新型杂交方法的描述, 参见下表1。

[0143] 表1

	新型方法	优点
[0144]	混合文库、阻断寡核苷酸、DNA诱饵和任选地 Cot-1 竞争者 DNA	DNA 生物素化诱饵被包含在混合物中。
	使用 AMPureXP 珠的浓缩混合物	避免污染; 在不到 1 小时内完成; 可以自动化并由机器人装置执行方法不需要热或真空。
	将混合物重悬于 5 μ l 2x 杂交缓冲液、2 μ l Hyb 组分 A (Roche) 和	提供所有组分的浓缩悬浮液。
[0145]	3 μ l 水中。	
	在 95°C 变性 10 分钟, 立即降到 47°C。	提供一步变性/杂交。提供更快的处理, 因为诱饵不在单独步骤中添加。
	47°C 孵育 16 小时。	使捕获孵育时间减少多达 48 小时。

[0146] 本文所述的新型方法的第一方面引入使用能够非特异性结合核酸的磁珠, 以将核酸材料快速浓缩到磁珠“小球”上, 然后将其以适合的浓度重悬在杂交缓冲液中。这种同样

改进的方法在减少的最终体积中保持相同量的诱饵,其提高了反应动力学,同时保持所有组分的浓度。

[0147] 本文所述的新型方法的第二方面将生物素化DNA捕获诱饵包含到浓缩反应中。

[0148] 新型方法的第三方面涉及改变纯化条件以确保有效纯化短单链DNA诱饵和短单链阻断寡聚物(除了更长的双链文库和Cot-1 DNA之外)。以2份珠:1份反应的改变的比率(制造商建议1.8:1)使用AMPureXP珠进行浓缩,其导致更有效结合短核酸(通常以1.8:1的比率去除)。

[0149] 新型方法的第四方面涉及使反应混合物以其存在所有组分的完整形式变性。传统方案建议DNA、Cot-1和阻断寡聚物在95℃变性,然后开启或打开板以允许另外转移生物素化DNA诱饵。从本文提出的新型方法获得的结果证实,除了在浓缩反应中包含诱饵(上文第二方面)之外,包含诱饵的整个反应混合物可以在95℃一起变性,然后立即降到杂交温度(47℃)。令人惊讶的是,该方法没有导致整体阻断效率损失,并且没有导致捕获杂交效率降低。相反,该方法导致所捕获的文库核酸的收率增加。这提供了令人吃惊的工作流程改进,因为它消除了与没有效率的热循环器的多个相互作用,并且消除了开启患者样品板引起的交叉污染的高风险。

[0150] 新型方法的第五方面进一步改进使用链霉抗生物素蛋白捕获珠和生物素化诱饵捕获文库核酸。传统方案需要从链霉抗生物素蛋白珠去除上清液,留下珠的小粒,其需要使用包含文库核酸的杂交反应(10-15μl)重悬。由于在相对小的体积中重悬的大量链霉抗生物素蛋白珠,这个方法已经证明是有问题的和困难的。为了克服与传统方法相关的问题并使得能够自动化,将链霉抗生物素蛋白珠的小粒重悬于10μl杂交缓冲液中(例如,在剧烈涡旋和移液下),然后加入到10-15μl的杂交反应中。这保持了最终反应的缓冲液组成并使得能够完全自动化。

[0151] 总而言之,本文提出的新型方法在比传统方案所需更少的时间内完成(例如,至少2天更短),需要更少的步骤和更少的样品处理,预料不到地产生更高质量的数据(中靶(on target)的百分比为>90%,相对于使用传统方案的70-80%),允许完全和高效的自动化,并提供更安全的患者DNA样品处理。此外,DNA文库不经受长时间的热,例如来自速度-真空干燥和长时间杂交,其可以导致文库核酸降解。

[0152] 实施例2:实施方式的实施例

[0153] 下述实施例说明某些实施方式,且不限本技术。

[0154] A1.一种用于大规模并行核酸测序的组合物,其包含:

[0155] a) 包含从一个或多个样品获得的多个文库插入物的核酸的文库和至少八个可区分核酸条形码,每个核酸条形码包含第一端和第二端,和文库的每个核酸包含(i)至少一个文库插入物,(ii)第一非天然核酸,(iii)第二非天然核酸和(iv)不超过所述可区分核酸条形码中的两个,其中

