

无偏倚全长转录本的价值

目录

| | |
|--------------------------------|----------|
| 转录组分析挑战 | 4 |
| 获取和分析全长转录本 | 4 |
| 准确识别各基因的多个同源异构体 | 5 |
| 识别和量化问题异构体 | 6 |
| RNA直接测序 - 无偏倚, 全长转录本, 以及碱基修饰分析 | 6 |
| 转录组分析中的RNA测序深度指南 | 7 |
| 快速识别和鉴定RNA病毒的需求 | 8 |

| | |
|---------------|----------|
| 数据分析工具 | 9 |
|---------------|----------|

| | |
|----------------------------|-----------|
| 个案研究 | 10 |
| 1. 纳米孔测序技术能精确解析选择性剪接转录本 | 10 |
| 2. 发育中的橄榄果蝇胚胎的转录动力学 | 11 |
| 3. 利用全长转录本进行癌症研究 | 12 |
| 4. 识别神经精神疾病基因中的新型异构体 | 14 |
| 5. 纳米孔测序精确量化任何大小的全长cDNA的丰度 | 15 |
| 6. 使用RNA直接测序进行全局性的转录组分析 | 15 |

| | |
|-----------|-----------|
| 总结 | 17 |
|-----------|-----------|

| | |
|--|-----------|
| 关于 Oxford Nanopore Technologies | 18 |
|--|-----------|

| | |
|-------------|-----------|
| 参考文献 | 18 |
|-------------|-----------|

引言

RNA测序（传统测序技术称为RNA-Seq）已经在生物学和医学的各个领域取得了前所未有的发展。它是研究人类遗传学和疾病相关病理学的宝贵工具。例如，在包括癌症在内的诸多疾病中，转录异构体的表达和用途是健康组织和患病组织之间变异的重要来源^{1, 2}。RNA测序也有助于识别不断增加的病症中出现的融合转录本³。

转录异构体的表达和用途是健康组织和患病组织之间变异的重要来源。

cDNA测序也极大地辅助了病毒性病原体的表征和及时检测^{4,5,6}，与“金标准”病毒分离培养法相比，大大提高了从采样到获得结果的效率。对被病毒感染的样本进行RNA测序，能够改善灵敏性⁷和特异性，并能对病毒的快速进化做出迅速反应，克服了依赖于ELISA和RT-PCR的其它替代方法中存在的问题。但是，RNA测序的众多应用并不局限于人类健康领域的研究。该方法已被应用于农业环境中，例如，用来研究干旱诱导的植物应激反应⁸。

最后，RNA测序在发育生物学中的应用已在协助阐明转录程序的变化与多种发育相关事件的联系⁹。

毫无疑问，传统的测序技术使全面转录组分析成为可能，并带来了很重要的科学发展。然而，这一领域仍有一些重要的局限性有待解决。在此，我们将重点介绍纳米孔技术如何灵活应对在全长转录本识别、异构体表达和定量、以及病毒检测中的挑战。



转录组分析挑战

获取和分析全长转录本

高通量的RNA测序方法在生成全长转录本时常常由于逆转录的过早终止而存在困难¹⁰。在使用oligo-dT引物时，这反映在以完整序列为代价的转录本3'末端的过度偏倚¹¹。该问题可通过用于全长转录本富集的反转录酶“模板转换法”缓解¹²。然而，由于大多数传统测序方法需要进行片段化，准确组装完整的转录本仍然存在困难，尤其是在一个读长可以比对到多个位点的情况下（例如高度保守的序列段）（图1）。

利用纳米孔测序技术，无论片段长度如何，都可以对整个片段进行测序。因此，全长cDNA可以作为单个片段来进行测序，大幅度减少了之后在多位点比对中的问题（图1）。Oikonomopoulos等人的一项研究证明¹³，纳米孔技术可发挥模板转换法的最大潜力，其中有90.8%的纳米孔读长为全长转录本。重要的是，使用模板转换法并不是利用纳米孔测序获取全长转录本的先决条件。如Hargreaves与Mulley的一项研究显示¹⁴，其中有88%的读长为全长转录本。纳米孔设备所提供的长读长范围介于1-100skb，使得全长转录本的识别和转录组从头（*de novo*）组装更易于传统测序平台。

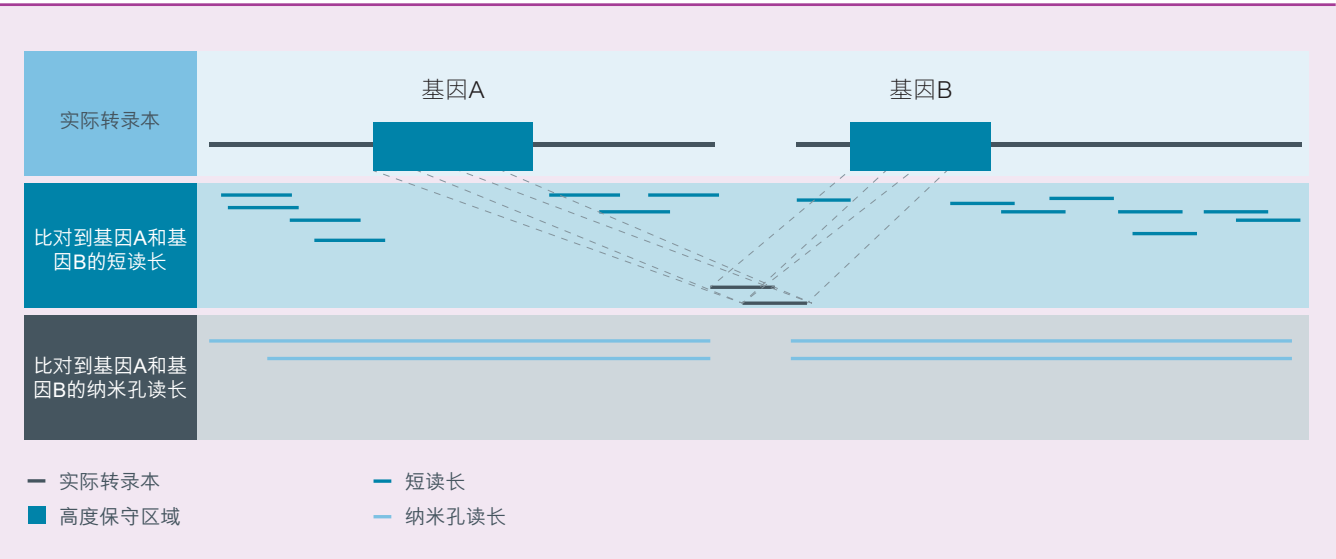


图1
利用短读长和纳米孔长读长组装全长转录本。

准确识别各基因的多个同源异构体

选择性剪接（或可变剪切）的过程中，前体mRNA分子可以产生多个外显子形式不同的成熟mRNA分子，从而允许多种蛋白质能被单基因编码¹⁵。为了能正确识别这些独特的mRNA同源异构体，则有必要对几个连续的外显子进行测序。但这些外显子的长度导致超出了短读长测序技术的能力（图2）。

长纳米孔读长，生成完整的转录异构体序列，避免组装中存在的问题。

而正因为此，计算机技术被用于进行异构体的重建。然而，在Steijger等人¹⁶的一项研究中表明，在超过半数以上所分析的转录本中，自动的转录本组装方法未能识别的其所含

的所有外显子组成。此外，在那些所有外显子都能被识别的转录本中，有超过一半的组装不正确。如果正在研究的读长来自高度相似的转录本，如旁系同源基因的读长，则这些情况会更加复杂，还有可能完全无法检测出罕见的同源异构体¹⁷。

