11：

根据Illumina二代测序手册，这种碱基分布会使光信号难以识别。因为 Illumina 测序仪是精密的光学仪器，依靠光信号的变化来识别碱基。平衡碱基表明在每个序列通道中信号大致相等，而不平衡碱基表明可能在通道中一个信号的数量较多，因此其他碱基的数量较少。当所有簇主要在一个通道中提供信号时，仪器可能难以识别簇的位置并进行碱基识别。难以识别的光信号会降低测序数据的准确性。在这种情况下，如果我们想要获得高质量的测序数据，就需要增加测序覆盖率。但这增加了时间和消耗品的成本。

14：

非有效载荷编码约束的阈值自适应生成过程如下。

Step 1:接收编码完成的有效载荷码集合。

Step 2:计算有效载荷编码集的GC内容和碱基连续性等特征值。

Step 3:计算生成非有效载荷GC内容的阈值，碱基连续性阈值，以及地址约束中是否需要不相关。

Step 4:输入编码该集合所需的地址位数，判断当前阈值条件是否超过非净荷编码集合的下界。

Step 5:超过则返回步骤4；否则，执行第 6 步；

Step 6:使用当前的非有效载荷编码阈值继续下一个编码过程，如图 8 所示。

15：

K-means Multi-Verse optimization (KMVO)算法是利用K-means聚类对Multi-Verse optimization (MVO)算法的改进。Wang等人在MVO算法的基础上，加入扰动因子提出了阻尼多元优化器(DMVO)算法。 Brown Multi-Verse Optimizer (BMVO) 算法是通过在 MVO 算法中加入布朗运动和单线方法得到的，CLGBO 算法通过采用 Cauchy 变异算子和 Levy 策略改进了基于梯度的优化器 (GBO) 算法。

16：

在这次实验中采用根据不同特征进行编码的策略，对自适应编码后的DNA序列进行组装，并将地址位信息组装成额外的DNA序列，存储在DNA池a(索引池)中。包含地址位、引物位和有效载荷位的完整DNA序列被存储在DNA池b中，也被称为存档池。当需要对数据进行随机访问时，首先对所有索引池进行排序，然后将结果解码并还原为索引。然后，将访问请求合成为地址位的互补序列，在归档池中通过PCR扩增纯化分离出要检索的DNA序列。最后对分离出来的DNA序列进行测序，解码成原始信息，得到访问内容。

18：

从线的波动和最大值和最小值可以明显看出，本文的有效载荷编码在DNA测序中具有更好的碱基平衡，Illuminate的测序手册指出，平衡文库可以让测序系统更好地保持焦点，提供更高质量的数据。

20：

连接反应过程：

1. T5核酸外切酶以5'到3'的方向消化DNA片段，每条DNA链形成一个单链DNA突出末端。
2. 两相邻片段的突出末端含同源区域，DNA片段可退火，互补序列配对。
3. 配对后仍存在缺口，利用DNA聚合酶补全缺口。
4. 最后DNA连接酶在切口处形成磷酸二酯键，完成连接。

21：

通过比较MFE和TM方差可知，本文组装结果具有良好的稳定性和热力学性能，减少了读写误差的产生。MFEAVE和TM方差分别降低了21-26%和10%。DNA编码在预期属性(GC平衡、MFE、连续碱基等)上优于Jeong，可以减少测序过程中的错误率偏差。一种高质量的编码组合策略可以改善编码间的耦合，其效果是避免加性存储器串扰，减少重复序列的数量，减少覆盖范围。

22：

表中的存储密度、存储容量和覆盖度分别是指测序过程中每个碱基的数据存储容量、存储数据的大小以及DNA序列的覆盖程度。值得一提的是，测序过程中覆盖率的增加会导致信息难以阅读。

从表4中可以看出，本文提出的自适应编码DNA存储系统即使在低覆盖率下也具有良好的存储密度，并且自适应过程从有效载荷位自适应地为非有效载荷的编码约束生成阈值，提高了编码耦合，增加了存储密度。由于难以合成更长的DNA链，该方案继续遵循长度为~100的DNA序列。在存储密度方面，本文提出的自适应编码DNA存储系统达到1.29 bit/nt(如果压缩前计算文件大小，存储密度可以达到1.87 bit/nt)，超过了Church等人的工作，使用更少的碱基存储更多的信息。虽然与choi等人的工作还有差距，但所提出的方法可以支持不需要任何外部存储器的随机访问。DNA存储系统的测序覆盖率对信息读取的复杂性起着决定性的作用，测序覆盖率越高，信息读取的难度越大。如表4所示，本文提出的存储系统的覆盖范围仅次于Erlich等人提出的，甚至比其他方案低一个数量级。虽然自适应编码DNA存储系统在存储密度方面略弱于Choi等人的研究，同时在测序覆盖率方面略高于Erlich等人的研究。但是，一个完整的存储系统需要实现对信息的随机访问，以降低读取的难度。与 Choi 和 Erlich 各自的存储方案相比，该存储系统可以实现对数据的随机访问。此外，该方案存储了 480 KB（689 KB 之前的压缩）数据，包括 Jifi、MP4 和 TXT。与Bornholt和Choi等人的图文简单存储相比，该方案具有更高的适用性和实际意义，适合信息时代各种形式的文件存储。虽然本文的存储容量与Blawat和Organick相比还有很大差距，但自适应编码DNA存储系统具有覆盖范围低、减少地址存储串扰的特点。此外，与以往简单的映射编码相比，本文采用了协调性更好的自适应编码;有效载荷位使用喷泉代码进行编码;包括RS纠错码;采用约束编码对非有效载荷位进行编码;非有效载荷位的编码约束阈值由有效载荷位自适应产生，提高了系统的整体耦合性。

需要注意的是，在表 4 中，我们给出了存储密度和存储容量的两个值。左边的是在不严格存储寻址表的情况下存储480 KB数据的存储密度，这是大多数现有工作所接受的。两种主流的存储密度计算方法包括是否包括引物。为了严格考虑，本文的存储密度计算包括引物。表中两个存储密度值分别代表480 KB数据(不含存储寻址表)和独立随机存储《哈利波特1》(含存储寻址表)的存储密度。虽然与没有引物的计算方法相比，左边的存储密度已经非常有说服力了。然而，在支持随机访问的存储系统中，访问数据所需的地址表通常存储在外部存储器中。因此，为了更好地说明本文提出的自适应方案在极端条件下的性能，如地震、海啸、电磁风暴等恶劣条件下，或地址表损坏时的性能。我们还进行了严格的随机访问实验，将《哈利波特1》的节选存储在索引池和存档池中，最终实现读取任意段落的访问要求。这种方法不需要任何外部内存，是真正的随机访问，并且存储密度记录在表4中斜线的右边。虽然DNA存储的发展才刚刚开始，特别是随机存取存储，但严格的标准随机存取对DNA存储的发展起到了很好的推动作用。