[0156] 第一非天然核酸和第二非天然核酸位于至少一个文库插入物的相对侧上,和

[0157] 第一非天然核酸和/或第二非天然核酸包含不超过两个可区分核酸条形码;和

[0158] b) 不超过四个U阻断核酸,其中(i)U阻断核酸各自与第一和/或第二非天然核酸的部分基本上互补,(ii)U阻断核酸中的至少一个配置为邻近可区分核酸条形码各自的第一端杂交,(iii)U阻断核酸中的至少一个配置为邻近可区分核酸条形码各自的第二端杂交,

和(iv)U阻断核酸各自基本上不与至少八个可区分核酸条形码的部分杂交。

[0159] A2.实施方式A1的组合物,其包含一个或多个捕获核酸,其中

[0160] (i) 捕获核酸包含结合对的成员,和

[0161] (ii) 捕获核酸各自配置为与一个或多个文库插入物的亚组特异性杂交。

[0162] A3.实施方式A1或A2的组合物,其中一个或多个样品从一种或多种物种获得。

[0163] A3.1.实施方式A1至A3中任一项的组合物,其包含四个或更多样品。

[0164] A3.2.实施方式A1至A3中任一项的组合物,其包含八个或更多样品。

[0165] A4.实施方式A3至A3.2中任一项的组合物,其中第一和第二非天然核酸对于一种或多种物种的基因组不是内源性的。

[0166] A5.实施方式A1至A4中任一项的组合物,其中一个或多个样品从一个或多个组织获得。

[0167] A6.实施方式A1至A4中任一项的组合物,其中一个或多个样品从一种或多种哺乳动物获得。

[0168] A7.实施方式A6的组合物,其中一种或多种哺乳动物是人。

[0169] A8.实施方式A6或A7的组合物,其中第一和第二非天然核酸对于一种或多种哺乳动物的基因组不是内源性的。

[0170] A9.实施方式A1至A8中任一项的组合物,其包含十个或更多个可区分核酸条形码。

[0171] A10.实施方式A1至A9中任一项的组合物,其中一个或多个文库插入物从八个或更多样品获得。

[0172] A11.实施方式A1至A10中任一项的组合物,其中文库的每个核酸包含可区分核酸条形码中的两个。

[0173] A12.实施方式A11的组合物,其中第一非天然核酸包含第一可区分核酸条形码和第二非天然核酸包含第二可区分核酸条形码。

[0174] A13.实施方式A12的组合物,其中U阻断核酸各自配置为阻断通过聚合酶进行延伸。

[0175] A14.实施方式A1至A13中任一项的组合物,其中第一和第二非天然核酸是合成核酸。

[0176] A15.实施方式A1至A14中任一项的组合物,其中第一和第二非天然核酸基本上不相同。

[0177] A16.实施方式A1至A15中任一项的组合物,其中第一和第二非天然核酸包含衔接子核酸。

[0178] A17.实施方式A1至A16中任一项的组合物,其中一至四个U阻断核酸包含10至40个核苷酸的长度。

[0179] A18.实施方式A1至A17中任一项的组合物,其中不超过四个U阻断核酸包含10至30个核苷酸的长度。

[0180] A19.实施方式A1至A18中任一项的组合物,其中不超过四个U阻断核酸包含10至20个核苷酸的长度。

[0181] A20.实施方式A1至A19中任一项的组合物,其中不超过四个U阻断核酸包含锁核酸。

- [0182] A21.实施方式A1至A20中任一项的组合物,其中不超过四个U阻断核酸包含桥连核酸。
- [0183] A22.实施方式A1至A21中任一项的组合物,其中不超过四个U阻断核酸包含约65℃至约90℃的熔融温度。
- [0184] A23.实施方式A1至A22中任一项的组合物,其中不超过四个U阻断核酸包含至少约65℃的熔融温度。
- [0185] A24.实施方式A1至A22中任一项的组合物,其中不超过四个U阻断核酸包含至少约75℃的熔融温度。
- [0186] A25.实施方式A1至A24中任一项的组合物,其中组合物包含四个U阻断核酸。
- [0187] A26.实施方式A1至A25中任一项的组合物,其中第一非天然核酸包含至少八个可区分核酸条形码之一、与第一U阻断核酸基本上互补的部分和与第二U阻断核酸基本上互补的部分,其中第一U阻断核酸配置为邻近一个可区分核酸条形码的第一端杂交,并且第二U阻断核酸配置为邻近一个可区分核酸条形码的第二端杂交。