长纳米孔读长能够生成完整的转录异构体序列，从而避免组装中存在的上述问题。其在剪接变体识别中的优势已被反复证明^{14,15,18,19}。值得注意的是，纳米孔测序能够识别唐氏综合征细胞粘附分子1(Dscam1)中的7000多个全长异构体—这是自然界中已知的最复杂的选择性剪接基因¹⁵。

纳米孔测序技术能够提供可靠的用于同源异构体的解决方案，使复杂选择性剪接转录组的从头（*de novo*）组装成为可能。

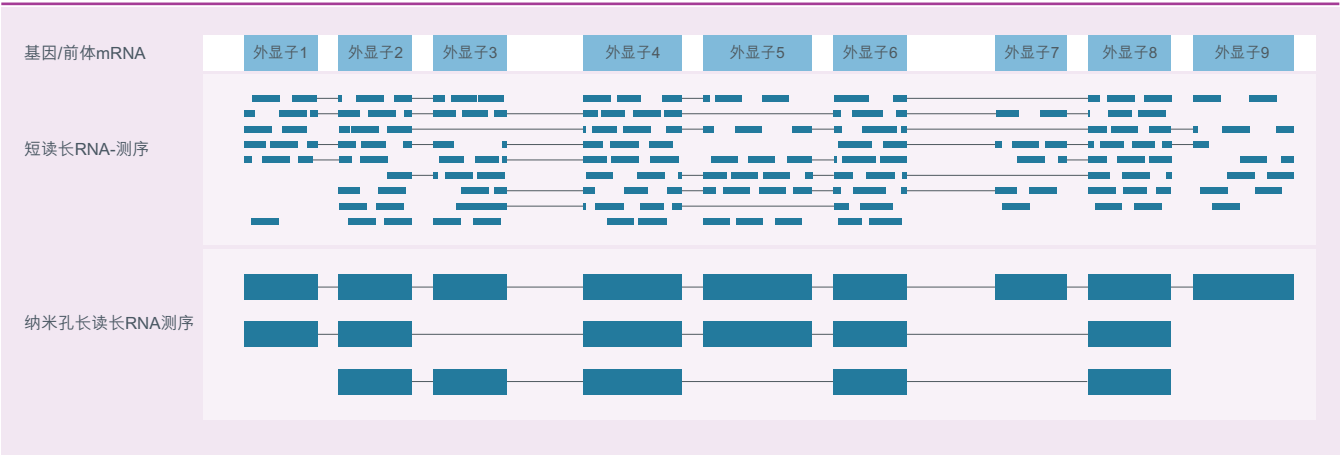


图2
选择性剪接可以让每个基因产生大量的mRNA异构体，这又能反过来改变蛋白质的组成和功能。传统RNA测序技术产生的短读长丢失了位置信息，使得正确组装选择性剪切mRNA异构体存在挑战性。长纳米孔读长能够横跨全长转录本，从而简化识别过程。

识别和量化问题异构体

由于无法准确识别同源异构体, 传统测序平台在量化异构体方面也面临诸多困难。短读长和纳米孔表达数据之间的强相关性表明²⁰, 纳米孔cDNA测序在单细胞基因表达分析中与其它成熟技术表现相当。此外, 纳米孔平台既能作为独立工具²⁰, 也能作为混合测序方法的一部分, 生成数据进行可靠的同源异构体量化¹⁸。

RNA直接测序 - 无偏倚, 全长转录本, 以及碱基修饰分析

直到最近, 基于测序的RNA分析均需要将RNA转化为互补DNA (cDNA) 链, 在这一过程中, 逆转录或扩增可引入偏倚。研究表明, 与总mRNA混合物相比, PCR扩增文库的复杂性有降低的趋势²¹。此外, 并非所有的转录物都以相同的效率扩增, 这就会导致一些种类的RNA的中断以及其它种类RNA的过度扩增²¹。短读长测序技术会产生出GC偏好性, 其中对GC含量低或含量高的序列代表性不足。使用传统的短读长测序技术会加剧这些问题。

基于cDNA研究的另一个局限性, 是在PCR扩增过程中会丢失有关碱基修饰的所有信息。已知这种碱基修饰具有调

节RNA活性和稳定性的作用, 因此也使得研究人员对它的兴趣日益增加。纳米孔测序能够直接进行RNA测序, 使其可以克服所有的上述挑战 - 提供无偏倚、全长、链特异性的RNA序列 (图3) ²²。

使用纳米孔测序技术, 读长长度仅受出现在孔中的RNA (或DNA) 片段大小的限制, 这使得它可以测序非常长的全长转录本。直接RNA测序能够处理的最长转录本目前长度超过20kb²³。

“直接RNA甲基化分析将彻底改变人们未来十年研究RNA的方式” ²⁴

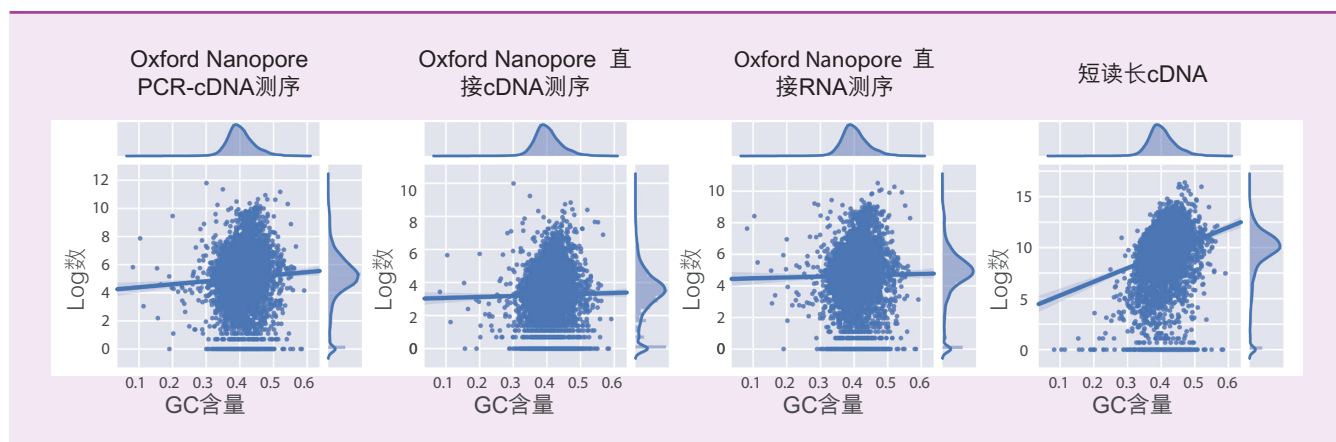


图3

包含扩增的测序工作流程易受序列特异性偏倚的影响。使用三种纳米孔测序技术(PCR-cDNA, 直接cDNA和直接RNA) 和典型的短读长cDNA技术构建酵母转录组文库。在所有情况下, 纳米孔数据集里的GC偏倚低于短读长数据集。

RNA直接测序的测序数据还可能用于检测表观遗传修饰。至目前为止, 使用纳米孔测序技术检测到的RNA修饰包括但不局限于假尿苷²³, N6-甲基腺苷(m6A)^{21,25}及5-甲基胞嘧啶(5mC)^{21,26}。圣克鲁兹加利福尼亚大学的Miten Jain博士认为: “纳米孔技术可以用于直接RNA甲基化分析的能力将彻底改变人们在未来十年研究RNA的方式”²⁴。

