



# (12)发明专利申请

(10)申请公布号 CN 108875312 A

(43)申请公布日 2018. 11. 23

(21)申请号 201810747946.2

G06N 3/12(2006.01)

(22)申请日 2013.07.17

(30)优先权数据

61/673,690 2012.07.19 US

61/676,081 2012.07.26 US

(62)分案原申请数据

201380038507.X 2013.07.17

(71)申请人 哈佛大学校长及研究员协会

地址 美国马萨诸塞州

(72)发明人 乔治·M·丘奇

(74)专利代理机构 上海专利商标事务所有限公

司 31100

代理人 陈扬扬

(51)Int. Cl.

G06F 19/22(2011.01)

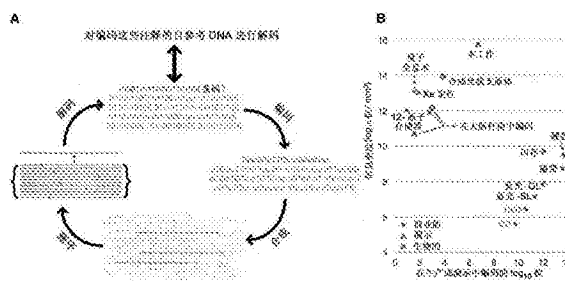
权利要求书2页 说明书23页 附图2页

(54)发明名称

利用核酸存储信息的方法

(57)摘要

本发明涉及利用一个或多个核酸存储数据的方法。



1. 一种利用核苷酸存储信息的方法,包括:  
将信息格式转换成位流的多个位序列,每个位序列具有对应的位条码,  
利用一个位/碱基编码将所述多个位序列转换成多个对应的寡核苷酸序列,  
合成所述多个对应的寡核苷酸序列,以及  
存储合成的所述多个对应的寡核苷酸序列。
2. 根据权利要求1所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在所述位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。
3. 一种从编码信息格式的位序列的多个合成的寡核苷酸序列检索所述信息格式的方法,包括:  
扩增多个寡核苷酸序列,  
对扩增的所述寡核苷酸序列进行测序,  
将所述寡核苷酸序列转换成位序列,以及  
将所述位序列转换成所述信息格式。
4. 根据权利要求3所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。
5. 一种从编码信息格式的位序列的多个合成的寡核苷酸序列存取所述信息格式的方法,包括:  
扩增多个寡核苷酸序列,  
对扩增的所述寡核苷酸序列进行测序,  
将所述寡核苷酸序列转换成位序列,  
将所述位序列转换成所述信息格式,以及  
使所述信息格式可视化。
6. 根据权利要求5所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。
7. 一种利用核苷酸存储信息的方法,包括将信息格式转换成位流,将第一位序列编码成对应的寡核苷酸序列,合成所述寡核苷酸序列,对所述寡核苷酸序列进行测序,将所述寡核苷酸序列解码成位序列,将所述位序列汇编成位流以及将所述位流转换成所述信息格式。
8. 根据权利要求7所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在所述位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。
9. 一种利用核苷酸存储信息的方法,包括将信息格式转换成第一位流,将第一位序列编码成对应的寡核苷酸序列,合成所述寡核苷酸序列,对所述寡核苷酸序列进行测序,将所述寡核苷酸序列解码成第二位序列,将所述第二位序列汇编成第二位流以及将所述第二位流转换成所述信息格式。
10. 根据权利要求9所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数

据块在所述位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

11. 一种利用核苷酸存储信息的方法,包括将第一信息格式转换成第一位流,将第一位序列编码成对应的寡核苷酸序列,合成所述寡核苷酸序列,对所述寡核苷酸序列进行测序,将所述寡核苷酸序列解码成第二位序列,将所述第二位序列汇编成第二位流以及将所述第二位流转换成第二信息格式。

12. 根据权利要求11所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在所述位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

13. 一种从编码信息格式的位序列的多个合成的寡核苷酸序列检索所述信息格式的方法,包括:

对扩增的所述寡核苷酸序列进行测序,  
将所述寡核苷酸序列转换成位序列,以及  
将所述位序列转换成所述信息格式。

14. 根据权利要求13所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

15. 一种从编码信息格式的位序列的多个合成的寡核苷酸序列存取所述信息格式的方法,包括:

对扩增的所述寡核苷酸序列进行测序,  
将所述寡核苷酸序列转换成位序列,  
将所述位序列转换成所述信息格式,以及  
使所述信息格式可视化。

16. 根据权利要求15所述的方法,其中,所述寡核苷酸序列包括数据块序列、指明所述数据块在位流中的位置的地址序列、或在所述寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

## 利用核酸存储信息的方法

[0001] 相关申请资料

[0002] 本申请要求于2012年7月19日提交的美国临时申请号61/673,690和于2012年7月26日提交的美国临时申请号61/676,081的优先权,出于所有目的在此通过引用将它们中的每一个以其整体合并于本文中。

[0003] 政府权益声明

[0004] 在美国海军研究办公室N000141010144的政府资助下作出本发明。政府对本发明具有一定的权利。

### 技术领域

[0005] 本发明总体上涉及利用一个或多个核酸序列存储信息的方法。

### 背景技术

[0006] 1.我们的数字总体正在迅速增长。例如,于2011年产生和复制了1.8泽字节(zettabyte) ( $10^{21}$ ) 的信息,并且该量每两年翻一番。参见1.“Extracting Value from Chaos” (IDC,Framingham,MA 2011);环球网站emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf。档案数据存储通常是硬盘驱动器、光学介质、和磁带的形式,其提供5-30年的推荐的使用寿命。参见2.J.Rothenberg,Scientific American 272,42-47 (1995)。随着数字信息不断积累,需要更高的密度和更长期的存储解决方案。作为用于信息存储的介质,DNA具有许多潜在的优点。参见3.C.Bancroft,T.Bowler,B.Bloom,C.T.Clelland,Science 293,1763-1765 (2001)。将一般的数字信息编码成DNA的第一个实例是1988年的35位。参见7.J.Davis,Art Journal 55,70-74 (1996)。近期工作涉及将数据压缩和加密成DNA (通常限于大写字母)、以及其在活细胞中的生存力。参见3、8、和9.C.Bancroft,T.Bowler,B.Bloom,C.T.Clelland,Science 293,1763-1765 (2001)、C.Gustafsson,Nature 458,703 (2009)、和D.G.Gibson et al.,Science 329,52-56 (2010)。

### 发明内容

[0007] 本公开内容的实施方式涉及利用核酸序列或包括核苷酸的序列作为用于信息存储的介质的方法。常见的核苷酸包括A、C、G、和T。本公开内容的方面涉及利用下一代测序和合成技术耐用、大规模读取和写入数字信息的方法,这样的方法的实例示意性地提供在图1A中。根据一个方面,将文本和/或图像转换成兆位。根据一个方面,转换成兆位的文本和/或图像包含位流(比特流,bit stream)。然后,将兆位编码成寡核苷酸。根据一个方面,寡核苷酸包括数据块序列(data block sequence)。根据一个方面,寡核苷酸包括指明数据块在位流(比特流)中的位置的地址序列(如条码序列)。根据一个方面,寡核苷酸包括在寡核苷酸的每个末端用于扩增和测序的旁侧共同序列(flanking common sequence)。根据一个方面,寡核苷酸包括数据块序列、指明数据块在位流中的位置的地址序列(如条码序列)、和在

寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

[0008] 根据本公开内容的一个方面,编码一位/碱基。根据该方面,可以以多种方式编码单条信息,即,对于零为A或C,对于数字1为G或T。因此,本公开内容的该方面避免难以读取或写入的序列特征如极端的GC含量(内容)、重复、或二级结构。

[0009] 根据一个方面,将位流分为寻址数据块。根据该方面,可以避免难以组装的长DNA构建物。

[0010] 根据一个方面,合成、存储和测序了每个单独寡核苷酸的许多拷贝。因为在合成和测序中的误差是很少一致的,所以每个分子拷贝纠正在其他拷贝中的误差。

[0011] 根据一个方面,体外进行所述方法。根据该方面,体外方法避免体内方法的克隆和稳定性问题。

[0012] 根据一个方面,高通量、下一代技术用于DNA合成和测序以允许对大量的信息进行编码和解码。

[0013] 根据一个方面,提供了一种利用核苷酸存储信息的方法,包括:将信息格式转换成位流的多个位序列(比特序列,bit sequence),每个位序列具有对应的位条码(比特条码,bit barcode),利用一个位/碱基编码将多个位序列转换成多个对应的寡核苷酸序列,合成多个对应的寡核苷酸序列,以及存储合成的多个对应的寡核苷酸序列。根据一个方面,寡核苷酸序列包括数据块序列、指明数据块在位流中的位置的地址序列、或在寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

[0014] 根据一个方面,提供了一种从编码信息格式的位序列的多个合成的寡核苷酸序列检索信息格式的方法,包括:扩增多个寡核苷酸序列,对扩增的寡核苷酸序列进行测序,将寡核苷酸序列转换成位序列,以及将位序列转换成信息格式。根据一个方面,寡核苷酸序列包括数据块序列、指明数据块在位流中的位置的地址序列、或在寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

[0015] 根据一个方面,提供了一种从编码信息格式的位序列的多个合成的寡核苷酸序列存取信息格式的方法,包括:扩增多个寡核苷酸序列,对扩增的寡核苷酸序列进行测序,将寡核苷酸序列转换成位序列,将位序列转换成信息格式,以及使信息格式可视化(可见,visualizing)。根据一个方面,寡核苷酸序列包括数据块序列、指明数据块在位流中的位置的地址序列、或在寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

[0016] 根据一个方面,提供了一种利用核苷酸存储信息的方法,包括:将信息格式转换成位流,将位序列编码成对应的寡核苷酸序列,合成寡核苷酸序列,对寡核苷酸序列进行测序,将寡核苷酸序列解码成位序列,将位序列汇编(组装,assembling)成位流以及将位流转换成信息格式。根据一个方面,寡核苷酸序列包括数据块序列、指明数据块在位流中的位置的地址序列、或在寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

[0017] 提供了一种利用核苷酸存储信息的方法,包括:将第一信息格式转换成第一位流,将第一位序列编码成对应的寡核苷酸序列,合成寡核苷酸序列,对寡核苷酸序列进行测序,将寡核苷酸序列解码成第二位序列,将第二位序列汇编成第二位流以及将第二位流转换成第二信息格式。根据一个方面,寡核苷酸序列包括数据块序列、指明数据块在位流中的位置