- [0188] A27.实施方式A1至A26中任一项的组合物,其中第二非天然核酸包含至少八个可区分核酸条形码之一、与第三U阻断核酸基本上互补的部分和与第四U阻断核酸基本上互补的部分,其中第三U阻断核酸配置为邻近一个可区分核酸条形码的第一端杂交,并且第四U阻断核酸配置为邻近一个可区分核酸条形码的第二端杂交。
- [0189] A28.实施方式A1至A27中任一项的组合物,其中不超过四个U阻断核酸与至少八个可区分核酸条形码基本上不互补。
- [0190] A29.实施方式A1至A28中任一项的组合物,其包含竞争者核酸。
- [0191] A29.1实施方式A29的组合物,其中竞争者核酸包含胎盘核酸。
- [0192] A30.实施方式A29的组合物,其中竞争者核酸包含重复核酸。
- [0193] A31.实施方式A30的组合物,其中重复核酸是人。
- [0194] A32.实施方式A30的组合物,其中竞争者核酸包含重复核酸的至少60%。
- [0195] A33.实施方式A30的组合物,其中竞争者核酸包含合成重复核酸。
- [0196] A34.实施方式A29的组合物,其中竞争者核酸包含C0t-1核酸。
- [0197] A35.实施方式A2至A34中任一项的组合物,其中结合对的成员包含生物素、抗原、半抗原、抗体或其部分。
- [0198] A36.实施方式A35的组合物,其中结合对的成员包含生物素。
- [0199] A37.实施方式A35的组合物,其中结合对的成员包含DNA结合蛋白识别序列或其部分。
- [0200] A38.实施方式A1至A37中任一项的组合物,其中不超过四个U阻断核酸是单链。
- [0201] A39.实施方式A1至A38中任一项的组合物,其中不超过四个U阻断核酸包含链终止子。
- [0202] A40.实施方式A1至A39中任一项的组合物,其中不超过四个U阻断核酸包含反向重复。
- [0203] A41.实施方式A1至A40中任一项的组合物,其中核酸的文库包含单链核酸。
- [0204] A42.实施方式A1至A41中任一项的组合物,其中核酸的文库包含扩增子。
- [0205] A43.实施方式A1至A42中任一项的组合物,其中多个文库插入物包含基因组核酸。

[0206] A44.实施方式A1至A43中任一项的组合物,其中不超过四个U阻断核酸不包含简并核苷酸碱基。

[0207] A45.实施方式A44的组合物,其中简并核苷酸碱基是3-硝基吡咯、5-硝基吡咯、其类似物或衍生物。

[0208] A46.实施方式A45的组合物,其中简并核苷酸碱基是肌苷、2'-脱氧肌苷、其类似物或衍生物。

[0209] A47.实施方式A1至A46中任一项的组合物,其包含第一、第二、第三和第四U阻断核酸,其中第一和第二U阻断核酸与第一非天然核酸的部分基本上互补,并且第三和第四U阻断核酸与第二非天然核酸的部分基本上互补。

[0210] A48.实施方式A1至A47中任一项的组合物,其中不超过四个U阻断核酸包含基本上不同的核酸序列。

[0211] B1.一种分析核酸文库的方法,其包括:

[0212] a) 获得包含从一个或多个样品获得的多个文库插入物的核酸的文库和至少八个可区分核酸条形码,每个核酸条形码包含第一端和第二端,和文库的每个核酸包含 (i) 至少一个文库插入物, (ii) 第一非天然核酸, (iii) 第二非天然核酸和 (iv) 可区分核酸条形码中的不超过两个,其中

[0213] 第一非天然核酸和第二非天然核酸位于至少一个文库插入物的相对侧上,和

[0214] 第一非天然核酸和/或第二非天然核酸包含不超过两个可区分核酸条形码;

[0215] b) 制备第一混合物,其包含使核酸的文库与不超过四个U阻断核酸接触,其中U阻断核酸各自与第一和/或第二非天然核酸的部分基本上互补,U阻断核酸中的至少一个配置为邻近可区分核酸条形码各自的第一端杂交和U阻断核酸中的至少一个配置为邻近可区分核酸条形码各自的第二端杂交;

[0216] c) 制备第二混合物,其包括使第一混合物与一个或多个捕获核酸接触,其中

[0217] (i) 捕获核酸包含结合对的第一成员,和

[0218] (ii) 捕获核酸各自配置为与一个或多个文库插入物的亚组特异性杂交;

[0219] d) 使第二混合物与结合对的第二成员接触,从而提供分离的核酸;

