

Agilent SureSelect XT HS RNA 文库制备：可兼容由新鲜样本或 FFPE 组织样品直接捕获构建 RNA 测序文库的简化和改进工作流程

作者

Carsten Carstens,
Katherine Felts, Sarah Johns
安捷伦科技有限公司

摘要

在本应用简报中，我们介绍了一种用于构建靶向 RNA 捕获测序文库的精简、改进工作流程。在 Agilent SureSelect XT RNA 直接捕获方案的基础上实施了四项重大改进：

1. 将过夜捕获杂交替换为 90 分钟快速杂交
2. 省去尿嘧啶去糖基化酶 (UDG) 处理（使用新酶保持 RNA 链特异性）
3. 将旧 SureSelect XT 接头替换为 SureSelect XT HS 接头，能够平行处理同一样品中的 DNA 和 RNA
4. 提供唯一分子标签 (MBC)，以改善 PCR 和片段重复的去冗余

这些改进将 2-3 天的周转时间缩短为 1 天。此外，通过使用多种起始样品类型（完整、新鲜冷冻、FFPE）分析融合数据发现，这一简化的工作流程可产生文库复杂度优异的 RNA 测序文库，并可使用低至 10 ng 的 RNA 起始量获得良好的测序性能。

前言

将高通量测序应用于转录组学（RNA 测序，RNA-seq），不仅能够实现全基因表达谱分析，而且增加了有关剪接变异、融合转录本、翻译后编辑和等位基因特异性表达的精确信息。RNA-seq 文库构建的标准方法包括：依赖于 Mg^{2+} 的酯交换反应对 RNA 起始材料的打断，采用随机六聚体引物进行第一链 cDNA 合成，以及随后通过 dUTP 链标记实现的第二链 cDNA 合成。为防止片段之间串接，将 cDNA 进行末端修复，并利用无模板 DNA 聚合反应将腺嘌呤添加至 3' 端。然后进行平台特异性测序接头的连接。接头连接后的所有步骤均与 DNA 测序工作流程相同。

尽管接头连接后 DNA 和 RNA 衍生的测序文库看起来完全相同，但 RNA 测序与 DNA 测序相比，仍然存在特定的差异。第一个区别在于需要保持测序片的方向性，以确定原始 RNA 所对应的基因组 DNA 链。这通常通过以下步骤实现：在第二链合成反应中加入尿嘧啶，继续进行文库制备，然后使用 UDG 去除第二链。第二个区别在于 RNA 测序中测得的转录本的动态范围宽，由于表达的编码转录本的相对丰度差异巨大，而核糖体 RNA (rRNA) 与编码 RNA 相比具有极高的丰度，因此该动态范围可跨越五个数量级以上。因此，RNA-seq 需要降低复杂度，以免将读出序列浪费在无信息转录本（如 rRNA）上。最常用的方法包括核糖体 RNA (rRNA) 去除法和 mRNA polyA 富集法。在一个典型样品中，即使去除 rRNA 后，所有表达基因的前 1% 仍将占有所有转

录本的 50% 左右。因此，如果研究目标需要研究低表达或中等表达的基因，则这种检测将获益于进一步降低的复杂度。

作为靶向去除方法的一种替代技术，也可以通过使用生物素化靶向序列捕获探针（又称“诱饵”）进行靶向序列捕获来降低复杂度，这是基因组测序中的常用方法。在 RNA 测序中，通过探针选择进行靶向序列捕获主要与 FFPE 所得的样品结合使用，众所周知，这种样品 rRNA 清除很不彻底，且由于 FFPE 样本降解而无法使用 poly(A) 富集^[1]。只需要检测少数几个或者某些特定的转录本时，靶向序列捕获也很有利。比如，检测指示潜在基因融合事件的融合转录本。另外，任何仅需获取转录本子集信息的情况，例如转录指纹分析或罕见的转录后编辑的检测，均可以获益于靶向序列捕获^[2]。

安捷伦提供了可用于构建靶向 RNA 测序文库的 SureSelect XT RNA 直接捕获文库制备试剂盒（货号 G7564A, G7564B）。我们证明该试剂盒可成功用于 FFPE 来源的样品^[3]。在本应用简报中，我们介绍了一种将 SureSelect XT RNA 直接捕获文库制备试剂盒的组分与 SureSelect XT HS 靶向序列捕获试剂盒（货号 G9706A）以及改良的方案相结合的简化改进工作流程。通过省去起始 RNA 的冻干，省去一个 SPRI 磁珠纯化步骤，以及将 24 小时探针杂交步骤替换为 1.5 小时杂交步骤，简化了工作流程。我们还省去了 UDG 处理步骤，通过在预捕获 PCR 期间使用可区分含尿嘧啶的模板 DNA 的 PCR 酶，来实现链特异性。此外，使用 SureSelect XT HS 建库流程，该流程可以同时兼容

