



Oxford Nanopore 靶向测序

入门指南

引言

靶向测序（通过富集目标 DNA/RNA 分子或去除不需要的分子）是一种非常有用的方法，用于为感兴趣区域生成足够的覆盖深度，从而进行具有成本效益的信息分析。

在诸多测序应用中，研究重点（单个基因或所选的基因组区域）仅包含一个全长基因组的很小一部分。在这些情况下，通过全基因组测序进行表征是低效、昂贵和耗时的。在癌症研究中，全基因组测序无法做到快速分析致癌融合基因；在宏基因组分析中，如果感兴趣的微生物在混合样本中代表性较差，则全基因组测序可能很难将其富集。

靶向富集是解决这些难题的有效策略：通过将更多的测序时间投入到感兴趣区域中，能够极大地增加它们的覆盖深度。这可以显著减少所需测序文库的数量和运行次数，减少数据分析负担，以实现更快且更具成本效益的工作流程。通常通过对所选序列进行基于聚合酶链式反应 (PCR) 的扩增来实现靶向富集，这为高覆盖深度区域测序提供了一种简单而经济的方法。

然而，并非所有序列都适用于 PCR：GC 含量高或高度重复的序列可能难以扩增或者甚至无法扩增。PCR 还会去除碱基修饰，使得无法对其进行分析。此外，通过传统短读长技术进行靶向测序限制了可用于分析的目标片段长度，妨碍对较大区域（诸如那些与结构变异 [SV] 相关联的区域）进行分析。相反，较大区域必须由较短片段重新组装而成；这可能导致误差，尤其是高度重复或序列数据中代表较差的区域。

纳米孔技术将创新富集方法与实时测序结合，与传统靶向测序方法相比较具备显著优势（[表 1](#)）。

表 1. 纳米孔技术用于靶向测序的优势

富集传统技术无法触及的区域

进行无 PCR 富集和测序，靶向重复区域或高 GC 含量的区域

对感兴趣区域进行表征，且读长序列长度不限

富集和测序任何读长序列长度的目标片段：从 20 bp 至超过 4 Mb

获取信息丰富的数据

在来自单个无扩增 DNA 测序运行的目标片段中，检测单核苷酸变异 (SNV)、SV、重复扩张序列以及修饰碱基

根据您的应用方向进行按需测序

适配您需求的测序方式 - 从用于现场测序的便携式平台，到高通量的按需台式系统

快速获取结果

通过快速富集和文库制备，外加实时测序和分析，将项目周转时间减少至数小时或更短

以最低的启动成本在您的实验室中进行测序

仅需耗材成本即可启动纳米孔测序；对混样建库中的目标片段进行测序，可进一步提高成本效益

无需在实验室进行富集即可开展靶向测序

使用适应性采样技术，在实时的纳米孔测序过程中对选定区域进行富集或去除

部署简单的端到端工作流程

根据从样本到获得答案指南，快速有效地进行测序和分析设置

基于化学试剂的纳米孔测序技术具有高度可扩展性，可在多种设备中部署，以满足不同应用方向的需要。便携式 MinION™ 和 MinION Mk1C 实现了不限地点测序。灵活扩展的 GridION™ 和超高通量 PromethION™ 台式设备则可对多个目标片段平行测序。

此外，无需等待分批处理样本：所有测序芯片均可单独寻址，因此可在一项或多项试验中按需进行测序。纳米孔测序能够实时呈现数据：测序开始后即可进行碱基识别和下游数据分析。将快速富集和文库制备选项相结合，仅需几小时即可完成从样本到获得答案的整个工作流程。

纳米孔技术可对任何长度的片段进行测序，无论是短片段还是超长片段 (>4 Mb)。长片段 PCR 可被用于对长度为数十 kb 的靶向片段进行富集和测序至高覆盖深度；基于 PCR 的靶向测序有各种灵活的方法。在单次运行中也可使用条形码选项，对混样建库中的多个目标片段进行测序。

然而，纳米孔工作流程的任一步骤均不需要进行扩增，且当前存在两种无 PCR 靶向测序的创新方法：适应性采样（见第 4-5 页）和 Cas9 靶向测序（见第 6 页）。

由此，通过纳米孔测序技术，可对先前通过高深度全基因组测序才能表征的大片段区域进行靶向测序，例如包含低复杂度且高 GC 含量序列、大片段 SV 以及重复扩张序列的大片段区域。通过对非扩增 DNA 进行测序，将碱基修饰保留在了目标片段中，并且可与核苷酸序列一同进行分析。适应性采样省去了对于任何基于实验室的富集步骤的需求：在测序同时实时完成富集。同时，采用简单的一站式方案，Cas9 的靶向富集可在不到两小时中完成，并且可被用于对非常大型的目标区域进行富集和测序：Cas9 富集序列的当前记录为 198 kb，从末端到末端跨越了整个 *BRCA1* 基因¹。

在本指南中，我们介绍了靶向纳米孔测序的不同方法和优势。



1. Iyer, S. Understanding genetic variation in cancer, using targeted nanopore sequencing. Presentation. Available at: nanoporetech.com/shrutiIyer [Accessed: 21 April 2023]

