

# Three decades of nanopore sequencing

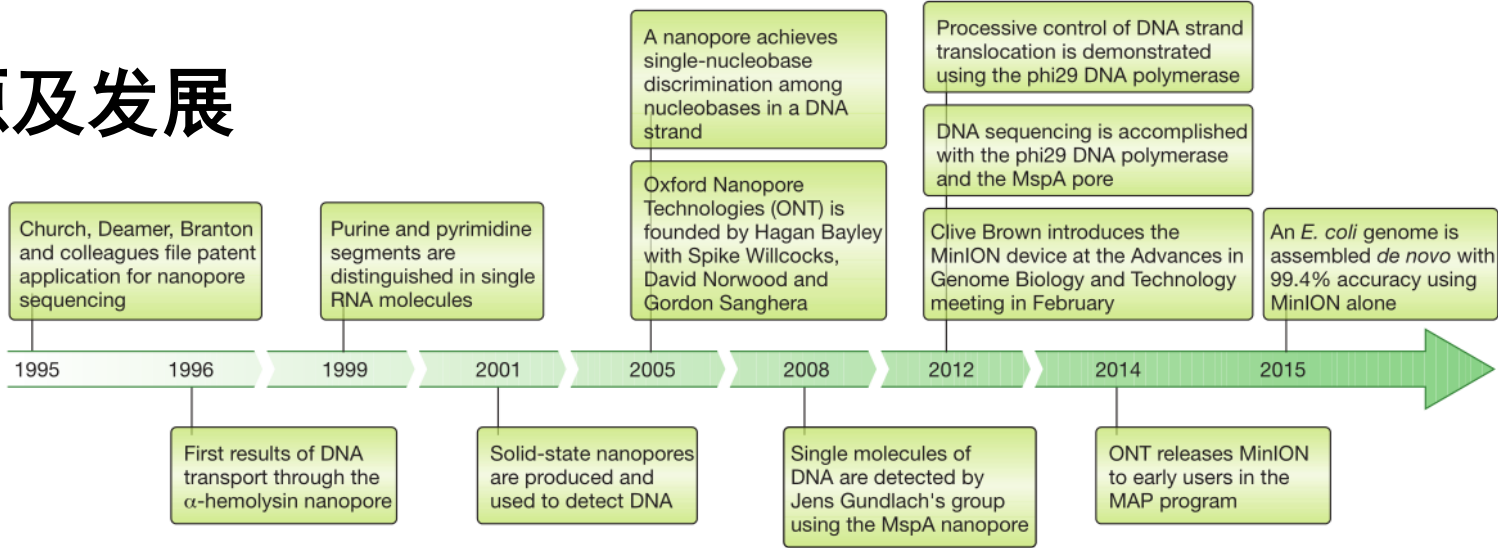


Tongqing Zhang, Laizhi zhang, Yunrui Pan  
BGI-Shenzhen

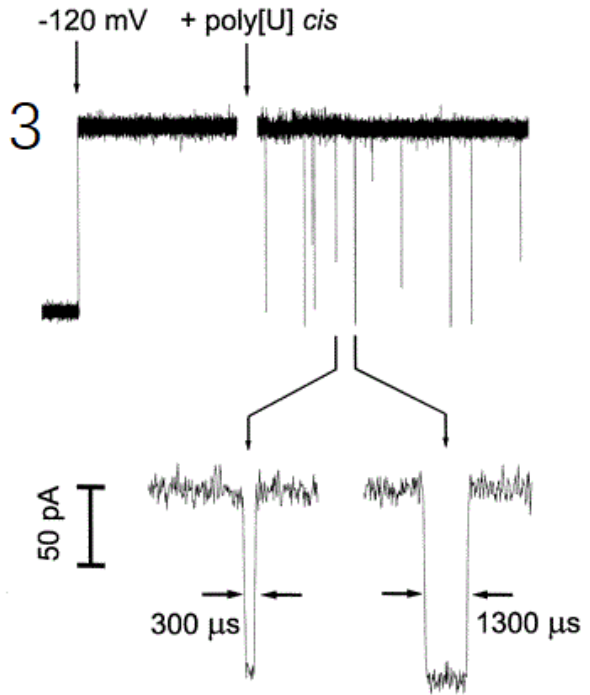
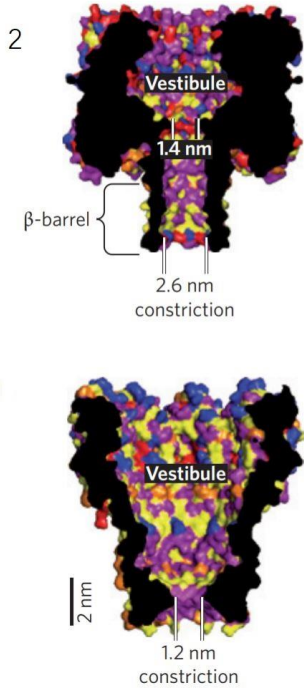
# Nanopore sequencing的起源及发展

## First decade

1. 1989年，有研究者提出可以利用电泳技术使DNA单链通过纳米孔；同时，George Church提出是否可以将这些纳米孔作为生物传感器。
2. 1994年4月，D. D. 和 D. B进一步实验证明了单链DNA可以通过溶血素孔（如2图，上： $\alpha$ -HL分子结构剖面图；下：MspA(B) 分子结构剖面图）。
3. 1995年，Church, Deamer, Branton and colleagues 申请了关于 Nanopore sequencing技术的专利。
4. 1996年，得到第一个通过  $\alpha$ -HL 颗粒的DNA转运结果。（3图：poly (U) 通过纳米孔记录的电流变化情况，证明了多核苷酸可以很容易的通过单个纳米孔，原理上，通过通道的单个嘌呤或嘧啶核苷酸应可以反映多核苷酸中每个核苷酸的分子大小和化学性质的方式阻断离子电流）



2图来源: <https://www.nature.com/articles/nnano.2011.129?message-global=remove&page=7>; 3图来源: <https://www.pnas.org/content/93/24/13770.full>;