的地址序列、或在寡核苷酸的每个末端用于扩增和测序的旁侧共同序列中的一种或多种或全部。

[0018] 根据一些方面,使用利用DNA密度的DNA存储信息。根据一些方面,在理论最大值下,在约4克的钠盐中,DNA可以编码2位/核苷酸(nt)或1.8泽字节。参见4。关于材料和方法的信息,可在科学在线(Science Online)上获得。根据一些方面,用商用寡核苷酸大小和本领域技术人员已知的测序技术(48位条码+128位有效负载)可获得毫克的DNA,并且对于100x覆盖,可实现拍字节(petabyte) ( $10^{15}$ )。可以将该量存储于1536孔板以提供艾字节(exabyte),其中一千的那些(0.5m,边立方(side cube))构成泽字节( $8 \times 10^{21}$ 位)。可以将存储数据的子集以混合物存储并且通过利用正交(即,最小交叉引发)旁侧引物分别检索。参见Church GM,Kieffer-Higgins S(1988) Multiplex Sequencing;和Kosuri S,Eroshenko N,LeProust E,Super M,Way J,Li JB,Church GM(2010) A Scalable Gene Synthesis Platform Using High-Fidelity DNA Microchips. Nature Biotech. 28(12):1295-9。

[0019] 根据一些方面,DNA不需要用于存储的平坦表面,容易保存,并且在几千年以后已经被恢复。参见5和6.5. J. Bonnet et al., Nucleic Acids Research 38,1531-1546(2010)和6. S. Pääbo et al., M.A. Uytterlinde et al., Eds. Annual Review of Genetics 38, 645-679(2004)。根据一些方面,DNA的基本的生物学作用提供进入自然的读取和写入酶,并且对于可预见的未来,确保DNA将保持可读标准。

## 附图说明

[0020] 从说明性实施方式的以下详细描述连同附图,将更全面地理解本发明的上述和其他特征和优点,其中:

[0021] 2.图1(A)是DNA信息存储的示意图。将在编码的html(超文本链接标示语言)书内的句子的12字节部分转换成具有确定编码位在整个书内的位置的19位条码(红色)的位(蓝色)。然后,利用1位/碱基编码( $ac=0; TG=1$ )将位序列编码成DNA,同时还避免4个或更多个核苷酸重复并且平衡GC含量。整个5.27兆位html书使用54,898个寡核苷酸,并且进行合成并从DNA微芯片中洗脱。在扩增(未示出所有寡核苷酸的共同引物序列)之后,利用下一代测序来对寡核苷酸库测序。针对共有序列(consensus)筛选借助于正确的条码和长度进行的单个读取,然后重新转换到位,获得原书。总计,写入、扩增、和读取产生在5.27兆位外的10位误差(比特误差,bit error)。(B)与其他技术的比较。我们绘制了如通过报告或商用装置中编码的 $\log_{10}$ 位测定的信息密度( $\log_{10}$ 位/ $\text{mm}^3$ )相对于目前的可量测性。参见4。关于材料和方法的信息,可在科学在线上获得。

[0022] 图2是设计的库的每个成员的观测数目的直方图。

## 具体实施方式

[0023] 本发明涉及利用低聚物存储信息的方法。这种低聚物可以由单体形成。示例性单体包括核苷酸。示例性低聚物包括寡核苷酸。根据一个方面,提供了一种编码信息的方法,其中,将位序列转换成核苷酸序列,其中,核苷酸序列是寡核苷酸。根据一个方面,使用核酸合成的商业可获得的方法。根据一个方面,使用核酸扩增的商业可获得的方法。根据一个方面,使用核酸测序的商业可获得的方法。

[0024] 根据一个方面,将一部分或多部分的信息格式,如html信息格式,如具有文本和/或图像的html书,转换成位,即0和1,并且添加位条码以形成位序列,即,如通常所理解的一系列的0和1。可以被转换成位的信息的其他格式是本领域技术人员已知的。根据一个方面,待转换成位的html信息格式的部分可以被称为字节部分。位条码可以确定编码位在整个html信息格式内的位置。然后,利用1位/碱基编码(ac=0;TG=1),将位序列转换(编码)成核苷酸序列,即寡核苷酸或DNA,以形成对应的编码寡核苷酸序列,即,寡核苷酸序列对应于或编码位序列。根据一个方面,避免了4个或更多个核苷酸重复并且平衡了GC含量。对应于部分的或整个html信息格式,产生了多个位序列。因此,产生多个对应的编码的寡核苷酸序列,其一起可以被称为库。编码的寡核苷酸序列的库表示html信息格式。然后,利用本领域技术人员已知的方法,如利用DNA微芯片,合成编码的寡核苷酸序列。然后,利用本领域技术人员已知的方法扩增合成的寡核苷酸,以形成寡核苷酸的库。然后,利用本领域技术人员已知的方法,如下一代测序方法,对寡核苷酸的库进行测序。然后,将测序的寡核苷酸转换成对应于html信息格式的位序列。利用本领域技术人员已知的方法,可以将位序列转换成信息格式。利用本领域技术人员已知的方法和装置,可以可视化或显示信息格式。

[0025] 本文中使用的核酸化学、生物化学、遗传学、和分子生物学的术语和符号遵循在本领域中标准专著和文章的那些术语和符号,例如,Komberg and Baker,DNA Replication, Second Edition (W.H.Freeman, New York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, New York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984) 等。

[0026] 如在本文中使用的,根据其对于本领域技术人员通常含义理解术语“位(比特)”。术语“位”可以是“二进制数字(binary digit)”的缩写并且可以是指指在计算和电信中信息的基本容量。“位”仅表示1或0(1或0)。可以在各种系统中并借助于两状态装置来实施该表示。

[0027] 如在本文中使用的,术语“核酸分子”、“核酸序列”、“核酸片段”和“低聚物”可互换使用并且旨在包括但不限于可以具有各种长度的核苷酸的聚合物形式,包括脱氧核糖核苷酸或核糖核苷酸、或其类似物。用于本发明中的低聚物可以是完全设计的、部分设计的(即,部分随机的)或完全随机的。在本发明的一些方面,核酸的池包含DNA的单链90聚体。

[0028] 一般而言,“扩增”包括经由引发的酶促合成的重复循环生产核酸分子的拷贝。“原位”扩增是指借助于定位在支撑物或珠子上而不是溶液中的模板核酸分子进行扩增。美国专利号6,432,360中描述了原位扩增方法。

[0029] 聚合酶的不同选择与不同性能如温度、链置换、和校对共存。如以上描述的,扩增可以是等温的并且以类似的适应如通过以下描述的多重置换扩增(MDA): Dean et al., Comprehensive human genome amplification using multiple displacement amplification, Proc. Natl. Acad. Sci. U.S.A., vol. 99, p. 5261-5266. 2002; 以及 Dean et al., Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification, Genome Res., vol. 11, p. 1095-

1099.2001;以及Aviel-Ronen et al.,Large fragment Bst DNA polymerase for whole genome amplification of DNA formalin-fixed paraffin-embedded tissues,BMC Genomics,vol.7,p.312.2006。扩增还可以通过不同的温度控制(temperature regiment)进行循环,如由Mullis et al.,Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction.Cold Spring Harbor Symp.Quant.Biol.,vole 51, p.263-273.1986所推广的传统的聚合酶链反应(PCR)。由Zhang et al.,Whole genome amplification from a single cell:implications for genetic analysis, Proc.Natl.Acad.Sci.U.S.A.,vol.89,p.5847-5851.1992;以及Telenius et al., Degenerate oligonucleotide-primed PCR:general amplification of target DNA by a single degenerate primer,Genomics,vol.13,p.718-725.1992描述了更适用于基因组扩增的变化。其他方法包括由Mitra and Church,In situ localized amplification and contact replication of many individual DNA molecules,Nuc.Acid.Res.,vole 27, pages e34.1999描述的Polony PCR;由Shendure et al.,Accurate multiplex polony sequencing of an evolved bacterial genome,Science,vol.309,p.1728-32.2005;以及Williams et al.,Amplification of complex gene libraries by emulsion PCR, Nat.Methods,vol.3,p.545-550.2006描述的乳液PCR(ePCR)。任何扩增方法可以先验地与反转录步骤结合,以允许RNA的扩增。根据一些方面,扩增不是绝对必需的,因为探针、报道基因(报道物,reporter)和具有足够的灵敏度的检测系统可以用来允许利用所描述的模板非杂交核酸结构检测单分子。适应系统中的灵敏度的方法包括选择激发源(例如,照明)和检测(例如,光检测器、光电倍增管)。适应信号水平的方法包括允许报道基因的堆积的探针,并且还可以使用高强度报道基因(例如,量子点)。

[0030] 可用于本公开内容的扩增方法可以包括使核酸与在促进杂交和链延伸的条件下特异性地杂交至核酸的一种或多种引物接触。用于扩增核酸的示例性方法包括聚合酶链反应(PCR)(参见,例如,Mullis et al.(1986) Cold Spring Harb.Symp.Quant.Biol.51 Pt 1:263和Cleary et al.(2004) Nature Methods 1:241;以及美国专利号4,683,195和4,683,202),锚定PCR、RACE PCR(cDNA末端快速扩增PCR)、连接链反应(LCR)(参见,例如,Landegran et al.(1988) Science 241:1077-1080;和Nakazawa et al.(1994) Proc.Natl.Acad.Sci.U.S.A.91:360-364),自主序列复制(Guatelli et al.(1990) Proc.Natl.Acad.Sci.U.S.A.87:1874),转录扩增系统(Kwoh et al.(1989) Proc.Natl.Acad.Sci.U.S.A.86:1173),Q- $\beta$ 复制酶(Lizardi et al.(1988) BioTechnology 6:1197),递推PCR(Jaffe et al.(2000) J.Biol.Chem.275:2619;和Williams et al.(2002) J.Biol.Chem.277:7790),美国专利号6,391,544、6,365,375、6,294,323、6,261,797、6,124,090和5,612,199中描述的扩增方法,或利用本领域技术人员众所周知的技术的任何其他核酸扩增方法。在示例性实施方式中,本文中公开的方法采用PCR扩增。