DNA 和 RNA 捕获文库构建，并在测序接头增加 MBC 以鉴定片段重复，这样可以大大降低对同一样品的 DNA 和 RNA 进行分管（又称“平行”）处理的实际障碍。

实验部分

RNA 来源

通用的人参比 RNA (UHRR) 以新鲜冷冻材料得自安捷伦科技公司（美国加利福尼亚州圣克拉拉市，部件号 750500-41）。一组匹配的乳腺肿瘤和正常相邻组织（新鲜冷冻材料和 FFPE 两种样本形式）来自 CureLine 人类生物样本 CRO（美国加利福尼亚州布里斯班，定制货号）。SeraSeq FFPE 肿瘤融合 RNA 参比物质 v2 购自 SeraCare（美国马里兰州盖瑟斯堡，货号 0710-0129）。

RNA 提取

如果需要，根据制造商的说明使用得自 Qiagen 的 RNeasy FFPE 试剂盒或 RNeasy mini 试剂盒（Qiagen USA，美国马里兰州日耳曼敦，货号分别为 73504 和 74104）分离 RNA。有关更详细的方案，请参见附录。

起始材料和测序文库的质量评估

在 Agilent 2100 生物分析仪系统（安捷伦科技公司，货号 G2939B）上，采用用于 RNA 质量评分的 Agilent RNA 6000 Pico 试剂盒（安捷伦科技公司，货号 5067-1513）或用于评估测序文库质量的 Agilent DNA 1000 试剂盒（安捷伦科技公司，货号 5067-1504）对核酸样品进行评价。

其他材料

放线霉素 D 购自 Sigma (美国密苏里州圣路易斯, 货号 A1410), 并在 DMSO 储备液中制成浓度为 4 $\mu\text{g}/\mu\text{L}$ 的溶液。使用 AMPure XP 磁珠 (Beckman Coulter, 美国乔治亚州亚特兰大, 货号 A63880) 进行 SPRI 磁珠纯化。使用 Dynabeads MyOne 链霉亲和素 T1 磁珠 (赛默飞世尔科技, 美国马萨诸塞州沃尔瑟姆, 货号 65601) 捕获生物素化探针。

SureSelect XT HS RNA 文库制备

使用 SureSelect XT RNA 直接捕获试剂盒 (安捷伦科技公司, 货号 G7564A) 和用于 Illumina 配对末端多重测序文库 (安捷伦科技公司, 货号 G9706A) 的 SureSelect XT HS 靶向序列捕获系统, 构建 RNA 测序文库。有关详细说明, 请参见附录。

SureSelect XT RNA 直接捕获文库制备

按照 SureSelect XT RNA 直接捕获文库制备方案 (手册) 中的说明, 生成、富集和测序 RNA 直接捕获文库。

靶向序列捕获

使用 SureSelect 人全外显子 V7 外显子组 (安捷伦科技公司, 货号 5191-4029) 靶向编码转录组来实施 SureSelect XT HS RNA 文库的靶向序列捕获。详细的探针捕获方案参见附录。

测序与数据分析

在 Illumina HiSeq 4000 上, 利用配对末端测序以 2×150 读出序列格式对测序文库进行分析。在表达分析中 (数据未显示), 使用带剪接识别的 STAR 2.6.0a 版软件包将 FASTQ 文件与转录组对齐, 并以 hg38 基因组作为参比。然后使用 RSEM 工具由 STAR 比对结果生成表达谱。使用 Picard RNA 分析方案生成常规文库统计数据 (链特异性、5'-3' 偏好性、MBC 去重、文库大小估计值), 并使用

.bam 文件标记重复, 这些文件下采样至 2×10^7 读出序列以生成归一化重复率。采用同一方案生成经 MBC 校正的重复统计数据 and 文库大小估计值, 而用 UmiAwareMarkDuplicatesWithMateCigar

去除片段重复。使用 STAR-Fusion 对融合转录本进行评分, 并使用作为 Trinity 癌症转录组分析工具包 (CTAT) 一部分的 FusionInspector 工具进行可视化^[4]。