[0220] e) 在扩增条件下使分离的核酸与引物组接触,从而提供扩增子;和

[0221] f) 分析扩增子。

[0222] B2.实施方式B1的方法,其中一个或多个样品从一种或多种物种获得。

[0223] B3.实施方式B2的方法,其中文库插入物从四个或更多个样品获得。

[0224] B4.实施方式B2的方法,其中文库插入物从八个或更多个样品获得。

[0225] B5.实施方式B1至B4中任一项的方法,其中一个或多个样品从一个或多个组织获得。

[0226] B6.实施方式B1至B5中任一项的方法,其中一个或多个样品从一种或多种哺乳动物获得。

[0227] B7.实施方式B6的方法,其中一种或多种哺乳动物是人。

[0228] B8.实施方式B6或B7的方法,其中第一和第二非天然核酸对于一种或多种哺乳动物的基因组不是内源性的。

[0229] B9.实施方式B1至B8中任一项的方法,插入物的文库包含十个或更多个可区分核

酸条形码。

[0230] B10.实施方式B1至B9中任一项的方法,其中第一非天然核酸包含第一可区分核酸条形码,和第二非天然核酸包含第二可区分核酸条形码。

[0231] B11.实施方式B1至B10中任一项的方法,其中文库的每个核酸包含可区分核酸条形码中的两个。

[0232] B12.实施方式B1至B11中任一项的方法,其中U阻断核酸各自配置为阻断通过聚合酶进行延伸。

[0233] B13.实施方式B1至B12中任一项的方法,其中第一和第二非天然核酸是合成核酸。

[0234] B14.实施方式B1至B13中任一项的方法,其中第一和第二非天然核酸基本上不相同。

[0235] B15.实施方式B1至B14中任一项的方法,其中第一和第二非天然核酸包含衔接子核酸。

[0236] B16.实施方式B1至B15中任一项的方法,其中一至四个U阻断核酸包含10至40个核苷酸的长度。

[0237] B17.实施方式B1至B16中任一项的方法,其中不超过四个U阻断核酸包含10至30个核苷酸的长度。

[0238] B18.实施方式B1至B17中任一项的方法,其中不超过四个U阻断核酸包含10至20个核苷酸的长度。

[0239] B19.实施方式B1至B18中任一项的方法,其中不超过四个U阻断核酸包含锁核酸。

[0240] B20.实施方式B1至B19中任一项的方法,其中不超过四个U阻断核酸包含桥连核酸。

[0241] B21.实施方式B1至B20中任一项的方法,其中不超过四个U阻断核酸包含至少约65℃的熔融温度。

[0242] B22.实施方式B1至B21中任一项的方法,其中不超过四个U阻断核酸包含至少约75℃的熔融温度。

[0243] B23.实施方式B1至B22中任一项的方法,其中不超过四个U阻断核酸包含约65℃至约90℃的熔融温度。

[0244] B24.实施方式B1至B23中任一项的方法,其中第一非天然核酸包含至少八个可区分核酸条形码之一、与第一U阻断核酸基本上互补的部分和与第二U阻断核酸基本上互补的部分,其中第一U阻断核酸配置为邻近一个可区分核酸条形码的第一端杂交,并且第二U阻断核酸配置为邻近一个可区分核酸条形码的第二端杂交。

[0245] B25.实施方式B1至B24中任一项的方法,其中第二非天然核酸包含至少八个可区分核酸条形码之一、与第三U阻断核酸基本上互补的部分和与第四U阻断核酸基本上互补的部分,其中第三U阻断核酸配置为邻近一个可区分核酸条形码的第一端杂交,并且第四U阻断核酸配置为邻近一个可区分核酸条形码的第二端杂交。

