

NGS 文库制备工作流程中质量控制的重要性：安捷伦自动化电泳产品系列应用综述

作者

Whitney Pike, Chava Pocerlich,
Steve Siembieda
安捷伦科技公司

摘要

二代测序 (NGS) 是分子生物学实验室的必要工具，可在多种基础、转化和临床研究环境中用于分析核酸样品。样品文库制备是 NGS 工作流程中的关键步骤，不仅耗时费力且成本较高。对制备过程中间步骤的起始样品，以及对测序之前的最终文库进行质量控制 (QC)，可识别质量较差或浓度不足的样品，确保获得成功的序列数据，有助于节省时间和资源。样品质量信息还可帮助排除故障或优化文库制备方案。安捷伦自动化电泳产品系列提供了多种用于核酸 QC 分析的仪器，包括生物分析仪系统、片段分析仪系统、TapeStation 系统和 Femto Pulse 系统。自动化电泳仪非常适合在整个 NGS 文库制备过程中测定样品的分子量大小、浓度和摩尔浓度，已在科学文献中广为引用。

前言

如今，NGS 平台通过使用多种基于不同化学方法的测序技术，可实现对复杂样品的测序。这些技术按测序读长可分为短读长和长读长。无论测序类型如何，所有 NGS 平台要确保成功的测序结果，都需要创建高质量文库。通常，NGS 文库从特定大小的 DNA 或 RNA 样品开始，然后使用针对特定文库制备方案和测序平台的酶进行处理。一般而言，制备 NGS 文库所需的步骤包括起始样品的片段化、平台特异性测序接头与样品的连接以及扩增文库以产生足够的测序材料。由于最终测序结果的质量高度依赖于文库的质量，因此在整个文库制备工作流程的各个节点进行 QC 检查有很大帮助。建议在开始文库制备方案之前以及片段化、接头连接、序列捕获后或得到最终文库之后进行 QC 检查（图 1）。QC 检查可提供有关样品大小、浓度和完整性的有价值的信息。高质量样品对于成功测序必不可少。通过 QC 检查可以确认样品的质量是否足以进行下游分析，并

识别质量较差，可能无法产生可靠结果的样品。因此，QC 检查点是文库制备工作流程中的关键步骤，可为研究人员节省宝贵的时间和资源。

安捷伦的自动化电泳产品系列提供了多种样品质量控制的专用仪器，并且配备专门用于整个 NGS 文库制备工作流程中各种样品类型的试剂盒。这些仪器包括生物分析仪系统、片段分析仪系统、TapeStation 系统和 Femto Pulse 系统。每种仪器都有相应的优势，可满足各个实验室的不同需求，例如不同的通量、灵敏度、速度和分辨率。自动化电泳仪器可提供出色的核酸定量、定性和分子量测定，确保在整个 NGS 工作流程中对样品进行可靠的评估。通过电泳图和数字化胶图可实现对每个样品的完整性、污染和降解情况的可视化。本综述讨论了在整个 NGS 文库制备工作流程中可以轻松执行 QC 步骤的环节，并提供了在每个特定 QC 检查点使用自动化电泳仪器的文献示例。

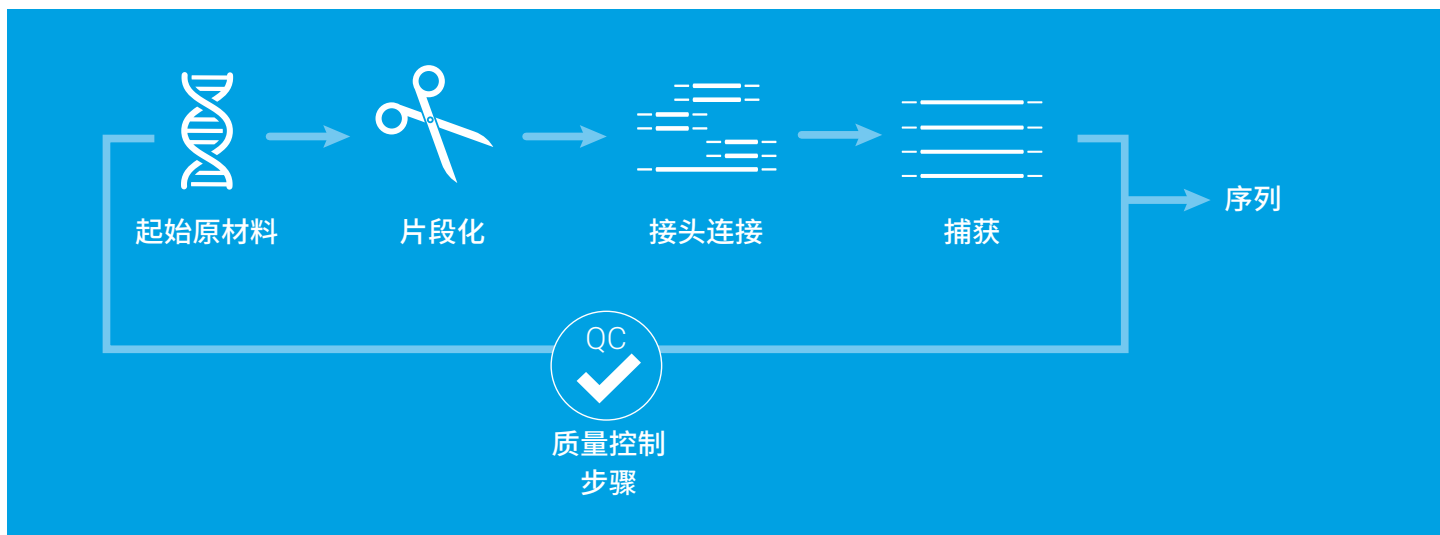


图 1. 在 NGS 文库制备过程中建议进行质量控制的步骤

NGS 文库制备中起始材料的质量控制

文库制备过程中最重要的 QC 步骤之一发生在 NGS 工作流程开始之前。这一关键 QC 步骤就是分析要使用的起始材料，无论是 DNA 还是 RNA。确保高质量的起始材料是制备高质量文库并获得成功测序结果的第一步。大多数 NGS 文库制备试剂盒建议使用特定的起始浓度和起始量，以获得最佳测序结果。如果起始材料降解或质量较差，可能会难以甚至无法利用样品创建文库。一些研究考察了样品质量和下游分析之间的相关性。

起始 DNA QC

起始基因组 DNA (gDNA) 材料的储存条件可能会影响其完整性。为了了解储存温度对 gDNA 的影响，Permenter 等^[1]比较了室温和 4 °C 下，在不同类型的试管中储存 1-130 天的 gDNA 样品的质量。为了分析降解情况，他们通过使用 TapeStation 系统获得的电泳图实现了对样品完整性的可视化，并在 150-10000 bp 范围内根据曲线下面积计算了降解程度。完整的样品显示大部分为高分子量物质（超过 10000 bp），表明为完整的 gDNA，而片段化或降解的样品则会在整个指定的降解范围内出现弥散条带（示例参见文献）。在此示例中，质量最高的样品是在 EDTA 试管中冷冻保存的样品。DNA 指纹图谱和 SNP 分析等下游分析均表明 gDNA 的质量关系着数据的质量。这为正确处理 gDNA 提供了有价值的信息，并表明了高质量起始材料对于 NGS 等高灵敏度应用的重要性。

在开发新方法时，了解各种程度的条件变化如何影响样品十分重要。例如，在 Zhong 等^[2]的研究中，用不同的方法和试剂对免疫沉淀 DNA 进行纯化，然后考察了所得 DNA 用于 ChIP 测序文库和测序时的质量。使用片段分析仪系统测定了不同纯化方法得到的 DNA 样品的分子量（图 2）。结果表明，使用 SPRI 磁珠类试剂进行纯化可消除 100 bp 以下的片段，而纯化柱提取和苯酚氯仿 (PC) 方法则会回收 35 bp 以下的片段。由于 ChIP 测序工作流程传统上会进行片段化并选择 100-300 bp 的 DNA 进行测序，因此基于纯化柱和 SPRI 磁珠的纯化方法不会影响其测序覆盖率。但是，作者指出，对于需要 100 bp 以下片段的应用，在计划实验时应考虑样品纯化方法和所得样品分子量。

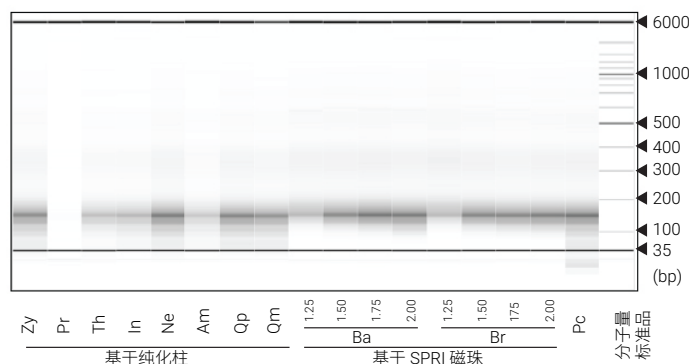


图 2. 使用不同的试剂和试剂盒纯化去交联染色质中的 DNA，然后使用安捷伦片段分析仪系统进行分析。此图转载自 Zhong 等^[2] 并经过修改。

缩写: Zy, ChIP DNA Clean & Concentrator (Zymo Research); Pr, Wizard SV 凝胶和 PCR 纯化系统 (Promega); Th, GeneJET PCR 纯化试剂盒 (赛默飞世尔科技公司); In, PureLink PCR 纯化试剂盒 (Invitrogen); Ne, Monarch PCR 和 DNA 纯化试剂盒 (New England Biolabs); Am, 染色质免疫沉淀 DNA 纯化试剂盒 (Active Motif); Qp, QIAquick PCR 纯化试剂盒 (Qiagen); Qm, MinElute PCR 纯化试剂盒 (Qiagen); Ba, Agencourt AMPure XP 试剂盒 (Beckman, 染色质:磁珠为 1:1.25 至 1:2); Br, RNAClean XP 试剂盒 (Beckman, 染色质:磁珠为 1:1.25 至 1:2); PC, 苯酚/氯仿提取。