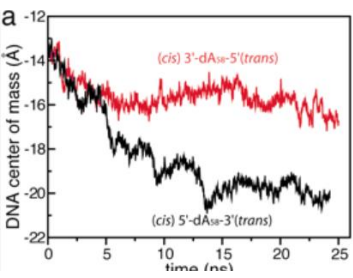




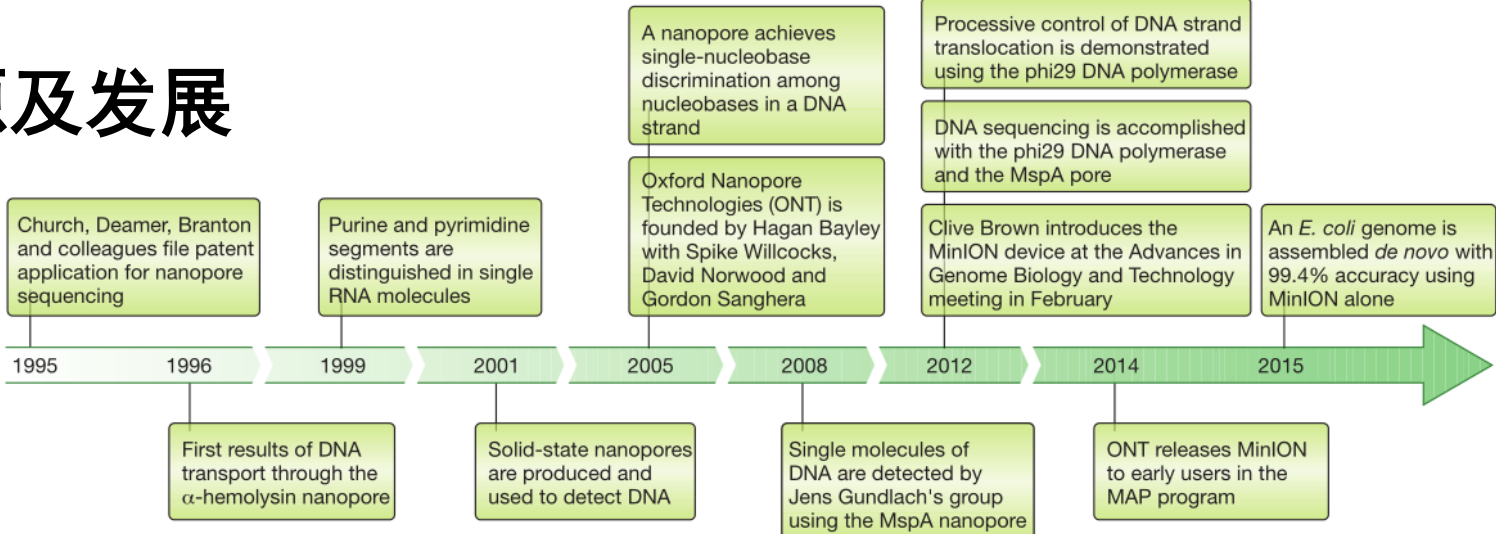
# Nanopore sequencing的起源及发展

## Second decade (1)

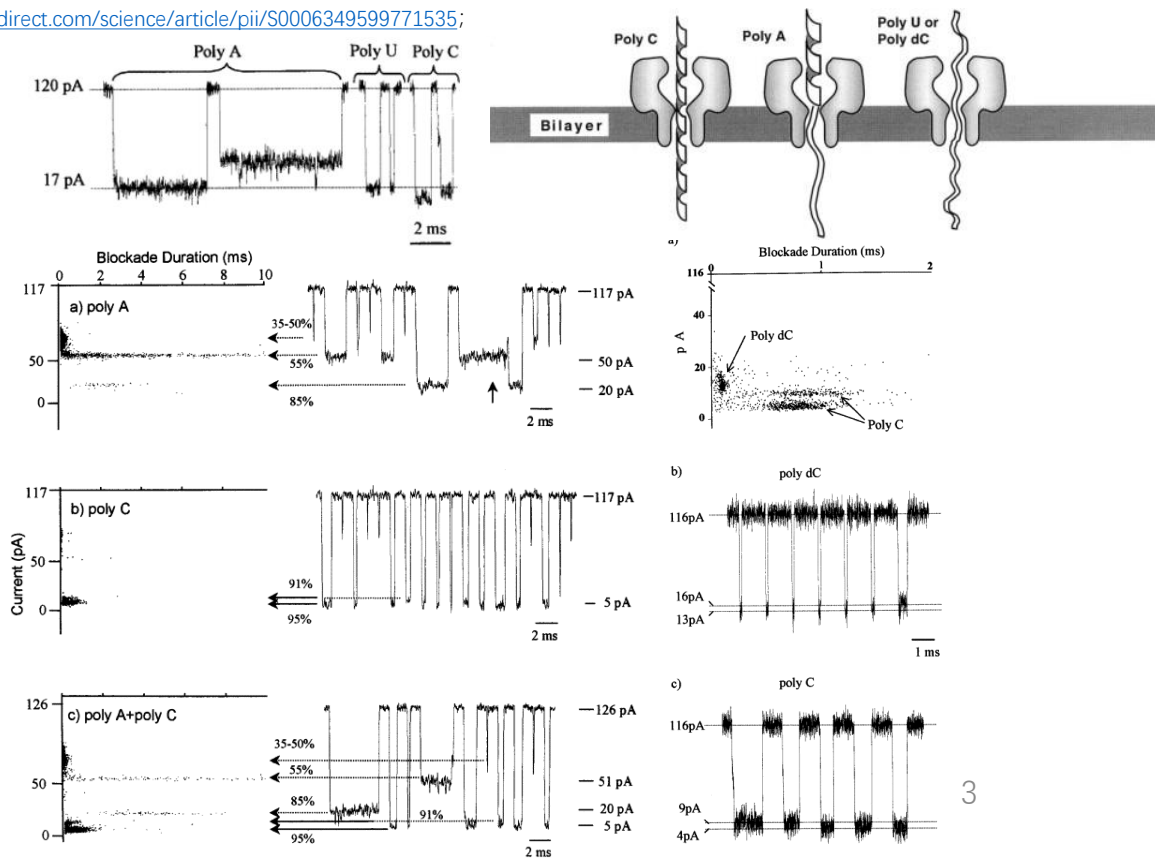
1. 在已经证明了ssDNA可以通过  $\alpha$ -HL并引起电流改变之后，下一个重要的问题是  $\alpha$ -HL孔是否可以区分嘌呤和嘧啶碱基。1997年，M. A. 开始研究这个问题。
2. M. A通过实验证明， $\alpha$ -HL孔可以提供寡聚结构或组成的信息。
3. 实验同时发现poly (C) 引起的电流阻断作用更大。进一步实验证明由于含有胞嘧啶的部分形成了一个狭窄的单链螺旋，可以通过 $\alpha$ -HL作为一个螺旋转位，而腺嘌呤部分形成一个直径较大的螺旋，必须解开才能使单链转位。
4. 分别以3' 和5' 段进入纳米孔的多聚核苷酸链其反应的电流图像也有不同。这就解释了为什么同样的序列可以得到两种结果。



左图来源: <https://www.pnas.org/content/102/35/12377>;



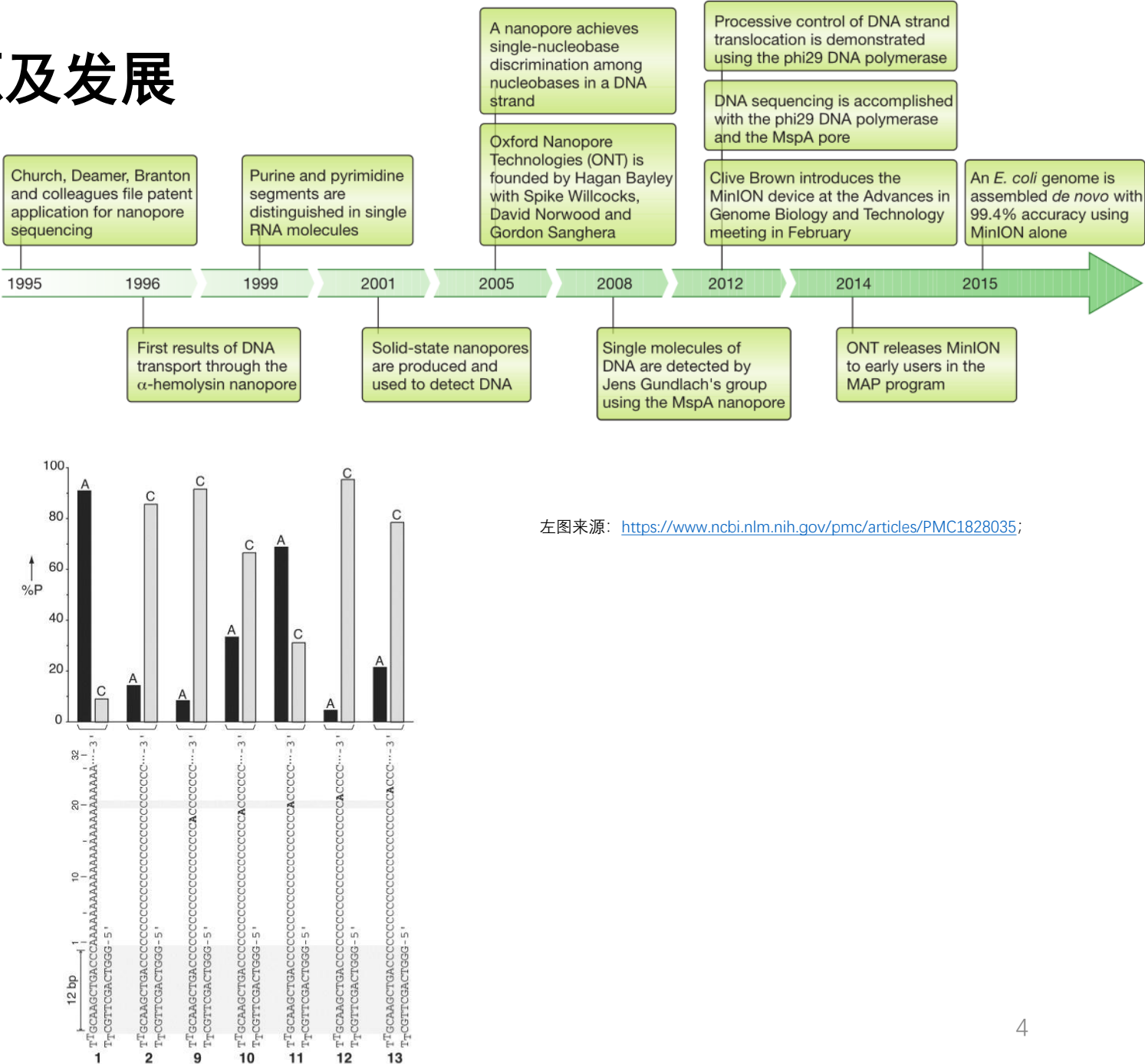
右4图均来自: <https://www.sciencedirect.com/science/article/pii/S0006349599771535>;



