



(12) 发明专利申请

(10) 申请公布号 CN 114774529 A

(43) 申请公布日 2022. 07. 22

(21) 申请号 202210458664.7

(51) Int. Cl.

(22) 申请日 2016.11.18

C12Q 1/6876 (2018.01)

(30) 优先权数据

C12Q 1/6869 (2018.01)

62/257,438 2015.11.19 US

C12N 15/11 (2006.01)

(62) 分案原申请数据

201680067419.6 2016.11.18

(71) 申请人 10X基因组学有限公司

地址 美国加利福尼亚州

(72) 发明人 保罗·雷夫金 詹森·安德伍德

迈克尔·史诺-莱文

塔尔耶伊·西格德·米克尔森

本杰明·辛德森

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

专利代理师 刘晓杰 武晶晶

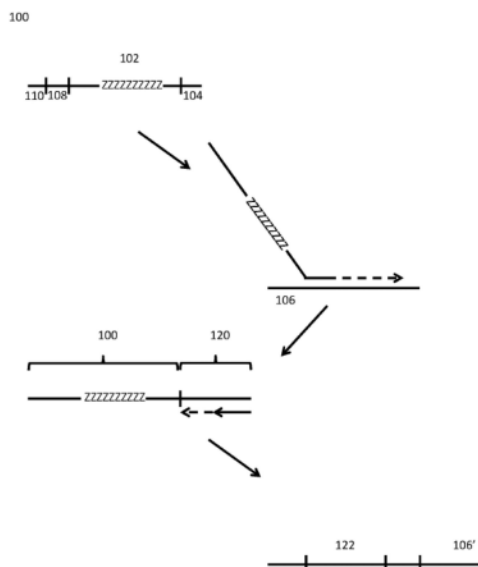
权利要求书1页 说明书24页 附图4页

(54) 发明名称

可转化标记组合物、方法及结合其的过程

(57) 摘要

本公开提供了提供用于分析操作,并且特别是分析生物系统,例如基于细胞的系统中的基因表达的分析中的可转化标记部分的方法、系统和组合物。



1. 一种差异化地标记多种分子物质中的个别成员的方法,其包括:
 - (a) 将第一标记部分附接至多个单独分子物质中的每一个,所述第一标记部分包含可转化标记部件;以及
 - (b) 将附接到所述多个单独分子中的每一个的所述可转化标记部件转化成转化的标记部件,以用不同的转化的标记部件清楚地标记所述多个分子物质中的多个不同成员。
2. 如权利要求1所述的方法,其中
所述多个单独分子物质包含多个单独核酸序列;
所述标记部分包含寡核苷酸区段;并且
所述标记部件包含可转化的寡核苷酸序列。
3. 如权利要求2所述的方法,其中所述可转化的寡核苷酸序列包含一个或多个可转化的核苷酸。
4. 如权利要求3所述的方法,其中一个或多个可转化的核苷酸包含简并核苷酸。
5. 如权利要求4所述的方法,其中所述一个或多个可转化的核苷酸中的一个或多个包含2-向简并性。
6. 如权利要求4所述的方法,其中所述一个或多个可转化的核苷酸中的一个或多个包含3-向简并性。
7. 一种分析核酸分子的方法,其包括:
 - (a) 将寡核苷酸区段附接至靶寡核苷酸分子以产生经标记寡核苷酸,其中所述寡核苷酸包含含有多个可变互补核苷酸的区域;
 - (b) 复制所述标记寡核苷酸以产生复制的经标记寡核苷酸,由此复制产生所述区域的随机或部分随机复制物;以及
 - (c) 分析所述复制的经标记寡核苷酸,其包括所述随机或部分随机复制物,以鉴定所述靶寡核苷酸分子。
8. 一种寡核苷酸组合物,其包含含有第一区域和第二区域的寡核苷酸,其中所述第二区域包含含有多个可变互补核苷酸的固定序列,所述多个可变互补核苷酸可转化以产生独特的分子标记。
9. 一种对相同核酸分子的群体中的核酸分子进行定量的方法,其包括:
 - (a) 以预期的诱变速率突变所述相同核酸分子的群体以产生不同突变核酸的群体;
 - (b) 对所述不同突变核酸分子进行测序;以及
 - (c) 基于多个不同突变核酸分子计算所述相同核酸分子的群体中的所述核酸分子的定量。
10. 一种区分来自两个或更多个相同核酸分子的扩增产物的方法,其包括:
 - (a) 使所述两个或更多个核酸分子经历诱变以产生两个或更多个突变的核酸分子;
 - (b) 扩增所述两个或更多个突变的核酸分子以产生扩增的突变核酸产物;以及
 - (c) 对所述扩增的突变核酸产物进行测序。

可转化标记组合物、方法及结合其的过程

[0001] 本申请是申请日为2016年11月18日、申请号为201680067419.6、发明名称为“可转化标记组合物、方法及结合其的过程”的中国专利申请(其对应PCT申请的申请日为2016年11月18日、申请号为PCT/US2016/062955)的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请请求2015年11月19日提交的美国临时申请号62/257,438的优先权,所述申请通过引用整体并入本文。

[0004] 背景

[0005] 生命科学领域在过去的二十年中发生了巨大的进步。从重组脱氧核糖核酸(DNA)技术取得的产品的广泛商业化到研究、开发和诊断的简化,通过发明和部署关键研究工具(例如聚合酶链式反应(PCR)、核酸阵列技术、强大的核酸测序技术),以及最近的高通量下一代测序技术的开发和商业化实现。所有这些改进已经组合在一起而推动了生物研究、医学、诊断学、农业生物技术,及无数其它相关领域的跨越式发展。

[0006] 化学反应的分析依赖于测量、定量和跟踪这些反应中所涉及的各种反应物和产物的消耗、产生、转变和转化的能力。尽管在某些情况下,这些反应物及其产物本身易于辨识和测量,但在许多情况下分析得益于使用与反应物和/或产物偶联的标记或标签部分以促进其测量和/或鉴定。

[0007] 在一些情况下,标签或标记部分包括更易辨识或检测的基团、分子或化学部分。这些可以包括组合物如荧光化学品、带电荷的化学基团、亲和结合基团、以及在某些情况下编码的分子或条形码,所述编码的分子或条形码在其结构内包含可变量的信息。特别有用的条形码分子的实例包括例如可以使用多种序列鉴定技术(例如,核酸测序、基于探针杂交的测定等)中的任何一种来读出的核酸条形码或标记。

[0008] 条形码策略已应用于多种标记和鉴定策略。例如,在一些情况下,在固体支持物(例如,珠粒)上逐步构建寡核苷酸已被用作在那些固体支持物上创建分子文库的特定化学合成操作(例如,在随机/组合合成过程)的指示物,其中所述寡核苷酸的结构单元各自反映给定固体支持物已经暴露于的特定化学合成操作(参见例如美国专利号5,708,153)。通过读出给定固体支持物上添加的核苷酸的序列,可以确定合成操作及其顺序,以鉴定在所述特定固体支持物上合成的化合物。

[0009] 在其它情况下,条形码寡核苷酸已在测序过程中用以将预先合成的具有已知序列的寡核苷酸附加到从不同样品创建的测序文库中,使得每个不同样品具有独特的条形码寡核苷酸,所述条形码寡核苷酸与来自所述样品的核酸的序列衔接并被读出。这可以允许对多个样本进行汇总分析,其中从汇总中得到的序列信息可以稍后归属于其起始样本。

[0010] 在另一测序应用中,寡核苷酸条形码已用于超高通量分配系统中,以共同分配样品核酸的长片段以及条形码携带颗粒,其中个别颗粒上的条形码是相同的,但颗粒文库代表了多样化的条形码文库。然后将这些条形码偶联到长起始片段的子区段上,使得在给定的分区内,每个长片段的所有子区段都承载相同的条形码序列。当使用例如短读段测序系统对子区段进行测序时,可以将具有相同条形码序列的子区段归属于相同的起始长分子。

这允许凭借包含的条形码序列来保留短序列读段的长程序列内容(参见例如公布的美国专利申请公开号2014-0378345,其全部公开内容出于所有目的通过引用整体并入本文)。

[0011] 在其它情况下,可以引入大量不同的条形码以与样品分子的集合接触,使得所述集合内的分子各自偶联至不同的条形码分子,从而无论所述分子在被鉴定之前被如何扩增、复制等,都允许将序列归属于特定的起始分子。在样品包括相同类型分子(例如,相同的核酸)的多个拷贝的情况下,对这些基础分子以及附接到每个分子的不同条形码的测序可以允许计数在标记时存在多少个个别成员,从而允许计数那些起始分子,例如以用于信使核糖核酸(mRNA)表达分析等。

发明概要

[0012] 本文认识到与当前可用的条形码策略相关的限制。例如,对于上述所有条形码策略,基本前提是需要大量不同的寡核苷酸条形码,从而允许区分大量不同的结果,例如样品、分区、分子等。在大量样品中准备、制造和分配这些不同的分子文库可能被证明在许多情况下具有挑战性。

[0013] 本公开提供了对这种方法的显著改进,所述改进也能给整个过程带来效率、成本节约和其它节约。本发明的装置、方法和系统为生命科学和其它领域的这些和其它挑战提供解决方案。

[0014] 本文提供了用于标记分子事件、反应、种类等但不需要复杂的、高度多样化的标记分子文库的组合物、系统和方法。具体而言,提供了标记部分,其可以具有较少数量、几个或甚至单个原始“标记”结构,所述“标记”结构可以原位转化或可原位转化转化为大量独特标记或“条形码”部分的集合。

[0015] 在一个方面,本公开提供了差异化地标记多个分子物质的个别成员的方法,其包括将第一标记部分附接到多个单独分子物质中的每一个,所述第一标记部分包含可转化标记部件;以及将附接到所述多个单独分子中的每一个的可转化标记部件转化成转化的标记部件,以用不同的转化标记部件清楚地标记所述多个分子物质中的多个不同成员。

[0016] 在一些实施方案中,所述多个单独分子物质包含多个单独核酸序列;所述标记部分包含寡核苷酸区段;并且所述标记部件包含可转化的寡核苷酸序列。

[0017] 在一些实施方案中,所述可转化的寡核苷酸序列包含一个或多个可转化的核苷酸。在一些实施方案中,所述一个或多个可转化的核苷酸包含简并核苷酸。在一些实施方案中,所述一个或多个可转化的核苷酸中的一个或多个包含2-向简并性。在一些实施方案中,所述一个或多个可转化的核苷酸中的一个或多个包含3-向简并性。在一些实施方案中,所述一个或多个可转化的核苷酸中的一个或多个包含4-向简并性。

[0018] 在一些实施方案中,所述一种或多种可转化核苷酸选自以下各项组成的组:肌苷、脱氧肌苷、脱氧黄嘌呤、2'-脱氧水粉菌素、2'-脱氧鸟苷、5-硝基吡啶、3-硝基吡啶、N6-甲氧基-2,6-二氨基嘌呤、6H,8H-3,4-二氢嘧啶并[4,5-c][1,2]噁嗪-7-酮,以及前述各种的非脱氧(或核糖)形式。在一些实施方案中,所述可转化的寡核苷酸序列包含1至20个可转化的核苷酸。

[0019] 在一些实施方案中,所述标记部分进一步包含一个或多个附加的寡核苷酸区段。在一些实施方案中,所述一个或多个附加的寡核苷酸区段选自引物序列区段、杂交序列区

段、连接序列区段、测序仪表面附接区段和条形码序列区段。在一些实施方案中,所述一个或多个附加的寡核苷酸序列包含选自随机引物序列和测序引物的引物序列。在一些实施方案中,所述一个或多个附加的寡核苷酸序列包含杂交序列。在一些实施方案中,所述杂交序列包含聚T序列。

[0020] 在一些实施方案中,所述方法进一步包括在附接之前将标记部分分配给包含待分析核酸的样品,并且其中所述附接包括将标记部分附接至待分析核酸。

[0021] 在一些实施方案中,所述标记部分包含聚T序列区段,并且待分析核酸包含mRNA分子。

[0022] 在一些实施方案中,所述分配包括将具有所述标记部分的个别细胞分配到分区中,并且其中这些待分析核酸包含在所述个别细胞内,并且其中在附接之前,将所述个别裂解以将待分析核酸释放至分区中。

[0023] 在一些实施方案中,所述可转化的寡核苷酸序列区段包含序列取代系统的靶序列。在一些实施方案中,所述序列取代系统包含CRISPR酶体系,并且所述靶序列包含靶向寡核苷酸的靶序列。

[0024] 在一些实施方案中,所述方法的转化是随机或半随机的。

[0025] 在另一方面,本公开提供了分析核酸分子的方法,其包括将寡核苷酸区段附接到靶寡核苷酸分子以产生经标记寡核苷酸,其中所述寡核苷酸包含含有多个可变互补核苷酸的区域;复制所述经标记寡核苷酸以产生复制的经标记寡核苷酸,由此复制产生所述区域的随机或部分随机复制物;以及分析复制的经标记寡核苷酸(包括随机或部分随机的复制物)以鉴定靶寡核苷酸分子。