成功的文库制备和测序结果需要高质量核酸。为了帮助鉴定合格质量的样品，安捷伦开发了多种质量评分，专门针对不同的自动化电泳平台。这些质量指标为用户提供了可靠的样品完整性评估，通过减少人为错误和用户评估之间的差异，节省了时间和成本，同时可轻松识别会导致不良结果的不合格样品。例如，片段分析仪和 Femto Pulse 系统可以利用基因组质量值 (GQN) 评估经过打断的 gDNA。对于这一评分，用户可以根据具体应用设定阈值并应用于样品。然后，根据测得的高于指定分子量阈值样品的总浓度分数计算 GQN 值。GQN 按 0 到 10 的等级对样品进行评分，0 表示样品均未超过分子量阈值。10 表示样品 100% 高于分子量阈值。此外，TapeStation 系统可提供有关 gDNA 质量的 DNA 完整值 (DIN)，这是一个不受用户影响的客观评分。DIN 按 1 到 10 对样品进行评分。高 DIN 代表高度完整的 gDNA，而低 DIN 则代表高度降解的 gDNA 样品。

对福尔马林固定石蜡包埋 (FFPE) 组织的处理尤为困难，因为为了长期保存，它们采用化学方法固定。可以从此类组织中提取核酸，但由于保存技术导致的交联和降解，其质量通常低于新鲜组织。低质量的核酸导致 FFPE 样品难以用于 NGS 等高

灵敏度的下游分析。但是，通过确定样品的质量，并对如何优化文库制备方案做出明智决策，可以实现对 FFPE 样品的成功测序。例如，Muscarella 等^[3] 评估了通过不同提取方法获得的 FFPE gDNA 的 TapeStation DIN 值，以帮助确定用于肿瘤学研究的可靠 NGS 工作流程。作者指出：“DIN 值使电泳图解析更容易，有利于样品间的比较，并为 NGS 技术提供了出色的实验重现性和高质量基因组 DNA 的定量。”^[3] 所有使用手动或自动试剂盒提取的样品均显示 DIN > 4.4，试剂盒之间的质量没有显著差异。苯酚-氯仿提取的样品质量明显偏低，其中一个样品的 DIN 值为 1.9，还有两个样品严重降解，无法计算 DIN 值。每种试剂盒的代表性示例以及每种方法的平均 DIN 值如图 3 所示。每种提取方法得到的样品均用于 NGS 文库制备，并使用 TapeStation 系统对所得文库的质量进行评估。每个文库表现出相似的扩增子模式，浓度分布均匀。尽管所有样品均提供了良好的测序结果，但作者指出，采用自动化方法提取 FFPE gDNA 可以节省时间和昂贵的试剂，且质量优于手动方法，并获得了更高质量的测序数据。

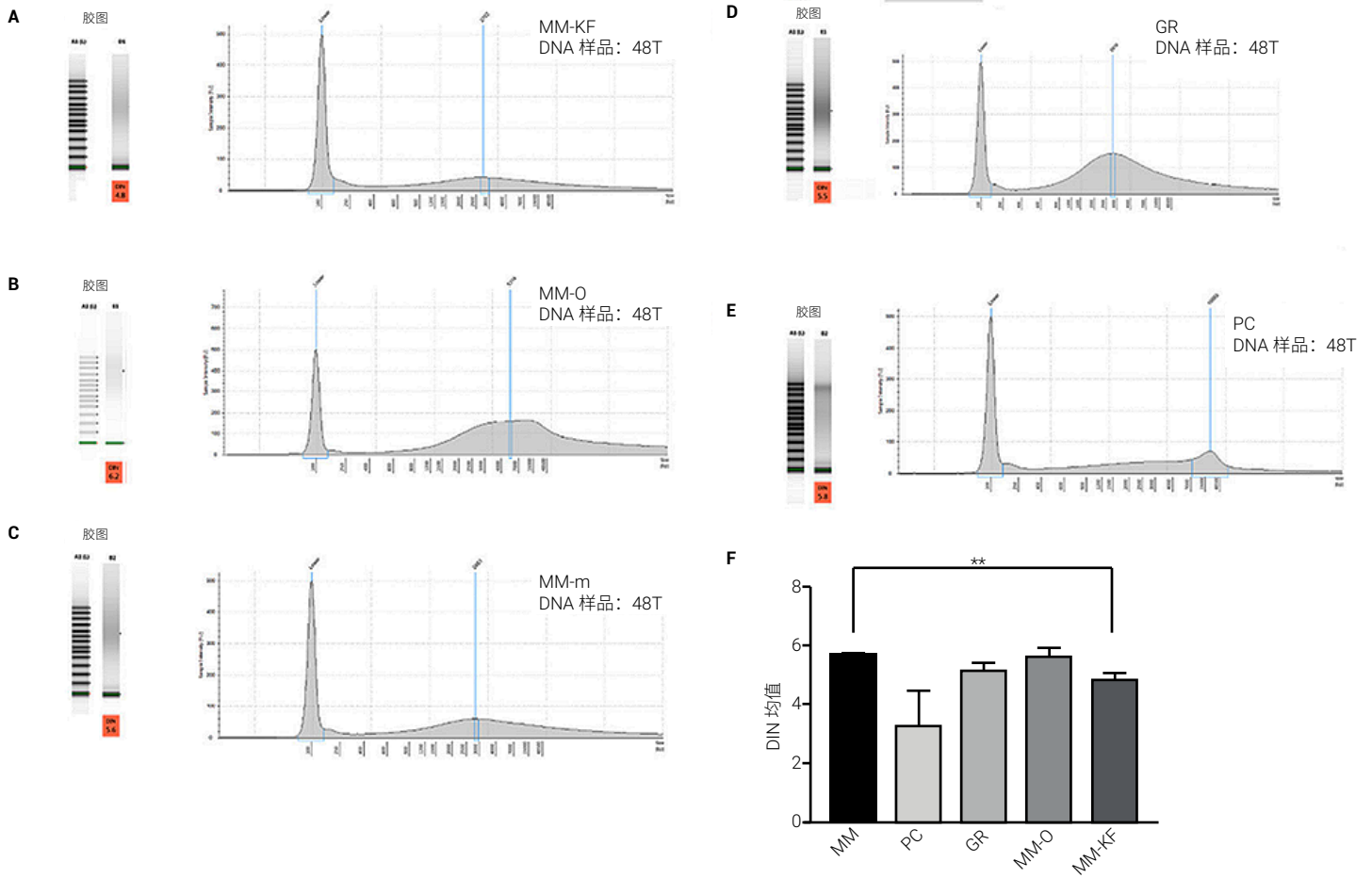


图 3. 使用 5 种提取方法从同一样品中提取 FFPE DNA，并在 Agilent TapeStation 系统上进行分析。(A-E) 从同一样品 48T 中提取的 DNA 的代表性谱图，其中：(A) MM-KF，MagMAX FFPE 总核酸分离试剂盒（赛默飞世尔），使用 King Fisher Duo（赛默飞世尔）的自动化流程；(B) MM-O，MagMAX FFPE 总核酸分离试剂盒（赛默飞世尔），使用 OMNIA Prima (MASMEC S.p.A) 的自动化流程；(C) MM-m，MagMAX FFPE 总核酸分离试剂盒（赛默飞世尔），手动流程；(D) GR，GeneRead DNA FFPE 试剂盒 (Qiagen)；(E) PC，苯酚-氯仿方法。(F) 从六个不同样品中提取的 FFPE DNA 的平均 DIN 值。 $** p < 0.001$, t 检验。此图转载自 Muscarella 等^[9] 并经过修改

近年来，长读长测序技术在基因组从头组装等应用中越来越常见。因此，高分子量 (HMW) gDNA 的提取和处理对于避免样品的意外片段化或降解至关重要。Femto Pulse 系统尤其适合分析高分子量 DNA 的质量，并能在短短几个小时内分离高达 165 kb 的样品。Klingström 等^[4] 讨论了高分子量 gDNA 的正确处理方法以及使用 Femto Pulse 系统分析大分子量样品以确保其完整性。在该示例中，高质量 gDNA 在大于 120 kb 处表现出一个尖锐的单峰。峰两侧出现小的弥散，表明一部分样品略微降解（左侧）或受保护未被片段化（右侧）（图 4）。

根据起始样品量，一些长读长测序方法能够对长片段进行测序，而无需经过可能引入偏差的片段化和扩增步骤。Kingan 等^[5] 使用 Femto Pulse 系统对起始 gDNA 材料和所得的 NGS 文库进行 QC 检查。对于长读文库制备方案，重点在于起始材料应大于 20 kb（通常约为 40 kb），并且不应含 5 kb 以下片段，以避免剪切和分子量选择步骤。由于 Femto Pulse 系统的分子量测定范围已扩展到 165 kb（图 5），因此可鉴定出满足上述要求的样品，从而省去了这些步骤，并以低起始浓度进行文库制备。