# Nanopore sequencing的起源及发展

## Second decade (2)

- 1. 2001年，纳米孔技术被用来检测DNA。
- 2. 2005年，证实纳米孔可以识别DNA中的单个碱基。同年，Oxford Nanopore Technologies (ONT) 公司创立。
- 3. 右图：纳米孔技术对单核苷酸检测的结果，其中黑柱代表灰色对应的核苷酸属于A的概率，而灰柱代表灰色对应的核苷酸属于C的概率，我们可以看到，所反映的结果与实际结果有很好的对应。

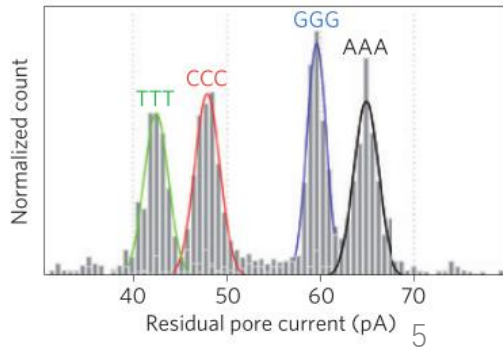
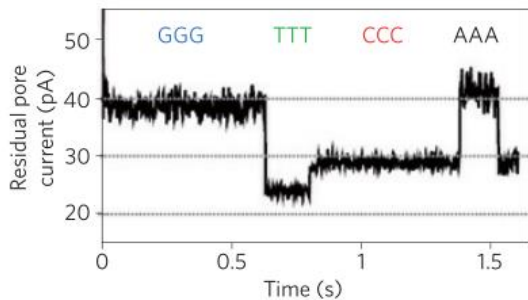
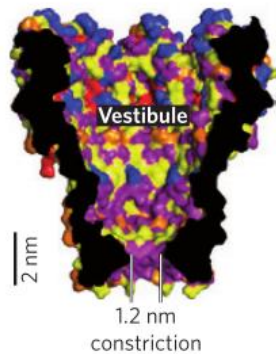
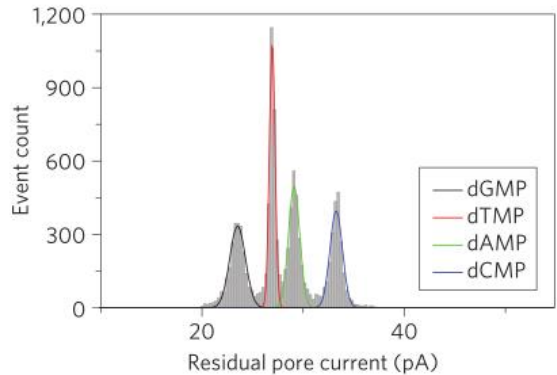
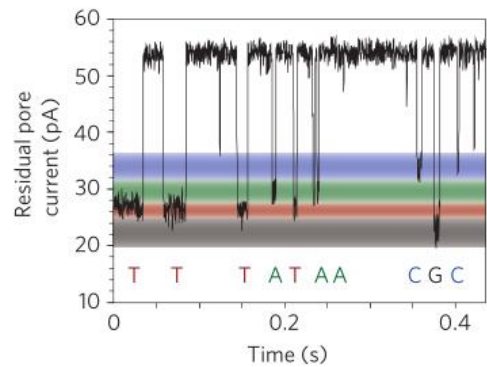
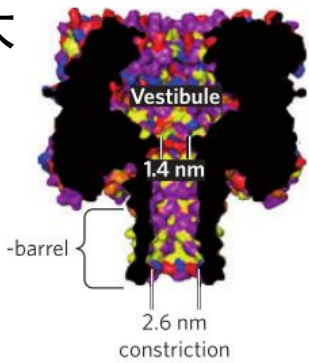
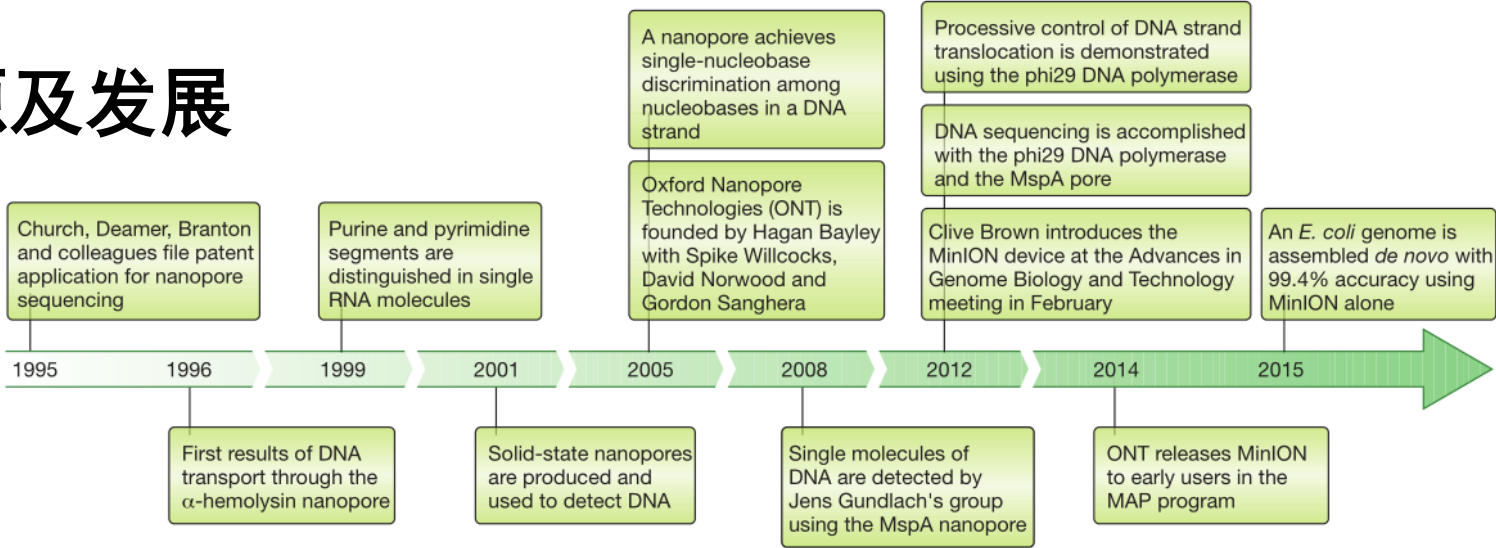
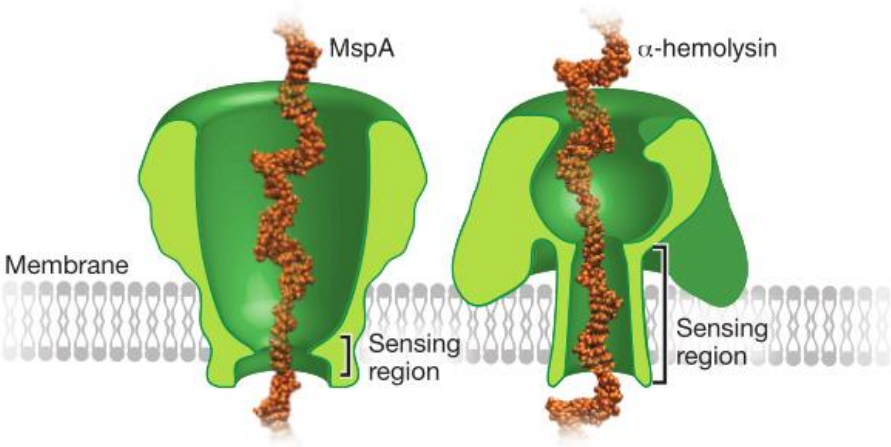