[0031] 在一些示例性实施方式中,提供了用于扩增核酸序列的方法。用于扩增核酸的示例性方法包括聚合酶链反应(PCR)(参见,例如,Mullis et al.(1986) Cold Spring Harb.Symp.Quant.Biol.51 Pt 1:263和Cleary et al.(2004) Nature Methods 1:241;以及美国专利号4,683,195和4,683,202),锚定PCR、RACE PCR、连接链反应(LCR)(参见,例如,Landegran et al.(1988) Science 241:1077-1080;和Nakazawa et al.(1994)

Proc.Natl.Acad.Sci.U.S.A.91:360-364),自主序列复制(Guatelli et al.(1990) Proc.Natl.Acad.Sci.U.S.A.87:1874),转录扩增系统(Kwoh et al.(1989) Proc.Natl.Acad.Sci.U.S.A.86:1173), $\alpha$ - $\beta$ 复制酶(Lizardi et al.(1988) BioTechnology 6:1197),递推PCR(Jaffe et al.(2000) J.Biol.Chem.275:2619;和Williams et al.(2002) J.Biol.Chem.277:7790),美国专利号6,391,544、6,365,375、6,294,323、6,261,797、6,124,090和5,612,199中描述的扩增方法,等温扩增(例如,滚环扩增(RCA)、超支化滚环扩增(HRCA)、链置换扩增(SDA)、解旋酶依赖性扩增(HDA)、PWGA),或利用本领域技术人员众所周知的技术的任何其他核酸扩增方法。

[0032] “聚合酶链反应”或“PCR”是指通过DNA的互补链的同时引物延伸体外扩增特定DNA序列的反应。换言之,PCR是用于制备由引物结合位点旁侧的靶核酸的多个拷贝或复制品的反应,这种反应包括以下步骤的一次或多次重复:(i)使靶核酸变性,(ii)将引物退火至引物结合位点,以及(iii)在三磷酸核苷的存在下,通过核酸聚合酶延伸引物。通常,在热循环仪中,通过对于每个步骤优化的不同的温度循环反应。特定温度、每个步骤下的持续时间、和在步骤之间的变化速率取决于本领域普通技术人员众所周知的许多因素,例如,通过参考文献举例说明:McPherson et al.,editors,PCR:A Practical Approach和PCR2:A Practical Approach (IRL Press,Oxford,分别为1991年和1995年)。例如,在利用Taq DNA聚合酶的常规PCR中,可以在高于90°C的温度下使双链靶核酸变性,在50-75°C范围内的温度下使引物退火,以及在68-78°C的温度下使引物延伸。

[0033] 术语“PCR”包括反应的衍生形式,包括但不限于RT-PCR、实时PCR、巢式PCR、定量PCR、多重PCR、装配PCR等。反应容积范围从几百纳升,例如,200nL,至几百微升,例如,200 $\mu$ L。“反转录PCR”或“RT-PCR”是指之前发生将靶RNA转换至随后扩增的互补的单链DNA的反转录反应的PCR,例如,Tecott等人,美国专利号5,168,038。“实时PCR”是指对于其,将反应产物即扩增子的量监测为反应进程的PCR。存在主要不同之处在于用于监测反应产物的检测化学物质的实时PCR的许多形式,例如,Gelfand等人,美国专利号5,210,015(“Taqman”);Wittwer等人,美国专利号6,174,670和6,569,627(嵌入染料(增补燃料));Tyagi等人,美国专利号5,925,517(分子信标)。Mackay et al.,Nucleic Acids Research,30:1292-1305(2002)中综述了用于实时PCR的检测化学物质。“巢式PCR”是指两步PCR,其中,利用其中的至少一种结合至第一扩增子的内部位置的一组新的引物,第一PCR的扩增子变成用于第二PCR的样品。如在本文中使用的,关于巢式扩增反应的“初始引物”是指用于产生第一扩增子的引物,以及“次级引物”是指用于产生第二或巢式扩增子的一种或多种引物。“多重PCR”是指其中在相同的反应混合物中同时进行多种靶序列(或单一靶序列和一种或多种参比序列)的PCR,例如,Bernard et al.(1999) Anal.Biochem.,273:221-228(双色实时PCR)。通常,不同组的引物用于被扩增的每种序列。“定量PCR”是指设计以测量样品或试样中的一种或多种特定靶序列的丰度的PCR。用于定量PCR的技术是本领域普通技术人员众所周知的,如以下参考文献中举例说明的:Freeman et al.,Biotechniques,26:112-126(1999);Becker-Andre et al.,Nucleic Acids Research,17:9437-9447(1989);Zimmerman et al.,Biotechniques,21:268-279(1996);Diviacco et al.,Gene,122:3013-3020(1992);Becker-Andre et al.,Nucleic Acids Research,17:9437-9446(1989)等。

[0034] 可用于本公开内容的测序方法包括Shendure et al.,Accurate multiplex

polony sequencing of an evolved bacterial genome, *Science*, vol.309, p.1728-32.2005; Drmanac et al., Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays, *Science*, vol.327, p.78-81.2009; McKernan et al., Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding, *Genome Res.*, vol.19, p.1527-41.2009; Rodrigue et al., Unlocking short read sequencing for metagenomics, *PLoS One*, vol.28, e11840.2010; Rothberg et al., An integrated semiconductor device enabling non-optical genome sequencing, *Nature*, vol.475, p.348-352.2011; Margulies et al., Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, vol.437, p.376-380.2005; Rasko et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany, *N. Engl. J. Med.*, Epub.2011; Hutter et al., Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups, *Nucleos. Nucleot. Nucl.*, vol.92, p.879-895.2010; Seo et al., Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides, *Proc. Natl. Acad. Sci. USA.*, Vol.102, P.5926-5931 (2005); Olejnik et al.; Photocleavable biotin derivatives: a versatile approach for the isolation of biomolecules, *Proc. Natl. Acad. Sci. U.S.A.*, vol.92, p.7590-7594.1995; US 5,750,34; US 2009/0062129和US 2009/0191553。

[0035] 通常,术语“核酸分子”、“核酸序列”、“核酸片段”、“寡核苷酸”和“多核苷酸”可互换使用并且旨在包括但不限于可以具有各种长度的核苷酸的聚合物形式,脱氧核糖核苷酸(DNA)或核糖核苷酸(RNA)、或其类似物。寡核苷酸通常由4个核苷酸碱基:腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、和胸腺嘧啶(T)(当多核苷酸是RNA时,尿嘧啶(U)替换胸腺嘧啶(T))的特定序列组成。因此,术语“寡核苷酸序列”是多核苷酸分子的字母表示;可替换地,术语可以应用于多核苷酸分子本身。可以将这种字母表示输入在具有中央处理单元的计算机中的数据库中并且用于生物信息学应用如功能基因组学和同源性搜索。寡核苷酸可以可选地包括一种或多种非标准的核苷酸、核苷酸类似物和/或经修饰的核苷酸。

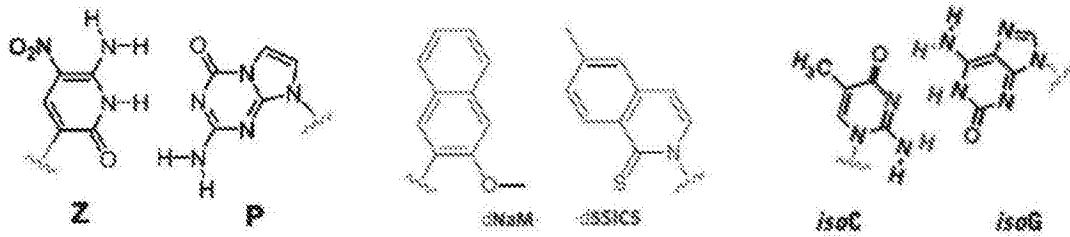
[0036] 本公开内容的实施方式包括本领域技术人员已知的天然存在的或合成的低聚物。这种低聚物包括寡核苷酸或多核苷酸(如DNA或RNA)、多肽(如胶原蛋白和万古霉素)、聚酮化合物(如脂肪和四环素)、多糖(如纤维素和淀粉)、多萜(如胆固醇和橡胶)、聚氨基酸(如木质素和聚生物碱)、多吡咯(如血红素和维生素B12)、以及聚酯(如PHA、PHV)。低聚物包括具有仿生类键的那些。参见US2008/0096253的表1。另外的低聚物包括非生物聚合物,如线性聚合物,包括聚硅氧烷、聚丙烯酰胺等。这种低聚物可以具有足够的热稳定性或易于在纳米孔或其他聚合物测序装置中检测。

[0037] 经修饰的核苷酸的实例包括但不限于,二氨基嘌呤、S<sup>2</sup>T、5-氟尿嘧啶、5-溴尿嘧啶、5-氯尿嘧啶、5-碘尿嘧啶、次黄嘌呤、黄嘌呤(xantine)、4-乙酰胞嘧啶、5-(羧基羟甲基)尿嘧啶、5-羧基甲基氨基甲基-2-硫代尿苷、5-羧基甲基氨基甲基尿嘧啶、二氢尿嘧啶、β-D-半乳糖Q核苷(beta-D-galactosylqueosine)、肌苷、N6-异戊烯基腺嘌呤、1-甲基鸟嘌呤、1-甲基肌苷、2,2-二甲基鸟嘌呤、2-甲基腺嘌呤、2-甲基鸟嘌呤、3-甲基胞嘧啶、5-甲基