起始 RNA QC

RNA 测序需要高质量 RNA 作为起始材料。但是，RNA 属于易降解的生物分子。储存温度、冻融操作、提取方法以及广泛存在的 RNA 酶都会影响 RNA 的完整性，这使得起始 RNA 的 QC 成为了高灵敏度下游应用之前的关键步骤。RNA 完整性可通过生物分析仪系统中的 RNA 完整值 (RIN) 以及为其他自动化电泳仪器开发的等效质量指标来评估：TapeStation 系统的 RNA 完整值当量 (RIN^e)^[6]，或片段分析仪和 Femto Pulse 系统的 RNA 质量值 (RQN)^[7]。每个 RNA 质量指标均考虑了 RNA 样品的整个电泳分离过程，包括核糖体片段的比例以及是否存在降解产物。

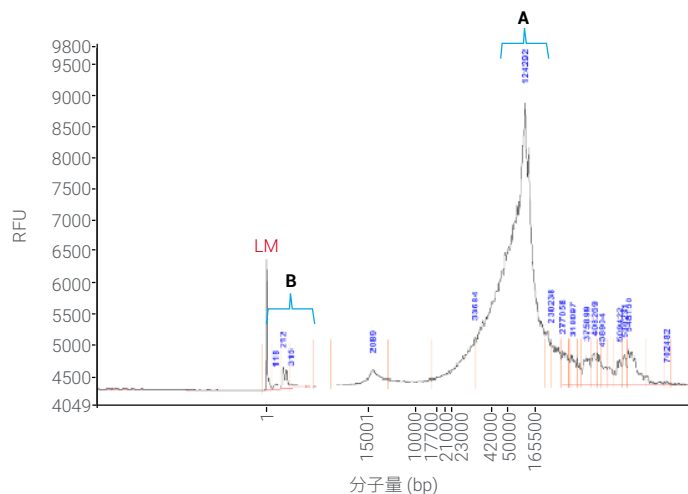


图 4. 在 Agilent Femto Pulse 系统上分析高分子量 DNA，确保样品的质量足以用于长读长测序。(A) 主要片段显示分子量大于 120 kb，表明 DNA 质量较高。(B) 主弥散峰左侧出现小片段表明样品仅发生轻微降解。此图转载自 Klingström 等^[4] 并经过修改

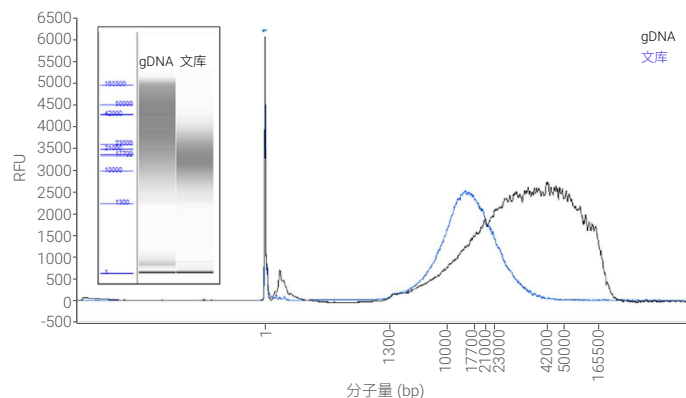


图 5. 从单个 *coluzzii* 按蚊中提取 gDNA，并使用经改良的 SMRTbell 文库制备方案进行 PacBio 单分子实时 (SMRT) 测序。测序前在 Agilent Femto Pulse 系统上评估了 (A) gDNA 和 (B) 最终文库的分子量和浓度。此图转载自 Klingström 等^[5] 并经过修改

许多测序实验室都认为高质量 RNA 的 RIN 评分不应低于 7.0，或者自行研究以建立样品质量标准，并用于工作流程。例如，Haller 等^[6] 使用多种方法评估了几种 RNA 样品。将样品分类为“通过”或“失败”，以表明样品是否能够用于下游微阵列芯片和 qRT-PCR 实验。在这些分析方法中，采用 RIN 评分作为“在不受用户影响的可靠自动化流程中，实现 RNA 质量控制标准化的工具。RIN 考虑了两个核糖体条带信号强度的降低，以及在两个峰之间和 18S 条带以下短片段数量的增加。”^[6] 每个“通过”等级的样品 RIN 均不低于 6.6，平均值为 7.5（参见文献）。随着 NGS 技术的发展，较高的 RIN 通过值也被用以确保成功的文库制备和测序结果。

尽管 RIN 是 RNA QC 最为成熟的标准，但已有部分研究证明了 RIN 与 RIN^e^[6] 和 RQN^[7] 的等效性。通过提供有关样品完整性的信息，RNA 质量指标使用户能够就如何处理不理想的样品做出明智的决策。这些决策包括如何针对略微降解的样品优化方案。Alpern 等^[9] 旨在改进一种独特 RNA 测序方法的现有方案：批量 RNA 加标签和测序 (BRB-seq)。在对方法进行改进后，作者使用片段分析仪系统评估了起始 RNA 完整性和得到的 NGS 文库。将 RNA 样品进行不同程度的片段化，并使用 BRB-seq 进行评估，以估计差异基因表达。如叠加图所示（图 6），RQN 随着 RNA 片段化程度的增加而下降。对于另一项质量评估辅助的方案更改，作者将市售试剂与实验室内部制备的试剂进行了比较。在一个示例中，他们通过片段分析仪系统观察到使用内部标记酶 (Tn5-A/B、Tn5-B/B) 时文库产量增加，这使他们能够降低文库制备成本（图 7）。这些 QC 检查点为他们提供了优化降解样品的 BRB-seq 方案所需的信息，使他们能够成功生成 NGS 文库，并成功检测出完整样品中 75% 以上的差异表达基因。总体而言，作者能够基于质量评估优化其 NGS 文库制备工作流程，这样即使对于 RQN 值较低的样品也能获得良好的测序结果。

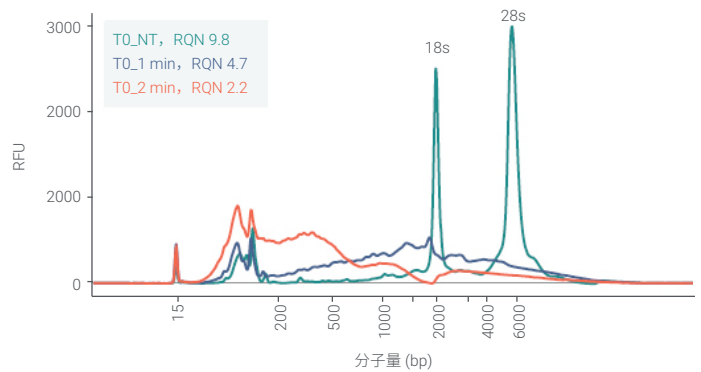


图 6. 为了评估片段化 RNA 的 BRB-seq 性能，有意将样品降解 1-2 分钟。使用安捷伦片段分析仪系统评估 RNA，并比较完整样品 (NT: 绿色迹线) 和降解样品 (1 分钟: 蓝色迹线; 2 分钟: 红色迹线) 的 RQN。T0: 前脂肪细胞。此图转载自 Alpern 等^[9] 并经过修改

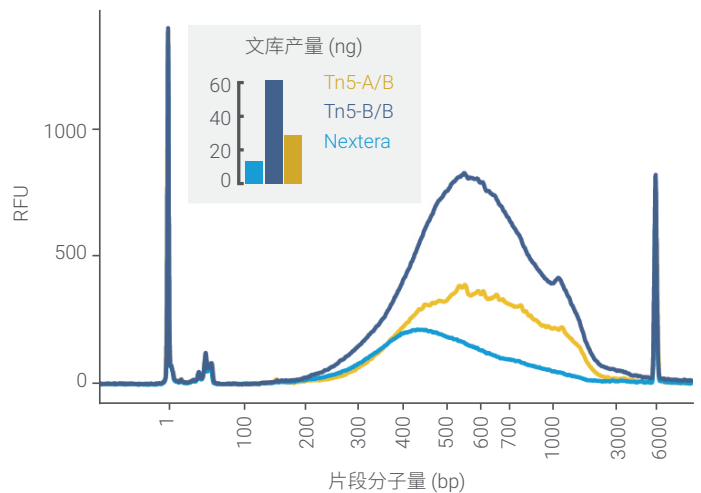


图 7. 使用不同的 Tn5 标记酶 (实验室内部制备的 Tn5-A/B 和 Tn5-B/B, 以及 Illumina Nextera) 制备 NGS 文库, 并使用安捷伦片段分析仪系统分析文库谱图和产量。此图转载自 Alpern 等^[9] 并经过修改

FFPE RNA QC

最近，越来越多的实验室开始对低质量样品或由于提取方法或储存条件而导致轻微降解的样品进行测序。FFPE 可以保护细胞结构和组织形态，是保存样品供后续使用的常见方法。但是，该方法也会带来复杂性（例如对核酸的化学修饰），使样品难以进行 NGS 分析。自动化电泳仪可通过 DV₂₀₀ 指标对 FFPE RNA 进行质量评估，该指标专门用于检测样品中大于 200 nt 的 RNA 百分比，以作为 RNA 降解程度的指标。根据 DV₂₀₀ 评分，研究人员可以决定是否需要改良以及如何改良其 NGS 工作流程。常见的调整包括增加或减少剪切时间以获得理想的分子量，或排除质量不佳而可能无法使用的样品。例如，Wimmer 等^[10] 使用生物分析仪系统和 DV₂₀₀ 评估了来自大鼠（图 8）、人和小鼠的几种新鲜冷冻（FF）和 FFPE RNA 样品的质量，并使用微阵列芯片和 RNA 测序对样品进行了比较。如电泳图所示，大鼠 FF 样品显示清晰的 18S 和 28S 峰，平均 RIN 为 6.1。相反，来自 FFPE RNA 的 rRNA 峰无法区分，因为样品表现为较大的平缓峰，RIN 为 2.3，表明 RNA 已降解。