左图来源: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1828035/>;

# Nanopore sequencing的起源及发展

## Third decade (1)

- 1. 2008年，单分子DNA被Jens Gundlach的团队用 MspA 纳米孔探测到。
- 2.  $\alpha$ -HL有一个圆柱形的 $\beta$ 管道可以容纳大约10个核苷酸，这些核苷酸都会对离子通道电流产生影响，降低了核苷酸分辨率，而MspA没有这种桶部，大大提高了对核苷酸的分辨能力。





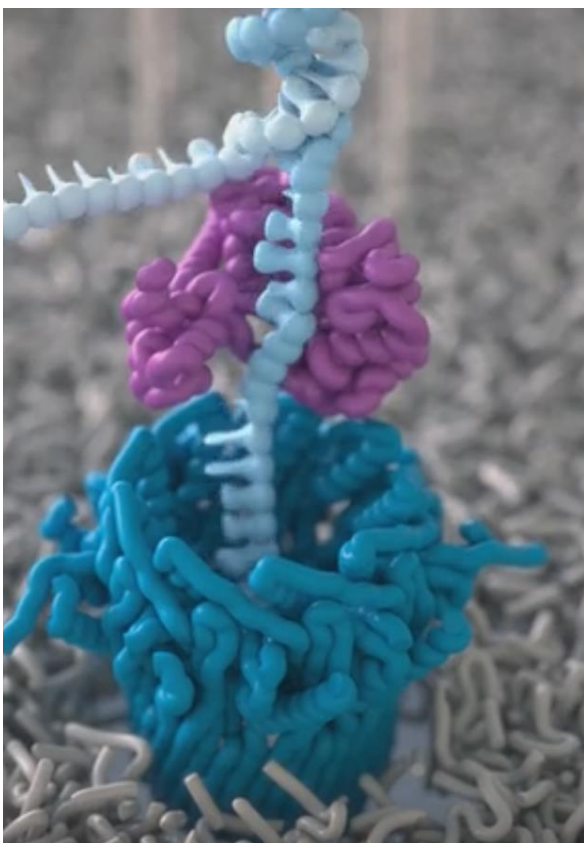
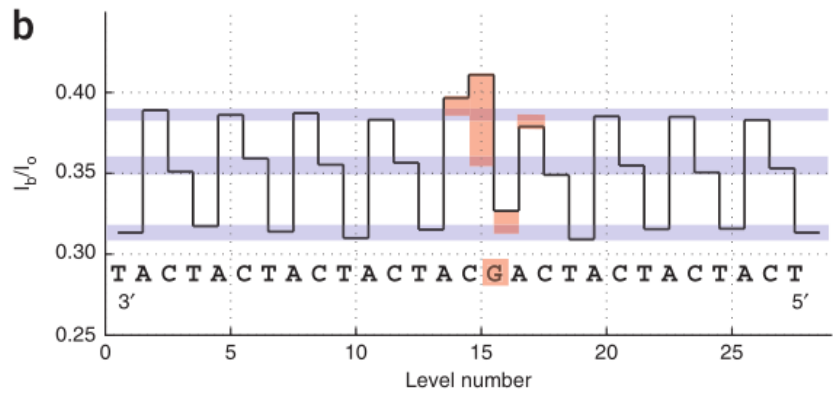
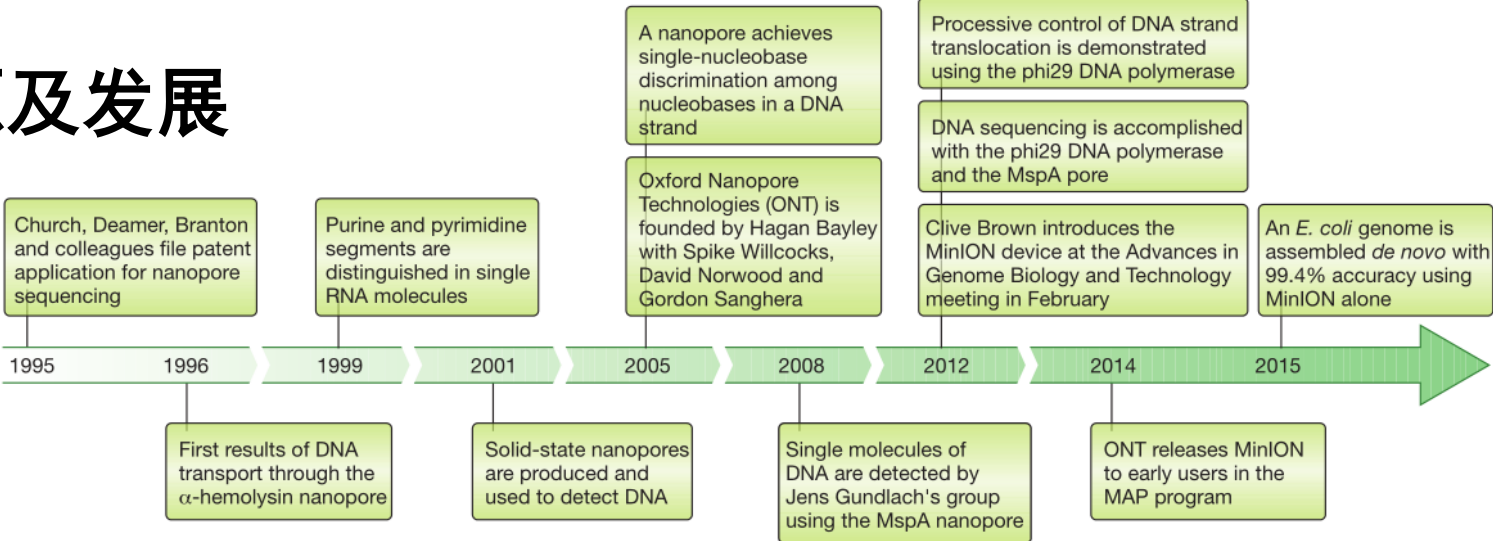
# Nanopore sequencing的起源及发展

## Third decade (2)

- 1. 2012年，MspA和phi29DNA聚合酶开始被用于Nanopore sequencing。
- 2. MspA的一个缺点就是ssDNA会以较快的速度移动，降低了测序的准确率。因此需要一种方法来降低核苷酸的移动速度；研究者发现，phi 29 DNA聚合酶进行实验取得了比较满意的结果。