适应性采样

通过利用实时纳米孔测序，以及完全由生物信息学驱动的策略，适应性采样拓宽了靶向富集的边界。这种方法来源于 Nanopore 社区的创新举措²⁻⁵，在测序期间同时开展靶向富集或去除，且样本制备期间不需要靶向富集。该方法已集成于 MinKNOW™ 中。MinKNOW™ 是驱动所有纳米孔测序设备的操作软件。

首先，对整个 DNA 文库进行制备以用于测序，无需任何扩增或富集步骤。随后将样本添加至测序芯片，并在 MinKNOW 中设置测序运行。在这里选择适应性采样，并上传 FASTA 格式的参考文件和包含富集或去除坐标的 BED 文件（图 1）。然后按照正常方式开始运行，且测序开始。

通过实时测序，当 DNA 链穿过纳米孔时可确定序列是否代表感兴趣区域。这是通过将测序链的开头快速定位至所提供的参考序列来实现的。如果序列位于要进行富集的目标区域内，或不是要被去除的序列，则软件允许继续测序。如果序列不是目标区域，或者要予以去除，链将选择性地从该纳米孔中弹出，以避免进一步测序，并且将孔占用时间释放以用于感兴趣区域。读长序列长度和目标区域数量不受限。通过这种方式，即便是非常大的基因组合，或是整个基因组，也可在测序期间对其进行完全富集或去除 — 这就省去了复杂而又耗时的基于实验室的传统方法，诸如杂交捕获。

MinION 和 GridION 设备提供适应性采样，PromethION 2、24 和 48 设备有 Beta 版本。

请前往以下链接观看适应性采样动画：

nanoporetech.com/resource-centre/adaptive-sampling-oxford-nanopore

2. Loose, M., Malla, S., and Stout, M. Real-time selective sequencing using nanopore technology. *Nat. Methods.* DOI: doi.org/10.1038/nmeth.3930 13:751-754 (2016)
3. Payne, A. et al. Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. *bioRxiv*. DOI: doi.org/10.1101/2020.02.03.926956 (2020)
4. Kovaka, S. et al. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat. Biotech.* DOI: doi.org/10.1038/s41587-020-0731-9 (2021)
5. Weilguny, L. et al. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nat. Biotech.* DOI: doi.org/10.1038/s41587-022-01580-z (2023)

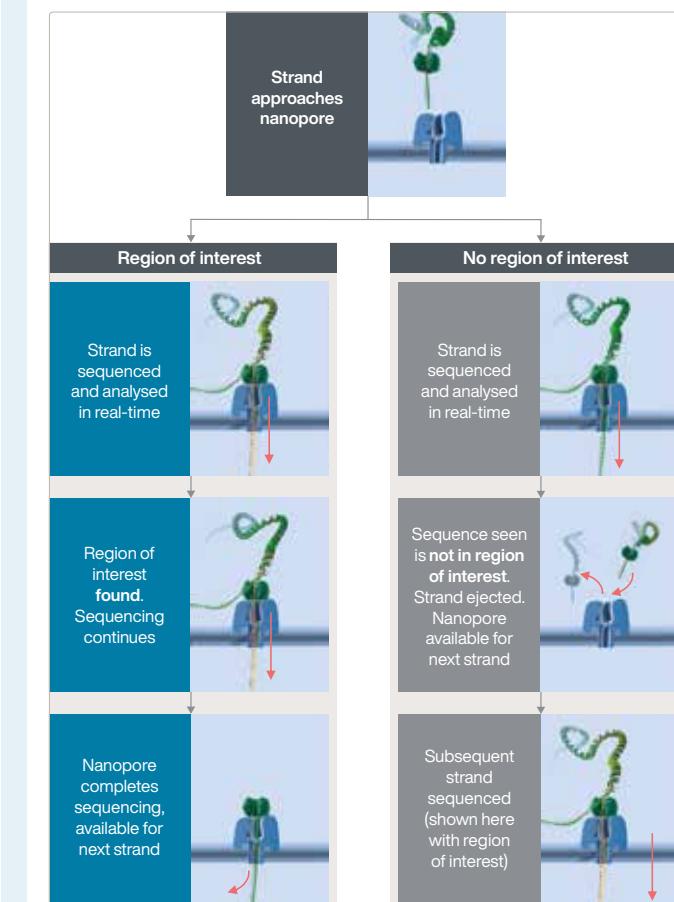


图 1. 使用适应性采样在实时测序期间对 DNA 链进行富集/去除

简并代表性 甲基化测序

亚硫酸氢盐转化和测序是识别和定量全基因组 DNA 甲基化的常用方法。然而，全基因组的亚硫酸氢盐测序成本高昂、需要投入大量劳力，并且是基于 PCR 的测序，会引入偏差，导致难以扩增的目标片段代表性较差。因此在亚硫酸氢盐测序下，人类基因组中只有约 75% CpG 的覆盖深度达到 50x。简并代表性亚硫酸氢盐测序 (RRBS) 更具成本效益，但仅能捕获哺乳动物基因组中 10–15% 的 CpG。