[0246] B26.实施方式B1至B25中任一项的方法,其中不超过四个U阻断核酸与至少八个可区分核酸条形码基本上不互补。

[0247] B27.实施方式B26的方法,其中不超过四个U阻断核酸基本上不与至少八个可区分核酸条形码的部分杂交。

- [0248] B28.实施方式B1至B27中任一项的方法,其包括在c)之前,将第一混合物与竞争者核酸接触。
- [0249] B28.1实施方式B28的方法,其中竞争者核酸包含胎盘核酸。
- [0250] B29.实施方式B28或B28.1的方法,其中竞争者核酸包含重复核酸。
- [0251] B30.实施方式B29的方法,其中重复核酸是人。
- [0252] B31.实施方式B29的方法,其中竞争者核酸包含重复核酸的至少60%。
- [0253] B32.实施方式B1的方法,其中竞争者核酸包含合成核酸。
- [0254] B33.实施方式B1的方法,其中竞争者核酸包含C0t-1核酸。
- [0255] B34.实施方式B1的方法,其中结合对的第一成员包含生物素、抗原、半抗原、抗体或其部分。
- [0256] B35.实施方式B34的方法,其中结合对的第一成员包含生物素。
- [0257] B36.实施方式B1的方法,其中结合对的第一成员包含DNA结合蛋白识别序列或其部分。
- [0258] B37.实施方式B1至B36中任一项的方法,其中不超过四个U阻断核酸是单链。
- [0259] B38.实施方式B1至B37中任一项的方法,其中不超过四个U阻断核酸包含链终止子。
- [0260] B39.实施方式B1至B38中任一项的方法,其中不超过四个U阻断核酸包含反向重复。
- [0261] B40.实施方式B1至B39中任一项的方法,其中核酸的文库包含单链核酸。
- [0262] B41.实施方式B1至B40中任一项的方法,其中核酸的文库包含扩增子。
- [0263] B42.实施方式B1至B41中任一项的方法,其中多个文库插入物包含基因组核酸。
- [0264] B43.实施方式B1至B42中任一项的方法,其中不超过四个U阻断核酸不包含简并核苷酸碱基。
- [0265] B44.实施方式B43的方法,其中简并核苷酸碱基是3-硝基吡咯、5-硝基吡咯、其类似物或衍生物。
- [0266] B45.实施方式B43的方法,其中简并核苷酸碱基是肌苷、2'-脱氧肌苷、其类似物或衍生物。
- [0267] B46.实施方式B1的方法,其中结合对的第一成员包含生物素、抗原、半抗原、抗体或其部分。
- [0268] B47.实施方式B1的方法,其中结合对的第一成员包含生物素。
- [0269] B48.实施方式B1的方法,其中结合对的第一成员包含DNA结合蛋白识别序列或其部分。
- [0270] B49.实施方式B1至B48中任一项的方法,其包括在杂交条件下杂交分离的核酸。
- [0271] B50.实施方式B1的方法,其中扩增条件包括热稳定聚合酶。
- [0272] B51.实施方式B1的方法,其中扩增条件包括聚合酶链式反应。
- [0273] B52.实施方式B1的方法,其中捕获核酸配置为与外显子的部分特异性杂交。
- [0274] B53.实施方式B1的方法,其中捕获核酸配置为与染色体的部分特异性杂交。
- [0275] B54.实施方式B1的方法,其中结合对的第二成员包含抗生物素蛋白、蛋白A、蛋白G、抗体或其结合部分。

- [0276] B55.实施方式B1的方法,其中结合对的第二成员包含抗生物素蛋白、或其部分。
- [0277] B56.实施方式B1的方法,其中结合对的第二成员包含基质。
- [0278] B57.实施方式B56的方法,其中基质包含磁性化合物。
- [0279] B58.实施方式B56的方法,其中基质包含珠。
- [0280] B59.实施方式B56的方法,其中基质包含聚苯乙烯、聚碳酸酯或琼脂糖。
- [0281] B60.实施方式B56的方法,其中基质包含磁珠。
- [0282] B61.实施方式B1的方法,其中(d)中的接触包括离心。
- [0283] B62.实施方式B1的方法,其中(d)中的接触包括使用磁体。
- [0284] B63.实施方式B1的方法,其中分析包括提供序列读数。
- [0285] B64.实施方式B63的方法,其中序列读数通过包括大规模并行测序的方法获得。
- [0286] B65.实施方式B63的方法,其中序列读数通过包括双端测序的方法获得。
- [0287] B66.实施方式B1至B65中任一项的方法,其中不超过四个U阻断核酸包含第一、第二、第三和第四U阻断核酸,其中第一和第二U阻断核酸与第一非天然核酸的部分基本上互补,并且第三和第四U阻断核酸与第二非天然核酸的部分基本上互补。
- [0288] B67.实施方式B1至B66中任一项的方法,其中不超过四个U阻断核酸包含基本上不同的核酸序列。
- [0289] C1.一种分析核酸文库的方法,其包括
- [0290] a) 获得包含第一组扩增子的核酸的文库,其中每个扩增子包含第一非天然核酸和第二非天然核酸、一个或多个可区分标识符和从一个或多个样品之一获得的文库插入物,其中文库插入物位于第一和第二非天然核酸之间;
- [0291] b) 制备混合物,其包括使核酸的文库与一种或多种阻断核酸和捕获核酸接触,其中
- [0292] (i) 一个或多个阻断核酸配置为与第一和第二非天然核酸的部分特异性杂交,
- [0293] (ii) 捕获核酸包含结合对的第一成员,和
- [0294] (iii) 捕获核酸配置为与第一组扩增子的亚组特异性杂交;
- [0295] c) 使混合物纯化,从而提供纯化核酸,其中纯化核酸包含核酸的文库,一个或多个阻断核酸和捕获核酸;
- [0296] d) 在杂交条件下杂交纯化核酸;
- [0297] e) 捕获捕获核酸,从而提供捕获的核酸;
- [0298] f) 在扩增条件下将捕获的核酸与引物组接触,从而提供第二组扩增子;以及
- [0299] g) 分析第二组扩增子。
- [0300] C2.实施方式C1的方法,其中一个或多个样品从人获得。
- [0301] C3.实施方式C2的方法,其中第一核酸和第二核酸对于人不是内源性的。
- [0302] C4.实施方式C1至C3中任一项的方法,其中一个或多个样品从选自乳腺组织、结肠组织、胰腺组织、胎盘或上皮组织的组织获得。
- [0303] C5.实施方式C1至C3中任一项的方法,其中一个或多个样品从血液获得。
- [0304] C6.实施方式C5的方法,其中一个或多个样品从循环血细胞获得。
- [0305] C7.实施方式C2至C6中任一项的方法,其中人是胎儿。
- [0306] C8.实施方式C5的方法,其中一个或多个样品包含无循环细胞核酸。