[0026] 在一些实施方案中,所述区域包含2至20个可变互补核苷酸。在一些实施方案中,所述区域包含两个或更多个连续的可变互补核苷酸。

[0027] 在一些实施方案中,所述可变互补核苷酸中的两个或更多个通过一个或多个非可变互补核苷酸彼此分开。

[0028] 在一些实施方案中,所述区域包含4至10个可变互补核苷酸。

[0029] 在一些实施方案中,第一寡核苷酸包含含有多个可变互补核苷酸的附加区域。

[0030] 在另一方面,本公开提供了寡核苷酸组合物,其包含含有第一区域和第二区域的寡核苷酸,其中所述第二区域包含含有多个可变互补核苷酸的固定序列,所述多个可变互补核苷酸可转化以产生独特的分子标记。

[0031] 在一些实施方案中,第一区域包含用于将所述寡核苷酸附接至待分析核酸分子的附接序列。在一些实施方案中,其中所述附接序列包含引物序列。在一些实施方案中,所述附接序列包含聚T序列。

[0032] 在一些实施方案中,第一区域包含条形码序列。在一些实施方案中,第一区域包括表面附接序列。

[0033] 在一些实施方案中,第二区域包含多个可变互补核苷酸和一个或多个非可变互补核苷酸。

[0034] 在另一方面,本公开提供了对相同核酸分子的群体中的核酸分子进行定量的方法,包括以预期诱变速率突变相同核酸分子的群体以产生不同突变核酸群体;对不同突变核酸分子进行测序;以及基于多个不同突变核酸分子计算相同核酸分子的群体中的核酸分

子的定量。

[0035] 在一些实施方案中,所述计算包括基于不同突变核酸分子的数目和所述诱变速率对相同核酸分子的群体中的核酸分子进行定量。

[0036] 在一些实施方案中,所述测序包括从不同突变核酸分子产生测序读段。

[0037] 在一些实施方案中,所述计算包括计算这些测序读段的比较以定量相同核酸分子的群体中的这些核酸分子。

[0038] 在另一方面,本公开提供了区分来自两个或更多个相同核酸分子的扩增产物的方法,其包括使两个或更多个核酸分子经历诱变以产生两个或更多个突变核酸分子;扩增所述两个或更多个突变核酸分子以产生扩增的突变核酸产物;并对所述扩增的突变核酸产物进行测序。

[0039] 根据下面的详细描述,本公开的其它方面和优点对于本领域技术人员来说将变得显而易见,其中仅示出和描述了本公开的说明性实施方案。如将认识到的,本公开能够具有其它和不同的实施方案,并且其若干个细节能够在各种明显的方面进行修改,所有这些修改都不脱离本公开内容。因此,这些附图和描述本质上被认为是说明性的,而不是限制性的。

[0040] 引用并入

[0041] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同每个单独的出版物、专利或专利申请被具体地和单独地指出为通过引用并入。在通过引用并入的出版物和专利或专利申请与包含在说明书中的公开内容相矛盾的情况下,所述说明书旨在取代和/或优先于任何此类相互矛盾的材料。

[0042] 附图简述

[0043] 本发明的新颖特征在所附权利要求书中具体阐述。通过参考以下详细描述和附图(在此也为“图式”和“图”)来获得对本发明的特征和优点的更好理解,所述详细描述阐述了利用本发明的原理的说明性实施方案,在附图中:

[0044] 图1提供了根据本公开的标记构建体及其实施的示意图;

[0045] 图2提供了本公开的示例标记过程的高级流程图;

[0046] 图3提供了用于在本公开的标记过程中将标记部分分配给个别细胞的分配系统和过程的示意图;以及

[0047] 图4提供了用于例如定量从细胞内的基因表达的信使核糖核酸(mRNA)的本公开的标记过程的示意图。

具体实施方式

[0048] 虽然本文已经示出和描述了本发明的各种实施方案,但是对于本领域技术人员而言显而易见的是,此类实施方案仅作为实例提供。在不脱离本发明的情况下,本领域技术人员可以想到许多变化、改变和替换。应该理解,可以采用在此描述的本发明实施方案的各种替代方案。

[0049] 如本文所用的术语“样品”通常指生物组织、细胞或液体。此类样品可以包括但不限于:痰、血液(例如,全血)、血清、血浆、血细胞(例如,白细胞)、组织、乳头抽出物、芯针或细针活检样品、含细胞的体液、自由漂浮的核酸、尿液、腹膜液和胸腔积液、或来自其的细

胞。样品可以是无细胞(或不含细胞)样品。样品可以包括一个或多个细胞。

[0050] 如本文所用的术语“核酸”通常指任何长度的核苷酸的单体或聚合形式,为脱氧核糖核苷酸或者核糖核苷酸,或者其类似物或者变体。核酸分子可以包括一个或多个未修饰或修饰的核苷酸。核酸可以具有任何三维结构,并且可以执行任何已知或未知的功能。以下是核酸的非限制性实例:核糖核酸(RNA)、脱氧核糖核酸(DNA)、基因或基因片段的编码或非编码区、由连锁分析定义的基因座(基因座)、外显子、内含子、信使RNA(mRNA)、转运核糖核酸(tRNA)、核糖体RNA、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微RNA(miRNA)、核酶、互补脱氧核糖核酸(cDNA)、重组多核苷酸、支链多核苷酸、质粒、载体、任何序列的分离DNA、任何序列的分离RNA、核酸探针和引物。核酸可以包含一个或多个修饰的核苷酸,例如甲基化的核苷酸和核苷酸类似物,例如肽核酸(PNA)、吗啉代和锁核酸(LNA)、乙二醇核酸(GNA)、苏糖核酸(TNA)、2'-氟,2'-OMe和硫代磷酸化DNA。核酸可以包括选自腺苷(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U)或其变体的一个或多个亚基。在一些实例中,核酸是DNA或RNA或其衍生物。核酸可以是单链或双链的。核酸可以是环形的。

[0051] 如本文所用的术语“核苷酸”通常指核酸亚基,其可以包括A、C、G、T或U或其变体或类似物。核苷酸可以包括可以并入生长核酸链中的任何亚基。此类亚基可以是A、C、G、T或U,或对一个或多个互补A、C、G、T或U特异的或与嘌呤(即A或G、或其变体或类似物)或嘧啶(即C、T或U、或其变体或类似物)互补的任何其它亚基。亚基可以使个别的核酸碱基或碱基组(例如,AA、TA、AT、GC、CG、CT、TC、GT、TG、AC、CA或其尿嘧啶对应物)能够分解。

[0052] 概要

[0053] 如本文所述的可转化标记组可以用于各种有用的情境中。例如,它们可用于原位赋予标记多样性水平,而不需要起始标记试剂中为此多样性水平。另外,它们可以用作复制循环的指示物,用作为随机区分标记,用作用于创建高度多样化的条形码文库的过程的一部分,用作某些类型的分析中的独特分子标识符分子,例如分子计数应用(例如,用于表达分析以提高对核酸中变体调用的置信度;例如,通过计数支持给定等位基因分子,并通过在具有共同分子标识符的短读段中取得共有区以提高测序准确度以及确定拷贝数变化),用作用于追踪群体中的谱系的追踪标记,例如用于系统发育重建的追踪标记,用作酶活性的指标,或多个分子之间的接近度或相互作用的指标。阅读本公开后,各种其它用途对于本领域技术人员而言将是显而易见的。

[0054] 在应用中,这些可转化标记部分可以用作用于分子定量过程中的个别分子标记。在许多应用中,独特的分子标识符标记已被用于标记个别分子,以便能够单独标识单独的起始分子以定量它们。在一个实例中,可能需要能够定量来自细胞或其它样品中给定基因的单信使核糖核酸(mRNA)分子的数量,以便能够测量所述基因一般地或者响应于一些刺激如候选药物或其它环境刺激的表达水平。在这种情况下,细胞内的从给定基因表达的mRNA分子的单独拷贝可以随机地用不同的核酸条形码分子标记,使得每个单独分子具有与其附接的独特标识符序列,或独特分子标识符(“UMI”)。因为从给定基因表达的每个起始mRNA分子现在具有与其附接的UMI序列,所以它可以经受多轮扩增,而不丢失关于起始分子数目的信息,例如与mRNA附接的每个不同UMI表示单独的起始分子。扩增允许例如使用靶向所关注基因或UMI的核酸阵列、核酸测序或其它方法来极大地简化检测。在扩增之后,对存在的不同UMI的检测允许推断给定基因的起始mRNA分子的数目,从而推断所述基因的表达。

UMI的这种类型的使用的实例在例如“Counting Absolute Numbers of Molecules Using Unique Molecular Identifiers”,Kivioja等人,Nature Methods 9,72-74(2012)中描述,该文献的全部内容出于所有目的以引用方式整体并入本文。

[0055] 尽管在某些情况下是有用的,但应理解的是,这些方法可以用于具有相对少量分子的样品,因为随着样品中分子数量的增加,必需的标记文库可能迅速增加复杂性和成本。换言之,随着要计数的分子数量增加,可能导致可能需要应用于所述样品的不同标记部分的数量必要和显著增加,以获得独特的分子标记。同样,随着待分析不同基因的数量增加,它增加了所述UMI文库所需的复杂性。此外,用于这些不同文库的创建、连接或其它附接、复制等的生物化学不能针对任何特定序列进行优化,而是可以针对平均序列进行优化,这通常会导致对所使用的实际序列都不进行优化。

[0056] 然而,如本文所述,可以使用如上所述在标记部分内引入可转化部分的相对简单且恒定的标记结构,以便将多样性原位赋予给标记分子。具体而言,可以采用具有单一但可转化的标记部分的标记部分,其中经标记的分子的后续处理将以随机或半随机方式转化所述标记分子,以将多样性赋予给样品中的标记组,其中这种多样性水平原来并不存在。这允许在分析中使用少量、几个或甚至单个可转变标记部分代替先前过程所需的更大量独特条形码分子,因为给定分析所需的多样性将在标记的随机或半随机变换时引入。

[0057] 在上述表达分析实例的情况中,代替单独且随机地附接到单独的mRNA分子的核酸UMI的不同文库,可以将单个、若干个或相对少量的可转化标记部分附接至不同的mRNA。在单轮复制后,给定基因的mRNA的每个拷贝可以与附接的随机或半随机序列标签一起复制,此凭借标记的复制过程的随机性或半随机性,可以产生每个起始分子的不同经标记的复制物。然后通过检测并计数不同的经转化标记部分的数量,可以推断起始mRNA分子的数量。

[0058] 可转化标记部分

[0059] 本公开提供了包括可转化元件的“标记”部分,所述可转化元件可以在它们与意图标记的部件相关联之后转换成期望的标记结构。在一些情况下,这些可转化结构可具有共同结构,但在转化过程后转化成不同的结构集合,例如,具有单一结构的标记的群体转化为不同结构的不同群体。在一些情况下,将这些可转化组转化成随机或半随机所得部分,以将多样性赋予给可经鉴定并用于表征例如反应、其反应物和/或其产物的经标记分子。

[0060] 尽管就核酸、多核苷酸等而言描述了可转化标记部分的具体实例,但应了解可采用其它可转化标记部分。例如,可转化部分可以包含核酸(例如,核苷酸、寡核苷酸、多核苷酸,包括核糖核苷酸和脱氧核糖核苷酸,以及它们的类似物,例如二脱氧核糖核苷酸、简并核苷酸等)、多肽(例如,蛋白质、酶、多肽、寡肽等)、碳水化合物(例如,葡聚糖、淀粉、纤维素等)、有机化合物、荧光基团、发色团、胶体元件、颗粒、珠粒等,其中第一结构可以在实施过程操作时转变为一个或多个第二结构,以获得反应中标记部分的多样性。

[0061] 在一个实例中,可转化部分可包含可转化寡核苷酸序列,其中在复制、翻译、转录或其它转化过程期间,所述序列中此类序列的这些核苷酸(在本文中也简称为“碱基”)可原位转化为经改变的或可变的所得物质。各种不同的机制可以用于原位转化序列中的核苷酸,包括例如使用简并碱基如互补碱基配对可以改变的碱基,基于序列区段的转化如去除和替换序列区段,以及个别碱基或碱基序列的化学转化如碱基的氧化脱氨基或其它化学修饰(例如用亚硝酸或烷化剂处理),暴露于电离辐射,用修饰碱基的酶(例如,腺苷脱氨酶、胞

嘧啶脱氨酶、黄嘌呤氧化酶、编辑体)处理,用改变引起模板驱动或非模板驱动插入或添加的碱基配对或过程的酶如M-MLV逆转录酶、末端脱氧核苷酸转移酶处理,或用催化它们自己插入的转座子进行处理。