纳米孔测序能够直接检测甲基化胞嘧啶（例如在 CpG 位点），而无需亚硫酸氢盐转化或 PCR。简并代表性甲基化测序 (RRMS) 与天然 DNA 纳米孔测序和适应性采样相结合，可以靶向 310 Mb 的人类基因组，包括 CpG 高度富集的区域。这种方法可以对 5-甲基胞嘧啶 (5mC) 等表观遗传修饰进行直接靶向分析。

基准研究显示，纳米孔测序数据的覆盖均一性高于亚硫酸氢盐测序数据。使用无 PCR 的纳米孔测序，有可能在整个基因组上产生更均匀的覆盖，包括其他测序方法无法触及的区域。纳米孔长读长和超长读长序列可保留甲基化和基因组变异的长范围信息，并实现有效的定相。

利用适应性采样，在测序期间进行灵活、实时的富集。RRMS 无需特殊的文库制备。从单个数据集中即可识别结构变异 (SV)、单核苷酸变异 (SNV) 和甲基化。

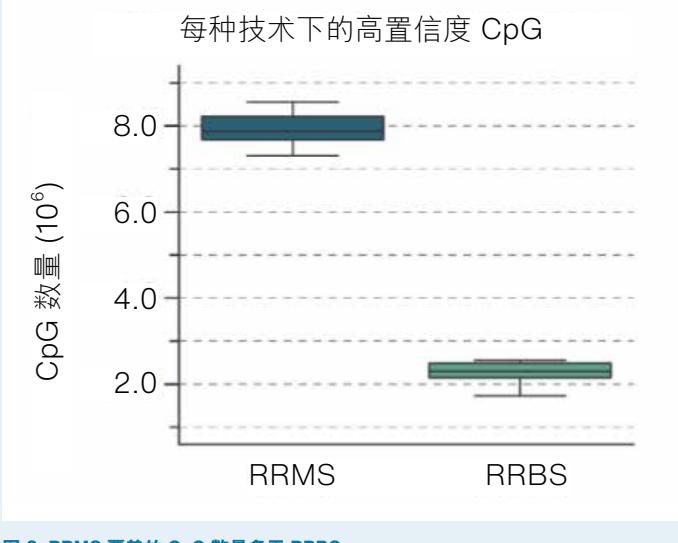


图 2. RRMS 覆盖的 CpG 数量多于 RRBS

有关 RRMS 的更多信息，请查看我们的甲基化入门指南：
nanoporetech.com/resource-centre/investigating-methylation-in-the-human-genome

Cas9 靶向测序

Cas9 靶向测序利用 Cas9 酶的可编程 DNA 切割能力来进行靶向富集。无 PCR 的文库制备工作流程非常简单（图 3），不到两个小时即可完成，并且可对长度为数十 kb 或以上的区域进行富集。

开始文库制备时，将 Cas9 与两种特异性 RNA 结合，形成核糖核酸蛋白复合物 (RNP)：

- **tracrRNA:** 一种催化活性所需的结构 RNA
- **crRNA:** 定制设计探针，与目标区域的临近区域互补

在文库制备中，首先避免连接测序接头，对样本中所有基因组 DNA 进行去磷酸化以“封闭”链末端。然后添加每个目标区域的 RNP；这些 RNP 与每个感兴趣区域的临近 DNA 序列（由 crRNA 探针决定）结合，使得 Cas9 能够对 DNA 进行切割。这会释放磷酸化末端，使得测序接头连接至暴露的靶标分子。将不需要的非磷酸化 DNA 去除，这节省了感兴趣区域的测序时间，产生数十倍或数百倍的富集。通过对天然 DNA 进行富集，这为靶向无法进行 PCR 扩增的大片段区域提供了一种有效方法，有助于深度表征 SV、重复扩张序列、SNV 和碱基修饰。

充分整合而精简实用的 Oxford Nanopore **Cas9 测序试剂盒**纳入了第三方试剂以及纳米孔文库制备所需的试剂，以实现更具成本效益的样本制备。

为了进行高保真切割，我们推荐使用 Alt-R 酿脓链球菌 HiFi Cas9 核酸酶 V3（可从 IDT 获取），不过也可将酶换成您所选择的酶。若需了解更多关于 crRNA 设计的信息，请参阅第 9 页。