右图：我们可以在ACT的重复模式中判断出单个G的扰动。

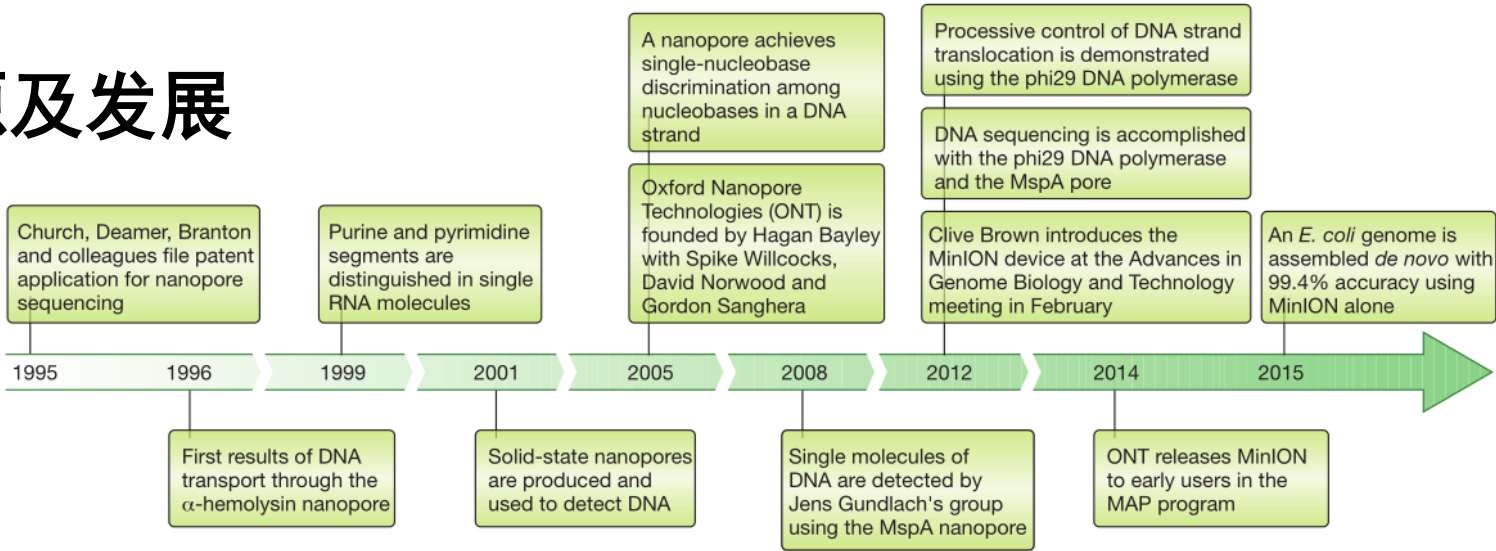
来源：[Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase \(nih.gov\)](#)



### Third decade (3)

1. 2012年，克莱夫·布朗在2月份的基因组生物学和技术进步会议上介绍了MinION。
2. 2014年，ONT开始向早期用户提供MinION服务
3. 2015年，使用MinION从头组装的大肠杆菌基因组序列达到了99.4%的准确率。

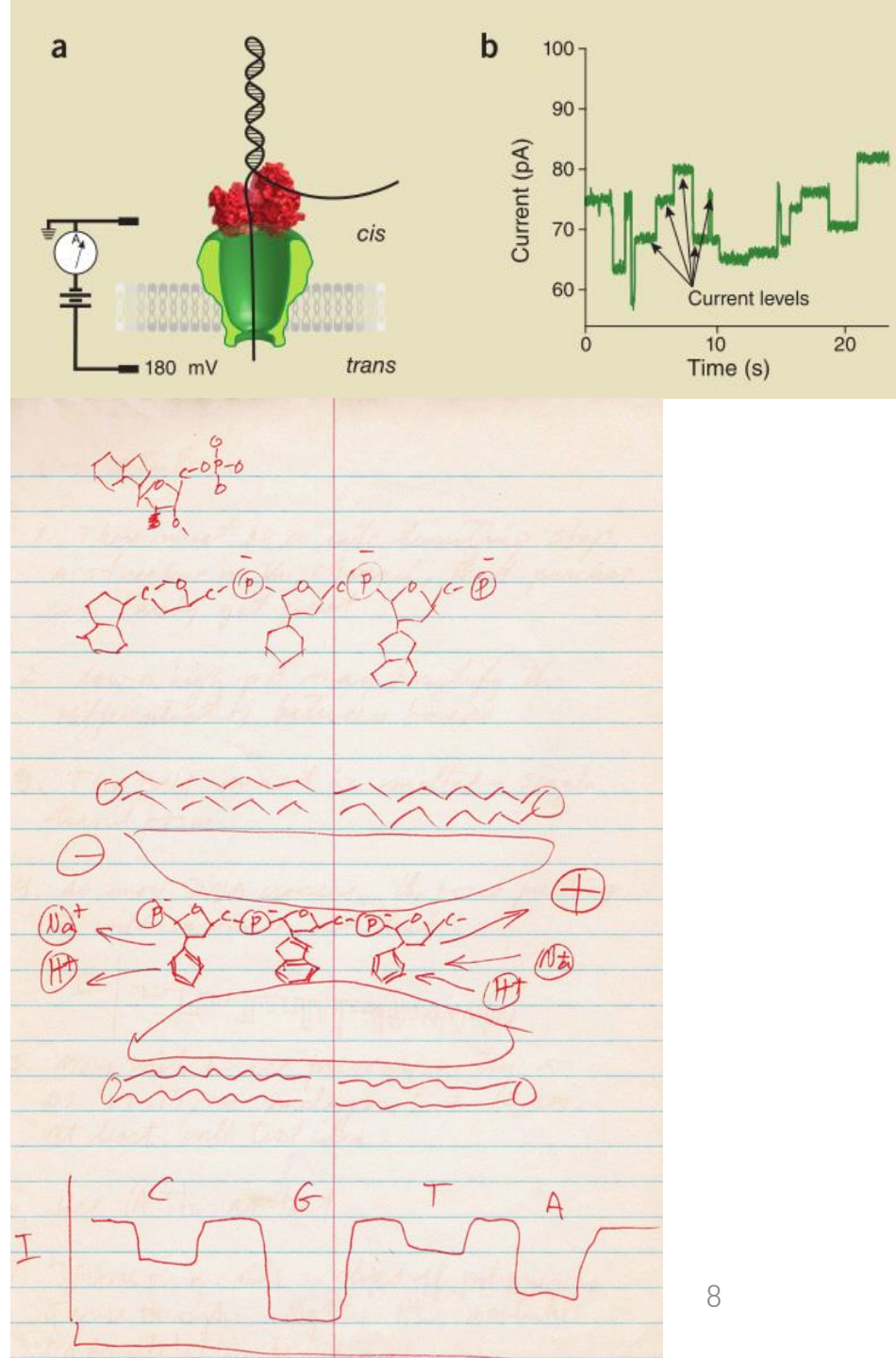
右图：来自nanopore官网的MinION介绍；MinION是一个有2048个蛋白质纳米孔的随身携带设备。



# Nanopore sequencing的起源及发展

## Nowadays (1)

1. 在2016年发表的文章中，作者主要提及了一些在当时难以克服的技术问题：DNA每次前进的距离不一致导致通过纳米孔产生的电流有细微差别，可能会有碱基产生的电流被当作背景噪音而忽略，这也是nanopore sequencing中最多出现的错误是deletion的原因；另一个问题是，尽管MspA的 $\beta$ 桶部很短，但在极限条件下仍然可以容纳4-6个碱基，这五个碱基对电流结果都会产生影响。
2. 作者也提出了一些构想：Nanopore技术也可以用于RNA或者蛋白质分析，而且也提出了一些减少误差的方法：可以通过重复读取同一段序列来减少，尽管在当时并不能实现；也可以采用更薄的单分子纳米孔，例如石墨烯薄膜，可以将同时通过的碱基数量减少到1-2个。