[0062] 在某些情况下,所述可转化核苷酸可以包括经过随机或半随机“补体”并入的核苷酸,所述核苷酸或碱基在本文中也可以被称为可变补体核苷酸或碱基。具体而言,在寡核苷酸复制、转录或翻译过程中,通过所涉及的酶或酶体系的正确处理通常响应于遇到给定核苷酸或核苷酸组而并入单一类型的互补结构单元。例如,使用典型的正确DNA聚合酶如复制给定DNA链的DNA聚合酶的模板驱动的组合酶介导的核酸复制在遇到一种类型的核苷酸时,将通常并入单一特定类型的互补核苷酸。例如,当在序列中遇到嘌呤腺苷(A)或嘧啶基(G)核苷酸时,聚合酶通常将分别在序列中并入嘧啶胸苷(T)或胞嘧啶(C)核苷酸作为互补碱基,并且反之亦然。因此,由这些碱基组成的典型条形码序列通常可以基本上每次都由正确的聚合酶复制到相同的互补结构中。然而,根据本公开的一些方面,条形码区段可以包含能够具有随机或半随机补体的一个或多个核苷酸,使得当复制时,它们响应地产生随机或半随机的复制物序列。如将理解,同样可以通过使用低保真聚合酶向常规碱基或例如具有大于0.1%的取代率以及在一些情况下大于1%、大于5%或甚至更高的取代率的非校正读码酶来驱动随机引入。此类低保真聚合酶的实例包括例如家族Y聚合酶、跨损伤合成聚合酶、大肠杆菌聚合酶IV和V、人聚合酶 ζ 、 η 、 τ 、 κ 和Rev1,以及具有降低或校正读码能力的聚合酶的修饰形式,例如phi29突变酶,例如phi29N62D和其它非校正读码突变体(例如,如Korlach等人,Methods in Enzymology,Real-Time DNA Sequencing from Single Polymerase Molecules, (2010) 472:431-455中所述),pfu-Pol的低保真突变体(参见例如Biles等人,Nucl.Acids Res.32(22):e176 2004),来自非洲猪瘟病毒(ASFV)的病毒聚合酶如DNA聚合酶X(Pol X)。此类聚合酶可以单独使用或与本文其它地方所述的可转化碱基组合使用,或者可以与这些聚合酶显示较高碱基取代速率的特定序列基元结合使用。在一些情况下,可以使用单一类型的聚合酶来实现标记序列的转化。相反,在其它情况下,对不同简并碱基具有不同反应的不同聚合酶的混合物可在单一反应混合物中组合以增加多样性或以其它方式更好地控制转化过程。

[0063] 已经描述了多种核苷酸或核苷酸样部分,其在聚合酶反应中复制时具有随机或半随机补体,例如在复制期间,它们可以与在所产生的或“复制物”链中的两个或更多个不同的核苷酸互补。为了便于讨论,这些碱基在本文中被称为简并碱基。例如,许多碱基能够以足够无偏差的方式与聚合酶相互作用,从而至少在所述碱基处提供复制的双向简并性,即,能够响应于和作为这些碱基的“互补”而并入两个或更多个不同的核苷酸。在一些情况下,可以使用导致聚合酶以至少2-向、至少3-向或甚至4-向简并性水平并入的碱基。通常,如本文所用,简并性通常是指在特定反应条件下(例如,使用特定聚合酶以及特定核苷酸、缓冲液和盐浓度等)的碱基将显示无偏并入,例如将响应于在至少1%的其遇到此类简并性碱基的情况中,至少5%的所述情况中,在一些情况下此时间的至少10%,在一些情况下此时间的至少20%,在一些情况下此时间的至少30%,在一些情况下此时间的至少40%,在一些情况下此时间的至少50%时的简并性碱基,而并入不同的核苷酸。例如,仅仅为了便于讨论,如果可转化核苷酸导致例如此时间的至少5%时的不同碱基并入,例如如果其此时间的至少5%时并入A,而在此时间的其余95%时并入G,则可转化核苷酸可表现出双向简并性。

[0064] 如上所述,在一些情况下,可以调节聚合反应混合物内各种核苷酸的浓度以在给定的反应中提供期望的简并率。例如,为了使不同碱基的并入均等,可以调节它们的相对浓度以增加一个碱基的并入率,同时降低另一个的相对并入率以响应给定的可转化碱基或简并碱基。因此,可以通过提供可产生各自在特定简并碱基处的等同并入率的浓度的各种核苷酸试剂,来在给定的简并碱基处甚至提供双向、三向或四向简并性。

[0065] 如将理解的那样,导致显示出上述简并性的“补体”并入的碱基将被表征为随机或半随机的。例如,在一些情况下,对补体碱基的子集(例如仅嘌呤或仅嘧啶)的限定的偏倚可以被鉴定为半随机的,其中给定碱基的完全无差别的互补配对可被视为完全随机的。

[0066] 此类可转化核苷酸的实例可以包括例如肌苷、脱氧肌苷、脱氧黄嘌呤、2'-脱氧水粉菌素、2'-脱氧鸟苷、5-硝基吡啶、3-硝基吡啶、N6-甲氧基-2,6-二氨基嘌呤、6H,8H-3,4-二氢嘧啶并[4,5-c][1,2]噁嗪-7-酮,以及前述各种的非脱氧(或核糖)形式。对于这些碱基,可以取决于所使用的可转化碱基而使用这些碱基中的一种或多种作为可转化标记序列内的可转化碱基来提供各种并入模式。例如,一些可转化碱基(例如肌苷)虽然显示简并性水平,但仍然表现出更强的与一种类型的核苷酸(如胞嘧啶核苷酸(C))互补的偏好并因此驱动所述类型的核苷酸的并入。其它可转化的碱基如5-硝基吡啶可显示更平衡的4-向简并性,例如,作为响应并入四种天然碱基如AGCT中的任一者的能力。在另一个实例中,脱氧黄嘌呤在显示4-向简并性的同时,在某些情况下表现出更强的与嘧啶核苷酸(例如T或C)互补的偏好。

[0067] 包含这些简并碱基中的一个或多个的标记或条形码序列可以通过将标记序列区段附加到所关注的靶核酸或核酸片段来采用。在聚合酶介导的靶核酸复制后,所述标记组也将被“复制”,但是所述复制可以将随机或半随机互补碱基并入简并碱基位置以产生附加于复制物分子的独特或半独特标记。因此,简并碱基的单个序列可以在聚合酶复制时产生一定量的和潜在大量的标记序列。

[0068] 如将理解,除非简并碱基之外,所述可转化寡核苷酸标记序列还可包括简并碱基,或者它们可包含所有简并碱基。同样,所包含的简并碱基可以具有双向简并性、三向简并性或四向简并性,和/或可以具有某些偏好而不论其简并性程度。在一些情况下,简并碱基可以散布有非简并碱基,或者双向简并碱基可以随机或甚至已知或预定模式散布有三向和/或4-向简并碱基。使用已知或预定的模式可以允许凭借标记序列对反映所包括的简并碱基和/或非简并碱基的模式已知或预定模式的反映来轻易鉴定标记序列。

[0069] 在一些情况下,并且取决于期望的可能多样性水平,给定标记序列中简并碱基的数量可以在1至100个或更多、1至20个可转化碱基、1至10个可转化碱基、1至5个可转化碱基,或前述范围内的任一者中的任何居间数量的可转化碱基。

[0070] 此外,如前所述,这些可转化碱基可以是在序列区段内连续的,或者它们可以散布有非简并碱基。此类散布的碱基可以将个别可转化碱基与标记序列内的其它可转化碱基分开,或者它们可以将可转化碱基的对、组或子集与可转化碱基的其它个体、对、组或子集分开。这些散布的可转化碱基同样可以存在于序列中的个别碱基中,或作为标记序列中不可转化碱基的连续对、组或子集。

[0071] 如将理解,可以通过选择标记序列中简并或可转化碱基的数量以及每种此类碱基的简并性水平来选择可转化标记序列的潜在多样性水平。此外,如上所述,可以通过提供具

有不同可转化核苷酸序列的可转化标记区段组,例如通过将顺序改组为标记分子文库中使用的简并碱基,来引入附加的多样性水平。此类选择可以通过许多要求或期望中的任何一个来激发,包括任何给定应用所需或期望的多样性水平,例如分子计数应用中待标记的预期分子的数量,以及希望能够从较高级别的签名(例如由标记序列的半随机性产生的签名)来辨识标记序列。例如,可以选择标记序列中的可转化碱基以反映所得序列的一般模式,例如将嘌呤或嘧啶特异的可转化碱基定位在某些位置,以及散布在其它可转化碱基之间的非简并碱基。通过并入半随机可转化碱基或整体序列的模式,可能能够更好地鉴定可能由标签序列产生的序列。

[0072] 在其它情况下,可转化标部分可以包括整体转化而不是通过构建块基础在结构单元上转化的区段。例如,在一些情况下,可以提供原始标记部分,其提供插入取代序列区段的靶标,所述取代序列区段在从共同原始标记区段开始时产生期望的多样性水平。此类方法的实例可以包括使用可转化序列区段可以被转化(例如全部或部分改变或替代)的定向诱变机制。例如,标记序列区段可以形成靶向序列替换系统的靶序列的基础。例如,可以使用例如与CRISPR相关的RNA指导的DNA内切核酸酶如Cas9来靶向可转化标记序列区段,所述DNA内切核酸酶能够通过所述指导RNA靶向特定序列并切除所述序列(参见例如,Genome Engineering Using the CRISPR-Cas9 System,Ran等人,Nature Protocol,(2013),8(11):2281-2308)。一旦被切除,就可以通过使用例如互补侧翼区域来容易地插入替代序列区段,所述互补侧翼区域允许使用例如常规连接生物化学物或采用例如非同源末端连接(NHEJ)或同源性定向修复(HDR)来在切除先前的可转化序列区段的时间点处连接新的序列区段。如将理解的,可以以类似的方式使用多种其它靶向编辑核酸酶,例如包括例如锌指核酸酶(ZFN)和转录激活因子样效应物核酸酶(TALEN),参见例如Porteus MH,Baltimore D.Chimeric nucleases stimulate gene targeting in human cells.Science.2003;300:763;Miller JC等人,An improved zinc-finger nuclease architecture for highly specific genome editing.Nat.Biotechnol.2007;25:778-785;Sander JD等人,Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA).Nat.Methods.2011;8:67-69;Wood AJ等人,Targeted genome editing across species using ZFNs and TALENs.Science.2011;333:307;Christian M等人,Targeting DNA double-strand breaks with TAL effector nucleases.Genetics.2010;186:757-761;Zhang F等人,Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription.Nat.Biotechnol.2011;29:149-153;Hockemeyer D等人,Genetic engineering of human pluripotent cells using TALE nucleases.Nat.Biotechnol.2011;29:731-734。

[0073] 在一些情况下,可转化标记部分(例如可转化序列)可以包含对化学诱变或UV诱变更敏感和/或经历化学诱变或UV诱变的序列,以驱动所述标记区段的转化或甚至所关注的所述序列区段的转化,使得此类诱变产生用于给定分析的足够多样性。例如,如果计数相同的序列,则可以以预期影响每个和每一分子的方式诱变这些序列。对这些序列的后续分析可以允许根据不同突变序列的数量来确定起始分子的数量。此类诱变在某些情况下可以再次使用例如靶向或引导寡核苷酸探针进行靶向,或者它可以是随机的,例如非靶向的。

[0074] 结构

[0075] 本文还提供了包含寡核苷酸的组合物,所述寡核苷酸包含在本文其它地方描述的标记寡核苷酸序列或区段作为其序列的一部分。这些组合物可包含单独的这些寡核苷酸,或这些寡核苷酸与其它组分(包括但不限于缓冲液、盐、反应物、酶、样品组分,例如细胞、组织或其它样品成分,固体支持物如颗粒、珠粒、水凝胶珠粒、阵列表面等)的组合。

[0076] 本文所述的这些标记部分可以在其较大结构内包含附加的元件,例如以赋予附加的功能给标记部分。例如,此类结构可以包括可以在标记部分的应用中提供功能的附加元件或者用于所得经标记的反应物的附加的元件。

[0077] 举例来说,所描述的标记部分可以提供在促进它们附接或附加到其它反应物的结构内。例如,它们可以包括可促进与其它基团、用于亲和附接的亲和结合部分(例如抗生物素蛋白,链霉抗生物素蛋白,生物素等)进行化学偶联的可激活化学基团,或者它们可以包括允许此偶联的其它机制。

[0078] 举例来说,如上所述的寡核苷酸标记部分可以包含附加的序列区段,所述区段允许它们附接或附加到其它序列区段,例如靶序列区段。如将理解,标记部分(例如标记寡核苷酸)与另一物质(例如靶核酸序列或其部分)的附接或附加包括多种不同的附接或附加方法。例如,在一些情况下,标记寡核苷酸与另一序列区段的附接可包括经由例如附接至另一序列区段的3'或5'端的连接,或通过与另一序列区段的共价交联或其它侧链附接进行的共价附接。