访问以下链接了解更多关于 Cas9 测序试剂盒的信息：

<https://store.nanoporetech.net/cas9-sequencing-kit.html>

Cas9 测序试剂盒受两份许可协议约束；完整声明请参阅上方 Cas9 测序试剂盒产品页面提供的链接。

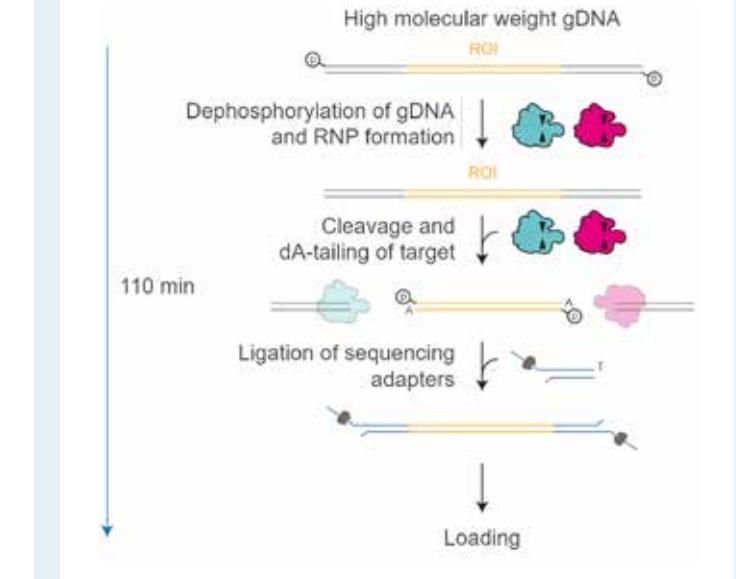
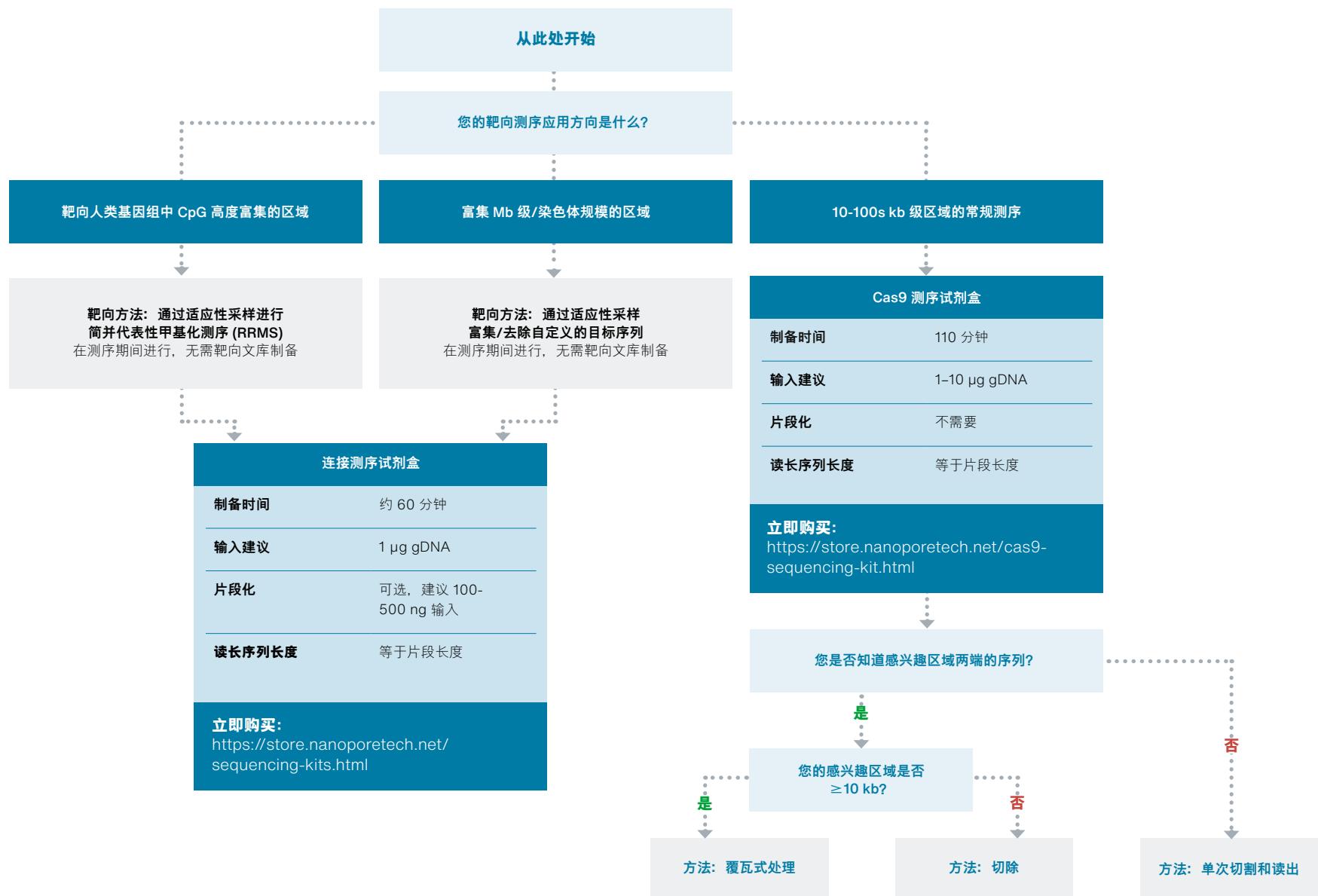