平缓峰范围较宽，覆盖 100–4000 bp，表明存在一些长 RNA 片段。作为替代方法，使用 DV₂₀₀ 评分衡量样品质量。尽管 FFPE 样品的平均 DV₂₀₀ 评分为 76%，远低于 FF 样品的 97%，但此评分表明该材料仍然可用于 Illumina 测序 (DV₂₀₀ > 70%)。作者指出：“基于 RIN（或等效数值）的 RNA 完整性分析虽然常用于新鲜样品和 FF 样品，但无法可靠地应用于 FFPE 材料，因为福尔马林固定会导致所有 RNA 发生片段化。相反，DV₂₀₀（长度超过 200 个核苷酸的 RNA 片段百分比）可能成为可行的替代方法。研究报告显示，DV₂₀₀ 大于 70% 表明为高质量 RNA，介于 50%–70% 之间为中等质量 RNA 样品，需要使用更高的起始量进行转录组分析。DV₂₀₀ 值低于 30% 表明 RNA 样品已严重降解，无法用于进一步实验。在我们的研究中，尽管组织起始量非常低，但 DV₂₀₀ 值仍然能够明显高于高质量 RNA 的建议阈值。”^[10] 生物分析仪系统提供的 DV₂₀₀ 评分使作者能够从原本可能被丢弃的样品中识别出可用于测序的样品。

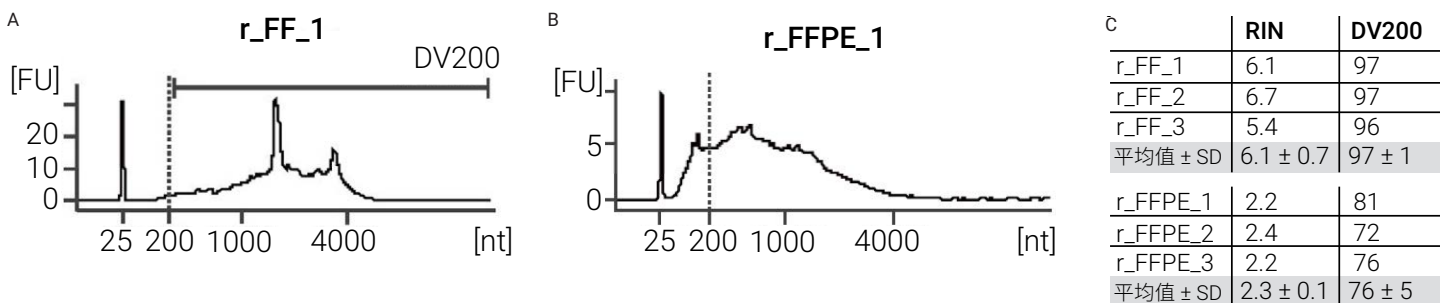


图 8. 在安捷伦生物分析仪系统上分析了由大鼠 (r) 显微切割样品制备的 (A) 新鲜冷冻 (FF) 和 (B) FFPE RNA。 (C) 报告了多个样品的 RIN 和 DV₂₀₀ 数值，以评估样品质量。此图转载自 Wimmer 等^[10] 并经过修改

游离 DNA QC

最近，游离 DNA (cfDNA) 越来越普遍地用作 NGS 的起始材料，因为在临床研究中能够通过微创方法采集所需核酸材料。每种自动化电泳仪器均可用于检查 cfDNA 谱图，包括分子量测定和定量分析。cfDNA 的电泳分离通常显示带有三个峰的谱图，表明存在核小体片段。Mendioroz 等^[11] 通过提取肌萎缩性侧索硬化症 (ALS) 患者的 cfDNA，研究了神经退行性疾病潜在的表观遗传生物标志物。在测序之前，使用片段分析仪系统 (图 9) 评估了 cfDNA 的浓度、质量和分子量分布。所有样品均表现出典型的片段化 cfDNA 模式，并成功用于下游焦磷酸测序和亚硫酸氢盐克隆测序，通过与对照组比较，鉴定出 ALS 患者的基因出现差异甲基化的区域。

虽然每种自动化电泳仪器均可用于 cfDNA 分析，但只有 TapeStation 是唯一具有专用 cfDNA 分析和经验证的质量指标的系统。%cfDNA 评分可应用于使用 cfDNA ScreenTape 分析法进行分析的 cfDNA 样品，从而定性检测样品中 cfDNA 的相对含量 (相对于任何 gDNA 污染，其可能会阻碍下游的测序分析)。

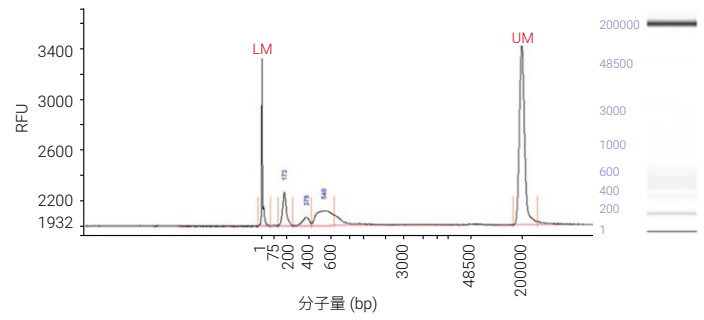


图 9. 在安捷伦片段分析仪系统上分析 cfDNA，在约 165、350 和 565 bp 处分别出现了峰，表明核小体片段化。此图转载自 Mendioroz 等^[11] 并经过修改

整个 NGS 文库制备过程的质量控制

确定了用于 NGS 文库制备的合适样品后，需采取多个步骤将其转变为用于测序的合适文库。为了获得良好可靠的测序数据，强烈建议在工作流程中多个环节设置 QC 检查点，以确保创建高质量文库。检查点中包括片段化和接头连接后的样品 QC。

片段化质量评估

片段化可将大分子量样品打断为小片段，在文库制备中起着关键作用。这些片段随后转换为可由特定测序技术读出的文库。文库制备所需的分子量分布因每种文库制备试剂盒和测序类型而异。因此，在进行文库制备之前，需要进行 QC 分析以确保样品长度正确。任何一种自动化电泳仪器均非常适合此用途。例如，对于 CHIP 测序文库，起始样品首先需片段化为 200–500 bp。但是，片段化方法并非总是最优化的，导致样品中存在过大或过小的片段。Blecher-Gonen 等^[12] 讨论了这些较大的片段如何干扰测序，并建议在文库混合后执行分子量选择步骤以消除这些片段。在分子量选择步骤前后使用 TapeStation 系统分析样品，确定了通过分子量选择可以排除较大的片段^[12]。测序结果表明，与未经过分子量选择的样品相比，进行分子量选择可改善簇生成，从而增加测序获得的峰数量。

可以通过多种方法进行样品的片段化，包括剪切和酶解。超声剪切采用声学探针，将高频短波长能量脉冲集中在样品上，以随机打断 DNA。但是，这种超声处理过程是多变的，需要仔细校准多个设置，包括声波的强度和频率，以产生一致的分子量分布。Tramontano 等^[13] 采用多种超声设置对样品进行剪切，以优化其剪切方案。每个经片段化的样品均采用片段分析仪分析并进行比较，目的是改进其用于扩增子测序的整体 NGS 文库制备工作流程。“通过 PCR 产物的片段化可以提高扩增子测序的灵活性和通量。可以对 PCR 产物中更多的碱基进行分析，并且打断为小片段后可以使用多种不同的短读长测序平台。”^[13] 因此，他们在 PCR 扩增子上测试了多种不同的超声处理参数，以确定产生最佳片段分布的方法（图 10）。随后将优化后的参数应用于供测序的混合扩增子（图 11）。比较片段化和未片段化的 PCR 扩增子的测序数据，片段化后覆盖率和数据一致性均有所增加。