[0079] 另外,附接可以包括例如通过亲和偶联(例如通过标记寡核苷酸的一部分与靶向序列区段的杂交)或通过其它分子物质的其它亲和机制(例如通过抗体/抗原偶联、抗生物素蛋白或链霉抗生物素蛋白/生物素偶联)或通过与特定缔合基团(例如,缔合肽等)缔合等而非共价附接到另一个序列区段。

[0080] 在其它方面,标记寡核苷酸的附接可以通过包含在标记寡核苷酸结构内的引物序列的引发和延伸,使得靶向序列区段的补体附接到延伸的引物/标记寡核苷酸。如将理解的是,标记寡核苷酸和序列区段在被称为附接时将互换地指任一或两个序列区段的补体或复制物。另外,并且如将理解的,附接将包括标记部分与序列区段的附接,以及标记寡核苷酸与序列区段的补体的附接,以及原始标记寡核苷酸的补体与序列区段的附接,其补体或此类补体的另一补体(即,原始序列区段的复制物)的附接。

[0081] 所述附加的序列区段可以包含用于通过杂交与靶序列附接的杂交探针,或者它们可以包括能够退火至靶序列区段的引物序列,使得引物区段的延伸将靶序列的补体复制成包括标记序列的延伸产物,其此出于本公开的目的而构成如本文所用的标记分子至靶标的附接或附加。

[0082] 在一些情况下,这些标记部分可以包括序列悬垂和/或桥接或夹板序列,以促进标记部分与给定序列的连接或其它偶联。

[0083] 可以构建引物或杂交序列以退火至靶序列内的特定序列,或者它们可以构建以退火至靶序列的随机部分,例如作为通用引物,例如随机n聚体序列,以使得不同的标记寡核苷酸上的不同引物序列可以在靶序列内的不同位置进行引发。

[0084] 图1中说明了标记寡核苷酸及其在标记靶序列的区段中的用途。如图所示,寡核苷酸100包含在序列内包含一个或多个简并碱基(Z)的标记区段102。一个或多个简并碱基可以位于寡核苷酸100的给定区域中。在一些情况下,寡核苷酸100包含各自具有一个或多个

简并碱基的至少2、3、4、5、6、7、8、9或10个区域。如上所述,虽然示出为包括许多连续的简并碱基,但是所述标记序列在一些情况下可以在其序列内包含一个或多个非简并碱基。类似地,虽然示出为10聚体标记序列或包含10个简并碱基的标记序列,但如本文其它地方所述,标记序列可以更长或更短,并且包含更多或更少的简并碱基。

[0085] 所述标记序列可以包括至少2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个简并碱基。所述标记序列可以包括至少2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个核苷酸。

[0086] 寡核苷酸100也被示出为在其3'末端包含引发序列104,以用于退火至靶样品核酸片段106并引发所述靶样品核酸片段的复制。如将认识到的,此引发序列可以特异于所关注的靶序列内的序列,其可以是随机引发序列,例如n聚体,或者其可以靶向特定类型的序列区段,例如以退火至mRNA分子的聚腺苷酸化末端(聚A尾)或其它常见序列类型。同样如图所示,寡核苷酸100可以包括附加的核酸区段,例如附加的条形码区段108、测序引物区段110以及测序仪特异性衔接区段(未示出)。例如,所述寡核苷酸100可以包括用于大规模平行测序(例如Illumina测序)的流式细胞序列。

[0087] 如图所示,标记寡核苷酸100退火至所关注的靶序列,并用于引发所述靶序列的延伸和互补复制,从而导致所述标记寡核苷酸100被附加至所述靶序列106的所述互补复制物区段120。在后续复制标记寡核苷酸100中的经标记区段102时,标记区段102的可转化性质复制到衔接至原始靶序列区段106的拷贝106'的随机或半随机序列区段122中。尽管源自标记寡核苷酸100中的相同标记寡核苷酸序列区段102,但所得到的随机或半随机区段122对于样品中的不同分子可以不同。

[0088] 如上所述,可以包括许多其它结构以及上述的可转化标记序列区段。例如,在一些情况下,可转化标记区段可以与其它标记或条形码结构一起被包括在寡核苷酸结构中。特别有用的条形码寡核苷酸的实例描述于例如公开的美国专利申请公开号2014/0378345、2014/0228255、2015/0376700、2015/0376605和2016/0122817中,该等专利的全部公开内容以引用方式整体并入本文。

[0089] 这些条形码可以有多种结构。在一些情况下,条形码是接头的一部分。通常,“接头”是用于使条形码能够附接到靶多核苷酸的结构。接头可包含例如条形码,可相容用于与靶多核苷酸连接的多核苷酸序列,以及功能性序列如引物结合部位和固定区。在一些情况下,接头是分叉接头。

[0090] 在一些情况下,这些条形码可以用于标记已经共分配为例如亚微米滴(纳升或皮升标度液滴)的序列区段片段。此类序列区段可以源自样品核酸的溶液或来自与用于标记的条形码共分区的单独细胞。另外地或可替代地,附加的标记或条形码结构可以包括反映核酸源自的特定样品的单独条形码,以便允许随后在汇集的测序运行中区分来自不同样品的核酸。

[0091] 在一些情况下,本文所述的标记寡核苷酸(包括任何附加的序列区段)可作为较大寡核苷酸文库的元件提供。例如,将标记寡核苷酸区段整合到条形码寡核苷酸文库中,例如在公开的美国专利申请公开号2014/0228255中描述的那些文库,该专利申请公开出于所有目的以引用方式整体并入本文。

[0092] 可以使用多核苷酸合成的随机方法,包括DNA合成的随机方法来产生条形码寡核

核苷酸文库。在随机DNA合成期间,可以将A、C、G和/或T的任意组合添加到偶联操作中,使得偶联操作中的每种碱基与所述产物的子集偶联。如果A、C、G和T以相同的浓度存在,则大约四分之一的所述产物将并入每种碱基。连续的偶联步骤和偶联反应的随机性使得能够产生 4^n 个可能的序列,其中n是多核苷酸中碱基的数目。例如,长度为6的随机多核苷酸文库可以具有 $4^6=4,096$ 个成员的多样性,而长度为10的文库可以具有1,048,576个成员的多样性。因此,可以生成非常大且复杂的文库。这些随机序列可以用作条形码。任何合适的合成碱也可以使用。在一些情况下,可以改变每个偶联操作中包含的碱基以合成优选的产物。例如,每个偶联操作中存在的碱基的数量可是1、2、3、4、5、6、7、8、9、10或更多。在一些情况下,每个偶联操作中存在的碱基的数量可是1、2、3、4、5、6、7、8、9、10或更多。在一些情况下,每个偶联操作中存在的碱基的数量可以是小于2、3、4、5、6、7、8、9或10。个别碱基的浓度也可以被改变以合成优选的产物。例如,任何碱基可以以另一种碱基浓度的约0.1、0.5、1、5或10倍的浓度存在。在一些情况下,任何碱基可以以另一种碱基浓度的至少约0.1、0.5、1、5或10倍的浓度存在。在一些情况下,任何碱基可以以另一种碱基浓度的小于约0.1、0.5、1、5或10倍的浓度存在。取决于应用,随机多核苷酸序列的长度可以是任何合适的长度。在一些情况下,随机多核苷酸序列的长度可以是4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个核苷酸。在一些情况下,所述随机多核苷酸序列的长度可以是至少4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个核苷酸。在一些情况下,所述随机多核苷酸序列的长度可以是小于4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在某些情况下,所述文库由成员的数量定义。在一些情况下,文库可包含约256、1024、4096、16384、65536、262144、1048576、4194304、16777216、67108864、268435456、1073741824、4294967296、17179869184、68719476736、 2.74878×10^{11} 或 1.09951×10^{12} 个成员。在一些情况下,文库可包含至少约256、1024、4096、16384、65536、262144、1048576、4194304、16777216、67108864、268435456、1073741824、4294967296、17179869184、68719476736、 2.74878×10^{10} 或 1.09951×10^{12} 个成员。在一些情况下,文库可包含小于约256、1024、4096、16384、65536、262144、1048576、4194304、16777216、67108864、268435456、1073741824、4294967296、17179869184、68719476736、 2.74878×10^{10} 或 1.09951×10^{12} 个成员。在某些情况下,所述文库是条形码文库。在一些情况下,条形码文库可以包含至少约1000、10000、100000、1000000、2500000、5000000、10000000、25000000、50000000或100000000个不同的条形码序列。

[0093] 这些随机条形码文库也可以包含其它多核苷酸序列。在一些情况下,这些其它多核苷酸序列在性质上是非随机的,并且包括例如引物结合位点、用于产生分支接头的退火位点、固定序列和使得能够用靶多核苷酸序列退火并因此对所述多核苷酸序列进行条形码编码的区域。

[0094] 在许多情况下,可以将此类文库提供为系至珠粒或颗粒以用于有效递送文库元件。例如,在一些情况下,珠粒被提供为具有衔接至它们的上述标记寡核苷酸结构。在此类情况下,个别珠粒可以包含包含第一区域的寡核苷酸,所述第一区域包含可转化的寡核苷酸序列,例如作为可转化序列或可转化核苷酸序列。如上所述,此第一区域对于衔接至给定珠粒或珠粒群体中的所有寡核苷酸可以是共同的。或者,第一区域可以在不同珠粒或不同珠粒群体之间变化。所述寡核苷酸还可以包含附加的区域或序列区段,例如第二、第三、第四等区域,其中此类附加的区域可以包括可变区,例如在不同珠粒上的寡核苷酸之间具有

序列变化的可变区。此类可变区可以包括在不同珠粒上的寡核苷酸之间不同的条形码序列。通过提供具有给定条形码序列区段的珠粒,但是其中此类条形码序列在其它珠粒上不同,可以通过仅在个体基础上分配珠粒而容易地将不同条形码划分成不同的分区。用于此类条形码序列的分配方法的实例在例如公开的美国专利申请公开号2014/0378345和2015/0292988中进行了描述,该等专利的每一者的全部公开内容出于所有目的以引用方式整体并入本文。

[0095] 分配方法可以包括使包含条形码序列悬浮液的水性流体流入包含分配流体的液滴生成结中。在液滴产生的窗口期间,这些条形码序列可以以变化小于30%的频率流入液滴生成结。所述方法还可以包括在液滴生成窗口期间分配分配流体中的条形码序列。另一种分配方法可以包括在水性流体中提供凝胶前体,并使具有凝胶前体的水性流体流过流体导管,所述流体导管流体地连接至包含分配流体的液滴产生结。所述分配流体可以包含凝胶活化剂。所述方法还可以包括在分配流体中形成水性流体的液滴,其中在这些液滴内所述凝胶活化剂接触所述凝胶前体以形成凝胶微囊。

[0096] 可以在这些寡核苷酸序列内提供其它可变区,例如用作随机 n 聚体引发序列,其中这种可变性可以在给定珠粒上的寡核苷酸序列之间、个别珠粒之间或珠粒群体之间存在。同样地,所述寡核苷酸序列可以包括与给定珠粒上、两个或更多个珠粒之间或在珠粒群体之间的寡核苷酸共同的其它共同区域,例如共同的引物序列,例如序列特异性引物序列、附接序列等。

[0097] 标记寡核苷酸可通过例如可逆或可切割的连接而附接至这些珠粒,使得可在施加刺激(例如化学、热、光学或机械刺激)时将寡核苷酸与珠粒分离。这种可切割连接的实例包括例如在公开的美国专利申请公开号2014/0228255和2014/0378345中所述的那些,该等专利中的每一个以引用方式整体并入本文。在一些情况下,这些珠粒本身可以包含可降解结构,例如可降解聚合物或可降解以进一步促进寡核苷酸从珠粒释放到基本均匀的反应混合物中的水凝胶。参见例如美国专利申请公开号2014/0228255和2014/0378345,该等专利中的每一个以引用方式整体并入本文。

[0098] 过程

[0099] 尽管以上详细描述为用于定量核酸分子(例如用于表达分析的mRNA),但下面的实例提供了用于在此类表达分析中采用可转化标记部分的一种类型的特定过程的一个详细实例。

[0100] 在用于评估一种或多种基因在细胞培养物中的表达的示例性过程中,可以使用过程并使用本文描述的可转化标记部分分别分析这些细胞的内容物,所述过程例如如在美国专利申请公开号2015/0376609中所述的过程,该专利出于所有目的以引用方式整体并入本文。