- [0307] C9.实施方式C1至C8中任一项的方法,其中扩增条件包含热稳定聚合酶。
- [0308] C10.实施方式C1至C9中任一项的方法,其中扩增条件包含聚合酶链式反应。
- [0309] C11.实施方式C1至C10中任一项的方法,其中(b)中的制备包括使文库的核酸与竞争者核酸接触。
- [0310] C12.实施方式C11的方法,其中竞争者核酸包含胎盘核酸。
- [0311] C13.实施方式C11的方法,其中竞争者核酸包含重复核酸。
- [0312] C14.实施方式C13的方法,其中重复核酸源自人。
- [0313] C15.实施方式C13的方法,其中竞争者核酸包含重复核酸的至少60%或更多。
- [0314] C16.实施方式C11的方法,其中竞争者核酸包含合成核酸。
- [0315] C17.实施方式C11的方法,其中竞争者核酸包含C0t-1核酸。
- [0316] C18.实施方式C1至C17中任一项的方法,其中捕获核酸配置为与文库插入物的部分特异性杂交。
- [0317] C19.实施方式C1至C18中任一项的方法,其中一个或多个阻断核酸配置为与第一非天然核酸和/或第二非天然核酸的部分特异性杂交。
- [0318] C20.实施方式C1至C18中任一项的方法,其中一个或多个阻断核酸配置为阻止阻断核酸通过聚合酶延伸。
- [0319] C21.实施方式C1至C20中任一项的方法,其中一个或多个阻断核酸包含链终止子。
- [0320] C22.实施方式C1至C21中任一项的方法,其中一个或多个阻断核酸包含反向重复。
- [0321] C23.实施方式C1至C22中任一项的方法,其中捕获核酸配置为与外显子的部分特异性杂交。
- [0322] C24.实施方式C1至C23中任一项的方法,其中捕获核酸配置为与染色体的部分特异性杂交。
- [0323] C25.实施方式C24的方法,其中捕获核酸配置为与包含遗传变异的文库插入物的部分特异性杂交。
- [0324] C26.实施方式C1至C26中任一项的方法,其中结合对的第一成员包含生物素、抗原、半抗原、抗体或其部分。
- [0325] C27.实施方式C26的方法,其中结合对的第一成员包含生物素。
- [0326] C28.实施方式C1至C27中任一项的方法,其中结合对的第一成员包含CNC结合蛋白识别序列或其部分。
- [0327] C29.实施方式C1至C28中任一项的方法,其中(e)中的捕获包括使混合物与结合对的第二成员接触。
- [0328] C30.实施方式C29的方法,其中结合对的第二成员包含抗生物素蛋白、蛋白A、蛋白G、抗体或其结合部分。
- [0329] C31.实施方式C30的方法,其中结合对的第二成员包含生物素蛋白、或其部分。
- [0330] C32.实施方式C29至C31中任一项的方法,其中结合对的第二成员包含基质。
- [0331] C33.实施方式C32的方法,其中基质包含磁性化合物。
- [0332] C34.实施方式C32的方法,其中基质包含珠。
- [0333] C35.实施方式C32的方法,其中基质包含聚苯乙烯、聚碳酸酯或琼脂糖。
- [0334] C36.实施方式C32的方法,其中基质包含金属。