RNA直接测序的更进一步好处是能够准确测量poly-A尾长。在真核生物中, 信使RNA (mRNA) 的3' 末端有一系列的腺苷碱基, 也被叫做poly-A尾。这些poly-A尾的长度大小可能不同, 其中最大的长度超过250个核苷酸, 因此超出了短读长测序技术的典型分析能力^{23, 27}。研究表明, poly-A尾的长度是转录后调控的重要因素, 进一步的研究可能为基因表达和疾病研究提供新的见解^{23, 27}。使用纳米孔测序, 研究人员已经鉴定了模式生物秀丽隐杆线虫整个发育过程中的动态poly-A尾分布²⁸。以及人类转录本异构体不同的poly-A尾长度研究²⁹。

与短读长测序相比, 利用纳米孔长读长直接进行RNA测序研究POLY-A的优势在于可以直接耦合剪切异构体和3'UTR异构体信息²⁸。

RNA直接测序技术已被临床研究人员所使用, 以此来开发病毒病原体检测的新方法, 大幅缩短得到检测结果时间。

在流感病毒的研究中, 美国疾病控制和预防中心的研究人员研究出了一种RNA直接测序方案, 将病原体检测和鉴定时间从几天减少到了几个小时³⁰。

转录组分析中的RNA测序深度指南

转录组分析中的RNA测序深度指南旨在比较两个样本之间转录谱的实验可能需要“适中的”测序深度, 而发现新型的转录本以及同源异构体定量通常需要更深的测序深度 (ENCODE指南和RNA测序最佳实践: 2016年12月修订^{31,32})。表1提供了关于不同RNA测序实验所需数据量的一般指导。

在单细胞测序研究中, Byrne等人证明利用少于短读长3-4倍数量的纳米孔读长可产生高度一致的结果。

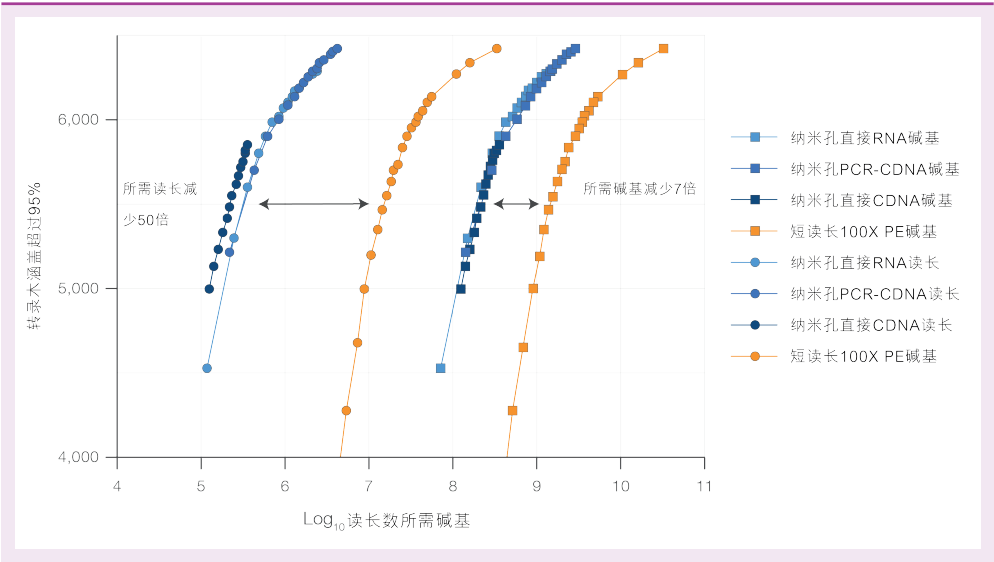
以下建议是基于短读长测序技术, 该技术对从头 (*de novo*) 组装和同源异构体的检测尤为困难。如果需获得类似结果, 需要的长读长可能少得多。例如, 在单细胞测序研究中, Byrne等人²⁰证明每个细胞获取少于短读长数量3-4倍的纳米孔读长, 足以在两个不同平台间产生高度一致的表达谱。此外, 70%以上的细胞达到了表达水平的量化饱和²⁰。将纳米孔PCR-cDNA, PCR-FREE直接cDNA和直接RNA测序与典型的短读长cDNA技术相比, 数据显示达到

表1
不同类型的RNA测序试验所需的读长数量指南。该信息是基于使用短读长测序的大约3Gb的哺乳动物基因组。所得数值由Liu等人^{33,34}、ENCODE指南以及RNA测序最佳实践提供^{31,32}。

| 检测或应用 | 每个样本推荐读长 (M) |
|--------------------------|---------------------------------|
| 差异表达谱 | 10-25 M 双端读长 (~1.5 Gb-3.75 Gb)* |
| 选择性剪接 | 50-100 M 双端读长 (~7.5 Gb-15 Gb) |
| 差异异构体表达 | >100 M 双端读长 (>15 Gb) |
| 从头 (<i>De novo</i>) 组装 | >100 M 双端读长 (>15 Gb) |

图4

与典型的短读长技术相比，覆盖相同数量的转录本，长读长纳米孔RNA测序技术需要的数据更少²¹。



相同转录本覆盖度，纳米孔测序所需读长和碱基数分别少50倍和7倍（图4）²²。

Oxford Nanopore为转录组研究提供的是可扩展的解决方案。利用Flongle™, MinION™, GridION™和PromethION™分别生成1.8Gb, 30Gb, 150Gb和8.6Tb的数据。

快速识别和鉴定RNA病毒的需求

测序对于检测临床样本中的病毒性病原体非常有用。它与其它方法不同的是能够监测病毒的快速变异。然而，大多数测序平台都难以运输，并且不适合用于现场作业。这意味着样本需要被送到具有相关设施的实验室，而这可能会在有快读检测需要的情况中导致不利的延迟，例如病毒爆发。此外，如果病毒滴度很低，为了获得可靠的检测结果，有时还需要进行病毒培养。

纳米孔测序能够大大提高病毒检测速度，使从采样到获得结果的时间缩短为约3小时³⁰。

Oxford Nanopore 提供手持便携式测序仪MinION™和MinIONMk1C™，适合现场和实验室作业。联合基于cDNA的RNA测序方案使用纳米孔测序，能够大大提高病毒检测速度，使从采样到获得结果的时间缩短为约3小时³⁵。此外，该技术可以直接从宏基因组样本中识别滴度低至10⁵拷贝/ml血液中的病毒⁵。

数据分析工具

越来越多的工具和分析流程可用于纳米孔直接RNA和cDNA长读长测序。根据研究目的, RNA测序读长通常被定位到转录组或基因组。转录组定位(Transcriptome mapping) 更易于转录本识别和定量, 但无法发现新型异构体。定位到基因组则允许更全面地分析剪接变体和剪接异构体, 但在生物信息学上挑战性更大。

除了这些社区开发的用于直接RNA和cDNA测序的分析工具和流程, Oxford Nanopore现在还提供三种经过验证的可重复的分析流程。

转录组工作流程³⁶利用转录组映射的方法, 通过纳米孔长读长实现差异基因表达 (DGE) 和差异转录使用 (DTU) 分析。该分析流程基于Love等人的研究方法³⁷。

对于基因组注释, Oxford Nanopore提供pinfish³⁸分析工作流程, 包含pinfish(结合转录组数据分析工具)和比对工具minimap2³⁹, 从纳米孔cDNA/直接RNA长读长中生成GFF2格式的注释文件。

最后, psychopper⁴⁰是一种通过识别具有反转录引物和链转换引物特征的序列来选择全长cDNA读长的工具。建议先运行psychopper获取cDNA读长, 再使用pinfish分析流程进行基因组注释。