[0101] 分析来自细胞的核酸的方法包括将来自个别细胞的核酸提供到单独分区中;产生来自单独分区内的核酸的一个或多个第一核酸序列,其中所述一个或多个第一核酸序列具有与其附接的寡核苷酸,所述寡核苷酸包含共同的核酸条形码序列;产生所述一个或多个第一核酸序列或衍生自所述一个或多个第一核酸序列的一个或多个第二核酸序列的表征,所述一个或多个第二核酸序列包含所述共同条形码序列;以及至少部分地根据在所生成的表征中存在所述共同核酸条形码序列而将所述一个或多个第一核酸序列或一个或多个第

二核酸序列鉴定为来源于所述个别细胞。

[0102] 例如,这些过程可用于分析和定量通常或者响应于某些刺激而在个别细胞内的基因表达。

[0103] 在至少一种方法中,可以采用一组标记寡核苷酸,其在它们的序列中包含共同的但可转化的标记部分。在标记一个或多个遗传元件(例如基因或基因片段,在本文中称为表达产物)的个别表达拷贝后,可以转化这些共同标记元件,使得对于每个个别表达的分子,在所表达的分子的标记拷贝中产生独特的或基本上独特的标记元件,以允许独特鉴定此起始表达拷贝。通过计数可独特鉴定的表达拷贝,可以推断给定基因的表达拷贝的数量/量。

[0104] 在图2中提供的流程图中示意性地示出了简化的过程。具体而言,在阶段202,细胞可以以信使RNA (mRNA) 的形式表达一种或多种基因。在阶段204,由给定细胞表达的各种个别mRNA分子经受用包含可转化标记区段的寡核苷酸序列进行标记。然后在阶段206处理个别经标记mRNA分子以将可转化标记区段转化为对于每个起始的经标记mRNA分子例如作为经标记的cDNA分子独特的新标记区段。然后可以在阶段208使这些所得cDNA分子经过扩增过程,同时将所得扩增材料的归属保存至其原始起始分子(基于转化的标记区段)。然后可以在阶段210对扩增的cDNA分子及其相关标记进行测序。根据所述序列数据,可以在阶段212处鉴定给定基因的序列,并且借助与所述基因的序列相关的不同标记的数量,来鉴定或至少推断所述基因的起始表达分子的数量。

[0105] 在示例性过程的更详细讨论中,使用例如微流体液滴或乳液生成系统将细胞悬浮液共分配成油乳液中的水性液滴,例如如在2015年6月26日提交的美国专利公开号2015-0376609中所述,该专利以引用方式整体并入本文。微囊也可以共分配到单独液滴中,其中这些微囊以可释放的方式携带与微囊偶联的所述标记寡核苷酸,例如以允许在对液滴的内容物施加刺激时释放标记寡核苷酸。如上所述,所述寡核苷酸通常将包含可转化标记序列区段以及其它功能序列区段,例如其它条形码区段、引发区段、衔接区段等。在此特定实例中,所述标记寡核苷酸还包含条形码区段,所述条形码区段对于给定微囊上的所有标记寡核苷酸将是共同的,但其可在不同微囊中变化。通过将单个细胞与单个微囊共分配,此条形码可以用作地址部分以标记和鉴定从个别细胞获得的所有核酸。

[0106] 在共分配之后,可以例如通过在液滴内包含裂解剂(例如去污剂、离液剂或其它裂解剂)来裂解这些液滴内的细胞,或者可以通过施加其它刺激(例如机械、热、电等)来裂解它们。一旦裂解,包括来自表达基因的信使RNA的细胞内容物将被释放到液滴中。存在于共分配的微囊上的标记寡核苷酸也被释放到液滴中并且可以被配置成与mRNA分子特异性相互作用,例如通过使用聚T序列区段作为针对mRNA的聚A尾部的捕获/引发区段(例如作为图1中的区段104)。

[0107] 整个过程在图3和图4中示意性地示出。如图3所示,在微流体通道网络300中,将个别所关注的细胞与本文所述的带有标记寡核苷酸的个别微囊一起共分配。细胞悬浮液320通过第一通道区段302到达第一混合接点,在第一混合接点处,其与具有标记寡核苷酸的微囊或珠粒322的流动悬浮液共同混合,以从另一通道区段306进入所述接点304。

[0108] 微囊悬浮液还可以包括裂解剂,一旦细胞被分配,所述裂解剂将与细胞混合作用于细胞。这些细胞和微囊以允许相邻微囊和相邻细胞之间有足够空间的速率流动,以便增加个别细胞与个别微囊共分配的可能性。将细胞和微囊324的共混合悬浮液驱入液滴生

成308接点或分配接点处,在那里它们通过从侧面通道310和312进入的油流同轴流动汇集,使得水性共混悬浮液的个别液滴326在出口通道区段314中的流动油流中形成。

[0109] 一旦共分配到液滴中,暴露于裂解剂的个别细胞将其内容物如mRNA分子328释放到液滴中。同样,所述微囊也将其标记寡核苷酸(例如,标记寡核苷酸330)的有效载荷释放到液滴中。一旦存在于给定液滴内的均匀混合物中,所述标记寡核苷酸可用于标记来自这些细胞的核酸片段,以及如本文特别描述且如图4示意性说明的mRNA分子。

[0110] 如图4所示,从微囊释放的标记寡核苷酸402,凭借它们包含聚T序列404,将与从液滴内的个别细胞释放的mRNA 408的聚A尾部406退火。如前所述,所述标记寡核苷酸包括可转化核苷酸的共同序列作为可转化标记区段410,以及从给定微囊释放的所有寡核苷酸所共有的寡核苷酸的条形码区段412。这些条形码用于将所得的所产生片段归属为当所有核酸随后测序时来源于相同的细胞。

[0111] 所述标记寡核苷酸402退火至mRNA分子408后,存在于水性液滴内(并且与细胞悬浮液和/或微囊悬浮液之一一起引入)的逆转录酶用于使标记寡核苷酸402沿着退火的mRNA 408(如虚线箭头所示)延伸,以将mRNA的表达的基因部分414复制作为具有附接的标记寡核苷酸402的cDNA片段416。在许多情况下,所使用的逆转录酶将包括末端转移酶活性,其将向经标记的cDNA分子416的3'末端添加一系列胞嘧啶残基418。然后将具有一组3'鸟苷残基的模板转换寡核苷酸420退火至cDNA分子的末端并通过逆转录酶扩增,以便将附加的引发序列422附加到所得经标记的cDNA分子424的末端。此时,所述标记寡核苷酸可以包含与给定分区内,或者甚至在许多或甚至全部分区内的其它经标记mRNA复制分子相同的可转化标记序列410。

[0112] 如将理解,随着可转化标记片段的每次连续复制,将产生新的和不同标记的复制物。因此,在许多情况下,希望将用可转化序列标记的mRNA的复制循环的次数控制为例如单次或几次复制循环,例如1-4个循环,优选单个循环复制。通常,可以使用热循环操作来控制循环操作的次数,例如使给定的标记操作仅暴露于单次解链、退火和延伸操作,以确保从每个起始mRNA分子仅产生一个转化的经标记mRNA复制物。

[0113] 如所指出的,然后通过使用在其针对标记区段410中的可转化核苷酸的并入时无偏差的DNA聚合酶引发从所述附加的引物区域422的DNA聚合延伸,使经标记的cDNA分子424经受单轮复制,(同时保持此类核苷酸的持续合成能力)。在单轮复制后,所得到的经标记的复制分子428将包含新的转化的标记区段426,该区段与存在于相同反应混合物中的其它复制标记区段相比将是基本上独特的,因此提供对原始分子的一定程度的独特性,而不管是否以与所述反应混合物中所有其它分子相同的方式处理。

[0114] 如将理解,给定复制标记片段的独特性水平将取决于给定过程内的简并性水平、可转化碱基的数量和分子数量。在一些情况下,这些参数可能处于给定分析内的这些分子被标记为具有完全独特性的水平,例如在给定的反应混合物中没有重复转化的标记元件,而在其它情况下,独特性水平将处于可能会预期给定基因的重复拷贝(例如表达产物)可以预期用独特的标记元件相对于彼此标记,但可能不存在绝对独特性的水平。在其它情况下,独特性水平将是给定标记区段可以在单轮复制之后产生至少10个不同的经转化标记区段(例如,具有至少10个不同的核苷酸序列),至少50个不同的经转化标记区段、至少100个不同的经转化的标记区段,至少200个不同的经转化标记区段,至少300个不同的经转化标记

区段,至少400个不同的经转化标记区段,至少500个不同的经转化标记区段,至少1000个不同的经转化标记区段,或者在一些情况下,来自共同的起始可转化标记区段的至少1000个不同的经转化标记区段。

[0115] 在单轮复制之后,可以处理样品以去除包含可转化标记区段的原始标记寡核苷酸(包括原始标记的cDNA分子),以便防止剩余的可转化标记寡核苷酸参与随后的扩增操作和注射新的经转化的分子进入分析。这些寡核苷酸的去除可以通过多种方法进行。例如,在一些情况下,这些原始标记寡核苷酸可以包括,例如通过亲和纯化促进将其从反应混合物中去除的“处理”部分。此类处理可包括例如可与固体支持物结合的互补探针序列杂交的特定核酸序列,以从反应混合物中去除那些部分。或者,这些处理可以包括其它亲和结合试剂,例如生物素、抗生物素蛋白、链霉抗生物素蛋白等,其可用于从反应混合物中取出原始可转化标记寡核苷酸。就此而言可以使用各种各样的亲和试剂,例如核酸、蛋白质、肽、抗原、抗体或任何前述的反应性部分。

[0116] 在许多情况下,可以使用消化去除过程,以避免可能伴随上述纯化过程的材料损失。具体而言,其中可转化标记寡核苷酸优先消化或降解的过程可以用于去除它们以避免参与随后的反应操作。在一个实例中,所述原始可转化标记寡核苷酸可包含允许其选择性消化或去除的特定区域或碱基。举例来说,这些标记寡核苷酸可以在整个寡核苷酸序列内的一个或多个位置包含含尿嘧啶的碱基。用尿嘧啶靶向消化过程(例如尿嘧啶DNA糖基化酶然后用DNA糖基化酶内切核酸酶VIII如USER处理)处理反应混合物,然后允许靶向消化原始标记寡核苷酸序列,而含有经转化的标记区段的复制物将不含有含尿嘧啶的碱基。或者,这些标记寡核苷酸可以包括特定的限制性内切核酸酶切割位点,其在与相关核酸内切酶接触时导致可转化标记寡核苷酸的切割。另外地或可替代地,可以使用包含5'保护基团如硫代磷酸酯基团的引物序列进行标记过程之后的复制过程,使得例如使用T7核酸外切酶保护在第一轮复制中产生的那些复制分子不受5'至3'外切核酸酶消化双链DNA底物的影响,而起始分子可以经历消化。类似地,标记寡核苷酸可以具有其它性质,使得它们易于消化,例如并入RNA碱基,使得所述标记寡核苷酸可以使用对RNA底物例如核糖核酸酶特异的核酸酶消化。

[0117] 在另一种方法中,所述标记寡核苷酸可并入阻止它们在第一轮复制后参与后续复制事件的序列部件。举例来说,所述原始标记寡核苷酸可包含序列元件,例如含尿嘧啶的碱基,其可防止它们通过例如存在于后续复制轮中的某些聚合酶复制。在产生经标记的cDNA分子424之后的第一轮复制中,使用热不稳定聚合酶,例如DNA Pol1, Klenow,其可以在标记区段410中对于可转化碱基无偏差,但能够通过包含尿嘧啶的碱基而被处理,被用于进行第一轮复制,从而导致产生不包含尿嘧啶碱基的经转化的标记寡核苷酸428。在第一轮复制之后,将反应温度升高至合适的解链温度,例如90C,将使复制链428从原始经标记的cDNA 424解链,同时也使第一聚合酶失活。第二种热稳定聚合酶也存在于反应混合物中,并且不能通过含尿嘧啶的碱基例如北方9度(9degrees north)、深孔(deep vent)等的仿古聚合酶进行复制,然后在随后的扩增操作中保持活性以选择性扩增复制的经转化标记寡核苷酸428,同时不复制任何原始标记寡核苷酸,例如经标记的cDNA 424或任何剩余但未并入的标记寡核苷酸402。在可替代地的安排中,这些聚合酶可以反复出现。例如,第一轮复制中存在第一个尿嘧啶加工聚合酶。在第一轮复制后,可以去除所述第一聚合酶,例如通过从核酸聚合酶中

纯化核酸,并且可以引入不能针对含尿嘧啶的碱基复制的第二聚合酶。所述原始标记部分中尿嘧啶的存在可以阻止原始标记部分的进一步复制,并随后产生新的转化的标记寡核苷酸。相反,在这些随后的复制轮中只可以产生转化的寡核苷酸的直接补体/复制物。

[0118] 虽然上面是根据经标记的cDNA分子424的单轮复制来描述的,但应该理解,可以通过允许对给定分析赋予附加标记的反褶积,来在本文描述的上下文内实施附加的复制轮,例如2、3、4、5或更多轮。例如,可以凭借附加的转化复制的轮次知道增加的预期的附加的独特标记分子的数量,可以考虑所得分子的附加的多样性水平,以推断起始分子的原始数量。