图 3. Cas9 靶向测序文库制备工作流程



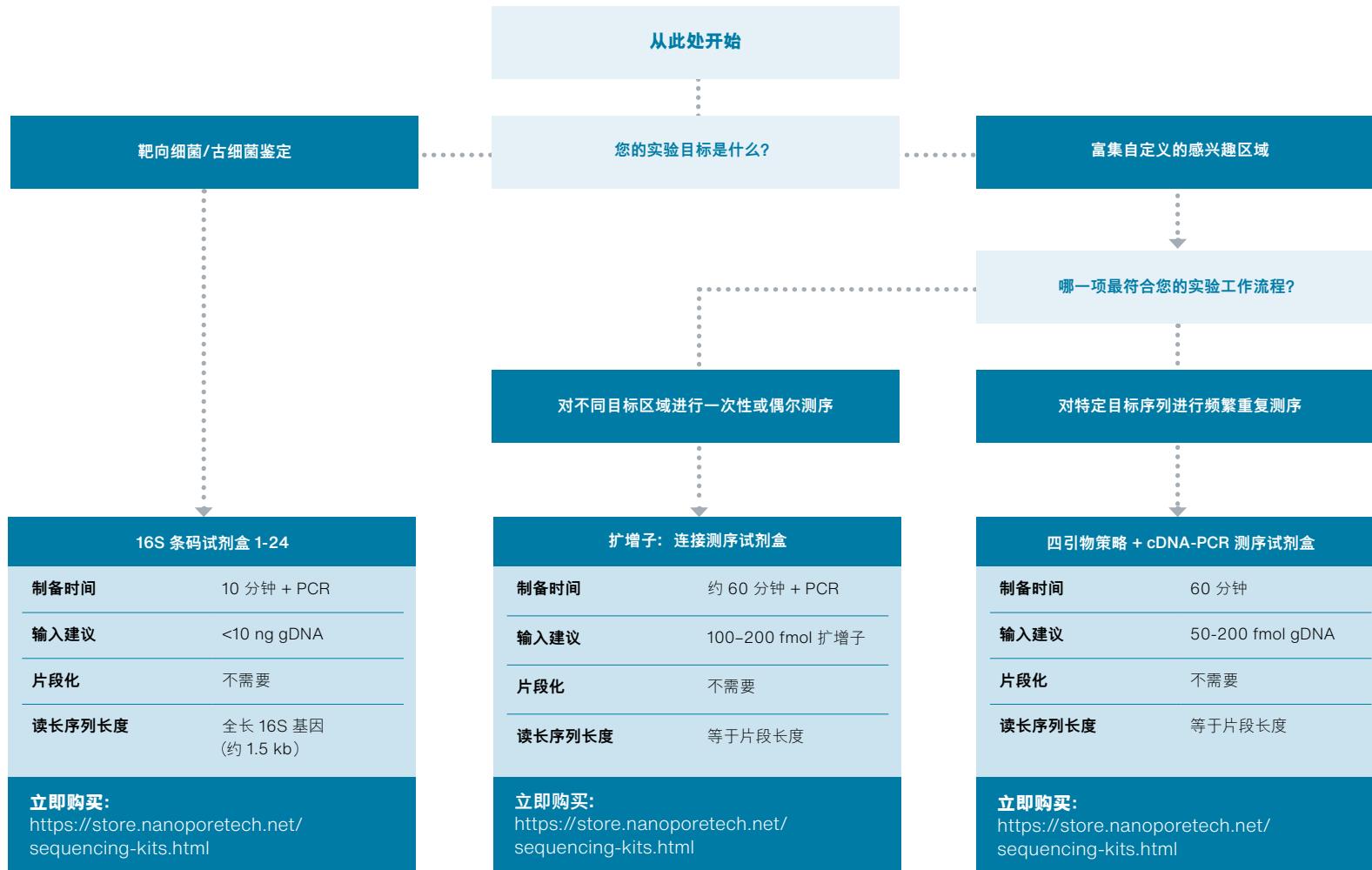
我该选择哪种方法?

无需 PCR 选项



我该选择哪种方法?

基于 PCR 的选项



从样本到获得答案

制备

我是否应当对样本进行扩增？

通过以下链接访问靶向测序页面：
[nanoporetech.com/applications/
targeted-sequencing](http://nanoporetech.com/applications/targeted-sequencing)

Oxford Nanopore 为您提供了一种无 PCR 的靶向富集解决方案：Cas9 靶向测序和适应性采样。通过这些方法，可以对无法通过 PCR 处理的高 GC 含量和 / 或低复杂度区域进行靶向测序，同时也保留了碱基修饰。然而，这些方法通过为目标区域的 DNA 分子获取测序时间来起作用，而不是在测序前增加样本中靶标分子的数量。因此，这些方法的输入要求更高，且可能无法呈现足够多的低频变体来生成用于稳健分析的优良覆盖深度。如果想要从低样本输入或检测低频变体开始，则可能需要借助 PCR 来确保对您的目标区域进行高深度测序。



试验设计

我如何设计用于进行 Cas9 靶向测序的 crRNA 探针？

使用 Cas9 和 crRNA 探针有三种方法来靶向感兴趣区域 (ROI)：

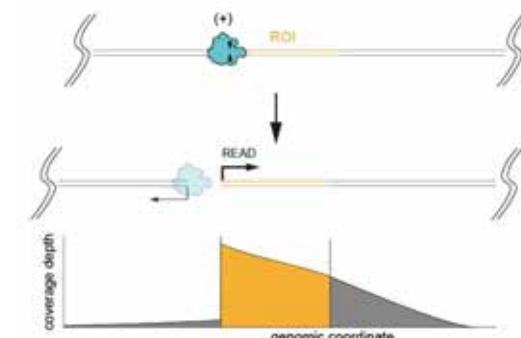
切除：将探针设计用于靶向临近 ROI 的每一个末端的区域，这使得 Cas9 能够对该部分进行切除；随后将测序接头连接至每一个末端。这种方法最适合至多 10 kb 的目标区域。