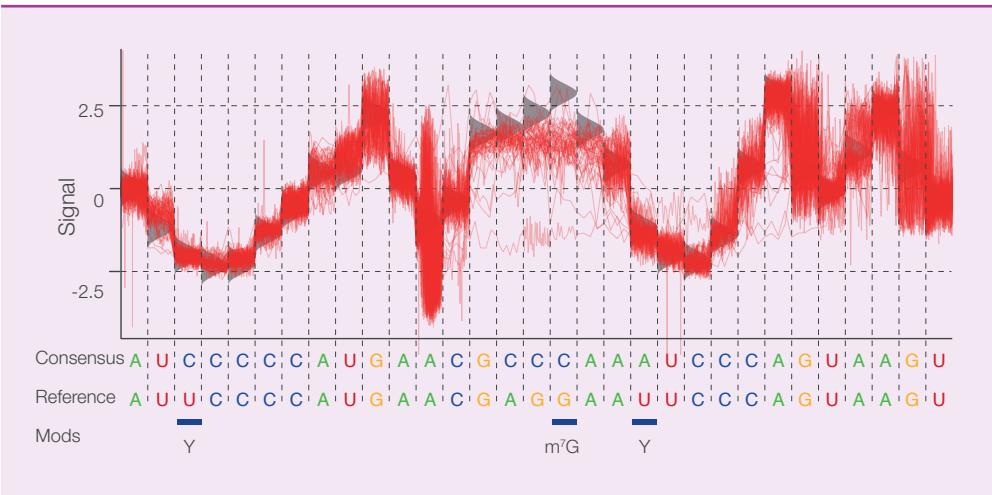
纳米孔直接RNA测序不需要扩增或链合成, 这意味着在测序过程中, 修饰碱基能穿过纳米孔, 并能提供可以在原始信号中检测到的独特特征。Oxford Nanopore开发了Tombo⁴¹分析工具, 可通过直接RNA或DNA测序检测碱基修饰(图5)。Tombo允许研究人员用不同的计算方法研究修饰的碱基, 针对不同应用类型和实验设计具有灵活性⁴¹。还可以通过训练该平台来识别更多的非标准碱基。

详细的逐步教程请参考纳米孔社区:

www.nanoporetech.com/knowledge.

图5

Tombo数据分析包, 从原始纳米孔信号中识别一系列修饰碱基。



个案研究1

纳米孔测序技术能精确解析选择性剪接转录本

短读长测序技术在解析有多种异构体的基因转录本时往往存在困难,这是由于当许多外显子彼此间的距离大于最大读长长度时,该技术就很难检测外显子的连接性^{15,42}。Bolisetty等人¹⁵使用纳米孔测序法和果蝇模型,对具有大量异构体(Dscam1, Rdl, MRP, Mhc)的基因进行了转录组分析。

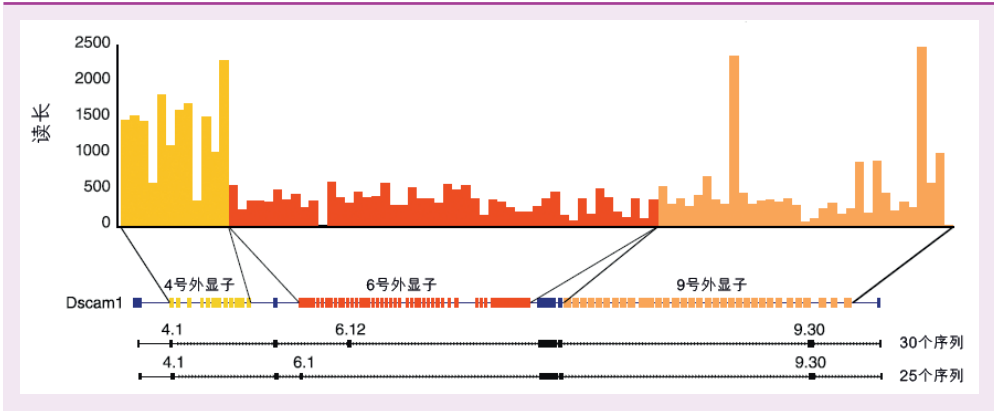
使用基因特异性RT-PCR,该团队检测了这四种基因的总共7899种全长异构体。Dscam1是已知的选择性剪接数量最丰富的基因,有可能产生总共38016种异构体(115个外显子,其中95种为选择性剪接,形成4个簇)⁴³。根据研究目的,作者聚焦于涵盖外显子3-10(19008种潜在异构体)的区域。尽管每个簇内的外显子彼此具有80%的核苷酸一致性(图6),但它们仍然能够从平均一致性>90%的全长比对中准确地解析出7874种Dscam1异构体。

构体的转录本尤为适用,即使在外显子间有显著的相似性的情况下也是如此。

准确解析出Dscam1异构体共7874种,全长比对的平均一致性达>90%。

Bolisetty等人¹⁵证明,长纳米孔读长技术对鉴定具有多种异

图6
Dscam1异构体的纳米孔测序。外显子簇变体(上部)以及两种高表达的Dscam1异构体(底部)的连接。图像经美国康涅狄格州法明顿康涅狄格州健康中心大学布伦顿·格拉维利教授给予许可;最初来源于BioMed Central出版的一篇文章¹⁵中的图表的一部分。



个案研究2

发育中的橄榄果蝇胚胎的转录动力学

橄榄果蝇 (*Bactrocera oleae*) 是橄榄树栽培最重要的害虫之一, 每年造成约8亿美元的损失。已知性别决定在胚胎发育的前6小时内开始, 但发育中胚胎的转录动力学尚未被探索, 雄性决定因素仍有待理解。为了研究发育中橄榄果蝇胚胎的转录组, Anthony Bayega及其同事使用MinION对来自混合性别胚胎的poly-A+RNA进行了全长cDNA测序, 这些胚胎在发育的前6小时中被每小时收集一次 (图7)^{44,45}

超过50%的表达基因具有至少一条全长读长。选择包含5'端接头、poly-A+尾和3'端接头的读长来聚焦于全长转录本。将此方法得到的序列与短读技术产生的序列进行比较, 表明检测相同数量的基因, 此方法所需的读长要少40倍。

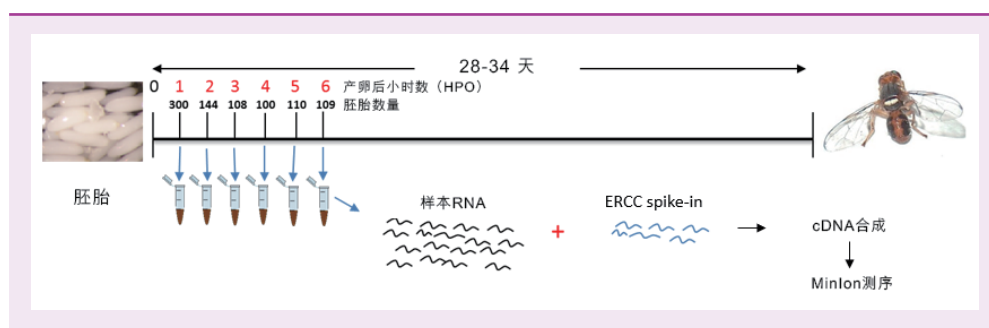
从头转录组组装识别出3553个新基因、8330个与NCBI预测匹配的基因、以及总共79810个转录本。总体而言, 与NCBI预测的转录组相比, 转录组多样性增加了四倍。此外, 用该数据集还校正了在NCBI中错误建模的38个基因。