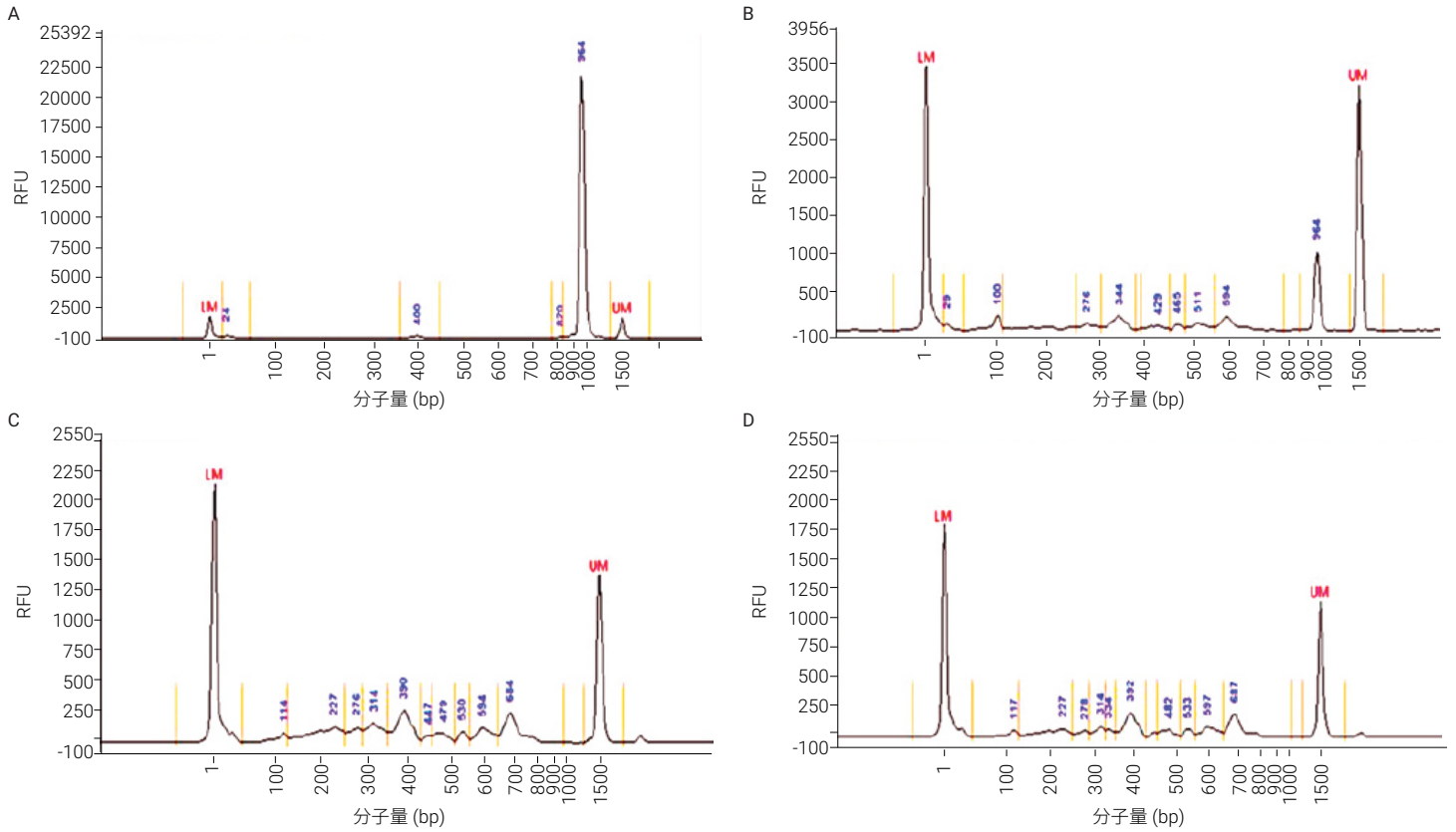


图 10. 使用各种超声处理参数将 PCR 扩增子片段化为 350–500 bp, 以获得合适测序的样品。使用安捷伦片段分析系统评估片段化程度。(A) 未片段化 PCR 扩增子, 长度为 964 bp。(B) 超声处理测试 1 获得的片段化模式。964 bp 处残留的大峰表明片段化不完全。(C 和 D) 超声处理测试 13 的技术重复样品, 显示出较宽的片段化模式, 在约 390 bp 处出现峰, 非常适合下游测序。此图转载自 Tramontano 等^[13] 并经过修改

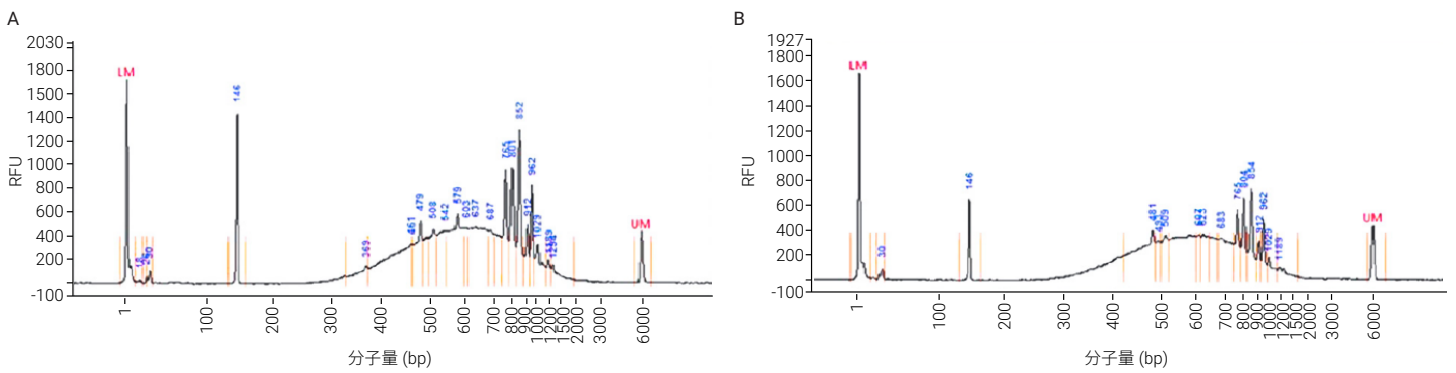


图 11. 在安捷伦片段分析系统上分析由混合 PCR 扩增子制备的两个 NGS 文库 (A、B) 的代表性示例。此图转载自 Tramontano 等^[13] 并经过修改

另外可使用酶解片段化，利用酶在 DNA 中产生双链断裂。尽管该方法易于放大为高通量实验，并且相对简便易用，但可能因 DNA 上的酶识别位点而引起测序偏倚问题。酶解方案还必须针对样品类型、缓冲液和浓度进行优化。例如，Aguirre 等^[14]将 ddRADseq（双酶切限制性酶切位点相关 DNA 测序）应用于非模式生物的基因分型。因为标准方案是针对模式生物开发的，所以需要在数个步骤上对方案进行优化以获得最可靠的数据。方案优化包括选择一对可为测序平台提供最佳分子量范围的限制性内切酶。片段分析仪系统用于酶解片段化前后的样品 QC，以及测序前的最终文库 QC。图 12 表明，用两组不同的酶进行片段化后分子量分布差异较大。片段分析仪显示的均质模式以及对 350–600 bp 弥散峰的分析有助于作者确定用于其实验方案的酶组。在此示例中，与 PstI-MspI 对（图 12B）相比，SphI-MboI 对（图 12A）产生的样品在测序所需的分子量范围内具有更高百分比。

随着专用测序类型的出现，一个常见的问题是选用哪种片段化方法。Lan 等^[15]利用 NGS 进行高通量 HLA 分型。为了优化方案，他们将传统的 TruSeq Nano（采用 Covaris 超声剪切）与基于转座酶（将酶解片段化和接头连接合并为单个反应的步骤）的 Nextera 文库制备方法进行了比较，并考察了测序覆盖范围的偏倚和基因分型检测的准确性。覆盖率或读出序列深度是基因分型中非常重要的指标，因为其表示了核苷酸在比对读出序列中出现的次数。因此，覆盖率与变异检测的特异性直接相关，从而影响基因分型的检测。通过 TapeStation 系统对使用两种技术准备的文库进行了验证和定量分析^[15]。Nextera 文库的分布比 TruSeq 文库更分散，导致电泳图上的分子量略宽。但是，在测序和剪切后，所有文库制备的比对读出序列的插入片段大小中值均很相近。总体而言，转座酶方法存在更大的覆盖偏倚，并且考察不同等位基因的覆盖率发现，

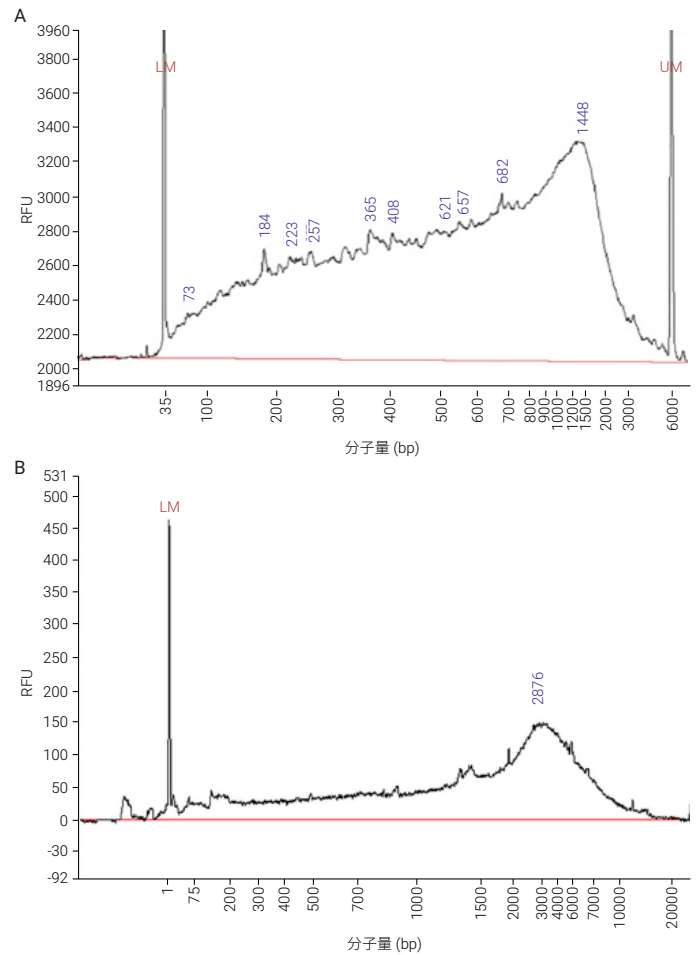


图 12. 使用安捷伦片段分析仪系统分析邓恩按 gDNA 在不同酶对下的体外酶解产物：(A) SphI-MboI 和 (B) PstI-MspI。SphI-MboI 酶解产生了大量的小分子量片段，处于分子量选择和测序所需的分子量范围内。此图转载自 Aguirre 等^[14]并经过修改