[0119] 在某些情况下,可以计算所需的多样性,从而可以计算可转化序列区段的组成。特别地,为了计算序列的有效多样性,可以将其确定为每个标记序列区段中可转化碱基的简并性水平和这种可转化碱基的数量的函数。使用所谓的生日问题,可以计算将共享相同标记的分子的预期数量。可以通过首先用检测器测量过程的输出(例如,测序包括简并碱基的序列区段的群体)并且计数每个可转化位点处观察到的碱基的频率来计算有效多样性。然后可以使用多样性指数来计算每个位点的有效多样性。这种值的实例可以是每个可转化碱基的香农熵的指数,乘以所述标记中这些碱基的数量。理想的无偏差4-向简并碱基具有多样性4。相反,正常的,规范的碱基可能具有多样性1(即它总会被视为本身)。一旦将这个(实验确定的)碱基数值和碱基数量组合,可以将其映射到从1到N的整数空间,其中N仅仅是(有效多样性)×(可转化碱基的数量)。应用以下公式计算从整数空间中采样时的预期碰撞次数,得出预期的重复序列数(注意,这与计算计算机科学中散列函数产生的碰撞次数的问题类似),其中从[1,d]中随机选择第k个整数将重复至少一个先前的选择的概率等于上面的 $q(k-1;d)$ 。选择重复先前选择的预期总次数为n个此类整数,等于:

$$[0120] \quad \sum_{k=1}^n q(k-1;d) = n - d + d \left(\frac{d-1}{d} \right)^n$$

[0121] 在创建原始序列区段的更独特的经标记的复制物之后,可以对经转化标记的寡核苷酸进行附加的处理操作以促进它们的分析。例如,在一些情况下,为了产生足够量的分子供分析,例如使用核酸阵列或核酸测序系统,可以对经转化标记的分子进行扩增(例如使用PCR)。

[0122] 在PCR扩增的情况下,可以处理经转化的经标记分子以将扩增引发序列添加到经标记的分子的一端或两端。在一些情况下,如上所述,所述标记部分可包括可用作扩增引物的引发序列。可以通过例如与随机引发序列偶联的扩增引物的连接或聚合酶延伸将附加的引发序列添加至经标记的片段的相对末端,从而提供用于扩增的重复序列。

[0123] 例如,在公开的美国专利申请公开号2014/0378345、2014/0228255中描述了使用各种方法来生产可扩增的经标记核酸分子,该等专利的全部公开内容出于所有目的以引用方式整体并入本文。

[0124] 核酸扩增是创建DNA的小或长区段的多个拷贝的方法。DNA扩增可用于将一个或多个期望的寡核苷酸序列连接至单个珠粒,例如条形码序列或随机N聚体序列。也可使用DNA扩增来利用随机N聚体序列来引发和延伸所关注的样品如基因组DNA,以产生样品序列的片段并将与引物相关的条形码偶联至所述片段。

[0125] 例如,核酸序列可通过将模板核酸序列和包含多个附接的寡核苷酸(例如,可释放

地附接的寡核苷酸)的珠粒共分配到分区(例如,乳剂液滴、微囊或任何其它合适类型的分区,包括本文其它地方描述的合适类型的分区)被扩增。所述附接的寡核苷酸可以包含与模板核酸序列的一个或多个区域互补的引物序列(例如,可变引物序列,例如,随机N聚体或靶向引物序列,例如靶向N聚体),另外还可以包含共同序列(例如,条形码序列)。所述引物序列可退火至模板核酸序列并延伸(例如,在引物延伸反应或任何其它合适的核酸扩增反应中)以产生模板核酸的至少一部分的一个或多个第一拷贝,使得一个或多个第一拷贝包含引物序列和共同序列。在包含所述引物序列的寡核苷酸可释放地附接至珠粒的情况下,所述寡核苷酸可在将引物序列退火至模板核酸序列之前从珠粒释放。此外,通常,所述引物序列可以通过也在分区中提供的聚合酶(例如本文其它地方描述的链置换聚合酶,如本文别处所述的外切核酸酶缺陷型聚合酶或任何其它类型的合适的聚合酶,包括本文其它地方描述的一种聚合酶)延伸。此外,可释放地附接至珠粒的这些寡核苷酸可以是外切核酸酶抗性的,并且因此可以包含一个或多个本文其它地方所述的硫代磷酸酯键。在一些情况下,所述一个或多个硫代磷酸酯键可以在所述寡核苷酸中的末端核苷酸间键处包含硫代磷酸酯键。

[0126] 在一些情况下,在产生一个或多个第一拷贝之后,可将所述引物序列退火至一个或多个第一拷贝,并再次延伸所述引物序列以产生一个或多个第二拷贝。所述一个或多个第二拷贝可以包含所述引物序列、共同序列,并且还可以包含与所述一个或多个第一拷贝的单个拷贝的至少一部分互补的序列,和/或与所述可变引物互补的序列。上述操作可以重复所需的循环次数以产生扩增的核酸。

[0127] 所描述的寡核苷酸可以包含在延伸反应(例如产生上述一个或多个第一或第二拷贝的延伸反应)期间不被拷贝的序列区段。如本文其它地方所述,此类序列区段可以包含一个或多个含尿嘧啶的核苷酸并且还可以导致在退火条件下形成发夹(或部分发夹)分子的扩增子的产生。

[0128] 可以通过将不同的核酸分配到各自包含第二分区(例如珠粒,包括本文别处描述的珠粒类型)的单独的第一分区(例如乳液中的液滴)来扩增多种不同的核酸。第二分区可以与多个寡核苷酸可释放地相关联。第二分区可包含任何合适数量的寡核苷酸(例如,大于1,000个寡核苷酸、大于10,000个寡核苷酸、大于100,000个寡核苷酸、大于1,000,000个寡核苷酸、大于10,000,000个寡核苷酸,或本文所述的每个分区的任何其它数目的寡核苷酸)。此外,第二分区可包含任何合适数量的不同条形码序列(例如,至少1,000个不同的条形码序列、至少10,000个不同的条形码序列、至少100,000个不同的条形码序列、至少1,000,000个不同的条形码序列、至少10,000,000个不同的条形码序列,或本文其它地方描述的任何其它数量的不同条形码序列)。

[0129] 此外,与给定第二分区相关的多个寡核苷酸可以包含引物序列(例如,可变引物序列、靶向引物序列)和共同序列(例如条形码序列)。而且,与不同第二分区相关的多个寡核苷酸可以包含不同的条形码序列。与多个第二分区相关联的寡核苷酸可以释放到第一分区中。释放后,第一分区内的引物序列可退火至第一分区内的核酸,然后所述引物序列可延伸以产生具有第一分区的至少一部分核酸的一个或多个拷贝。通常,所述一个或多个拷贝可以包括释放到第一分区中的条形码序列。

[0130] 可以在流体液滴内的内容物上进行核酸(例如DNA)扩增。流体液滴可以包含附接至珠粒的寡核苷酸。流体液滴可以进一步包含样品。流体液滴还可以包含适合于扩增反应

的试剂,其可以包括Kapa HiFi Uracil Plus、修饰的核苷酸、天然核苷酸、含尿嘧啶的核苷酸、dTTP、dUTP、dCTP、dGTP、dATP、DNA聚合酶、Taq聚合酶、突变型校正读码聚合酶、北向9度、经修饰的(NEB)、exo(-)、exo(-)Pfu、深孔exo(-)、孔exo(-)和无环核苷酸(acyNTPS)。

[0131] 附接到流体的液滴内的珠粒的寡核苷酸可用于扩增样品核酸,使得寡核苷酸附接到样品核酸上。所述样品核酸实际上可以包含任何寻求分析的核酸,包括例如全基因组、外显子组、扩增子、靶向基因组区段,例如基因或基因家族、细胞核酸、循环核酸等,并且如上所述,可以包括DNA(包括gDNA、cDNA、mtDNA等)、RNA(例如,mRNA、rRNA、总RNA等)。用于条形码的这些核酸的制备通常可以通过容易获得的方法来完成,例如富集或下拉方法、分离方法、扩增方法等。为了扩增期望的样品例如gDNA,流体的液滴内的寡核苷酸的随机N聚体序列可用于引发期望的靶序列并作为靶序列的互补序列延伸。在一些情况下,所述寡核苷酸可以在引发之前如本文别处所述从液滴中的珠粒释放。对于这些引发和延伸过程,可以使用任何合适的DNA扩增方法,包括聚合酶链式反应(PCR)、数字PCR、逆转录PCR、多重PCR、巢式PCR、重叠延伸PCR、定量PCR、多重置换扩增(MDA)或连接酶链式反应(LCR)。在一些情况下,可以进行流体的液滴内的扩增直至可以产生一定量的包含条形码的样品核酸。在一些情况下,扩增可进行约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个循环。在一些情况下,扩增可进行大于约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个循环或更多个循环。在一些情况下,扩增可进行小于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个循环。

[0132] 在初始处理操作之后,所得到的经标记的核酸分子文库可以进行测序以确定文库分子的总体序列。通过识别不同的经转化的标记区段的数量,可以推断原始起始分子的数量,包括确定预测的或预期的起始分子的数量。如本文所用,定量(或定量)起始分子是指一般定量,而不是特定和限定的定量。这种一般定量可以通常用作相对度量,例如用于比较来自两个或更多个样品、但多个时间点的相同或不同样品,响应刺激的样品等的数量度量,或者可以用作起始分子的近似数量的一般指示,而不需要确切且绝对准确地确定分子的精确数量。

[0133] 试剂盒

[0134] 本文还提供用于实施上述方法和过程的试剂盒和系统。如将了解的,试剂盒通常可以包括用于实施这些方法的各种试剂。例如,用于实践所述过程的试剂盒通常可以包括上述标记组合物,例如包含上述可转化标记区段的寡核苷酸。在一些情况下,试剂盒可以包括此类组合物的不同文库,其包括大量的不同的寡核苷酸,所述寡核苷酸包含与可转化的标记区段结合的不同的条形码区段,这些可转化的标记区段在一些或全部文库中可能是共同的,但是这会在转化时将产生这种标记区段的多样性。在一些情况下,这些寡核苷酸文库可以被结合至颗粒,例如凝胶珠粒或微囊,并且在一些情况下,可以包括寡核苷酸内的附加的序列元件,例如,如测序仪特异性引发和/或附接序列,例如如在公开的美国专利申请公开号2014/0378345、2014/0228255、2015/0376700、2015/0376605和2016/0122817中所描述的,该等专利的全部公开内容出于所有目的以引用方式整体并入本文。

[0135] 并入条形码序列区段的寡核苷酸,其作为独特的标识符,还可以包括附加的序列区段。此类附加的序列区段可以包括功能序列,例如引物序列、引物退火位点序列、固定序列或其它用于后续处理的识别或结合序列,例如用于测序连接包含寡核苷酸的条形码的样

品的测序引物或引物结合位点。进一步,如本文所用,提及包含在包含序列的条形码内的特定功能序列也设想将所述互补序列包含在任何此类序列中,使得在互补复制时将产生具体描述的序列。

[0136] 另外,这些试剂盒还可以包括用于实施本文所述过程的其它试剂,例如酶,包括例如逆转录酶、DNA聚合酶,例如Klenow、DNA Pol1、Phi29和/或仿古聚合酶,例如北向9度(9degrees north),深孔(deep vent)等。同样可以包括其它酶,例如连接酶、USER酶、CRISPR-Cas9相关酶、PCR扩增酶,例如,Taq聚合酶等,等等。

[0137] 在一些情况下,本文所述的试剂盒还可以包括用于将样品材料(例如细胞、核酸等)分配成单独的分区(例如乳液中的液滴)的试剂和部件。这些试剂和部件可以包括例如分配油例如氟化油、氟化表面活性剂和微流体装置,用于产生如本文所述的分配的样品材料、试剂和标记寡核苷酸的乳液。这些部件可以与设计用于驱动流体通过微流体装置的合适的仪器系统结合和/或在其上使用,以便产生如上所述的分配的试剂乳液。分配试剂、微流体装置和仪器系统的实例在例如公开的美国专利申请公布号2010/0105112、2015/0292988、2014/0378345、2014/0228255中进行了描述,该等专利的全部公开内容出于所有目的以引用的方式整体并入本文。