通过在cDNA合成阶段加入RNA标准 (ERCC) 确定绝对转录本数值, 随后使用这些数值生成标准曲线, 用来后续将每个胚胎转录本丰度的相对读长数转换为绝对读长数。发现每个胚胎的转录本总数在发育的第1小时和第2小时之间下降到一半; 而与第2小时的水平相比, 第3小时增加了143%。类似的转录本丰度模式曾在双翅目黑腹果蝇 (*Drosophila melanogaster*) 中报道过。

利用从成年雄性和雌性橄榄果蝇头部获得的长读长序列, 研究人员得以进一步探索性别决定所涉及的基因异构体的复杂性。在一些基因中, 如double sex (dsx) 和 transformer (tra) 这两个决定性别的主调控因子中, 可能出现各种各样的选择性剪接事件。dsx的异构体复杂性在胚胎和成虫组织之间显著不同; 较长的异构体在胚胎发育早期最为突出, 而成虫脑中较短的异构体更为普遍。特别值得注意的是, 包含外显子4的情况仅在成年雌性组织中观察到。该研究首次提供了关于橄榄果蝇胚胎转录组复杂性的信息, 并展示了更好地理解性别决定在控制对经济有严重影响的农业害虫种群数量中具有潜力。

图7

Bayega等人从样本采集到测序的实验过程示意图⁴⁴。



个案研究3

利用全长转录本进行癌症研究

纳米孔测序能够测序全长转录本并精确识别和量化多种异构体, 使其对癌症研究尤为有用。该技术已被应用于许多癌症类型的研究, 包括白血病⁴⁶、乳腺癌⁴⁷、卵巢癌⁴⁷及肺癌⁴⁸。

Suzuki等人⁴⁸采用纳米孔cDNA测序法对与肺癌有关的一组基因中(EGFR, KRAS, NRAS, NF1, ALK 和RET)的多种常见突变类型, 包括单个碱基替换、短片段缺失、外显子跳跃和融合进行了测序。此外, 该团队利用纳米孔测序获得的长读长对两个抗癌药物疗效相关的基因组DNA突变进行了定相。在研究取得成功之后, 研究人员评论说到: “简单的MinION测序的方法使中小规模的研究中心和医院可能通过驱动基因的分型来开展研究, 并选择合适的治疗方法”⁴⁸。

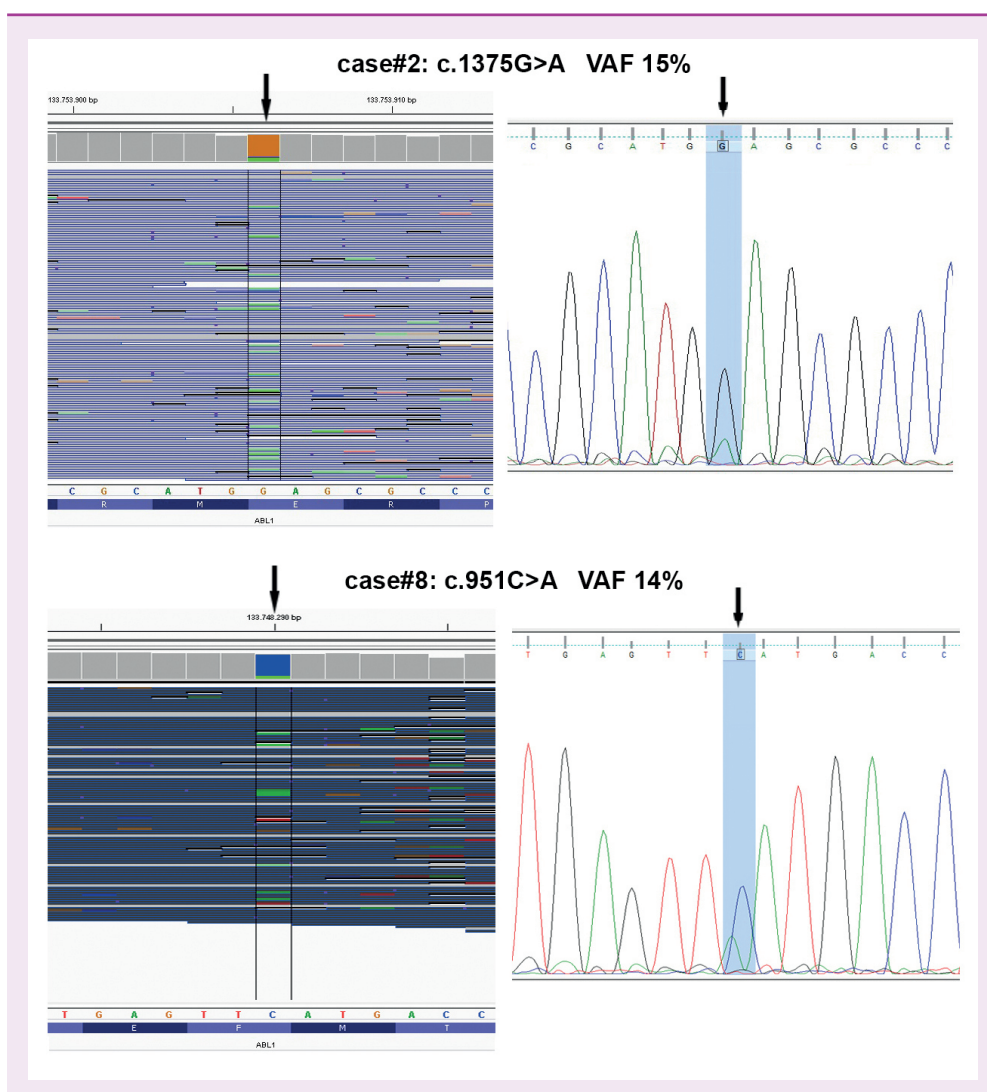
在意大利巴里大学, 研究人员正研究利用纳米孔测序检测BCR-ABL1激酶结构域(KD) 突变的潜力⁴⁶。BCR-ABL1基因融合发生在95%的慢性粒细胞白血病患者中⁴⁹, 该基因编码非调节酪氨酸激酶蛋白。BCR-ABL1基因中的KD突变引起并表明其对一线治疗药物酪氨酸激酶抑制剂产生了抗药性⁴⁶。目前Sanger测序是用于BCR-ABL1中变体检测的金标准, 但由于其灵敏度仅为20%, 因此并不适用于识别低水平的变体。

MINION在灵敏度、成本和省时性方面明显优于SANGER测序⁴⁶。

在一项涉及24个样本的研究中, 使用100x和1000x测序深度, 研究小组评估了纳米孔测序在鉴定大小为1.7kb的cDNA扩增子中的KD突变的能力。在2个样本中, 纳米孔测序数据鉴定出了具有临床意义的突变, 而使用Sanger测序结果最初并不明显 (图8)。该团队还指出, 使用低于100x的纳米孔测序深度足以识别所有突变。通过纳米孔测序技术获得的长读长更能检测到在同一克隆中的突变 (in cis; 复合突变体), 而这对于Sanger测序或短读长测序极具挑战。研究人员总结说到: “我们的数据表明MinION在灵敏度、成本和省时方面明显优于Sanger测序, 它的优点还在于可以确定多个突变的克隆组成”⁴⁶。

图8

纳米孔测序（左侧）在2个样本中检测到了起初在Sanger测序数据（右侧）中并不明显的突变，这表明纳米孔技术增强了灵敏度。图片由意大利巴里大学Crescenzo Minervini博士提供。