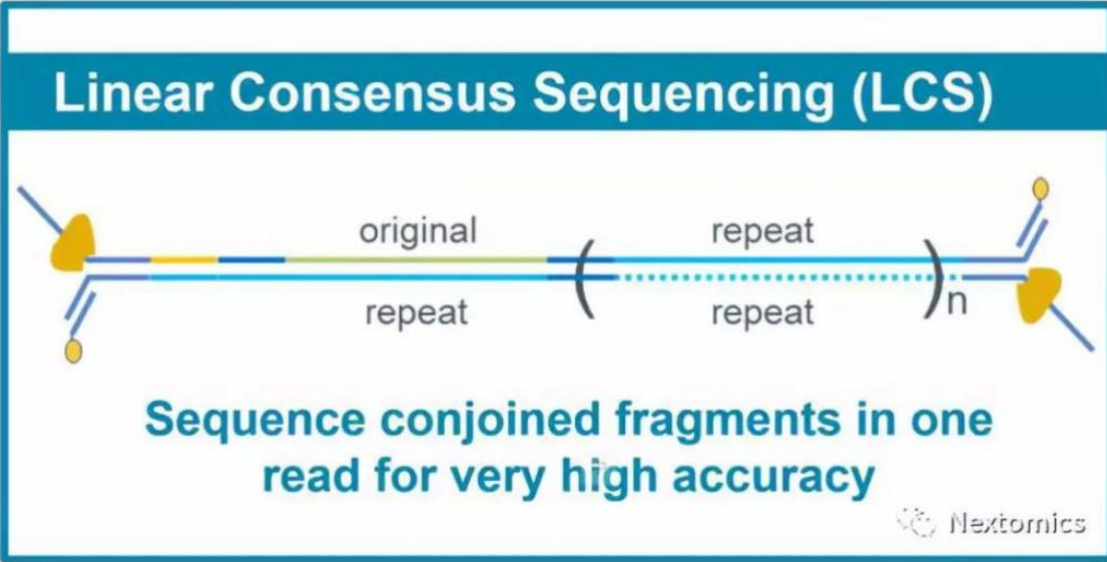
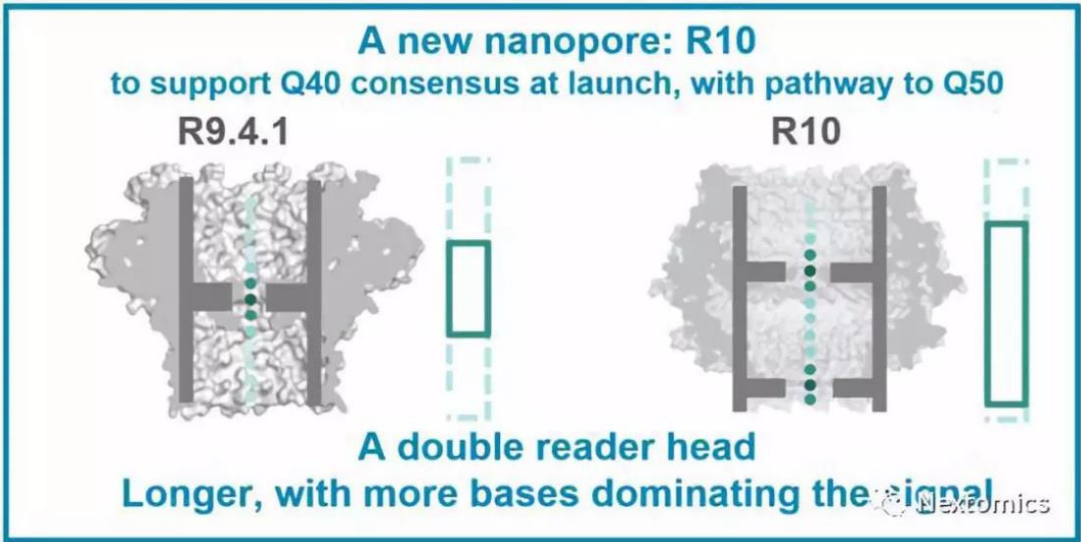




# Nanopore sequencing的起源及发展

## Nowadays (2)

1. 目前我们已经发展了足够的技术来克服其中的一些问题。
2. LCS技术：是将一条链的数条拷贝结合在一起进行测序，以获得更高的准确度。
3. 在最新的R10版纳米孔中采用了双读取头，同时采用了最新的Flip-flop算法，以最大程度的增加homopolymer区域（也是Nanopore技术最容易出错的区域）的正确率。
4. 目前采取基于RNN的碱基识别技术，识别电流信号并分析碱基信息。