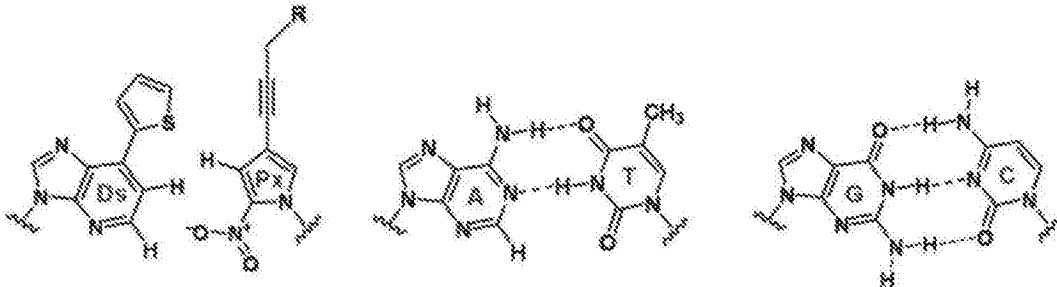
胞嘧啶、N6-腺嘌呤、7-甲基鸟嘌呤、5-甲基氨基甲基尿嘧啶、5-甲氧基氨基甲基-2-硫代尿嘧啶、β-D-甘露糖Q核苷、5'-甲氧基羧基甲基尿嘧啶、5-甲氧基尿嘧啶、2-甲硫基-D46-异戊烯基腺嘌呤、尿嘧啶-5-羟乙酸(v)、怀丁昔(wybutoxosine)、假尿嘧啶、Q核苷、2-硫代胞嘧啶、5-甲基-2-硫代尿嘧啶、2-硫代尿嘧啶、4-硫代尿嘧啶、5-甲基尿嘧啶、尿嘧啶-5-羟乙酸甲酯、尿嘧啶-5-羟乙酸(v)、5-甲基-2-硫代尿嘧啶、3-(3-氨基-3-N-2-羧基丙基)尿嘧啶、(acp3)w、2,6-二氨基嘌呤等。还可以在碱基部分(例如,在通常可用于与互补核苷酸形成氢键的一个或多个原子处和/或在通常不能够与互补核苷酸形成氢键的一个或多个原子处)、糖部分或磷酸酯骨架处,修饰核酸分子。核酸分子还可以包含胺修饰基团,如氨基烯丙基-dUTP(aa-dUTP)和氨基己基丙烯酰胺-dCTP(aha-dCTP),以允许胺活性部分,如N-羟基琥珀酰亚胺酯(NHS)的共价连接。

[0038] 在本公开内容的寡核苷酸中的标准DNA碱基对或RNA碱基对的替代物可以以位/立方毫米提供较高的密度、更高的安全性(耐天然毒素的偶然或有目的的合成)、更容易辨别光程序性聚合酶(photo-programmed polymerases)、或较低的二级结构。在Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A (2012) KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry, *Nature Chem. Biol.* 8:612-614; Seo YJ, Malyshev DA, Lavergne T, Ordoukhanian P, Romesberg FE. *J Am Chem Soc.* 2011 Dec 14; 133(49): 19878-88, Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs; Switzer CY, Moroney SE, Benner SA. (1993) *Biochemistry.* 32(39):10489-96. Enzymatic recognition of the base pair between isocytidine and isoguanosine; Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S, Hirao I. *Nucleic Acids Res.* 2012 Mar; 40(6):2793-806. Highly specific unnatural base pair systems as a third base pair for PCR amplification; 以及 Yang Z, Chen F, Alvarado JB, Benner SA. *J Am Chem Soc.* 2011 Sep 28; 133(38):15105-12, Amplification, mutation, and sequencing of a six-letter synthetic genetic system 中描述了与用于从头和/或扩增合成的天然和突变聚合酶相容的这种替代碱基对。

[0039] 已经表明以下6对(A-T、G-C、Z-P、Ds-Px、NAM-SSICS、isoC-isoG(异胞嘧啶-异鸟嘌呤))与聚合酶相容并且彼此正交(即,低水平的交叉配对)。



[0040]



[0041] 在一些示例性实施方式中,可以利用一种或多种亚磷酰胺连接物和/或通过本领域技术人员已知的连接法测序,制备寡核苷酸序列。还可以通过任何合适的方法制备寡核苷酸序列,例如,标准亚磷酰胺方法如以下本文中描述的那些以及由Beaucage和Carruthers((1981) *Tetrahedron Lett.* 22:1859)描述的那些或根据Matteucci et al. (1981) *J. Am. Chem. Soc.* 103:3185)的三酯方法,或通过本领域中已知的利用商用自动化寡核苷酸合成仪或高通量、高密度阵列方法的其他化学方法(参见美国专利号5,602,244、5,574,146、5,554,744、5,428,148、5,264,566、5,141,813、5,959,463、4,861,571和4,659,774,出于所有目的通过引用将其全部内容结合于本文中)。还可以从多种供应商商业上获得预合成的寡核苷酸。

[0042] 在一些示例性实施方式中,可以利用本领域中已知的多种微阵技术制备寡核苷酸序列。预合成的寡核苷酸和/或多核苷酸序列可以连接至载体或利用以下参考文献中陈述的光引导方法、流道和点样方法、喷墨方法、基于针的方法和基于珠子方法原位合成: McGall et al. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:13555; *Synthetic DNA Arrays In Genetic Engineering*, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) *Nat. Genet.* S21:10; *Microarrays: Making Them and Using Them In Microarray Bioinformatics*, Cambridge University Press, 2003; 美国专利申请公开号2003/0068633和2002/0081582; 美国专利号6,833,450、6,830,890、6,824,866、6,800,439、6,375,903和5,700,637; 以及PCT申请号WO 04/031399、WO 04/031351、WO 04/029586、WO 03/100012、WO 03/066212、WO 03/065038、WO 03/064699、WO 03/064027、WO 03/064026、WO 03/046223、WO 03/040410和WO 02/24597。

[0043] 在一些示例性实施方式中,将本文描述的一种或多种寡核苷酸序列固定在载体(例如,固体和/或半固体载体)上。在一些方面,可以利用本文描述的一种或多种亚磷酰胺连接物将寡核苷酸序列连接至载体。合适的载体包括但不限于,载玻片、珠子、芯片、颗粒、线(strand)、凝胶、片、管、球体、容器、毛细管、垫、薄片、膜、板等。在各种实施方式中,固相载体可以是生物的、非生物的、有机的、无机的、或它们的任何组合。当使用基本上是平面的载体时,载体可以被物理地分成区域,例如,借助于沟、沟槽、孔、或化学障碍物(例如,疏水性涂层等)。在一些示例性实施方式中,载体是微阵列。如在本文中使用的,在一个实施方式

中,术语“微阵列”是指包括具有其上存在空间上限定的各自包括固定的杂交探针的非重叠区域或部位的阵列的基本上平坦表面的固相载体的类型的阵列。“基本上平坦”是指表面上的感兴趣的特征或目的,如探针部位,可以占据在表面上方或下方延伸并且其尺寸相对于表面的尺寸较小的体积。例如,设置在光纤束的面上的珠子产生探针部位的基本上平坦的表面,或设置或合成在多孔平坦基底(基质,基板,substrate)上的寡核苷酸产生基本上平坦的表面。空间上限定的部位可以另外地是“可寻址的”,因为其位置和在该位置处固定的探针的识别是已知的或可确定的。

[0044] 固定在微阵列上的寡核苷酸包括在测定反应中生成或来自测定反应的核酸。通常,在微阵列上的寡核苷酸或多核苷酸是单链的并且通常通过5'-端或3'-端共价连接至固相载体。在一些示例性实施方式中,探针经由本文描述的一种或多种可切割连接物固定。以微阵列包含核酸的非重叠区域的密度通常大于 $100/\text{cm}^2$ ,并且更通常地,大于 $1000/\text{cm}^2$ 。与核酸探针相关的微阵列技术综述于以下示例性参考文献中:Schena,Editor,Microarrays:A Practical Approach (IRL Press,Oxford,2000);Southern,Current Opin.Chem.Biol.,2:404-410(1998);Nature Genetics Supplement,21:1-60(1999);以及Fodor等人,美国专利号5,424,186、5,445,934、和5,744,305。

[0045] 将寡核苷酸固定至载体的方法是本领域中已知的(珠子:Dressman et al.(2003) Proc.Natl.Acad.Sci.USA 100:8817,Brenner et al.(2000) Nat.Biotech.18:630,Albretsen et al.(1990) Anal.Biochem.189:40,and Lang et al.Nucleic Acids Res.(1988) 16:10861;硝化纤维素:Ranki et al.(1983) Gene 21:77;纤维素:Goldkorn(1986) Nucleic Acids Res.14:9171;聚苯乙烯:Ruth et al.(1987) Conference of Therapeutic and Diagnostic Applications of Synthetic Nucleic Acids,Cambridge U.K.;特氟隆-丙烯酸酰胺:Duncan et al.(1988) Anal.Biochem.169:104;聚丙烯:Polisky-Cynkin et al.(1985) Clin.Chem.31:1438;尼龙:Van Ness et al.(1991) Nucleic Acids Res.19:3345;琼脂糖:Polisky-Cynkin et al.,Clin.Chem.(1985) 31:1438;以及聚丙烯酰胺葡聚糖(sephacryl):Langdale et al.(1985) Gene 36:201;胶乳:Wolf et al.(1987) Nucleic Acids Res.15:2911)。

[0046] 如在本文中使用的,术语“连接”是指共价相互作用和非共价相互作用两者。共价相互作用是在两个原子或基团之间通过共享一对电子(即,单键)、两对电子(即,双键)或三对电子(即,三键)形成的化学键。共价相互作用在本领域中还称作电子对相互作用或电子对键。非共价相互作用包括但不限于,范德华相互作用、氢键、弱化学键(即,经由短程的非共价力)、疏水相互作用、离子键等。非共价相互作用的综述可以在Alberts et al.,in Molecular Biology of the Cell,3d edition,Garland Publishing,1994中找到。

[0047] 根据本公开内容的测序引物是能够结合至靶多核苷酸的已知结合区并且促进本公开内容的寡核苷酸探针的连接的那些。可以借助于计算机程序如,例如,DNAWorks、或Gene201igo,设计测序引物。结合区的长度可以变化,但其应足够长以杂交测序引物。靶多核苷酸可以具有多个不同的结合区,从而允许对靶多核苷酸的不同部分进行测序。选择测序引物以形成高度稳定的双链体,使得在连接的连续循环中它们保持杂交。可以选择测序引物,使得可以在5'至3'方向或3'至5'方向或在两个方向上进行连接。测序引物可以包含经修饰的核苷酸或键以增强它们的杂交效率,或改善它们的稳定性,或防止自一个末端或

另一个末端的延伸。

[0048] 根据一个方面,通过如以上描述的与测序引物一起使用的RCA制备单链DNA模板(ssDNA)。可替换地,将单链模板连接至乳液中的珠子或纳米颗粒,并且通过ePCR扩增。结果是具有单一扩增的ssDNA模板的克隆珠子。

[0049] 为了识别多个平行模板核苷酸序列的目的,将模板在pH 7.4的PBS缓冲液中稀释,并且利用多种连接方法,如生物素-链霉亲和素、叠氮化物-烷基化物(azide-alkyle)(例如,点击化学)、NHS-酯或硅烷化(例如醛硅烷、环氧硅烷、氨基硅烷)结合至图案化或非图案化基底。根据一个方面,将克隆(rolonies)连接至图案化表面,如SiO<sub>2</sub>固体表面,用1%的氨基硅烷(v/v)处理并且允许相互作用一段时间(通常为5分钟至2小时之间)。然后,使用洗涤1缓冲液洗去任何未结合的模板。