所有测试的文库制备方案之间存在较大差异，表明所有方法均存在一定的覆盖偏倚。本文说明了为什么不同的测序类型需要使用不同的片段化方法，并强调了方法优化对于高灵敏度 NGS 应用的重要性。

用于 RNA 测序文库的 cDNA 制备

在 RNA 测序文库的制备过程中，一个重要步骤是从总 RNA 中分离出 mRNA，并通过逆转录将其转化为 cDNA（互补 DNA），以用于后续的文库制备步骤和测序。由于 cDNA 是由 mRNA 转录而来，文库仅包含收集 RNA 的生物体或组织的表达基因的编码区。因此，该过程中常见的 QC 步骤包括在逆转录前检查总 RNA 的质量，以及在进行剩余文库制备步骤之前检查得到的 cDNA 的质量。Misra 等^[16]详细介绍了使用 FACS（荧光激活细胞分选）分离拟南芥精细胞的单细胞 RNA 测序 (scRNA-seq) 方案。简单而言，将单细胞分选的精细胞 RNA 进行逆转录、预扩增和 cDNA 纯化。除其他确认样品质量的测试外，还在进行后续文库制备步骤和测序之前，使用高灵敏度 NGS 试剂盒在片段分析仪系统上确认了 cDNA 的分子量分布。cDNA 的电泳图显示两个底部较宽的峰，平均长度为 1500–2000 bp（图 13A，13B）。在测序之前，在片段分析仪系统上使用相同的试剂盒来确认最终的单细胞文库质量（图 13D）。由于从植物中获得单个细胞并从极少的细胞内容物中分离 cDNA 非常困难，因此这些 QC 步骤有助于选择最佳 cDNA 样品和文库用于测序。

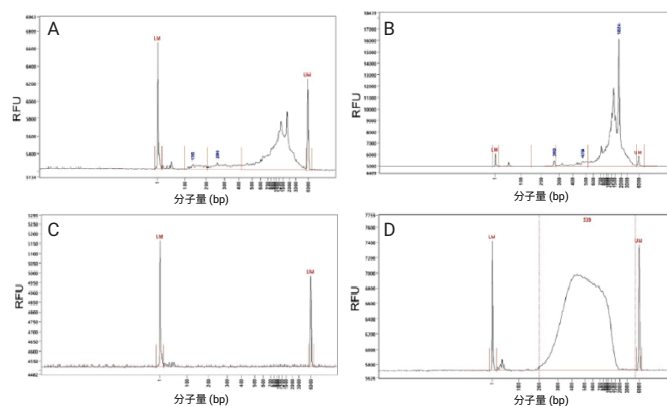


图 13. 使用安捷伦片段分析仪系统对各种单细胞 RNA 测序文库的制备步骤进行 QC 检查。图中所示为代表性步骤的电泳图：(A) 单个细胞的 cDNA 扩增；(B) 大量样品的 cDNA 扩增，作为阳性对照；(C) 不含细胞的阴性对照；(D) 最终 NGS 文库。此图转载自 Misra 等^[16] 并经过修改

接头连接

样品前处理后，NGS 文库制备的下一步通常包括末端修复、腺苷酸化和接头连接。接头为平台特异性序列，可在测序过程中识别各个片段。接头连接后进行 QC 可确保接头与样品成功连接。当使用自动化电泳仪进行分析时，连接成功表现为分子量分布与片段化样品相比有轻微偏移，连接接头后文库的平均分子量更大。如果连接效率低，则会导致样品损失，这对于起始材料有限（例如 cfDNA）的高灵敏度测序实验非常不利。因此，已经对 NGS 文库制备过程中的接头连接步骤进行了多种优化尝试。Nix 等^[17] 在一项研究中考察了 cfDNA 与单链和双链接头的连接效率，以优化文库制备方案，减少样品损失。图 14 展示了 TapeStation 系统如何应用于工作流程中的所有 QC 检查点：接头 (B)、起始材料 (C)、连接后 (D) 和 PCR 后的样品 (E)。如电泳图所示，TapeStation 的灵敏度可观察到连接后样品中的多个峰。经鉴定，这些峰分别为游离接头、未连接的产物、单端连接产物和双端连接产物。通过定量每个峰区域内的摩尔浓度来确定连接后的连接效率。然后计算所有连接产物中双端连接产物的百分比，作为连接效率。将 PCR 后的样品与起始 DNA 进行比较，结果显示分子量略有偏移，接头成功连接使 PCR 后的峰分子量增加。总体而言，作者利用 TapeStation 测得其方案中双链接头的连接效率高于单链接头。

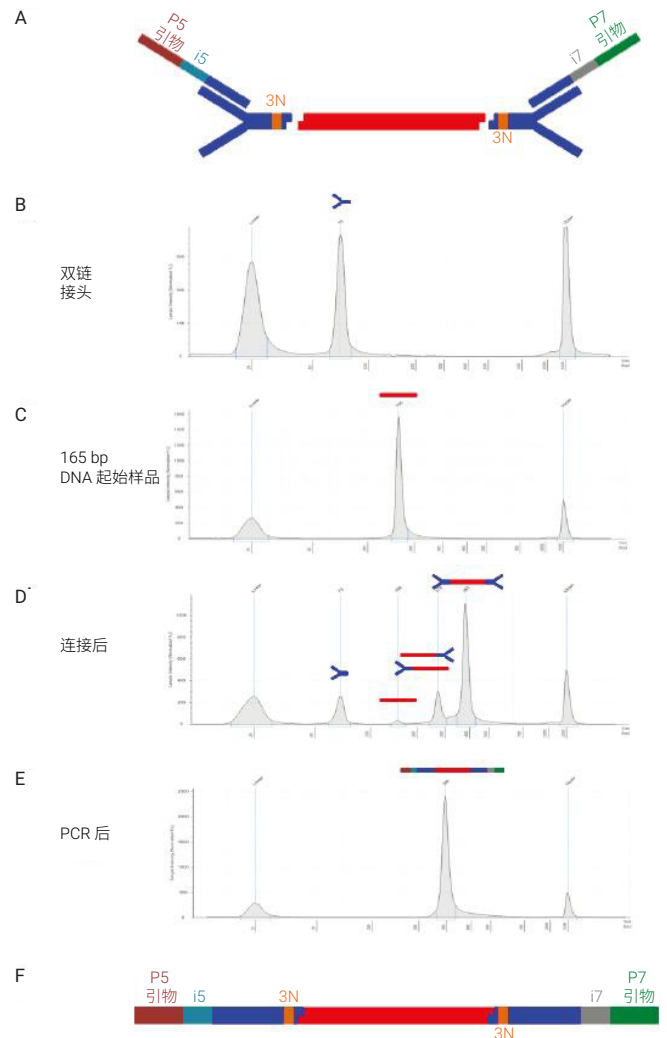


图 14. 使用安捷伦 TapeStation 系统对接头连接进行 QC。(A) 插入片段（红色）和双链接头（蓝色）的示意图；(B) 由单链 DNA 组成的双链接头的电泳图，在 75 bp 处显示有一个峰；(C) 165 bp 为起始 DNA；(D) 使用 B 和 C 作为参照，连接后 QC 中显示四个峰经鉴定为未连接接头、未连接的插入片段，单端连接产物和双端连接产物。PCR 后带有双端接头的 165 bp 插入片段的样品谱图 (E) 和 (F) 示意图。此图转载自 Nix 等^[17] 并经过修改

最终 NGS 文库的质量控制

文库制备的最后一步是 PCR 扩增，可捕获连接接头的片段，并生成足够的文库拷贝用于测序。此步骤还可以用于添加标签，以在单次测序运行中对多个样品进行多重检测。尽管最终文库的分子量和形状取决于所使用的文库制备试剂盒，但使用任何一种自动化电泳平台均可对所有类型的文库进行可视化和分析。除文库弥散峰以外的任何其他峰均表明存在过量的引物或接头二聚体。此阶段的分析还有助于确定文库扩增是否过度或不足，并利于混合文库以进行测序。最终文库的 QC 可确保高质量的制备，并可确定文库的分子量、浓度和摩尔浓度。

文库扩增

在 NGS 文库制备的最后步骤中，通过 PCR 将第一步制备的接头连接片段扩增。在靶向扩增子测序中，进行了两个 PCR 反应。第一个反应是仅扩增目标片段，第二个捕获反应是向扩增子添加标签。在测序过程中，仅包含标签的片段会退火至流通池。因此，整个扩增子测序的成功与否取决于 PCR 反应性能。