个案研究4

识别神经精神疾病基因中的新型异构体

双相情感障碍和精神分裂症有很强的遗传成分, 全基因组关联研究 (GWAS) 确定了很多风险基因座。其中一个基因座CACNA1C——编码电压门控性钙通道亚基, 是一个潜在的治疗靶点。然而, 它的超长长度 (>13kb), 外显子数量(>50) 以及高水平的选择性剪接意味着我们对CACNA1C亚型结构和表达知之甚少⁴²。

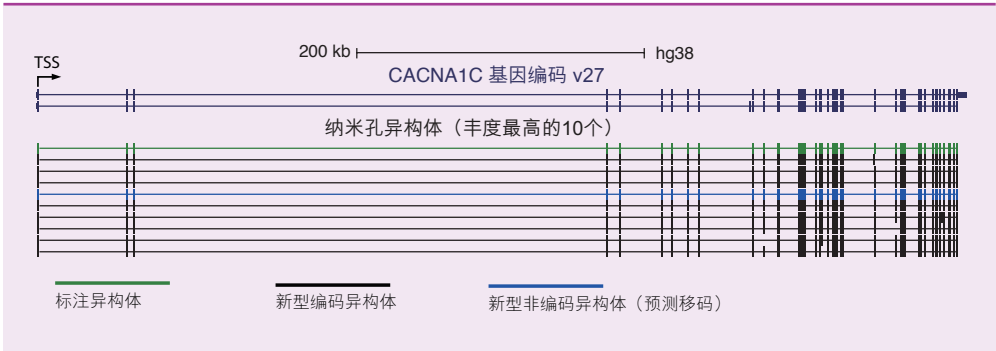
来自牛津大学和Earlham研究所的研究人员利用纳米孔测序来研究死亡后人脑的六个区域中 (小脑, 纹状体和4个皮层子区域) CACNA1C的表达⁴²。纳米孔测序的长读长完整测序了CACNA1C(6.5kb) 的外显子区域, 并发现了38个可能的新外显子和90个异构体⁵⁰。在40个早前注释的异构体中, 只有7个被识别出来。研究小组认为, 在以往的数据中存在假阳性是造成这种差异的原因之一, 而这是由于其他测序方法在转录组组装时的固有复杂性造成的。

有趣的是, 10种表达最多的异构体中有9种为新型转录本(

图9)⁵⁰。对83种新异构体进行初步研究后, 还发现了许多会改变蛋白质及其功能的异构体。该团队还证明了异构体分布情况在脑区域之间有所不同, 并且CACNA1C的剪接比目前认识的更为复杂⁵⁰。

首席研究员Michael Clark博士强调: “这些结果是评估CACNA1C作为潜在治疗靶点的第一步, 并证明了长读长纳米孔测序对阐明所表达的基因本质的能力⁴²。”

图9
在脑组织10个最丰富的CACNA1C异构体中, 有9个是新型异构体, 突出显示长读长有利于进行异构体表征。图片由英国牛津大学Michael Clark博士提供。



个案研究5

纳米孔测序精确量化任何大小的全长cDNA的丰度

Oikonomopoulos等人¹³比较了纳米孔技术和其他测序方法在进行基于cDNA的转录组研究中的优劣。该团队使用表定量数据质量控制联盟(External RNA Controls Consortium, ERCC)的 RNA Spike-in混合物和从人类 HEK-293细胞制备的cDNA库, 对不同测序仪的性能进行了比较。所有测试平台均使用模板转换cDNA实验方案。大多数(90.8%)的纳米孔读长为转录本全长, 且未观察到转录长度偏倚。

基于纳米孔数据的转录本丰度估计值与ERCC混合物的已知浓度之间存在很强的相关性(皮尔逊相关系数 $r=0.98$)。

对于HEK-293样本, 短读长数据产生的转录本丰度估计值与基于纳米孔数据的估计值之间也存在很强的相关性(皮尔逊相关系数 $r=0.75$)。对于同一样本, 纳米孔数据丰度估计值也与基于另一长读长测序平台数据的估计值高度一致(皮尔逊相关系数 $r=0.82$)。这些结果表明, 纳米孔技术可用于可靠地量化任何大小的全长cDNA分子, 进而有可能减少繁重的分析²²。

个案研究6

使用RNA直接测序进行全局性的转录组分析

酿酒酵母菌株CEN.PK113-7D因具有简单的遗传操作性和快速的生长速率而被广泛用于学术和工业研究。为了对其进行深入了解, 阿肯色州大学的研究人员首先使用纳米孔测序创建了一个完整的基因组组装, 使用长读长来改进现有的片段化的短读长组装⁵¹。随后该团队进行了RNA直接测序, 以确定其在两种不同生长条件下的基因表达模式(双向生长)。

数据揭示了在葡萄糖和乙醇上生长的酵母复制品上独特的基因表达谱(图10)。正如预期的那样, 在使用乙醇作为生长基质的有机物中, 与乙醇代谢和细胞应激反应有关的转录本显著增加; 而葡萄糖组中上调基因与快速生长过程和乙醇发酵相关。

研究小组发现, 约70%通过直接RNA测序获得的序列与全长转录本相对应, 并且可以检测到5kb以上的全长转录本。使用纳米孔技术的RNA直接测序获得数据与短读长测序技术获得的数据相比较后显示, 虽然数据总量大约是使用短读长数据的一半, 但两者的平均覆盖深度相当 (图11)。此外, 有证据表明纳米孔读长比短读长RNA分析技术⁵⁰表现出更低的GC偏倚⁵²。

长读长RNA直接测序还可以鉴定许多聚腺苷酸化非编码RNA, 包括rRNA, 端粒酶RNA和长链非编码RNA (lncRNA)⁵¹。

该小组在总结时指出: “我们相信RNA直接测序将成为未来 ‘完整基因组时代’ 进行转录组分析的一项多功能的工具”。

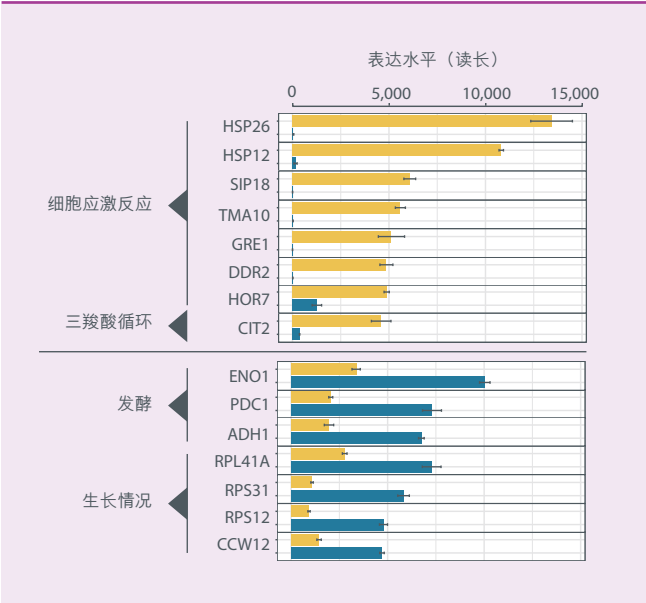


图10
RNA直接测序能够鉴定在葡萄糖 (蓝色) 或乙醇 (黄色) 上生长的酵母复制品的独特的基因表达谱。图片由美国阿肯色州大学Intawat Nookaew博士提供