[0138] 本发明的乳液可以使用本领域普通技术人员已知的任何合适的乳化程序形成。就此而言,可以理解的是,可以使用微流体系统、超声波、高压均化、振荡、搅拌、喷雾过程、膜技术或任何其它适当的方法来形成乳液。在一个特定实施方案中,使用微毛细管或微流体装置来形成乳液。通过所述方法产生的液滴的尺寸和稳定性可以根据例如毛细管尖端直径、流体速度、连续相和不连续相的粘度比以及两相的界面张力而变化。微流体系统内可能会产生不同大小和体积的液滴。这些尺寸和体积可以根据例如流体粘度、输注速率和喷嘴尺寸/构造的因素而变化。取决于具体的应用,可以选择不同体积的液滴。例如,液滴可具有小于1. μ L(微升)、小于0.1. μ L(微升)、小于10mL、小于1mL、小于0.1mL或小于10pL的体积。

[0139] 这些试剂盒还可以包括使用所提供的试剂和部件以执行本文描述的过程的说明书,以及用于分析所得数据的说明书和软件。这些说明书可以打印在一个或多个文档中或者以电子方式提供,例如在电子文件中或在用户的电子设备上的用户界面(UI)如图形用户界面(GUI)中。

[0140] 本公开的方法和系统可以由包括一个或多个计算机处理器和计算机存储器的计算机系统执行。本文提供的系统和方法的各方面可以在编程中体现。所述技术的各个方面可以被认为是典型地以机器(或处理器)可执行代码和/或相关数据的形式“产品”或“制品”,这些代码和/或相关数据被携带或体现在一种机器可读介质中。机器可执行代码可以存储在电子存储单元如存储器(例如,只读存储器、随机存取存储器、闪存)或硬盘中。“存储”型介质可以包括计算机、处理器等的有形存储器或其相关模块(例如,各种半导体存储器、磁带驱动器、磁盘驱动器等)中的任何或全部,其可以随时为软件编程提供非临时性存储。所有或部分软件有时可以通过互联网或各种其它电信网络进行通信。例如,此类通信可以使软件能够从一台计算机或处理器加载到另一台,例如从管理服务器或主计算机加载到应用服务器的计算机平台中。因此,可以承载软件元件的另一种类型的介质包括光学、电学和电磁波,例如跨本地装置之间的物理接口使用,通过有线和光学陆地线网络以及各种空

中链路使用。承载这种波的物理元件,例如有线或无线链路、光链路等也可以被认为是承载所述软件的介质。如本文所使用的,除非限于非临时性的有形“存储”介质,否则例如计算机或机器“可读介质”之类的术语是指参与向处理器提供指令以供执行的任何介质。

[0141] 因此,例如计算机可执行代码的机器可读介质可以采取许多形式,包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质包括例如光盘或磁盘,例如任何计算机等中的任何存储设备,例如可用于实现附图中所示的数据库等。易失性存储介质包括动态存储器,例如这种计算机平台的主存储器。有形的传输介质包括同轴电缆;铜线和光纤,包括构成计算机系统内总线的导线。载波传输介质可以采取电信号或电磁信号的形式,或者例如在射频(RF)和红外(IR)数据通信期间产生的声波或光波的形式。因此,计算机可读介质的常见形式包括例如:软盘、软磁盘、硬盘、磁带、任何其它磁介质、CD-ROM、DVD或DVD-ROM、任何其它光介质、穿孔卡片、纸带、具有孔图案的任何其它物理存储介质、RAM、ROM、PROM和EPROM、FLASH-EPROM、任何其它存储器芯片或盒式磁带、传输数据或指令的载波、传输这种载波的电缆或链路,或者计算机可以从其读取编程代码和/或数据的任何其它介质。这些形式的计算机可读介质中的许多形式可能涉及将一个或多个指令的一个或多个序列携带到处理器以供执行。

[0142] 本发明提供了包括但不限于以下实施方式:

[0143] 1. 一种差异化地标记多种分子物质中的个别成员的方法,其包括:

[0144] (a) 将第一标记部分附接至多个单独分子物质中的每一个,所述第一标记部分包含可转化标记部件;以及

[0145] (b) 将附接到所述多个单独分子中的每一个的所述可转化标记部件转化成转化的标记部件,以用不同的转化的标记部件清楚地标记所述多个分子物质中的多个不同成员。

[0146] 2. 如实施方式1所述的方法,其中

[0147] 所述多个单独分子物质包含多个单独核酸序列;

[0148] 所述标记部分包含寡核苷酸区段;并且

[0149] 所述标记部件包含可转化的寡核苷酸序列。

[0150] 3. 如实施方式2所述的方法,其中所述可转化的寡核苷酸序列包含一个或多个可转化的核苷酸。

[0151] 4. 如实施方式3所述的方法,其中一个或多个可转化的核苷酸包含简并核苷酸。

[0152] 5. 如实施方式4所述的方法,其中所述一个或多个可转化的核苷酸中的一个或多个包含2-向简并性。

[0153] 6. 如实施方式4所述的方法,其中所述一个或多个可转化的核苷酸中的一个或多个包含3-向简并性。

[0154] 7. 如实施方式4所述的方法,其中所述一个或多个可转化的核苷酸中的一个或多个包含4-向简并性。

[0155] 8. 如实施方式4所述的方法,其中所述一个或多个可转化的核苷酸选自以下各项组成的组:肌苷、脱氧肌苷、脱氧黄嘌呤、2'-脱氧水粉菌素、2'-脱氧鸟苷、5-硝基吡啶、3-硝基吡啶、N6-甲氧基-2,6-二氨基嘌呤、6H,8H-3,4-二氢嘧啶并[4,5-c][1,2]噁嗪-7-酮,以及前述各者的非脱氧(或核糖)形式。

[0156] 9. 如实施方式4所述的方法,其中所述可转化的寡核苷酸序列包含1个至20个可转

化的核苷酸。

[0157] 10. 如实施方式2所述的方法,其中所述标记部分进一步包含一个或多个附加的寡核苷酸区段。

[0158] 11. 如实施方式10所述的方法,其中所述一个或多个附加的寡核苷酸区段选自引物序列区段、杂交序列区段、连接序列区段、测序仪表面附接区段和条形码序列区段。

[0159] 12. 如实施方式10所述的方法,其中所述一个或多个附加的寡核苷酸序列包含选自随机引物序列和测序引物的引物序列。

[0160] 13. 如实施方式10所述的方法,其中所述一个或多个附加的寡核苷酸序列包含杂交序列。

[0161] 14. 如实施方式13所述的方法,其中所述杂交序列包含聚T序列。

[0162] 15. 如实施方式2所述的方法,其包括在所述附接之前将标记部分分配给包含待分析核酸的样品,并且其中所述附接包括将所述标记部分附接至待分析核酸。

[0163] 16. 如实施方式15所述的方法,其中所述标记部分包含聚T序列区段,并且所述待分析核酸包含mRNA分子。

[0164] 17. 如实施方式15所述的方法,其中所述分配包括将具有所述标记部分的个别细胞分配到分区中,并且其中所述待分析核酸包含在所述个别细胞内,并且其中在所述附接之前,将所述个别细胞裂解以将所述待分析核酸释放至所述分区中。

[0165] 18. 如实施方式2所述的方法,其中所述可转化的寡核苷酸序列区段包含序列取代系统的靶序列。

[0166] 19. 如实施方式18所述的方法,其中所述序列取代系统包含CRISPR酶体系,并且所述靶序列包含靶向寡核苷酸的靶序列。

[0167] 20. 如实施方式1所述的方法,由此所述转化是随机的或半随机的。

[0168] 21. 一种分析核酸分子的方法,其包括:

[0169] (a) 将寡核苷酸区段附接至靶寡核苷酸分子以产生经标记寡核苷酸,其中所述寡核苷酸包含含有多个可变互补核苷酸的区域;

[0170] (b) 复制所述标记寡核苷酸以产生复制的经标记寡核苷酸,由此复制产生所述区域的随机或部分随机复制物;以及

[0171] (c) 分析所述复制的经标记寡核苷酸,其包括所述随机或部分随机复制物,以鉴定所述靶寡核苷酸分子。

[0172] 22. 如实施方式21所述的方法,其中所述区域包含2至20个可变互补核苷酸。

[0173] 23. 如实施方式21所述的方法,其中所述区域包含两个或更多个连续的可变互补核苷酸。

[0174] 24. 如实施方式23所述的方法,其中所述可变互补核苷酸中的两个或更多个通过一个或多个非可变互补核苷酸彼此分开。

[0175] 25. 如实施方式21所述的方法,其中所述区域包含4至10个可变互补核苷酸。

[0176] 26. 如实施方式21所述的方法,其中所述第一寡核苷酸包含含有多个可变互补核苷酸的附加区域。

[0177] 27. 一种寡核苷酸组合物,其包含含有第一区域和第二区域的寡核苷酸,其中所述第二区域包含含有多个可变互补核苷酸的固定序列,所述多个可变互补核苷酸可转化以产

生独特的分子标记。

[0178] 28. 如实施方式27所述的组合物,其中所述第一区域包含用于将所述寡核苷酸附接至待分析核酸分子的附接序列。

[0179] 29. 如实施方式28所述的组合物,其中所述附接序列包含引物序列。

[0180] 30. 如实施方式28所述的组合物,其中所述附接序列包含聚T序列。

[0181] 31. 如实施方式27所述的组合物,其中所述第一区域包含条形码序列。

[0182] 32. 如实施方式27所述的组合物,其中所述第一区域包含表面附接序列。

[0183] 33. 如实施方式27所述的组合物,其中所述第二区域包含多个可变互补核苷酸和一个或多个非可变互补核苷酸。

[0184] 34. 一种对相同核酸分子的群体中的核酸分子进行定量的方法,其包括:

[0185] (a) 以预期的诱变速率突变所述相同核酸分子的群体以产生不同突变核酸的群体;

[0186] (b) 对所述不同突变核酸分子进行测序;以及

[0187] (c) 基于多个不同突变核酸分子计算所述相同核酸分子的群体中的所述核酸分子的定量。

[0188] 35. 如实施方式34所述的方法,其中(c)包括基于所述不同突变核酸分子的数目和所述诱变速率对所述相同核酸分子的群体中的所述核酸分子进行定量。

[0189] 36. 如实施方式34所述的方法,其中(b)包括从所述不同突变核酸分子产生测序读段。

[0190] 37. 如实施方式36所述的方法,其中(c)包括计算所述测序读段的比较以对所述相同核酸分子的群体中的所述核酸分子进行定量。

[0191] 38. 一种区分来自两个或更多个相同核酸分子的扩增产物的方法,其包括:

[0192] (a) 使所述两个或更多个核酸分子经历诱变以产生两个或更多个突变的核酸分子;

[0193] (b) 扩增所述两个或更多个突变的核酸分子以产生扩增的突变核酸产物;以及

[0194] (c) 对所述扩增的突变核酸产物进行测序。

[0195] 实施例

[0196] 进行了一个总结实验,其中在包含7个可转化或简并碱基中的1个的序列上合成第一条链,其中合成通过3种不同的聚合酶中的1种进行。第一条链是使用在延伸引物的5'末端含有4个硫代磷酸酯的引物合成的,因此T7外切核酸酶可用于降解含有可转化碱基的模板链,使合成的第一链保持完整。测序结果显示了跨七种碱基和三种酶的广泛并入模式和聚合酶效率。通过将可转化碱基包埋在随机聚合物中,我们能够鉴定使可转化碱基位点处的有效多样性最大化的侧翼碱基的组合,例如通过影响聚合酶接近可转化碱基时的动力学或通过影响可转化碱基位点处模板合成链双链体中的堆积相互作用。在某些情况下,尽管不同配置产生不同水平的多样性,但是一种最佳组合似乎包括Taq聚合酶与5-硝基吡啶或脱氧异鸟嘌呤核苷中的一种或两种。

[0197] 虽然本文已经示出和描述了本发明的优选实施方案,但是对于本领域技术人员而言显而易见的是,此类实施方案仅以举例方式提供。不希望本发明受到说明书中提供的具体实例的限制。尽管已经参考前述说明书描述了本发明,但是本文的实施方案的描述和说

明不意味着以限制的含义来解释。在不脱离本发明的情况下,本领域技术人员现在将想到许多变化、改变和替换。此外,应该理解,本发明的所有方面不限于依赖于各种条件和变量的在此阐述的具体描述、配置或相对比例。应该理解,可以在实践本发明时采用在此描述的本发明实施方案的各种替代方案。因此预期本发明还应涵盖任何此类替代、修改、变化或等同物。以下权利要求旨在限定本发明的范围,并且由此涵盖这些权利要求范围内的方法和结构及其等同物。