[0050] 接下来,制备测序引物并且杂交至测序引物杂交部位。根据一些方面,可以制备可以杂交至模板的已知序列的测序引物。可替换地,在模板制备期间,根据本领域技术人员已知的并且在本文中描述的方法,通过连接、扩增、转位或重组将具有已知核酸序列的连接物加入至未知的核酸序列。仍然可替换地,具有一定水平的简并的测序引物可以用于沿着模板杂交至一些位置。根据一个方面,引物简并用于允许引物沿着模板半随机地杂交。基于本领域技术人员已知的统计方法选择引物简并以促进引物沿着模板的长度以一定的间隔进行杂交。根据该方面,可以设计具有促进结合每N个碱基,如每100个碱基、每200个碱基、每2000个碱基、每100,000个碱基的一定简并性的引物。引物沿着模板的长度的结合基于引物的设计和引物设计将沿着模板的长度结合约每N个碱基的统计可能性。因为测序引物P1将通过连接延伸,所以通常合成测序引物的末端基团以准备好通过DNA连接酶共价连接至寡核苷酸探针。如果在测序引物的5'端和寡核苷酸探针的3'端之间发生连接,则磷酸基团(5'-P<sub>04</sub>)必须存在于测序引物上同时羟基基团(3'-OH)必须存在于寡核苷酸探针上,并且反之亦然。为了将测序引物杂交至测序引物杂交部位,使用稀释在5X SSPE缓冲液中的1μM的测序引物。然后,在室温以上温育混合物几分钟以促进适当退火(在25°C和55°C之间的温度下,通常在1至5分钟之间)。

[0051] 根据一些方面,可以利用本领域技术人员已知的喷墨技术、本领域技术人员已知的电化学技术、本领域技术人员已知的微流技术、本领域技术人员已知的光生酸、或本领域技术人员已知的光脱保护单体,制备寡核苷酸序列。这种技术具有以下优势:以高速、低成本、更少的有毒化学物质、增强的便携性和交错DNA生物化学(例如修饰、聚合酶、杂交等)与从头(数字或模拟)合成的能力,制备寡核苷酸。例如,直接来自照相机光学或来自数字显微镜显示装置(DMD)的空间上图案化的光,可以与水溶液化学物质一起使用。参见US2003/0228611。例如,通过将偶氮苯氨基酸(参见Hoppmann C,Schmieder P,Heinrich N,Beyermann M.(2011)Chembiochem.12(17):2555-9.doi:10.1002/cbic.201100578.Epub 2011 Oct 13.Photoswitchable click amino acids:light control of conformation and bioactivity)结合在聚合酶或5'→3'外切核酸酶域(如果存在)的活性位点,模板非依赖性聚合酶如末端转脱氧核苷酰酶(TdT)或多(A)聚合酶,可替换地,模板依赖性聚合酶如Taq或Phi29衍生物,可以具有它们的基本聚合酶功能、碱基特异性或通过光可编程的保真度(精确度)。

[0052] 光敏神经元(光遗传学)可以触发离子敏感聚合酶(参见Zamft B,Marblestone A,

Kording K, Schmidt D, Martin-Alarcon D, Tyo K, Boyden E, Church GM (2012) Measuring Cation Dependent DNA Polymerase Fidelity Landscapes by Deep Sequencing. PLoS One, 出版中), 或, 对于一些应用, 离子流模式本身可以构成存储的数据集。

[0053] 可以在具有或不具有聚合酶扩增的情况下存储和读取从头聚合物。扩增可以经由热循环或等温循环。由于对目前的化学合成是方便的, 所以扩增子可以是短的 (100至200聚体), 或长达1Mbp, 如借助于聚合酶可实现的。

[0054] 通过以下可以确定结合的核苷酸类型: a) 在dNTP溶液的循环模式 (图案) 中, 与在那个时间点下存在的特定dNTP (或rNTP或其他单体类) 一致的光脉冲的交叉, b) ‘笼中的’ (即, 可光活化的或可光灭活的) dNTP、rNTP或阳离子, c) 碱基特异性、光调节位阻或构象的选择性 (参见Hoppmann C, Schmieder P, Heinrich N, Beyermann M. (2011) Chembiochem. 12 (17): 2555-9. doi:10.1002/cbic.201100578. Epub 2011 Oct 13. Photoswitchable click amino acids: light control of conformation and bioactivity)。多(A)聚合酶是特别有用的, 因为相对于其他rNTP, 其对于ATp的特异性是由于可以通过光敏氨基酸键 (如偶氮苯, 具有或不具有交联) 模拟的构象变化。

[0055] 通过不应当解释为限制性的以下实施例进一步说明本发明。出于所有目的, 通过参照, 将贯穿本申请引用的所有参考文献、专利和公开的专利申请的内容以它们的全部相结合于此。

[0056] 实施例I

[0057] 利用DNA存储写入信息

[0058] 选择由Basic Books (纽约, NY) 出版的称作Regenesis (Church GM and Regis E) 的草稿书的HTML版本以表明现代格式、图像、和Java脚本。与典型的网页一样, 使用了通用字集转换格式、8位 (UTF-8)、与用于特殊字符和字体的ASCII和UNICODE向后兼容的可变宽度编码。存在黑白混合的并且JPEG编码的11个图像 (在很少的质量损失下, 通常为10:1数据压缩)。以base64格式, 将这些“内联”嵌入 (即, 不分开文件) 在html中。在任何这些JPEG段的中部的一致位误差 (比特误差) 将仅影响在该段内的数据下游。文本中的位误差将最多影响包含误差的寡核苷酸中的12个字符。Java脚本是可以动态弯曲以跟随光标位置的37字节文本字符串 (遗传密码的助记符编码) 的简单显示。根据本公开内容, DNA (如其他数字媒介) 可以编码用于数字计算机的可执行指令。

[0059] 实施例II

[0060] 将写入信息编码成DNA

[0061] 将书转换到html格式 (具有嵌入式jpg图像)。以位形式阅读书, 并且然后将单个位转换成对于0为A或C以及对于1为T或G。随机选择碱基, 同时不允许均聚物运行大于3次。位流的地址是19位长并且连续编号, 从000000000000000001开始。脚本Bits2DNA.pl (参见代码部分) 是用于将html文件编码成DNA片段的程序。

[0062] 实施例III

[0063] 合成和扩增

[0064] 在安捷伦寡核苷酸库合成微阵列平台 (Agilent Oligo Library Synthesis microarray platform) 上合成了54,898个寡核苷酸。通过安捷伦洗脱DNA以产生100μL TE (10mM Tris-Cl pH 7.5, 0.1mM EDTA) 中的~10皮摩尔寡核苷酸的池。

[0065] 如下扩增库。利用BioRad CFX96实时PCR机器中的Sybr Fast Master Mix (Kapa Biosystems),使用各自200nM的引物MD-测试-1F和MD-测试-1R用于6次循环,将1 $\mu$ L(预期~10飞摩尔)的库用于50 $\mu$ L的PCR扩增反应中,并且监测扩增期间的Sybr绿色通道。

[0066] 95 $^{\circ}$ C,3分钟

[0067] 95 $^{\circ}$ C,10秒

[0068] 60 $^{\circ}$ C,30秒

[0069] 阅读(读取) Sybr绿色通道

[0070] 转到步骤2,总共10个循环

[0071] 68 $^{\circ}$ C,30秒

[0072] 保持在4 $^{\circ}$ C下

[0073] 根据制造商的说明,利用Qiagen MinElute PCR净化柱将得到的PCR产物纯化进入10 $\mu$ L的缓冲液EB (10mM Tris-Cl, pH 8.5)中。洗脱的DNA产生如通过NanoDrop 2000c(纳米微滴2000c)分光光度计测定的36.8ng/ $\mu$ L (A260/A280=1.85)的浓度。

[0074] 使用相同条件但这次使用200nM的PE-PCR引物1.0-F和PE-PCR引物2.0-R,扩增两管的1 $\mu$ L的1:11稀释的(在水中)扩增反应9个循环。根据制造商的建议(Agencourt),利用Ampure珠子净化PCR反应以除去残余引物并且再悬浮在50 $\mu$ L的TE中。最终产物是~22ng/ $\mu$ L,如通过纳米微滴和琼脂糖凝胶成像两者所量化的。使用的引物如下,其中\*表示硫代磷酸酯键。

[0075] >MD测试1-F

[0076] ACACTCTTCCCTACACGACGCTCTTCCGATC\*T

[0077] >MD测试1-R

[0078] CTCGGCATTCTGCTGAACCGCTCTTCCGATC\*T

[0079] >PE PCR引物1.0-F

[0080] AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATC\*T

[0081] >PE PCR引物2.0-R

[0082] CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATC\*T

[0083] 实施例IV

[0084] 测序和处理

[0085] 使用配对末端100次读数,通过将1mL的14pM库(14飞摩尔;1:1000的扩增库)负载在HiSeq 2000的单泳道上对扩增库进行测序。从泳道,我们得到具有87.14%  $\geq$  Q30和34.16的平均Q得分的346,151,426百万个配对读数。因为正在对具有配对100bp读数的115bp构建物进行测序,所以SeqPrep,参见11 J.St.John,SeqPrep <https://github.com/jstjohn/SeqPrep> (2011),用于利用以下命令(对于HiSeq泳道的单个板(tile))将重叠读取结合成单个重叠群:SeqPrep -f MTMC2\_NoIndex\_L006\_R1\_002.fastq.gz -r MTMC2\_NoIndex\_L006\_R2\_002.fastq.gz -1 tile2r1.fastq.gz -2 tile2r2.fastq.gz -s tile2-merged.fastq.gz -E tile2-align.txt.gz -o 50 -m 0.1 -n 0.8。

[0086] 在SeqPrep以后,形成292,371,030个重叠群。利用Bowtie2版本2.0.0- $\beta$ 5参见17 B.Langmead,S.L.Salzberg,Nature Methods 9,357-360 (2012)和SamTools版本0.1.18参见18 H.Li et al.,Bioinformatics 25,2078-2079 (2009),利用以下命令:zcat\*merged\*|

bowtie2 -p 10 --end-to-end-x../.../agilentlib-U-|samtools view-bS->alltiles-merged.bam,将重叠群与参比对准(align)。