单次切割和读出：将探针设计用于仅靶向 ROI 的一个末端，Cas9 将在此处进行切割。仅将测序接头连接至该末端，这使得能够从该位置仅在一个方向上开始测序。这种方法能够在仅知道一个末端的序列的情况下进行靶向富集：对于标识未知基因融合伴侣或定位移动原件而言，这是一种理想的方法。

覆盖式处理：探针设计跨越大型 ROI，产生多个重叠的片段。这确保了对于超大目标区域进行良好且均匀的覆盖，并且对于 >10 kb 的区域而言是最佳方法。

在您选择了最符合试验要求的方法后，您可以在线进行探针设计：我们推荐使用工具 CHOPCHOP 来搜索候选探针序列。有关设计和运行 Cas9 靶向测序试验的完整流程指南，请查看我们的信息表，在继续执行方案之前请务必阅读该信息表。

Cas9 靶向富集的单次切割和读出方法



前往以下链接查看 Cas9 信息表：
nanoporetech.com/cas-enrichment

从样本到获得答案

试验设计

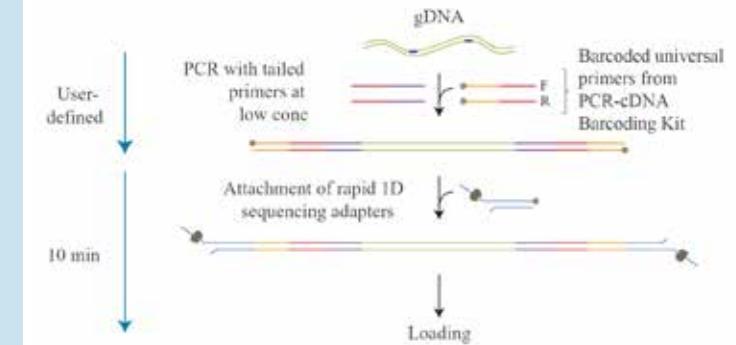
我应该为自定义靶向选择哪种基于 PCR 的方法？

我们拥有一系列可用于基于 PCR 的富集方法，以进行调整来满足您的试验目标。在对单个靶标进行扩增和测序时，您只需使用自定义引物对您的目标区域进行扩增，然后使用连接测序试剂盒对样本进行制备，以供测序。该试剂盒也可用于制备用于测序的预扩增样本。为了在不需要进一步扩增的情况下对混样建库中的许多样本进行测序，您可以在单个测序芯片上进行混合和测序之前，使用无扩增条码试剂盒将独特的条码连接至每一个扩增的样本。

如果需要快速项目周转时间，则可使用四引物 PCR 方法。此处，尾添加引物被用于对来自 gDNA 的目标区域进行扩增，由此得以使用 cDNA-PCR 测序试剂盒，通过我们的快速连接引物进行后续扩增。可在短短十分钟内制备用于测序的扩增样本。如果您希望使用该方法对您的样本进行混样建库，PCR-cDNA 条码试剂盒中有可供使用的条码化快速连接引物。四引物 PCR 通常需要进行一些优化工作，所以对于需要重复多次的工作流程而言是理想之选。

如需了解更多关于我们基于 PCR 的测序试剂盒的信息，请访问以下链接：
<https://store.nanoporetech.net/sample-prep.html>

具有自定义尾添加引物和条码化通用引物的目标区域的四引物 PCR，用于混样建库靶向测序



从样本到获得答案

提取

我如何提取高分子量 (HMW) DNA?

访问以下链接了解更多关于 HMW DNA 提取的信息:

community.nanoporetech.com/docs/prepare/extraction_protocols

保留提取物中的 HMW DNA 将有助于确保目标区域完好无损，这样就能在单个序列中对其进行端到端测序；在使用长序列富集方法（诸如 Cas9 靶向测序）时，这一点尤为重要。如要为您的样本类型查找和比较针对片段长度优化的方案，或是想要获取包括污染和 DNA 储存在内的话题，请访问 Nanopore 社区的文档部分。

The screenshot shows a search results page for "HMW DNA extraction protocols". The top navigation bar includes links for Home, Plan, Prepare, Sequence, and Analyse. Below the search bar, there are filters for All, Sample Stories, Extraction Protocols, DNA/RNA Handling, and Library Prep Protocols. The search results list two items: "Chemistry Technical Document" and "Ligation Sequencing gDNA V14 - human sample (N50 > 30 kb) on PromethION (SQK-LSK114)".