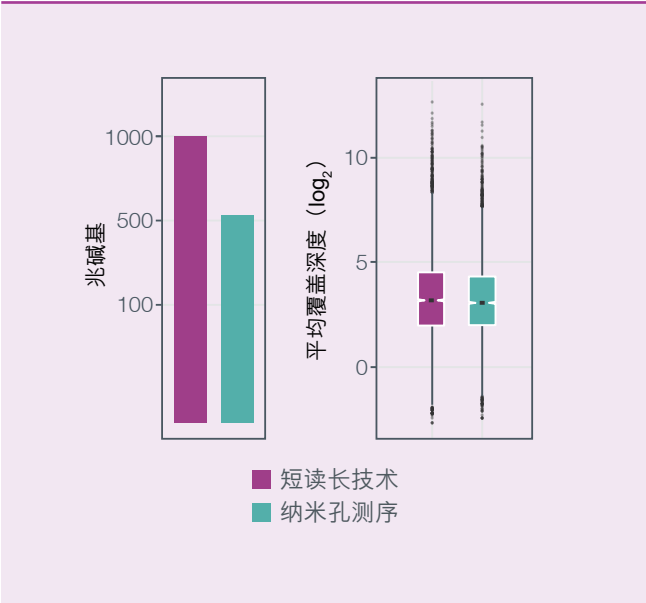


图11
如平均覆盖深度方框图所示, 由500Mb直接RNA测序数据产生的转录物的动态范围与从1000Mb短读长测序技术数据获得的转录物的动态范围相似。图片由美国阿肯色州大学Intawat Nookaew博士提供

总结

使用传统测序方法进行RNA测序已协助解答了许多生物问题。但这种方法存在局限性，例如在获得全长转录物、问题异构体的识别和量化、文库构建相关的偏倚（PCR、逆转录等）等方面均能力有限。其中的一部分问题已经通过其他商业上可用的长读长测序设备解决，但是对于大规模实验来说，使用该方法往往成本高昂。

纳米孔测序可提供适用于异构体及其丰度鉴定的全长转录本。

纳米孔测序能提供所有长读长在转录组分析中所能提供的优势。纳米孔技术可生成全长转录本，适用于异构体以及丰度的精确鉴定。便携、快速的实验室制备步骤和实时分析流程，也使得纳米孔测序在病毒性病原体研究领域，特别是在实地作业的环境中具有优势⁶。无逆转录和PCR扩增偏倚的RNA直接测序的出现更是进一步推动了转录组学的分析。最后，RNA直接测序还可以通过单个文库制备鉴定多种RNA修饰，无需经过冗长的亚硫酸氢盐转化或抗体沉淀操作²¹。这些特点，再结合纳米孔平台和技术所提供的扩展性，使纳米孔测序成为了（用印第安纳大学伯明顿分校Marinov博士的话说）“转录组研究中最令人兴奋的发展”⁵³。

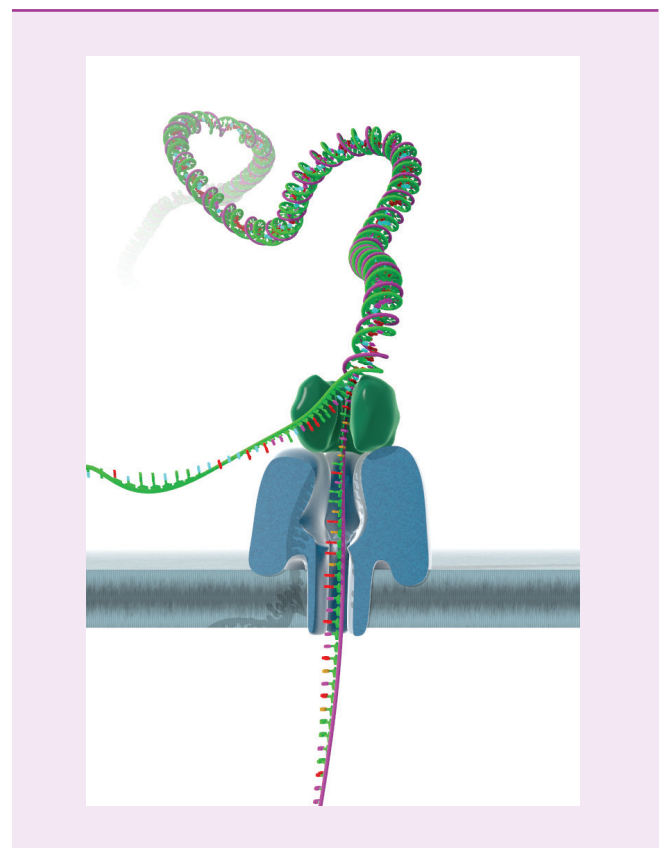


图12

一个RNA分子正在通过纳米孔。cDNA第一链（可选步骤）能提供最大的测序效率，但是不穿过纳米孔。

关于 Oxford Nanopore Technologies Technologies

Oxford Nanopore Technologies公司开发的第一款纳米孔测序设备MinION——一种便携式、实时的、长读长和低成本设备, 对测序技术产生了革命性的改变。

纳米孔技术是唯一可以直接测序RNA的测序技术, 消除了能够造成潜在的偏倚的因素。由于长读长能够对全长转录本进行测序, 它也被越来越多地用于cDNA的分析 (表2), 且可以在测序核苷酸的同时读取碱基修饰。

MinION Mk1C和Flongle是纳米孔测序产品集中的最新产品。MinION Mk1C将MinION的便携性和高产量与全集成的触摸屏、用于机载分析的高性能计算机以及内置Wi-Fi、蓝牙和移动数据的功能完美结合。Flongle是一种用于MinION和GridION的测序芯片, 旨在为较小的测序实验和分析提供最经济有效的方法。

使用GridION Mk1, PromethION24和PromethION48能够满足更高通量的需求。紧凑型台式系统采用了与MinION相同的纳米孔技术, 可分别同时运行多达5个, 24个和48个测序芯片。每个测序芯片可以独立使用, 研究人员可以选择一次使用多少测序芯片, 同时进行不同的实验。

关于无偏倚的长读长RNA测序的最新信息, 请访问 www.nanoporetech.com/RNA

表2
Oxford Nanopore公司的RNA测序仪的应用领域概述

| 应用 | PCR-cDNA测序试剂盒SQK-PCS109 | Direct-cDNA测序试剂盒SQK-DCS109 | Direct RNA测序试剂盒SQK-RNA003 |
|------------------|-------------------------|----------------------------|---------------------------|
| 基因表达 | *** | *** | * |
| 剪接变体和融合蛋白 | *** | *** | ** |
| 病毒RNA | *** | *** | *** |
| 测序天然RNA允许探索RNA修饰 | | | *** |

参考文献

1. Beyer, K. et al. Differential expression of alpha-synuclein, parkin, and synphilin-1 isoforms in Lewy body disease. *Neurogenetics* 9, 163-172 (2008).

2. Gonzalez, E. & McGraw, T. E. The Akt kinases: isoform specificity in metabolism and cancer. *Cell Cycle* 8, 2502-2508 (2009).

3. Bao, Z.-S. et al. RNA sequencing of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res.* 24, 1765-1773 (2014)

4. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228-232 (2016).

5. Greninger, A. L. et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 7, 99, doi:10.1186/s13073-015-0220-9 (2015).

6. Batovska, J., Lynch, S.E., Rodoni, B.C., Sawbridge, T.I., & Cogan, N.O. Metagenomic arbovirus detection using MinION nanopore sequencing. *J. Virol. Methods* 249:79-84 (2017).

7. Broadhurst, M. J., Brooks, T. J. & Pollock, N. R. Diagnosis of ebola virus disease: Past, present, and future. *Clin. Microbiol. Rev.* 29(4):773-93 (2016).