- [0335] C37.实施方式C1至C36中任一项的方法,其中(e)中的捕获包括通过包括离心的方法回收捕获的核酸。
- [0336] C38.实施方式C1至C37中任一项的方法,其中(e)中的捕获包括通过包括使用磁体的方法回收捕获的核酸。
- [0337] C39.实施方式C1至C38中任一项的方法,其中杂交条件包括变性。
- [0338] C40.实施方式C1至C39中任一项的方法,其中(d)中的杂交包括将捕获的核酸与第一组扩增子中的一个或多个的部分杂交。
- [0339] C41.实施方式C1至C40中任一项的方法,其中杂交条件包括在约25℃至约70℃的温度下孵育捕获的核酸。
- [0340] C42.实施方式C41的方法,其中孵育是在约35℃至约60℃的温度下。
- [0341] C43.实施方式C41至C42中任一项的方法,其中孵育约1小时至约24小时的时间量。
- [0342] C44.实施方式C41至C43中任一项的方法,其中孵育约12小时至约20小时的时间量。
- [0343] C45.实施方式C1至C44中任一项的方法,其中(d)中的杂交包括使混合物与杂交缓冲液接触。
- [0344] C46.实施方式C1至C45中任一项的方法,其中(d)中的杂交包括(i)使混合物与杂交缓冲液接触,(ii)变性和(iii)杂交的顺序步骤。
- [0345] C47.实施方式C1至C46中任一项的方法,其中方法不包括干燥步骤。
- [0346] C48.实施方式C1至C47中任一项的方法,其中杂交条件不包括聚合酶。
- [0347] C49.实施方式C1至C48中任一项的方法,其中分析包括提供序列读数。
- [0348] C50.实施方式C49的方法,其中序列读数通过包括大规模并行测序的方法获得。
- [0349] C51.实施方式C49的方法,其中序列读数通过包括双端测序的方法获得。
- [0350] C52.实施方式C1至C51中任一项的方法,其中第一非天然核酸包含一个或多个可区分标识符。
- [0351] C53.实施方式C1至C52中任一项的方法,其中第二非天然核酸包含一个或多个可区分标识符。
- [0352] C54.实施方式C1至C53中任一项的方法,其中一个或多个可区分标识符包含核酸条形码。
- [0353] C55.实施方式C1至C54中任一项的方法,其中一个或多个阻断核酸包含锁核酸。
- [0354] C56.实施方式C1至C55中任一项的方法,其中一个或多个阻断核酸包含桥连核酸。
- [0355] C57.实施方式C1至C56中任一项的方法,其中方法不包括(c)之前的变性步骤。
- [0356] C58.实施方式C1至C57中任一项的方法,其中方法不包括(d)之前的变性步骤。
- [0357] C59.实施方式C1至C58中任一项的方法,其中方法不包括在(d)之前加热到高于80℃的温度。
- [0358] C60.实施方式C1至C59中任一项的方法,其中方法不包括在(d)之前加热到高于90℃的温度。
- [0359] C61.实施方式C1至C60中任一项的方法,其中捕获的核酸包含文库的核酸的亚组。
- [0360] C62.实施方式C1至C61中任一项的方法,其中一个或多个样品包含10个或更多个样品。

[0361] C63.实施方式C1至C62中任一项的方法,其中一个或多个可区分标识符由10个或更多个可区分标识符组成。

[0362] C64.实施方式C1至C63中任一项的方法,其中第一核酸和第二核酸包含合成核酸。

[0363] C65.实施方式C1至C64中任一项的方法,其中文库插入物包含基因组核酸的部分。

[0364] C66.实施方式C1至C65中任一项的方法,其中(c)中的纯化不包括加入配置为结合结合对的第一成员的结合对的第二成员。

[0365] C67.实施方式C1的方法,其中(c)中的纯化包括使核酸与基质非特异性结合的方法。

[0366] C68.实施方式C1至C67中任一项的方法,其中(c)中的纯化包括使用阴离子交换树脂。

[0367] C69.实施方式C1至C68中任一项的方法,其中(c)中的捕获包括加入配置为结合结合对的第一成员的结合对的第二成员。

[0368] C70.实施方式C1至C69中任一项的方法,其中在(e)之前,混合物不固定在流动池或阵列的基质上。

[0369] D1.一种分析基因组DNA文库的方法,其包括:

[0370] a)获得包含第一组单链扩增子的基因组DNA文库,其中每个扩增子包含第一非天然核酸和第二非天然核酸,一个或两个核酸条形码,和从十个或更多个人受试者之一的基因组获得的文库插入物,其中文库插入物位于第一和第二非天然核酸之间,和其中第一组扩增子包含来自十个或更多个人受试者的多个文库插入物;

[0371] b)制备混合物,其包括使第一组扩增子与一至四个阻断核酸、C0t-1DNA和捕获核酸接触,其中

[0372] (i)一至四个阻断核酸配置为与第一和/或第二非天然核酸特异性杂交,

[0373] (ii)一至四个阻断核酸包含锁核酸和包含10至30个核苷酸的长度,

[0374] (iii)捕获核酸包含生物素,和

[0375] (iv)捕获核酸配置为与多个文库插入物的亚组特异性杂交;

[0376] c)使混合物与包含非特异性核酸结合基质的磁珠接触,从而提供纯化核酸,其中纯化核酸包含第一组扩增子、一至四个阻断核酸、C0t-1DNA和捕获核酸;

[0377] d)杂交纯化核酸,其中杂交包括(i)使混合物与杂交缓冲液接触,(ii)加热纯化核酸至至少95°C约10分钟和(iii)通过在约40°C至约50°C孵育12至20小时杂交纯化核酸的顺序步骤;

[0378] e)捕获捕获核酸,其中捕获包括使纯化核酸与配置为特异性结合捕获核酸的抗生物素蛋白涂布磁珠接触,并使用磁体固定捕获的核酸,从而提供捕获的核酸;

[0379] f)在扩增条件下使捕获的核酸与引物组接触,从而提供第二组扩增子;和

[0380] g)通过包括双端测序的方法从第二组扩增子获得序列读数,其中所述方法不包括干燥步骤。

[0381] 本文引用的每个专利、专利申请、出版物和文件的全部内容通过引用并入本文。上述专利、专利申请、出版物和文件的引用并非承认上述任一者是相关现有技术,也不构成对这些出版物或文件的内容或日期的任何承认。

[0382] 在不脱离本技术的基本方面的情况下可以对前述进行修改。虽然已经参照一个或

多个具体实施方式详细描述本技术,但是本领域普通技术人员将认识到可以对本申请中具体公开的实施方案进行改变,但是这些修改和改进是在本技术的范围和精神内。

[0383] 本文说明性描述的技术适合地可以在不存在本文未具体公开的任何要素的情况下实施。因此,在本文的每个实例中,术语“包含”、“基本上由……组成”和“由……组成”中的任一个可以用其他两个术语中的任一个替代。已经使用的术语和表达是用作描述而非限制的术语,并且使用这样的术语和表达不排除所示出和描述的特征或其部分的任何等同物,并且各种修改在要求保护的本技术的范围内是可能的。术语“一个(a)”或“一种(an)”可以指其修饰的要素中的一个或多个(例如“试剂”可以指一种或多种试剂),除非上下文明显地,描述了要素中的任一个或要素中的多于一个。本文所用的术语“约”指基础参数的10%内的值(即,加或减10%),并且在一串值的开头使用术语“约”修饰每个值(例如,“约1、2和3”指约1、约2和约3)。例如,“约100克”的重量可以包括90克与110克之间的重量。此外,当本文描述值列表(例如,约50%、60%、70%、80%、85%或86%)时,该列表包括其所有中间值和分数值(例如,54%、85.4%)。因此,应理解虽然本技术已经通过代表性实施方式和任选特征具体公开,但是本领域技术人员可以采用本文公开的概念的修改和变化,并且这样的修改和变化被认为在该技术的范围内。

[0384] 本技术的某些实施方式在随后的权利要求中阐述。

四寡聚物
阻断策略

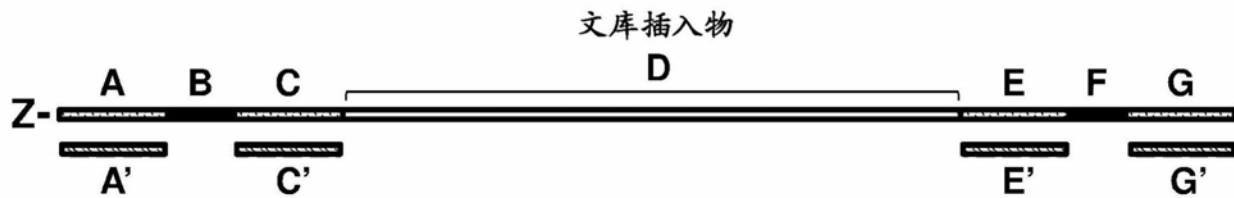


图1