[0087] 在对准之后,将267,993,389 (92%) 对准合成库的一个成员,产生 $4882 \pm 1261$ 的平均覆盖( $\pm 1$ 标准偏差)。过滤产生全长115bp重叠群的SeqPrepped重叠群,导致190,284,472个读数和 $3419 \pm 998$ 的平均覆盖。平均而言,对于库的每个成员, $\sim 69.5\% \pm 0.4$ 的读数具有全长。具有最少读数的构建物是寡核苷酸md-37545,其在115bp过滤前后具有94和9个读数;得到的共有序列对于该寡核苷酸(oligo)仍然是正确的。

[0088] 图2是设计的库的每个成员的观测数目的直方图。利用箱中的并且绘制的(红色)的Bowtie2,将由SeqPrep形成的重叠群(即,在读取之间具有重叠)的所有读数对准合成的库。仅对于长度为115bp的重叠群,以绿色显示相同的信息。由分布尾来看,插入显示为放大。

[0089] 实施例V

[0090] 误差

[0091] 从共有序列库,在总结在以下表1中的设计和读取序列之间发现差异(偏差)。如示出的,发现22个差异,其中的10个导致位误差(比特误差)(粗体)。大多数的误差(20/22)位于其中在测序期间仅存在单覆盖的序列的最后15个碱基内。此外,大多数的误差(18/22)导致至少3个连续重复核苷酸的运行。筛出4个或更多个重复核苷酸的均聚物读数(灰色框)将导致12个差异,其中7个是位误差。

[0092] 表1

[0093]

条码	误差位置	误差类型	位误差	参比上下文	读取上下文	均聚物
AACTGTCGCTATTC ACTCA	115	A->G	是	CAC-A-	CAC-G-	否
ACTAACGCACCTG GAATCA	106	A->C	否	CTT-A-C CT	CTT-C-C CT	否
ACTTTCGGCATGA GTACCC	103	A->C	否	CGC-A-C CC	CGC-C- CCC	是
AGCGGCTCGTCGG TGTC	40	T->C	是	GGT-T-C CG	GGT-C-C CG	否
ATACGGCTCATTAC AAACC	105	G->C	是	TCT-G-C CC	TCT-C-C CC	是
ATGCGGGCAAATC ACAGCA	106	A->C	否	AAC-A-C CT	AAC-C- CCT	是
ATGGCCGTAATGG AGAAAC	102	C->A	否	TAG-C-A AG	TAG-A- AAG	否

[0094]

ATGTTCTGAATTAG CGCCC	108	C->G	是	CAA-C-G AG	CAA-G- GAG	否
CAATGTAGATCCTC GAAAC	106	A->C	否	CAG-A-C CC	CAG-C- CCC	是
CCGGCCTAAACGG CACGCC	106	A->C	否	CTC-A-C CT	CTC-C-C CT	是
CGATATTCGGGAA CACCCA	102	G->C	是	AAC-G-C CC	AAC-C- CCC	是
CGATATTCGGGAA CACCCA	106	A->C	否	CCC-A-C CT	CCC-C- CCT	是
CGGCGGAGCGGAG ACGCCA	106	C->A	否	AGG-C-A AG	AGG-A- AAG	否
CTGCTCTTCAACC GCTACA	115	T->G	否	CGA-T- GCA-A-C	CGA-G- GCA-C-	否
GGTAATTTCTAGTA CAGCC	105	A->C	否	CC	CCC	是
GGTCGCATAAACT TGACCC	105	A->G	是	CGC-A-C GA	CGC-G- GGA	否
GGTCGCATAAACT TGACCC	106	C->G	是	GCA-C-G AG	GCG-G- GAG	否
GTGCCAATAAAGT GGTCCC	102	T->C	是	TCG-T-C CG	TCG-C- CCG	否
GTGCCAATAAAGT GGTCCC	106	C->A	否	CCG-C-A AG	CCG-A- AAG	否
GTGTCCCACCCAC CCACCC	83	A->G	是	ACA-A-C TG	ACA-G- CTG	否
TCCCAGGCAGCTA CCCGCA	102	T->C	是	GCG-T-C CC	GCG-C- CCC	是
TGACGCGCCGGTT GGGCCC	106	A->C	否	ACC-A-C CT	ACC-C- CCT	是

[0095] 表1示出在设计 and 读取库之间的差异。每个误差是一行，显示与误差相关的条码、在寡核苷酸(在115之外)中的位置、误差类型，是否误差导致位变化、初始上下文、和新上下文(误差位置在短线的中部)，以及最后是否误差导致可以已经被滤出的4个碱基的运行。导致位误差的线被粗体化，并且基于4个连续碱基的运行的可以已经被过滤的线被阴影化。

[0096]

类型	标记	日期	参比	位	位 /mm <sup>3</sup>	注释
商用	CD	1982	(19)	5.6e9	4.13e5	1.2 mm 厚 CD; 120mm 直径
商用	DVD-SL	1996	(20)	3.76e10	2.77e6	1.2 mm 厚 DVD-SS-SL; 120mm 直径
商用	DVD-QL	2000	(20)	1.37e11	1.01e7	1.2 mm 厚 DVD-DS-DL; 120mm 直径
商用	蓝光 SL	2002	(21)	2.00e11	1.47e7	1.2 mm 厚蓝光光盘(1层)
商用	蓝光 QL	2010	(21)	1.02e12	7.52e7	1.2 mm 厚蓝光光盘(XL 4层)

[0097]

商用	磁带	2010	(22)	4.00e13	5.59e8	Oracle StorageTek T10000 T2 - 5TB 5.2 $\mu$ m 厚度, 1147m 长度, ~12 mm 宽
商用	闪存	2012	(23,24)	1.02e12	5.02e9	NAND 闪存; 用于单芯片的密度的 Sandisk (22) 在 170mm <sup>2</sup> 中, 128 吉位; 取自 Toshiba 芯片堆积的 150 $\mu$ m 深度 (23)
商用	硬盘	2012	(25)	4.80e13	3.10e9	硬驱->希捷 (Seagate) 1 兆兆位/英寸 <sup>2</sup> =1.55e9 位/mm <sup>2</sup> =1.55e9 位/mm <sup>2</sup> , 假设 ds1mm 盘
演示	12 原子存储器	2012	(26)	8	1.11e12	9 nm <sup>2</sup> /位 (假设 100 $\mu$ m 厚度) 低温非易失性内存
演示	Xe 定位	1991	(27)	70	1E13	具有在 14 x 5 nm <sup>2</sup> 晶格上分开的间隔 1nm 的 Xe 原子的拼写 IBM; 1 位/nm <sup>2</sup> ; 假设 100 $\mu$ m 厚度
演示	量子全息术	2008	(28)	3.5E+1	1.38e13	35 位图像对, 17x17 nm <sup>2</sup> 架空原子和 4x5 读取空间 =((4x5)/(17x17)) *20 位/nm <sup>2</sup> =1.38 位/nm <sup>2</sup> =1.38e12 位/mm <sup>2</sup> ; 假设 100 $\mu$ m 厚度
生物的	超分辨率 GFP	2011	(29)	27	4.0E10	9 3x3 位字段 (81 位); 250nm 中心-中心间隔; 1 位/250nm <sup>2</sup> ; 假设 100 $\mu$ m 厚度
生物的	在大肠杆菌中的 DNA	1988	(7,30)	35	5.0e10	大肠杆菌, 0.7 $\mu$ m <sup>3</sup> , 来自 (29)
生物的	在大肠杆菌中的 DNA	2001	(3,30)	561	8.01e11	大肠杆菌, 0.7 $\mu$ m <sup>3</sup> - 118 字符 (27 种可能性) =27118=2x; x=561
生物的	在大肠杆菌中的 DNA	2005	(8,30)	1007	1.44e12	大肠杆菌, 0.7 m <sup>3</sup> - 233 字符 (20 种可能性) =20233=2x; x=1007
生物的	支原体的	2010	(9)	7920	8.80E+13	支原体, ~0.09 $\mu$ m <sup>3</sup> 的体积;
生物的	本工作的	2012		5.27e6	5.49e15	假设 1e-3g/mm <sup>3</sup> ; 330.95 g/mol/核苷酸; 96 位/159bp; 100x 倍覆盖; 330.95*2*159=105242.1 g/mol=1.748e-19 g/分子 =1.748e-16 g/1000 个分子 =1.748e-13 mm <sup>3</sup>

[0098] 实施例VI

[0099] 关于数据密度的计算

[0100] 通过利用2位/单链DNA的核苷酸计算理论DNA密度。我们使用的DNA的分子量基于ATGC平衡库的钠盐的无水重量的330.95g/mol/核苷酸的平均值。这导致1位/ $2.75 \times 10^{-22}$ g的重量密度,并且因此可以将 $1.8 \times 10^{21}$ 字节存储在3.96g中。当然,根据冗余的类型、条形编码、和所期望的编码方案,实际最大值将是较少稠密的多个数量级。这种理论上的最大值计算没有用于图1B中。

[0101] 图1B上绘制的数据是在非常不同的技术之间的比较。在其中未记录厚度的平面密度计算的情况下,选择100 $\mu$ m作为深度(这比硬驱盘小 $\sim 10$ x,并且比目前的闪存堆积小33%)。假定了当前的信息编码密度(96位/159bp)、和在存储器中DNA的100x合成的覆盖。假定了 $1\text{g}/\text{cm}^3$ (纯水的密度)的近似体积,其可能是干燥的DNA的轻微低估。

[0102] 表2

[0103] 表2示出了在图1B中使用的数据,以将在不同介质中的数据密度进行比较。为了将用于信息编码的截然不同的技术进行比较,通过做出各种假设,将所有数据密度信息转换成体积数据密度。对于商用技术,使用其中可获得的关于基底厚度的可获得信息。在闪存的情况下,使在类数据密度方面最好的与来自不同制造商的芯片堆积厚度结合。为了证明(演示, demonstration)数据存储技术,未报道基底厚度。因此,假定了100 $\mu$ m厚度,其是堆叠闪存技术的目前的厚度的1/3。对于与所有演示在4°K下和在真空中进行一样公布的那些技术,这可能是不切实际的。干燥的DNA的密度接近于水的密度,其可以是低估数据。对于使用克隆的DNA的其他生物演示,单个细胞的体积用作体积。最后,灰色行未示于图1B中,因为它们被其他数据点所遮盖,但为了完整性,包括在此。