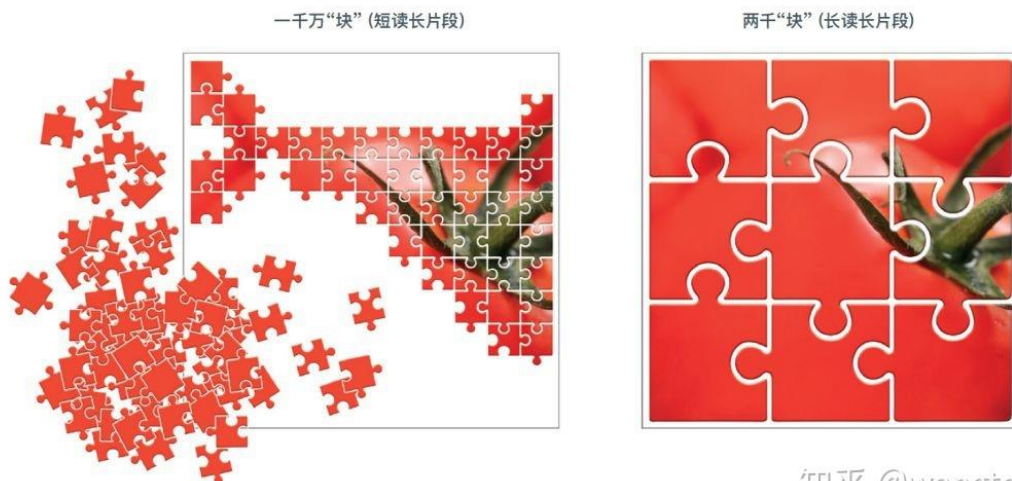


# Nanopore sequencing的优势

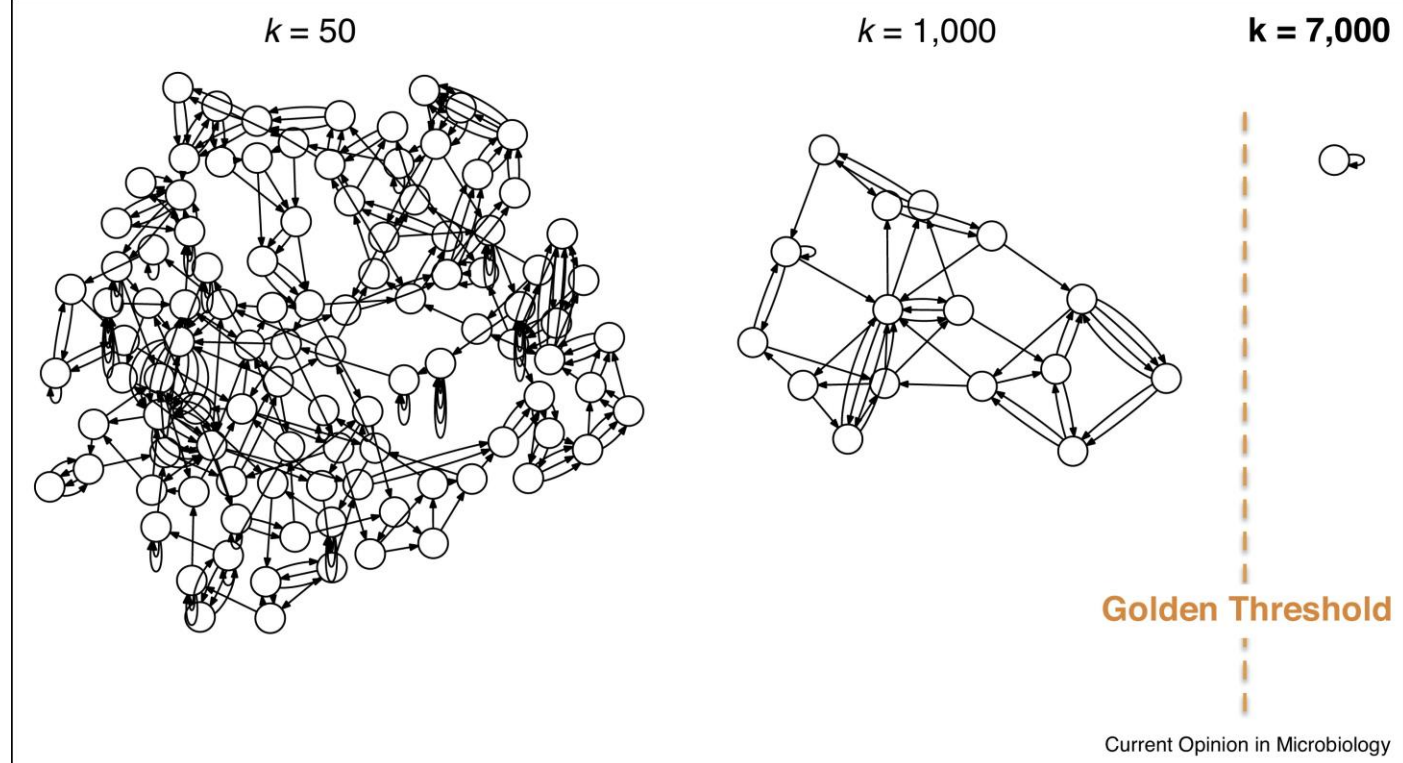
- High accuracy
- Long-read sequencing
- Uniform coverage
- Epigenetics
- RNA
- Single-Molecule Resolution
- SMRT

# Long reads

- 相比于NGS, Long-read sequencing可以更加准确的测量含有重复序列的 DNA 序列; 这些重复序列, 或是拷贝数变异的正常序列在长序列测序中更容易发现, 这种测序技术也可以更准确地检测出更大规模的突变, 长段的DNA 被删除或移动。这些结构变异常常导致一些遗传性疾病, 但由于缺乏可用的技术, 过去我们并没有对他们没有被广泛研究。
- 读取长度的增加也极大的简化了contigs的组装。一旦测序长度达到了“golden threshold”, 我们便可以消除contigs组装中可能会引起的歧义



知乎 @wangtong

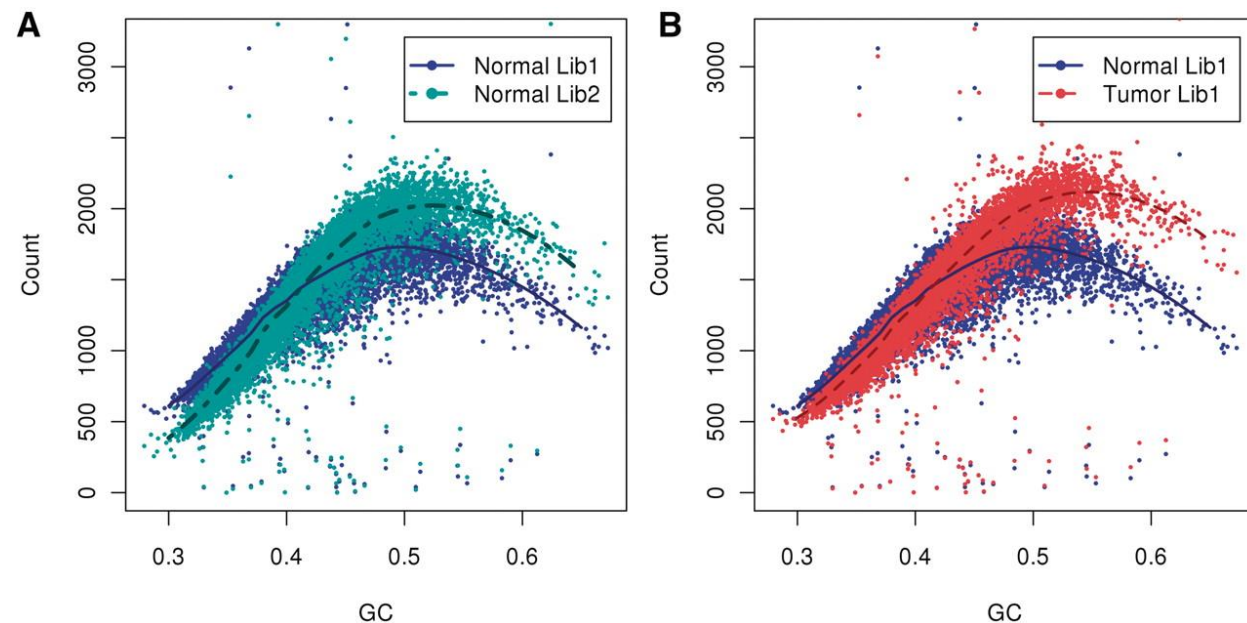


上图来源: <https://www.sciencedirect.com/science/article/pii/S1369527414001817?via%3Dihub>;



# Uniform coverage

- 富集GC的区域，NGS技术测得的覆盖率会偏低，而SMRT则没有这个缺点；换句话说，SMRT可以对其他技术很难涉及的区域进行排序和组装。



上图来源: <https://academic.oup.com/nar/article/40/10/e72/2411059>;

# Single-Molecule Resolution

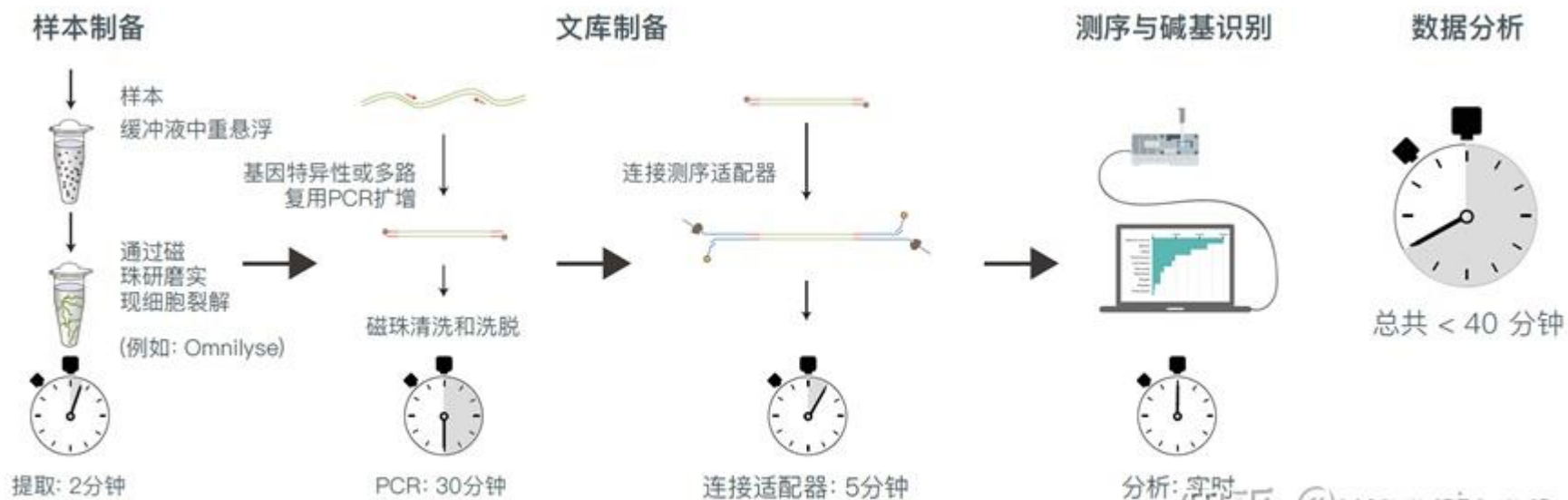
- 单分子分辨率：因为测序模板为单独的DNA链，因此可以用来区分相似序列之间的区别。