测序

我应选择哪种设备？

纳米孔测序高度可扩展且十分灵活，配备了多个平台来满足各种应用的需求。便携式 MinION 可在笔记本电脑上运行，而 MinION Mk1C 则结合了 LCD 显示器和自带的计算设备，以用于在实验室内外进行一体化测序和分析。这些方法适用于对单个目标区域或较小组合进行测序。在较小规模的应用方面，可使用 Flongle™（一种“测序芯片适配器”），将较小的 Flongle 测序芯片适配到这些设备上——非常适用于仅需要小型数据集的试验，例如对单个已扩增目标序列进行测序。在较大规模的应用方面，GridION 能够运行最多五张独立可寻址 MinION 测序芯片，从而实现灵活的按需测序，或者用于对多个目标序列进行大规模混样建库。

最后，功能强大的 PromethION 可提供 TB 级测序数据。PromethION 2 配备两张独立可寻址的测序芯片，可为小型实验提供高产量测序，而 PromethION 24 和 PromethION 48 分别配备 24 和 48 张独立可寻址的测序芯片，能够对数千个目标序列进行平行测序。

如需详细比较纳米孔测序设备，请访问以下链接：

nanoporetech.com/products/comparison

纳米孔测序设备的最大产出量*



从样本到获得答案

测序

我如何使用适应性采样在测序期间对样本进行富集？

访问适应性采样目录：

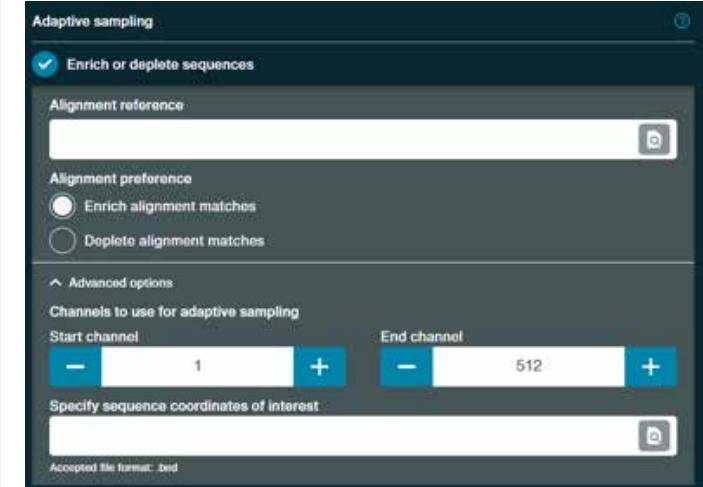
community.nanoporetech.com/adaptive_sampling_catalogue/

适应性采样已整合进 MinKNOW 中。在 MinKNOW 中，您可以选择是否希望对序列进行富集或去除。这些序列可包括所选区域（例如，一个或多个感兴趣基因）或整个基因组（例如，要在宏基因组样本中去除冗余的有机体中基因组）。这些序列以参考基因组文件和 BED 文件（可选）的形式上传，其中指明了基因组间隔的坐标。如果未提供 BED 文件，则将富集或去除整个参考基因组。

在富集试验期间，适应性采样将拒绝所有在参考基因组/BED 文件中不存在的序列；而在去除试验中，仅会拒绝存在于参考基因组/BED 文件中的序列。运行所得的富集倍数将取决于各种因素，包括读长序列长度和您希望测序或拒绝的序列比例。

如需浏览 Nanopore 社区成员提交的适应性采样 BED 文件，或提交您自己的文件，请访问适应性采样目录。

适应性采样选项



分析

我如何对数据进行分析？

Oxford Nanopore 和社区开发了诸多工具，用于分析纳米孔长读长序列。EPI2ME™ 提供的分析工作流程可作为云计算或本地运行解决方案，用于实时分析或运行后分析。它具有直观的用户界面，无需命令行经验即可使用。其工作流程包括分析微生物和抗微生物耐药性基因，以及人类基因变异（包括 SNV、SV 和甲基化）。生物信息学流程适用于许多应用，诸如 Cas9 靶向测序数据评估。

如需更多有关目标区域分析的信息，请访问以下链接：
nanoporetech.com/data-analysis



从样本到获得答案

分析

我如何对目标区域中的甲基化进行评估？

使用 Cas9 靶向测序或适应性采样，可在不进行 PCR 的情况下对目标序列进行富集，同时保留甲基化，而甲基化可在未进行任何特别的文库制备步骤的情况下检测到。

若要分析甲基化，我们建议使用 Remora。Remora 模型已整合进测序设备和单独的碱基识别工具（例如 Guppy、Dorado），其将典型碱基的识别与甲基化识别区分开来，从而在单次运行中实现最高质量的典型碱基识别和甲基化识别，并最大限度地减少计算费用。Remora 模型可用于检测 5mC 和 5-羟甲基胞嘧啶 (5hmC)。您也可使用 Remora 训练碱基识别模型，以检测其他感兴趣的修饰。