[0104] 实施例VII

[0105] 代码

[0106]

```
Bits2DNA.pl
# cd "\Perl\gmc\Bin_DNA"
#\Perl\bin\perl Bits2DNA.pl GMC Jul-2011 & 27-May-2012
# docstore.mik.ua/oreilly/perl/cookbook/ch02\_05.htm (bin) ch01\_05.htm (char)
```

[0107]

```

# http://perldoc.perl.org/functions/pack.html rand.html
# Each oligo is L(19)+8N(12)= 115 bp, long flanked by 22-mer
amplification primers.
# DNA Encoded Artifacts Registry (DEAR) to coordinate global standards.

open IN,"in.html"; open OUT,">Bits2DNA.txt"; binmode IN;
${"0"}="a"; ${"1"}="G"; # lowercase a,c = zero bit.
${"a"}="c"; ${"G"}="T"; ${"c"}="a"; ${"T"}="G";
$u1=""; $u2=""; $u3=""; # Initialize; keep homopolymer runs < 4
$N=12; # Length of segment in bytes (not including segment number)
$L=19; # 2^19 = 524,288 = max number of oligos L=00010011
$seed=2; srand($seed); # remove this line to get a random seed
print int2bp(262144)," ",int2bp(262145);
$f="CTACACGACGCTCTTCCGATCT"; # forward 'universal'
sequencing & amplification primer
$r="AGATCGGAAGAGCGGTTCAGCA"; # reverse 22-mer primer

$sn=0; print OUT $f,int2bp(0),""; ###
while (read (IN, $text, 65536)) {
    @ascii_num = unpack("C*", $text);
    foreach $val (@ascii_num) {
        print OUT byt2bp($val); ###
        $sn++;
        if($sn%$N==0){
            print OUT $r,"\n",$f,int2bp($sn/$N),""; ###
        } # N bases per output line
        } # each byte
    } # 65 Kbytes
for ($k=$sn%$N; $k<$N; $k++){
    print OUT byt2bp(int(rand(256))); ###
}

```

```
} # pad last data line to keep all oligos same size.
  print OUT "$r\n"; ###

sub byt2bp { # convert rightmost 8 bits (MSB first byte) to 8 bp
  my $b = unpack("B32", pack("N", shift));
  $p="";
  for ($i=24; $i<=31; $i++){
    $x=substr($b,$i,1); # bits 24 to 31 inclusive
    $u=$t{$x};
    if(rand(2)<1){$u=$t{$u};} # pick synonym a=c; G=T
    if(($u eq $u1) && ($u eq $u2) && ($u eq $u3)){ $u=$t{$u};}
    $u1=$u2; $u2=$u3; $u3=$u; # Shift previous base string
    $p = $p.$u;
  }
[0108] return $p;
}

sub int2bp { # convert rightmost $L bits of 32 bit integers to $L bp
  my $b = unpack("B32", pack("N", shift));
  $p="";
  for ($i=31; $i>=32-$L; $i--){
    $x=substr($b,$i,1); # bits 31 to SL
    $u=$t{$x};
    if(rand(2)<1){$u=$t{$u};} # pick synonym a=c; G=T
    if(($u eq $u1) && ($u eq $u2) && ($u eq $u3)){ $u=$t{$u};}
    $u1=$u2; $u2=$u3; $u3=$u; # Shift previous base string
    $p = $p.$u;
  }
  return $p;
}
```

```
buildConsensus.py
import sys

#builds consensus sequence from individual base counts
def getConsensus(finalbuckets):
    sequence = ""
    for i in range(len(finalbuckets)):
        letterindex = finalbuckets[i].index(max(finalbuckets[i]))
        if letterindex == 0:
            sequence += 'A'
        elif letterindex == 1:
            sequence += 'C'
        elif letterindex == 2:
            sequence += 'G'
        elif letterindex == 3:
            sequence += 'T'
    return sequence

oligolength = 115
currentbarcode = ""
#initialize vector to building consensus
buckets = [[0 for col in range(4)] for row in range(oligolength)]

for line in sys.stdin:
    splitline = line.split()
    count = int(splitline[0])
    barcode = splitline[1]
```

[0109]

```
sequence = splitline[2]
if not barcode == currentbarcode:
    if not currentbarcode == "":
        print getConsensus(buckets)

buckets = [[0 for col in range(4)] for row in range(oligolength)]
currentbarcode = barcode
for i in range(oligolength):
    if sequence[i] == 'A':
[0110]         buckets[i][0] += count
    elif sequence[i] == 'C':
        buckets[i][1] += count
    elif sequence[i] == 'G':
        buckets[i][2] += count
    elif sequence[i] == 'T':
        buckets[i][3] += count

#print final consensus
print getConsensus(buckets)
```

[0111] 实施例VIII

[0112] 方法总结

[0113] 根据一个方面,将包括53,426个字、11个JPG图像和1个Java脚本程序的书的html编码的草稿转换成5.27兆位,参见4。关于材料和方法的信息在科学在线上可获得。然后将位编码于54,898 159nt寡核苷酸上,其各自编码96位数据块(96nt),指明数据块在位流中的位置的19位地址(19nt),以及用于扩增和测序的旁侧22nt共同序列。通过喷墨印刷、高保真度DNA微芯片合成寡核苷酸库。参见10.E.M.LeProust et al.,*Nucleic Acids Research* 38,2522-2540 (2010)。为了阅读编码的书,通过有限循环PCR扩增库,并且然后在Illumina HiSeq的单泳道上测序。连接重叠的配对末端100nt读数以降低测序误差的影响。参见11.J.St.John,SeqPrep <https://github.com/jstjohn/SeqPrep> (2011)。然后,仅利用产生预期的115nt长度和完美的条码序列的读数,在每个数据块的每个碱基(base)处以平均约3000倍覆盖生成共有序列。用总计10位误差(在5.27百万之外)恢复所有数据块,其主要位于在寡核苷酸的末端处的均聚物运行内,其中我们仅具有单个序列覆盖,参见4。关于材料和方法的信息在科学在线上可获得。

[0114] 在100x合成覆盖下,本文公开的实施方式的密度是约5.5拍位/mm<sup>3</sup>。本文公开的实施方式的尺度是约5.27兆位。利用商业可获得的材料和仪器进行本文描述的方法步骤。图1B是示出了与其他储存介质相比,本公开内容(“本工作”)的实施方式提供了较高的信息密度的信息密度比较。根据一些方面,实施方式包括使用压缩、冗余编码、奇偶检验、和误差校正以改善密度、分布、和误差率。根据一些方面,在不存在物理隔离的情况下,还可以通过使用允许正交扩增的独特的引发位点存储DNA库。参见12.S.Kosuri et al.,*Nature Biotechnology* 28,1295-1299 (2010)。根据一些方面,除了DNA或DNA修饰之外的聚合物也可以用于最大化读取、写入、和存储能力。参见13.S.A.Benner,Z.Yang,F.Chen,*Comptes Rendus Chimie* 14,372-387 (2011)。根据一个方面,手持式、单分子DNA测序仪可以用于读取DNA编码信息。参见15.E.Pennissi,*Science* 336,534-537 (2012)。根据一个方面,利用与库合成和共有序列测序结合的寻址数据块的一般方法与商业可获得的DNA测序和合成技术兼容。根据一个方面,DNA的可替换的消费级应用如信息存储,可以加快用于改善合成和测序技术的新技术的发展。参见16.S.Kosuri,A.M.Sismour,*ACS Synth Biol* 1,109-110 (2012)。

[0115] 出于所有目的,通过引用将本文确定的参考文献和以下参考文献以它们的全部内容合并于此,并且在整个本公开内容中通过对应的数字来提及。

[0116] 3.“Extracting Value from Chaos”(IDC,Framingham,MA 2011);world wide web site emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.

[0117] 4.J.Rothenberg,*Scientific American* 272,42-47 (1995) .

[0118] 5.C.Bancroft,T.Bowler,B.Bloom,C.T.Clelland,*Science* 293,1763-1765 (2001) .

[0119] 6.Information on materials and methods is available on Science Online

[0120] 7.J.Bonnet et al.,*Nucleic Acids Research* 38,1531-1546 (2010) .

[0121] 8.S.Pääbo et al.,M.A.Uyterlinde et al.,Eds.*Annual Review of Genetics* 38,645-679 (2004) .

[0122] 9.J.Davis,*Art Journal* 55,70-74 (1996) .

[0123] 10.C.Gustafsson,*Nature* 458,703 (2009) .

[0124] 11.D.G.Gibson et al.,*Science* 329,52-56 (2010) .

[0125] 12.E.M.LeProust et al.,*Nucleic Acids Research* 38,2522-2540 (2010) .

[0126] 13.J.St.John,SeqPrep <https://github.com/jstjohn/SeqPrep> (2011)

[0127] 14.S.Kosuri et al.,*Nature Biotechnology* 28,1295-1299 (2010) .

[0128] 15.S.A.Benner,Z.Yang,F.Chen,*Comptes Rendus Chimie* 14,372-387 (2011) .

[0129] 16.P.A.Carr,G.M.Church,*Nature Biotechnology* 27,1151-62 (2009) .

[0130] 17.E.Pennissi,*Science* 336,534-537 (2012)

[0131] 18.S.Kosuri,A.M.Sismour,*ACS Synth Biol* 1,109-110 (2012)

[0132] 19.B.Langmead,S.L.Salzberg,*Nature Methods* 9,357-360 (2012) .

[0133] 20.H.Li et al.,*Bioinformatics* 25,2078-2079 (2009) .

[0134] 21.Ecma International,Data interchange on read-only 120mm optical

data disks (CD-ROM), (ECMA Standard 130, Geneva, Switzerland 1996, world wide website [ecma-international.org/publications/files/ECMA-ST/Ecma-130.pdf](http://ecma-international.org/publications/files/ECMA-ST/Ecma-130.pdf).)

[0135] 22. Ecma International, 120mm DVD-Read-Only Disk, (ECMA Standard 267, Geneva, Switzerland 2001, world wide website [ecma-international.org/publications/files/ECMA-ST/Ecma-267.pdf](http://ecma-international.org/publications/files/ECMA-ST/Ecma-267.pdf).)