每个单独的扩增子必须得到充分扩增，使单独的标签连接到每个扩增子，以便在生物信息学过程中进行区分。在处理复杂样品（例如低浓度或受污染物的样品）时，必须优化 PCR 过程以确保最佳反应，以获得良好的测序结果。Sidstedt 等^[16]利用靶向 MPS（大规模并行测序）技术考察了两种已知的 PCR 抑制剂对法医学 STR（短串联重复序列）分析的影响。法医学样品中可能包含这类抑制剂等阻碍 DNA 聚合的分子，导致下游分析中模板的可用性降低。因此，优化 PCR 方案对于这类样品的成功测序至关重要。通过使用片段分析仪系统分析扩增子的分子量、浓度以及生成的 NGS 文库，作者发现抑制剂会降低起始 PCR 的效率。这导致测序时的读出序列计数降低。但是，通过添加 BSA 可以改善 PCR 化学反应，从而抵消效率的下降。使用自动化电泳仪分析复杂样品有助于改善 NGS 文库制备工作流程，帮助研究人员节省时间、成本和珍贵样品。

二聚体

二聚体是 PCR 反应的不良产物，即引物或接头与自身的拷贝而非模板发生退火。发生这种情况的原因很多，例如模板浓度低、引物过多、反应温度不理想或扩增过度。当使用自动化电泳仪分析包含二聚体的样品时，得到的电泳图会在主文库弥散峰的左侧显示一个峰。该峰的大小取决于引物的长度，但通常接近或小于 100 bp。仪器提供的高灵敏度可轻松检测到很少量的二聚体，并且可轻松计算出样品中所含二聚体百分比。二聚体是一个常见问题，如果其出现在电泳图中，则表明在测序之前应进行额外的净化步骤以消除二聚体。核心实验室的建议指出，除非文库所含接头二聚体低于 0.5%，否则不适合进行测序。这是因为引物二聚体等较短的片段与流通池结合的效率要高于文库的较大片段。即使文库只含有 5% 的二聚体，也可能会消耗 50%–60% 以上的测序读出序列，导致测序运行失败^[19,20]。

Marchal 等^[21] 提出了一种用于新合成 DNA 的测序方案。他们在方案中强调了 PCR 不良产物对测序结果的影响，并指出“引物耗尽导致的过度扩增会产生 PCR 不良产物，应避免此情况的发生。” 作者认为，如果浓度太低，必须重新扩增，但应当避免这种情况，因为这会增加 PCR 偏倚的风险。在生物分析仪系统上分析了几个“优质和不良文库”。通常，如果文库在要求的分子量范围内显示为钟型且分布紧密的弥散峰时，则认为是“优质”文库（图 15A）。而“不良”文库在文库弥散峰之外还会存在其他峰（图 15B），表明存在接头二聚体。如果发生过度扩增，主文库的峰会加倍，并且在文库后会出现多余弥散峰^[22]（图 15C）。由于二聚体会消耗大部分的测序读出序列，因此从文库中尽可能将其去除非常重要。这可以通过净化步骤和方案改进来实现，可以降低所用接头和引物的浓度，或优化用于后续文库制备的扩增循环数。

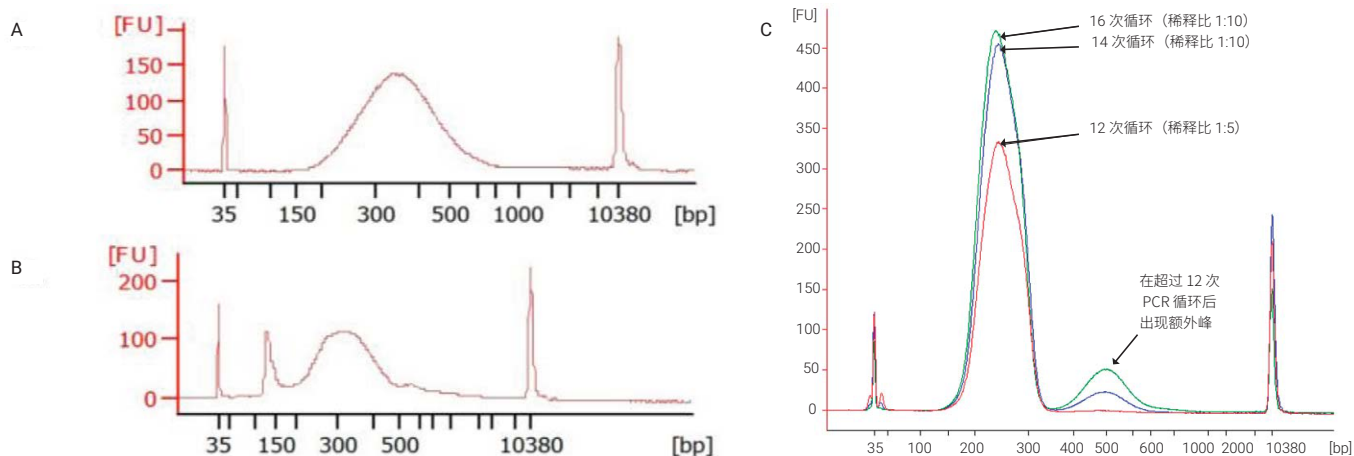


图 15. 使用安捷伦生物分析仪系统对 NGS 文库进行 QC 检查。(A) 优质文库示例；(B) 存在接头二聚体的文库示例，二聚体约 150 bp；(C) 有意过度扩增 NGS 文库，在主文库峰的两倍长度处产生了一个多余的小峰。此图转载自 Marchal 等^[21] 并经过修改

分子量测定、定量和摩尔浓度

摩尔浓度测定是最终文库 QC 的重要组成部分，因为在测序仪上样前，需要将样品稀释至适当的摩尔浓度，并尽可能与其他文库混合。分子量和浓度均用于计算摩尔浓度，以确保准确的定量和分子量测定，并对文库进行质量检查，是成功测序的关键。通过自动化电泳仪可轻松准确地测得样品分子量，样品分布表明了文库质量。样品的浓度也可以使用与每种仪器对应的分析软件进行计算。在一项旨在改善文库制备方案的研究中，Quail 等^[23]表示，与传统的分光光度法相比，使用生物分析仪定量文库时获得了更好的簇密度。作者指出，准确的流通池载样对测序结果至关重要：“进入流通池的 DNA 浓度决定了产生的簇数量……存在一个最佳 DNA 浓度范围，以在最佳的密度范围内产生簇，从而获得尽可能多的数据。”^[23] 高估文库

浓度会导致簇过少，增加测序成本。低估文库浓度则会使簇密度过高，导致簇重叠，减少了从测序中获得的数据量。例如，Quail 等^[23]表明，使用分光光度法进行定量分析会导致簇密度不一致，这可能是由于存在接头二聚体和无法与文库区分的未扩展二聚体。由于自动化电泳仪可根据分子量分离样品，因此使用生物分析仪 DNA 1000 试剂盒进行分析可区分预期的扩增子和二聚体，从而更准确地定量文库。借助分子量分析的额外优势，作者能够在单个 QC 检查点计算文库的摩尔浓度，并在流通池载样时产生更一致的簇密度。每种自动化电泳仪及相应的分析试剂盒均可提供准确而可靠的分子量测定、定量和摩尔浓度，使它们成为 NGS 文库 QC 分析的理想选择。

整个 NGS 文库制备过程的自动化电泳分析

安捷伦提供的自动化电泳仪包括生物分析仪、片段分析仪、TapeStation 和 Femto Pulse 系统，是整个 NGS 文库制备工作流程中质量控制的理想选择。它们可通过分析起始样品、在整个过程中的多个检查点进行分析以及分析最终文库，从而确保良好的样品质量，实现成功测序。

自动化电泳仪在整个 NGS 工作流程的应用具有充分的文献依据。例如，Marosy 等^[24] 在其用于从 FFPE 样品生成外显子组富集文库的整个实验方案过程中，在三个检查点使用了生物分析仪系统。第一个检查点在片段化后，以便在进行后续的文库制备步骤前确保分子量分布和样品浓度。第二个检查点在捕获前扩增反应，以确认浓度和由接头连接导致的分子量偏移。最后，作者在捕获后扩增步骤之后使用生物分析仪测定了最终文库的摩尔浓度，以便适当稀释样品用于流通池载样。这些关键步骤的 QC 检查使作者能够决定如何进行后续的文库制备工作流程，避免损失珍贵样品、试剂和时间。

每种自动化电泳仪均为各种类型样品提供了不同的试剂盒，以确保准确可靠的分子量测定、定量和摩尔浓度，并利用质量指标对样品完整性进行客观评估。这些仪器的优势在于可根据样品的类型和分子量，使用不同的试剂盒在应用之间轻松切换。例如，如前所述，Muscarella 等^[3] 在进行文库制备之前，使用 TapeStation 与基因组 DNA ScreenTape 来确定样品的 DIN 值。然后在测序前使用高灵敏度 D5000 ScreenTape 分析对最终文库进行 QC。通过在不同步骤执行质量检查，可以确保只有最佳样品和文库用于测序。

自动化电泳仪还可以用作帮助研究人员优化工作流程的工具，或为无法使用通用试剂盒处理的样品制定定制方案。Aguirre 等^[14] 对数种方案进行多方面整合，并进行了优化，从而实现了一种独特物种的成功测序。为了优化方案，他们利用片段分析仪系统对起始样品进行 QC 分析，并在整个文库制备工作流程中对样品质量、分子量和浓度进行了 QC (图 12)。