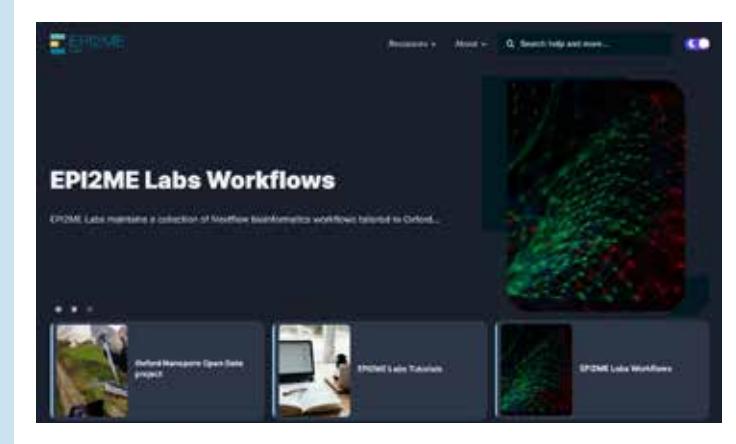
Oxford Nanopore 通过使用简并代表性甲基化测序，并将适应性采样与 Remora 相结合，可对甲基化胞嘧啶进行直接检测。

[点击此处了解有关 RRMS 的更多信息：](#)

labs.epi2me.io/rrms2022.07

[在 GitHub 上查找 Remora:](#)

github.com/nanoporetech/remora



案例研究

案例研究 1：使用适应性采样富集微生物混合样本中的低丰度菌种

对来自宏基因组混合样本的感兴趣基因组进行测序，可能存在一定难度，尤其是当一些物种在样本中的代表性不及其他物种时，可能导致测序对样本中较罕见基因组的覆盖深度不足，并且结果偏向于丰度较高的基因组。适应性采样可以保留或拒绝来自特定基因组和感兴趣区域的读长序列，而动态适应性采样不仅如此，还可以连续评估实验参数，例如覆盖度或基因型不确定性，并且基于来自测序运行的实时信息，对链进行保留或拒绝。

BOSS-RUNS（纳米孔测序前读长序列的获益-优化短期策略）是一项动态适应性采样工具，由欧洲分子生物学实验室（EMBL）和诺丁汉大学开发，可实时评估片段的前瞻性测序价值。为了确定动态适应性采样是否可用于富集混合样本中较为罕见的基因组，Weilguny 等人使用 BOSS-RUNS，对包含八个不同丰度物种的 Zymobiomics 微生物群落进行了测序（Zymo 研究）⁶。对于丰度最高的菌种，仅需几分钟即可完成解析。最初接受来自任何基因组的任何读长序列，之后便可拒绝来自丰度最高菌种的读长序列（这些序列已

得到解析）。这样，就可以在运行期间对较罕见的物种进行富集，使覆盖深度从丰度最高的物种重新分布到丰度最低的物种。例如，与对照测序运行相比，使用 BOSS-RUNS 时的大肠杆菌（仅占起始样本 DNA 的 0.1%）测序产量增加至 3.9 倍。

作者展望未来，总结道：“该方法减少了获得答案的时间，并增加了获取的信息量，这在临床环境或病原体监测中可能具有重要意义”。

阅读发表文章（2023 年 1 月）：

nanoporetech.com/dynamic-adaptive-sampling

6. Weilguny, L. et al. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nat. Biotech.* DOI: doi.org/10.1038/s41587-022-01580-z (2023)

案例研究

案例研究 2：利用 Cas9 靶向纳米孔测序技术，表征 2 型肌强直性营养不良的重复扩张序列

2 型肌强直性营养不良 (DM2) 是一种重复扩张序列引起的疾病，其特征是 *CNBP* 基因的内含子 1 中存在 CCTG 重复序列。该疾病可导致进行性肌肉无力、心脏缺陷和内分泌功能障碍。非致病性等位基因可能包含最多 26 个 CCTG 重复序列，但 DM2 患者中可能存在 75-11,000 个重复序列，这是迄今为止发现的最大型重复扩张序列之一。目前用于 DM2 基因检测的方法包括 PCR 和 Southern 印迹杂交分析，其成本高昂、耗时较久、缺乏灵敏度，并且不能将重复扩张序列区块解析至单核苷酸水平。

无 PCR 纳米孔测序通常可生成数百 kb 长度的读长序列，因此可以对 DM2 致病性扩张序列（单个读长序列的长度可能在 20 kb 至 50 kb 之间）进行端到端测序。Alfano 等人使用 Cas9 靶向测序，表征了来自 9 份 DM2 临床研究样本的全长扩张序列⁷。通过该方法，该团队观察到 >500x 富集的 *CNBP* 等位基因，从而可以表征长度为 344 bp 至 46.6 kb 的重复扩张

序列，其中 46.6 kb 重复扩张序列是迄今为止在单核苷酸解析度下已分析的最长重复扩张序列。

作者指出，Cas9 富集后进行纳米孔测序有助于“通过单次分析准确报告体细胞嵌合体的重复序列长度、结构/基序和程度，从而深入表征 *CNBP* 扩张序列”。

阅读发表文章（2022 年 5 月）：
nanoporetech.com/characterisation-myotonic-dystrophy

7. Alfano M. et al. Characterization of full-length *CNBP* expanded alleles in myotonic dystrophy type 2 patients by Cas9-mediated enrichment and nanopore sequencing. *eLife*. DOI: doi.org/10.7554/eLife.80229 (2022)

Oxford Nanopore Technologies

电话: +44 (0)845 034 7900

电邮: support@nanoporetech.com

微信公众号: nanoporetech

www.nanoporetech.net