[0136] 23. Blu-Ray Disc Association, White Paper - Blu-Ray Disc Format (2<sup>nd</sup> Edition, Universal City, CA 2010, world wide website [blu-raydisc.com/Assets/Downloadablefile/general\\_bluraydiscformat-15263.pdf](http://blu-raydisc.com/Assets/Downloadablefile/general_bluraydiscformat-15263.pdf).)

[0137] 24. Oracle, StorageTek T10000 Family Tape Cartridge (Oracle, Redwood Shores, CA 2010, world wide website [oracle.com/us/products/servers-storage/storage/tape-storage/033617.pdf](http://oracle.com/us/products/servers-storage/storage/tape-storage/033617.pdf).)

[0138] 25. SanDisk, SanDisk Develops Smallest 128Gb NAND Flash Memory Chip (SanDisk, Milipitas, CA 2012, world wide website [www.sandisk.com/about-sandisk/press-room/press-releases/2012/sandisk-develops-worlds-smallest-128gb-nand-flash-memory-chip](http://www.sandisk.com/about-sandisk/press-room/press-releases/2012/sandisk-develops-worlds-smallest-128gb-nand-flash-memory-chip).)

[0139] 26. Toshiba, NAND Flash Memory in Multi Chip Package (Toshiba, Tokyo, Japan, 2011, world wide website [toshiba-components.com/memory/mcp.html](http://toshiba-components.com/memory/mcp.html).)

[0140] 27. Seagate, Seagate Reaches 1 Terabit Per Square Inch Milestone In Hard Drive Storage With New Technology Demonstration (Seagate, Cupertino, CA 2012, world wide website [seagate.com/about/newsroom/press-releases/terabit-milestone-storage-seagate-pr/?paramChannelName=newsroom](http://seagate.com/about/newsroom/press-releases/terabit-milestone-storage-seagate-pr/?paramChannelName=newsroom).)

[0141] 28. S. Loth et al., Science 335, 196-199 (2010).

[0142] 29. D. M. Eigler, E. K. Schweizer, Nature 344, 524-526 (1990).

[0143] 30. C. R. Moon, L. S. Mattos, B. K. Foster, G. Zeltzer, H. C. Manoharan, Nature Nanotechnology 4, 167-172 (2009).

[0144] 31. T. Grotjohann et al., Nature 478, 204-208 (2011).

[0145] 32. H. E. Kubitschek, J Bacteriol 172, 94-101 (1990).

[0146] 33. "Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA" Federal Registrar 75, 62820-62832 (2010) FR Doc No: 2010-25728.

[0147] 等效物

[0148] 对于本领域技术人员而言,其他实施方式将是显而易见的。应当理解的是,提供上述描述仅为了清楚并且仅是示例性的。本发明的精神和范围并不限于以上实施例,而是由权利要求所涵盖。出于所有目的,通过引用将以上引用的所有出版物、专利和专利申请以它们整体合并于此,至如同通过引用将每个单独的出版物或专利申请具体指定如此合并的相同的程度。

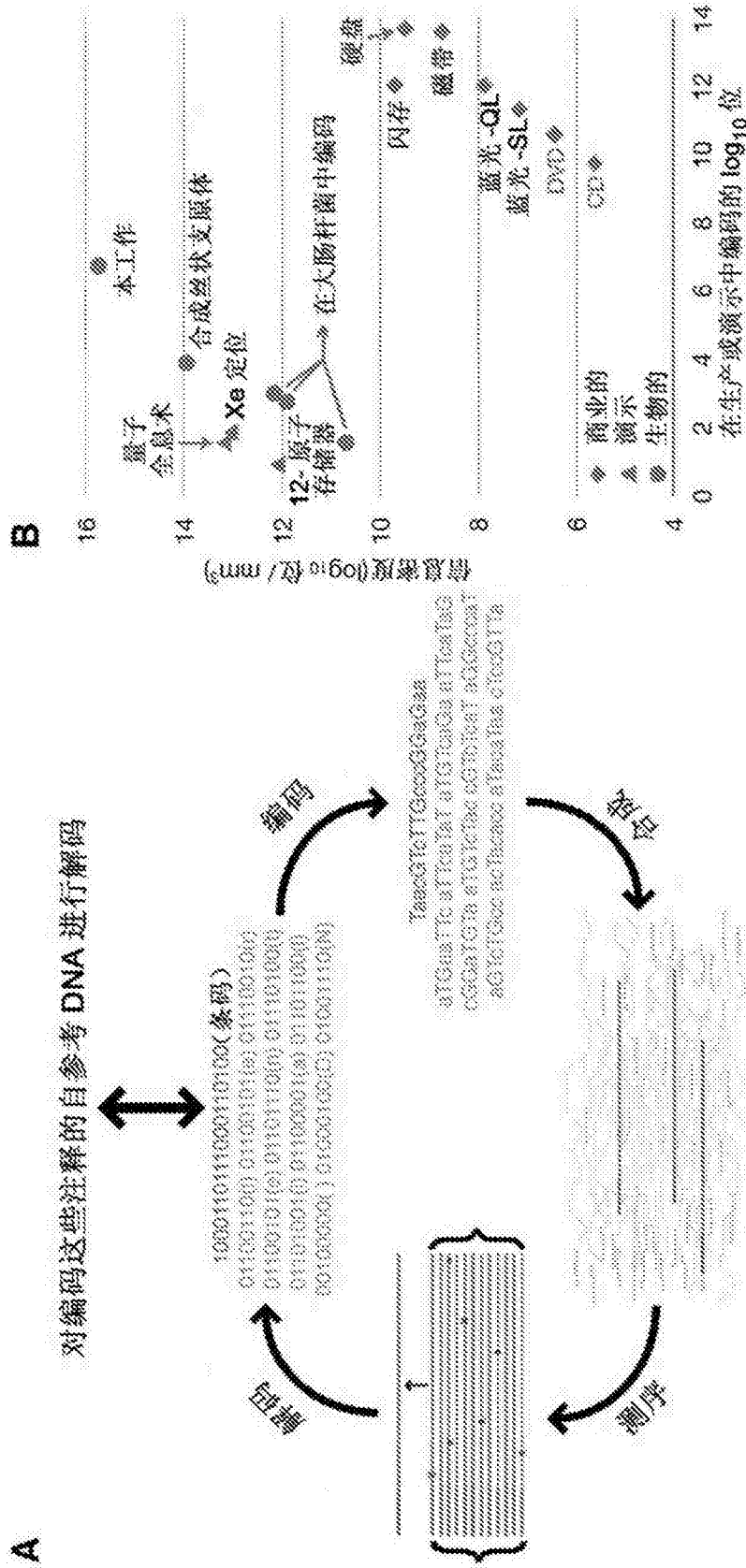


图1

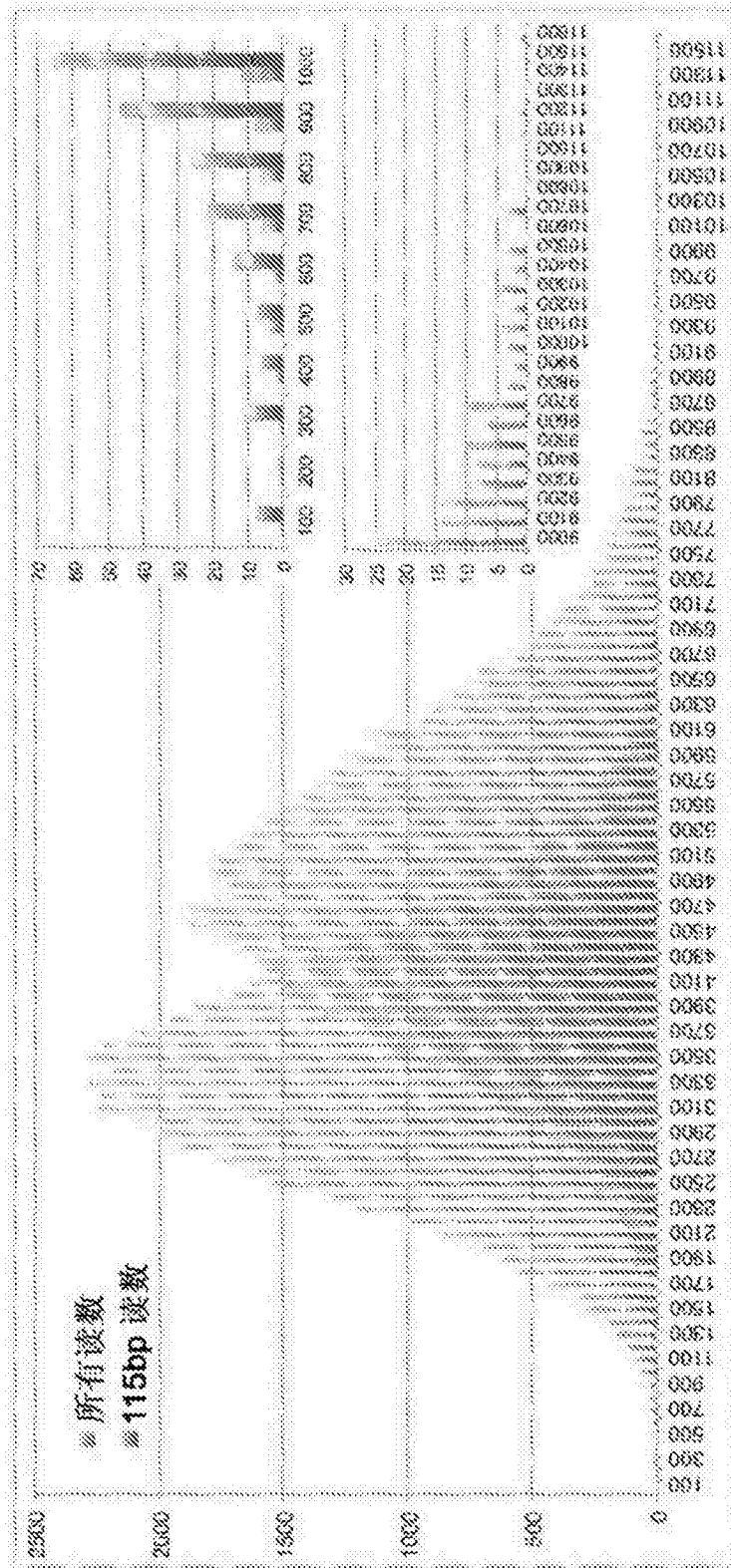


图2