100

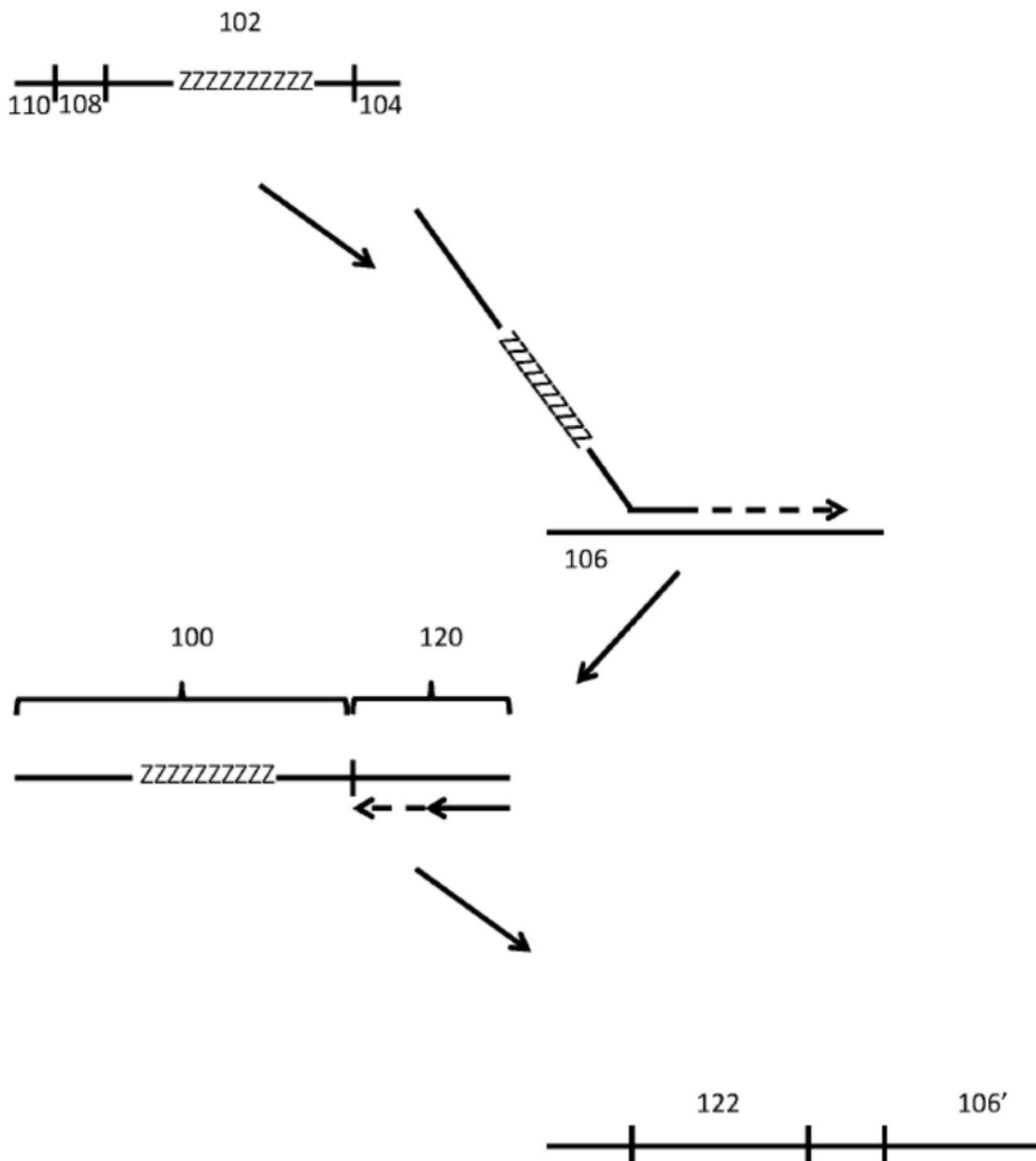


图1

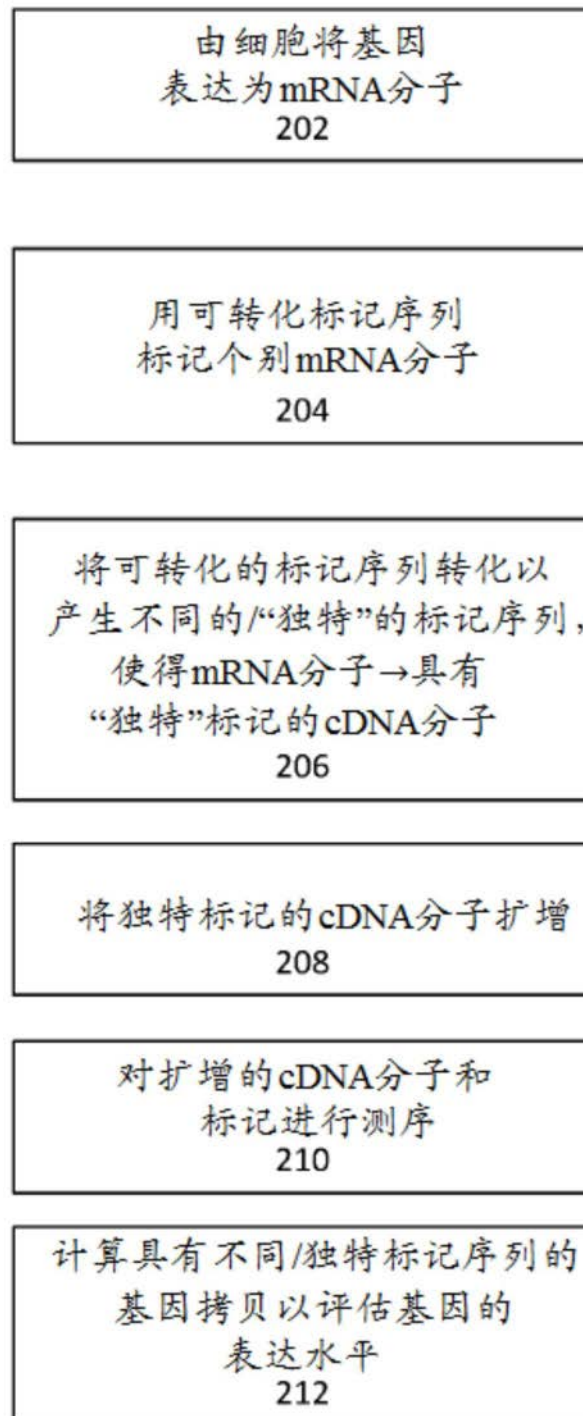


图2

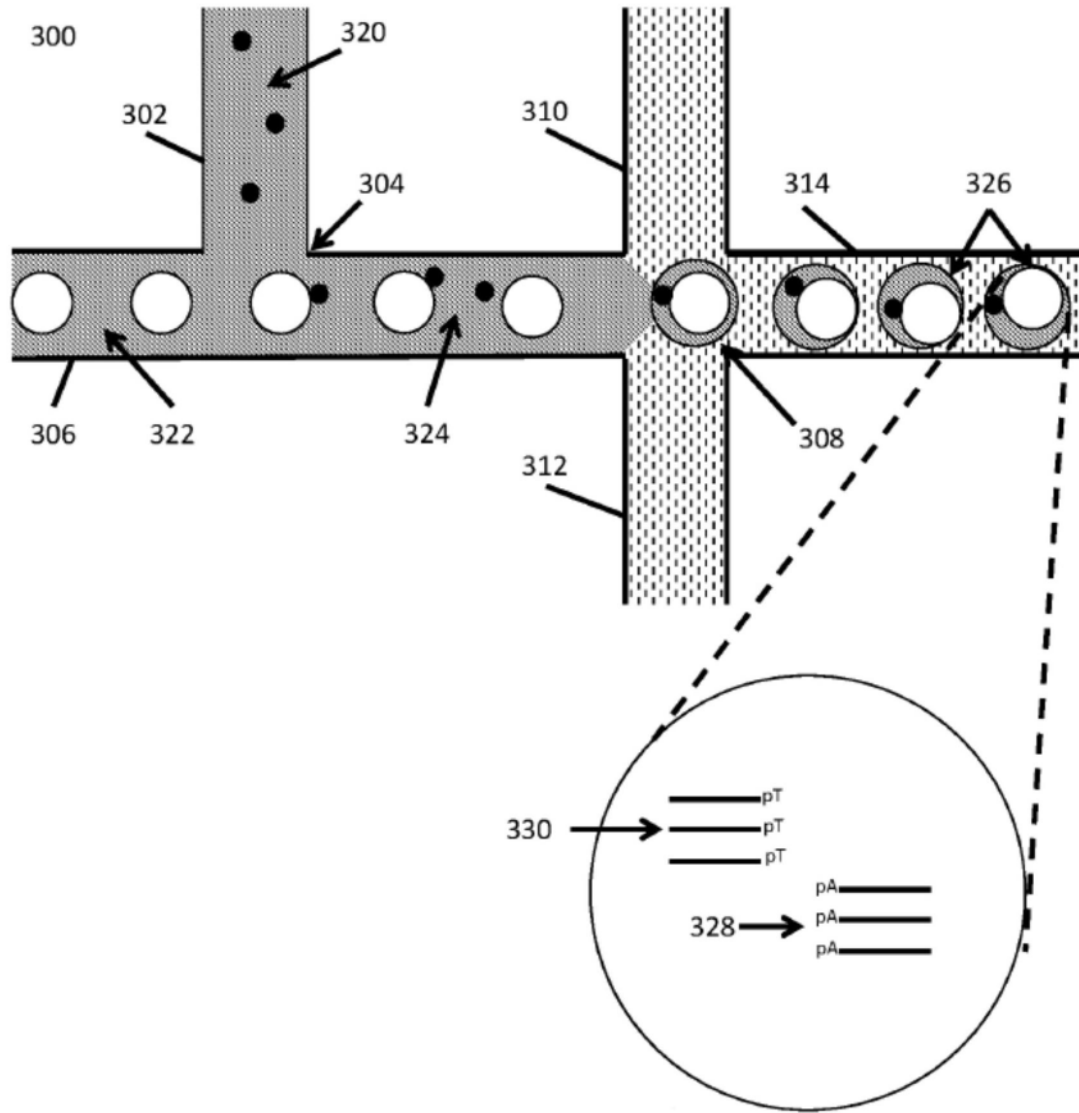
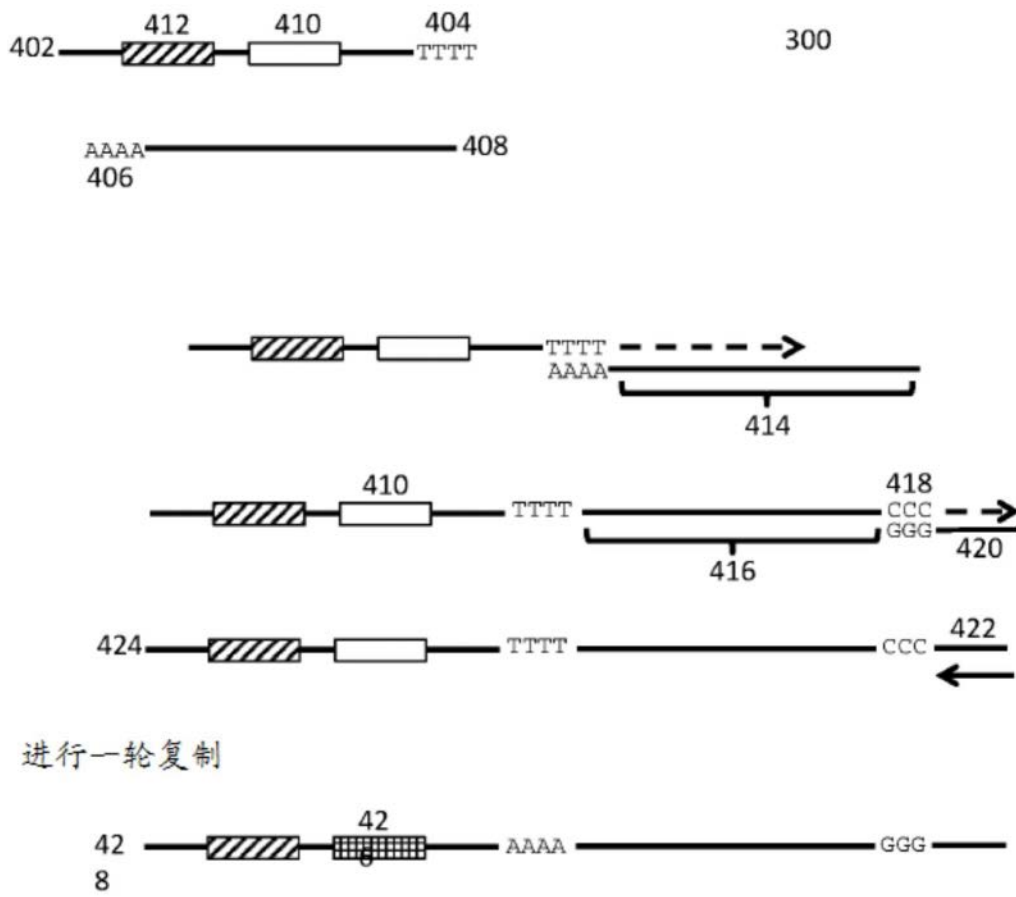


图3



非摇摆碱基复制物从具有摇摆碱基的寡核苷酸富集或纯化并扩增/测序

图4