总结

整个文库制备工作流程中的质量控制步骤对于 NGS 文库制备至关重要，并能确保成功的测序结果。安捷伦自动化电泳仪提供了一系列试剂盒，可兼容多种样品类型和分子量，可提供准确可靠的分子量测定、定量和摩尔浓度。此外，每种仪器均采用专为平台优化的特定质量指标，可对样品完整性进行客观评估。尽管不同类型的仪器质量指标的计算有所不同，但最终的质量评分在不同平台间具有很高的可比性。多篇文献均证明自动化电泳仪非常适合这类重要的 QC 步骤，不仅适用于传统的 NGS 文库制备，还可用于故障排除和方案优化，以及为复杂或独特样品开发定制工作流程。

参考文献

1. Permenter, J.; Ishwar, A.; Rounsavall, A.; Smith, M.; Faske, J.; Sailey, C. J.; and Alfaro, M. P. Quantitative Analysis of Genomic DNA Degradation in Whole Blood Under Various Storage Conditions for Molecular Diagnostic Testing. *Mol. Cell Probes* **2015**, 29, 449–453
2. Zhong, J.; Ye, Z.; Lenz, S. W.; Clark, C. R.; Bharucha, A.; Farrugia, G.; Robertson, K. D.; Zhang, Z.; Ordog, T.; and Lee, J. H. Purification of Nanogram-Range Immunoprecipitated DNA in ChIP-seq Application. *BMC Genomics* **2017**, 18, 985. <https://creativecommons.org/licenses/by/4.0/>
3. Muscarella, L. A.; Fabrizio, F. P.; De Bonis, M.; Mancini, M. T.; Balsamo, T.; Graziano, P.; Centra, F.; Sparaneo, A.; Trombetta, D.; Bonfitto, A.; Scagliusi, V.; Larizza, P.; Capoluongo, E. D.; and Fazio, V. M. Automated Workflow for Somatic and Germline Next Generation Sequencing Analysis in Routine Clinical Cancer Diagnostics. *Cancers* **2019**, 11, 1691. <https://creativecommons.org/licenses/by/4.0/>
4. Klingström, T.; Bongcam-Rudloff, E.; and Pettersson, O. V. A Comprehensive Model of DNA Fragmentation for the Preservation of High Molecular Weight DNA. *bioRxiv online* **2018**. <https://creativecommons.org/licenses/by/4.0/>
5. Kingan, S. B.; Heaton, H.; Cudini, J.; Lambert, C. C.; Baybayan, P.; Galvin, B. D.; Durbin, R.; Korlach, J.; and Lawniczak, M. A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing. *Genes* **2019**, 10, 62. <https://creativecommons.org/licenses/by/4.0/>
6. Performance Characteristics of the RNA and the High Sensitivity RNA ScreenTape Assays for the 4150 TapeStation system (Agilent 4150 TapeStation 系统的 RNA 和高灵敏度 RNA ScreenTape 分析的性能特征), 安捷伦科技公司技术概述, 出版号 5994-1038EN, **2019**
7. Comparison of RIN and RQN for the Agilent Bioanalyzer and the Fragment Analyzer Systems (比较安捷伦生物分析仪和片段分析仪系统的 RIN 和 RQN), 安捷伦科技公司技术概述, 出版号 5994-1860EN, **2020**
8. Haller, A. C.; Kanakapalli, D.; Walter, R.; Alhasan, S.; Eliason, J. F.; and Everson, R. B. Transcriptional Profiling of Degraded RNA in Cryopreserved and Fixed Tissue Samples Obtained at Autopsy. *BMC Clin. Pathol.* **2006**, 6
9. Alpern, D.; Gardeux, V.; Russeil, J.; Mangeat, B.; Meireles-Filho, A. C. A.; Breysse, R.; Hacker, D.; and Deplancke, B. BRB-seq: Ultra-Affordable High-Throughput Transcriptomics Enabled by Bulk RNA Barcoding and Sequencing. *Genome Biol.* **2019**, 20, 71. <https://creativecommons.org/licenses/by/4.0/>
10. Wimmer, I.; Tröscher, A. R.; Brunner, F.; Rubino, S. J.; Bien, C. G.; Weiner, H. L.; Lassmann, H.; and Bauer, J. Systematic Evaluation of RNA Quality, Microarray Data Reliability and Pathway Analysis in Fresh, Fresh Frozen and Formalin-Fixed Paraffin-Embedded Tissue Samples. *Sci. Rep.* **2018**, 8, 6351. <https://creativecommons.org/licenses/by/4.0/>
11. Mendioroz, M.; Martínez-Merino, L.; Blanco-Luquin, I.; Urdániz, A.; Roldán, M.; and Jericó, I. Liquid Biopsy: A New Source of Candidate Biomarkers in Amyotrophic Lateral Sclerosis. *Ann. Clin. Transl. Neurol.* **2018**, 5, 763-768
12. Blecher-Gonen, R.; Barnett-Itzhaki, Z.; Jaitin, D.; Amann-Zalcenstein, D.; Lara-Astiaso, D.; and Amit, I. High-Throughput Chromatin Immunoprecipitation for Genome-Wide Mapping of in vivo Protein-DNA Interactions and Epigenomic States. *Nat. Protoc.* **2013**, 8, 539-554
13. Tramontano, A.; Jarc, L.; Jankowicz-Cieslak, J.; Hofinger, B. J.; Gajek, K.; Szurman-Zubrzycka, M.; Szarejko, I.; Ingelbrecht, I.; and Till, B. J. Fragmentation of Pooled PCR Products for Highly Multiplexed TILLING. G3: *Genes, Genomes, Genetics* **2019**, 9, 2657. <https://creativecommons.org/licenses/by/4.0/>
14. Aguirre, N. C.; Filippi, C. V.; Zaina, G.; Rivas, J. G.; Acuña, C. V.; Villalba, P. V.; García, M. N.; González, S.; Rivarola, M.; Martínez, M. C.; Puebla, A. F.; Morgante, M.; Hopp, H. E.; Paniago, N. B.; Marcucci Poltri, S. N. Optimizing ddRADseq in Non-Model Species: A Case Study in *Eucalyptus dunnii* Maiden. *Agronomy* **2019**, 9, 484. <https://creativecommons.org/licenses/by/4.0/>
15. Lan, J. H.; Yin, Y.; Reed, E. F.; Moua, K.; Thomas, K.; and Zhang, Q. Impact of Three Illumina Library Construction Methods on GC Bias and HLA Genotype Calling. *Human Immunol.* **2015**, 76, 166-175

16. Misra, C. S.; Santos, M. R.; Rafael-Fernandes, M.; Martins, N. P.; Monteiro, M.; and Becker, J. D. Transcriptomics of Arabidopsis Sperm Cells at Single-Cell Resolution. *Plant Reprod.* **2019**, *32*, 29-38
17. Nix, D. A.; Hellwig, S.; Conley, C.; Thomas, A.; Fuertes, C. L.; Hamil, C. L.; Bhetariya, P. J.; Garrido-Laguna, I.; Marth, G. T.; Bronner, M. P.; and Underhill, H. R. The Stochastic Nature of Errors in Next-Generation Sequencing of Circulating Cell-Free DNA. *PLoS One* **2020**, *15*.
<https://creativecommons.org/licenses/by/4.0/>
18. Sidstedt, M.; Steffen, C. R.; Kiesler, K. M.; Vallone, P. M.; Rådström, P.; and Hedman, J. The Impact of Common PCR Inhibitors on Forensic MPS Analysis. *Forensic Sci. Int. Gen.* **2019**, *40*, 182-191
19. Illumina Library Sequencing Services. <https://dnatech.genomecenter.ucdavis.edu/illumina-library-sequencing> (2020年6月17日访问)
20. Genome Sequencing Services Center. <http://med.stanford.edu/gssc/faq.html> (2020年6月17日访问)
21. Marchal, C.; Sasaki, T.; Vera, D.; Wilson, K.; Sima, J.; Rivera-Mulia J. C.; Trevilla-García, C.; Nogues, C.; Nafie, E.; and Gilbert, D. M. Genome-Wide Analysis of Replication Timing by Next-Generation Sequencing with E/L Repli-seq. *Nat. Protoc.* **2018**, *13*, 819-839
22. Quality Control of NGS Libraries with Daisy Chain Molecules (NGS文库中雏菊链分子的质量控制), *安捷伦科技公司应用简报*, 出版号 5994-2233EN, **2020**
23. Quail, M. A.; Kozarewa, I.; Smith, F.; Scally, A.; Stephens, P. J.; Durbin, R.; Swerdlow, H.; and Turner, D. J. A Large Genome Centre's Improvements to the Illumina Sequencing System. *Nat. Methods* **2008**, *5*, 1005-1010
24. Marosy, B. A.; Craig, B. D.; Hetrick, K. N.; Witmer, P. D.; Ling, H.; Griffith, S. M.; Myers, B.; Ostrander, E. A.; Stanford, J. L.; Brody, L. C.; and Doheny, K. F. Generating Exome Enriched Sequencing Libraries from Formalin-Fixed, Paraffin-Embedded Tissue DNA for Next-Generation Sequencing. *Curr. Protoc. Hum. Genet.* **2017**, *92*(1), 18.10.1-18.10.25

www.agilent.com/automated-electrophoresis

仅限研究使用。不可用于诊断目的。
PR7000-7429

本文中的信息、说明和指标如有变更,恕不另行通知。

© 安捷伦科技(中国)有限公司, 2020
2020年9月30日, 中国出版
5994-2281ZHCN

查找当地的安捷伦客户中心:

www.agilent.com/chem/contactus-cn

免费专线:

800-820-3278, 400-820-3278 (手机用户)

联系我们:

LSCA-China_800@agilent.com

在线询价:

www.agilent.com/chem/erfq-cn