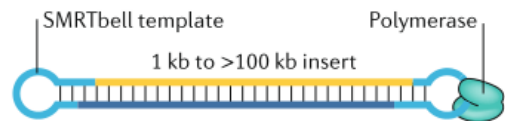
# Epigenetics

- 测序的同时就可以对修饰的碱基进行检测（在通过纳米孔的时候，有化学修饰的碱基会产生不同的电流信息）。

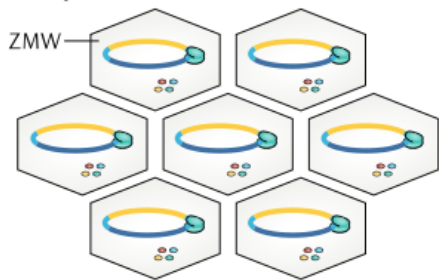
# SMRT

- 纳米孔技术提供的是动态、实时的测序，支持在几分钟内就获得病原体鉴定等时间关键型应用的检测结果。测序时，DNA快速通过纳米孔。DNA片段通过纳米孔的速率已从推出时的每秒35个碱基提升到了现在每秒450个碱基。每一个在阵列上完成的读长，在数秒后就可以用来开始数据分析，而不是几小时甚至几天。用户可以在测序早期了解样本的质量和状态，也可以在获得足够的数据后停止测序。快速的采样到结果周转时间在未来感染性疾病、实地动态监测疫情爆发以及其它诊断方面具有巨大潜力。

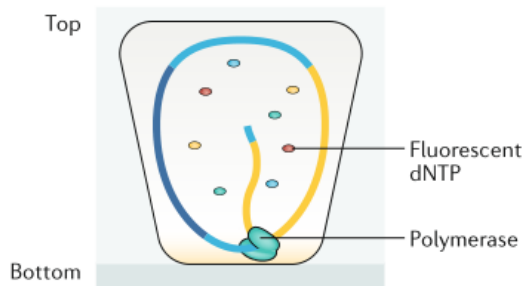




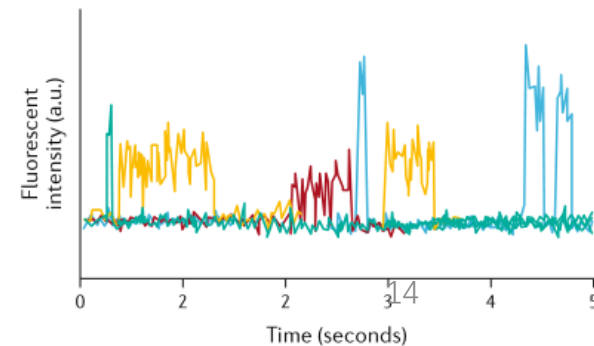
Flow cell (top view)



Single ZMW (cross section)



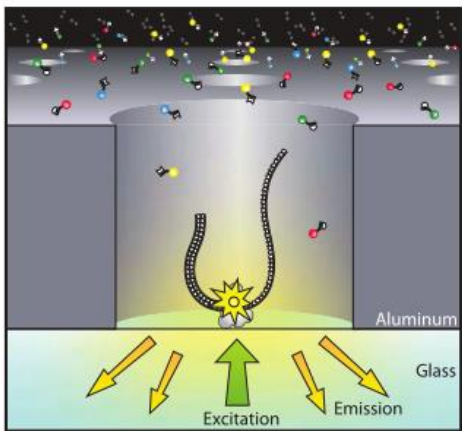
Readout



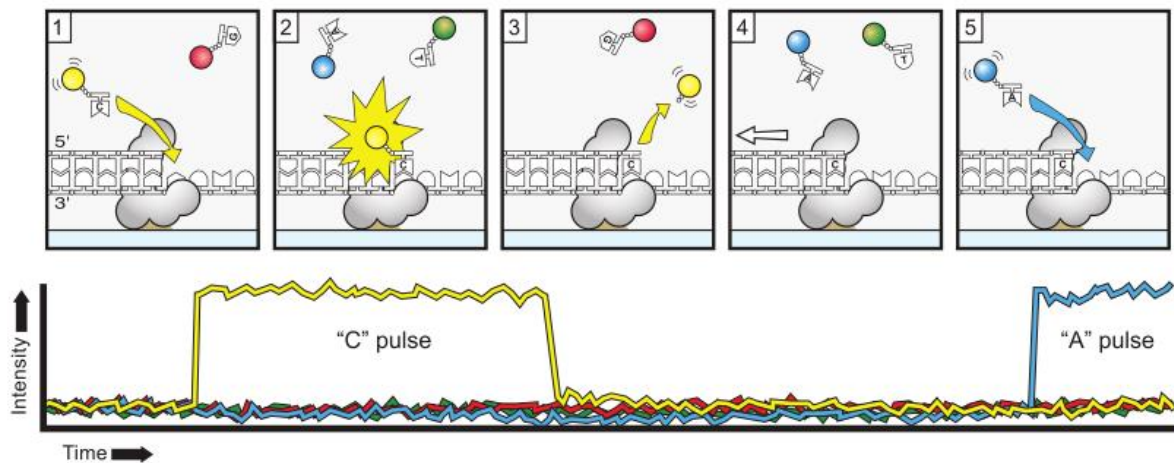
# 与另一种SMRT技术 (Pacbio) 的对比

- Pacbio 测序使用的Cell 是一张厚度为100 nm的金属片，一面带有几十上百万个直径为几十纳米的小孔，称为零模波导孔 (zero-mode waveguide, ZMW)。
- 在纳米孔底部，锚定着测序模板 (DNA单链) 和DNA聚合酶，同时包含着四种被不同荧光基团修饰的dNTP。由于每次添加的dNTP所携带的荧光颜色是不同的，在激光的激发下可以发出不同的荧光，根据散射出的荧光信号可以判断添加的碱基类型。
- 激光从ZMW的下方进入，由于ZMW的直径小于激光的波长，检测激光会被限制在纳米孔内部，不会进入小孔上方的溶液区，干扰临近ZMW的测序；被激发的荧光也只会从ZMW下方的玻璃散发，被检测器检测。
- DNA聚合酶介导的延伸反应会沿着一个方向进行，在下一个dNTP添加之前，前一个dNTP上的荧光基团会从复合物上脱落下来，所以单独的一个碱基检测到的荧光信号只会持续很短的一段时间，根据检测到的不同波长和峰值以及他们之间的间隔，就可以得到和模板序列配对的序列信息。

A



B





# 主要优缺点对比

## Pacbio SMRT

1. 使用光信号记录碱基信息。
2. 采取边合成边测序的方式，但因为激光破坏酶和DNA结构，也限制了它测序的最大长度。
3. 测序误差为完全随机误差，可以通过采用HIFI技术来提高覆盖率将误差降低至0.001%以内。

## Nanopore

1. 使用电信号记录碱基信息
2. 不涉及序列合成和高能激光，因此测序长度可以达到极长。
3. 测序误差除了一些随机误差之外多在同聚体区域，可以通过上述所说的Flip-flop技术减少同聚体区域的误差。对于随机误差，无法多次重测序将误差缩小到同Pacbio SMRT技术的水平