8. Fracasso, A., Trindade, L. M. & Amaducci, S. Drought stress tolerance strategies revealed by RNA sequencing in two sorghum genotypes with contrasting WUE. *BMC Plant Biology* 16, 115 (2016).

9. Trapnell, C. et al. Transcript assembly and quantification by RNA sequencing reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511-515 (2010).

10. Kazuo, M. & Sumio, S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* 138, 171-174 (1994).

11. Finotello, F. et al. Reducing bias in RNA sequencing data: a novel approach to compute counts. *BMC Bioinformatics* 15, S7 (2014).
12. Zhu, Y., Machleder, E., Chenchik, A., Li, R. & Siebert, P. Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *Biotechniques* 30, 892-897 (2001).
13. Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D. & Ragoussis, J. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci. Rep.* 6 (2016).
14. Hargreaves, A. D. & Mulley, J. F. Assessing the utility of the Oxford Nanopore MinION for snake venom gland cDNA sequencing. *PeerJ* 3: e1441 (2015).
15. Bolisetty, M. T., Rajadinakaran, G. & Graveley, B. R. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* 16, 204 (2015).
16. Steijger, T. et al. Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* 10, 1177-1184 (2013).
17. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12, 671-682 (2011).
18. Weirather, J. L. et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* 6 (2017).
19. Hardwick, S.A. Targeted, high-resolution RNA sequencing of non-coding genomic regions associated with neuropsychiatric functions. *Front. Genet.* 10:309 (2019).
20. Byrne, A. et al. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* 8 (2017).
21. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15(3):201-206 (2018).
22. Oxford Nanopore Technologies. Quantitative RNA-seq: PCR-cDNA, PCR-free Direct cDNA and Direct RNA sequencing. Poster. Available at: <<https://nanoporetech.com/resource-centre/quantitative-rna-seq-pcr-cdna-pcr-free-direct-cdna-and-direct-rna-sequencing>> [Accessed: 22 October 2019]
23. Timp, W. & Jain, M. Direct RNA cDNA sequencing of the human transcriptome. Presentation. Available at: <<https://nanoporetech.com/resource-centre/direct-rna-cdna-sequencing-human-transcriptome>> [Accessed: 22 October 2019]
24. Jain, M. Personal communication with Oxford Nanopore Technologies on 27 February 2018.
25. Liu, H. et al. Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* 10(1):4079 (2019).
26. Viehweger, A. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res.* 29(9):1545-1554 (2019).
27. Jalkanen, A.L., Coleman, S.J. & Wilusz, J. Determinants and implications of mRNA poly(A) tail size - Does this protein make my tail look big? *Semin. Cell Dev. Biol.* 34: 24-32 (2014).
28. Roach, N.P. et al. The full-length transcriptome of *C. elegans* using direct RNA sequencing. *bioRxiv* 598763 (2019).
29. Workman, R.E. Nanopore native RNA sequencing of a human poly(A) transcriptome. *bioRxiv* 459529 (2018).
30. Keller, M. Direct RNA sequencing of Influenza viral RNA using the MinION nanopore sequencer. Presentation. Available at: <<https://nanoporetech.com/resource-centre/direct-rna-sequencing-influenza-viral-rna-using-minion-nanopore-sequencer>> [Accessed: 22 October 2019]
31. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74 (2012).
32. Sloan, C. A. et al. ENCODE data at the ENCODE portal. *Nucleic Acids Research* 44, D726-D732 (2016).
33. Liu, Y. et al. Evaluating the impact of sequencing depth on transcriptome profiling in human adipose. *PLoS One* 8, e66883 (2013).
34. Liu, Y., Zhou, J. & White, K. P. RNA sequencing differential expression studies: more sequence or more replication? *Bioinformatics* 30, 301-304 (2014).
35. Chiu, C. Personal communication with Oxford Nanopore Technologies on 1 December 2017.
36. GitHub. Pipeline-transcriptome-de. Available at: <<https://github.com/nanoporetech/pipeline-transcriptome-de>> [Accessed: 03 October 2019]
37. Love, M.I., Soneson, C. and Patro, R. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Research* 7:952 (2018).
38. GitHub. Pinfish. Available at: <<https://github.com/nanoporetech/pinfish>> [Accessed: 03 October 2019]
39. GitHub. Minimap2. Available at: <<https://github.com/lh3/minimap2>> [Accessed: 03 October 2019]
40. GitHub. Pychopper. Available at: <<https://github.com/nanoporetech/pychopper>> [Accessed: 03 October 2019]
41. GitHub. Tombo. Available at: <<https://nanoporetech.github.io/tombo>> [Accessed: 22 October 2019]
42. Clark, M. Elucidating the expression and splicing patterns of neuropsychiatric disease genes in human brain. Presentation. Available at: <<https://nanoporetech.com/resource-centre/elucidating-expression-and-splicing-patterns-neuropsychiatric-disease-genes-human>> [Accessed: 22 October 2019]
43. Schmucker, D. et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101(6):671-84 (2000).
44. Bayega, A. et al. Transcriptome landscape of the developing olive fruit fly embryo delineated by Oxford Nanopore long-read RNA-seq. *bioRxiv* 478172 (2018).
45. Bayega, A. Transcriptome landscape of the developing olive fruit fly embryo delineated by Oxford Nanopore long-read RNA-seq. Presentation. Available at: <<https://nanoporetech.com/resource-centre/anthony-bayega-transcriptome-landscape-developing-olive-fruit-fly-embryo-delineated>> [Accessed: 21 October 2019]
46. Minervini, C.F. et al. Mutational analysis in BCR-ABL1 positive leukemia by deep sequencing based on nanopore MinION technology. *Exp. Mol. Pathol.* 103(1):33-37 (2017).
47. de Jong, L.C. et al. Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res.* 19:127 (2017).
48. Suzuki, A. et al. Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer. *DNA Res.* 24(6): 585-596 (2017).
49. Moore, F.R., Rempfer, C.B., and Press, R.D. Quantitative BCR-ABL1 RQ-PCR fusion transcript monitoring in chronic myelogenous leukemia. *Methods Mol. Biol.* 999:1-23 (2013).
50. Clark, M.B. et al. Long-read sequencing reveals the splicing profile of the calcium channel gene CACNA1C in human brain. *bioRxiv* 260562 (2018).
51. Jenjaroenpun, P. et al. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Res.* 6(7):e38 (2018).
52. Nookaew, I. Transcriptional landscapes analysis through direct RNA sequencing. Presentation. Available at: <<https://nanoporetech.com/resource-centre/transcriptional-landscapes-analysis-through-direct-rna-sequencing>> [Accessed: 21 October 2019]
53. Marinov, G. K. On the design and prospects of direct RNA sequencing. *Brief. Funct. Genomics* 16(6):326-335 (2017).

Oxford Nanopore Technologies

phone +44 (0)845 034 7900

email sales@nanoporetech.com

twitter @nanopore

www.nanoporetech.com



Oxford Nanopore Technologies, 风轮图标, EPI2ME, Flongle, GridION, Metrichor, MinION, MinKNOW, PromethION, SmidgION, Ubik 和 VolTRAX 是 Oxford Nanopore Technologies plc 在不同国家的注册商标。包含的所有其他品牌和名称均为其各自所有者的财产。© 2021 Oxford Nanopore Technologies plc。版权所有。Oxford Nanopore Technologies 产品并非旨在用于健康评估或诊断、治疗、缓解、治愈或预防任何疾病或状况。

WP_1047(CN)_V1_24Aug2021