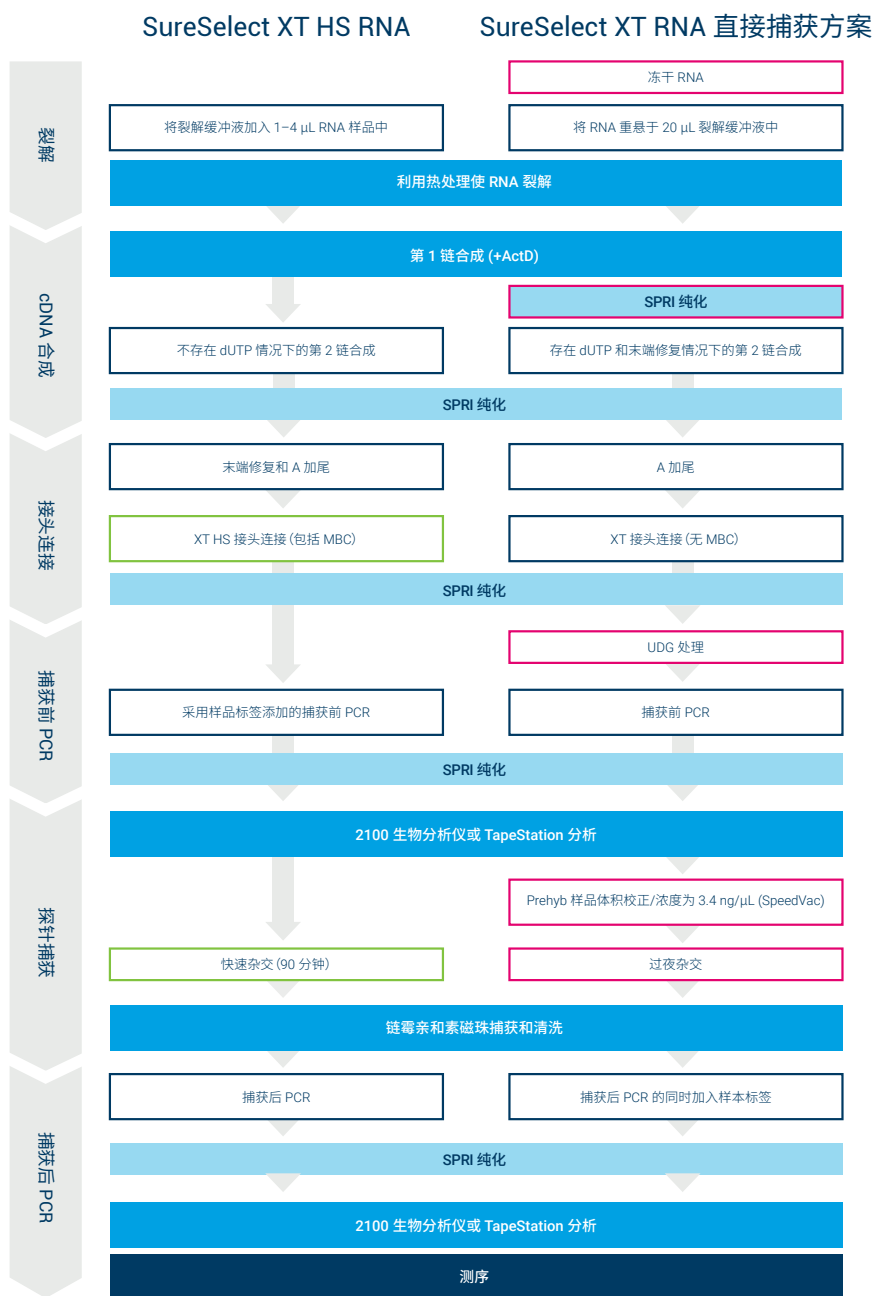


图 1. Agilent SureSelect XT HS RNA 工作流程 (RNA XT HS) 与 Agilent SureSelect XT RNA 直接捕获文库制备试剂盒 (RNA 直接捕获) 的比较。粉色边框指示从 RNA 直接捕获工作流程中删除的步骤。绿色边框指示可改善 RNA XT HS 方案功能的步骤

结果与讨论

RNA XT HS 和 RNA 直接捕获工作流程概述

我们希望将 SureSelect XT RNA 直接捕获文库制备试剂盒 (RNA 直接捕获) 中的组分与 SureSelect XT HS 靶向序列捕获试剂盒相结合, 开发出一种简化的改进工作流程, 以生成高质量 RNA 测序文库。由此得到的这一工作流程称为 SureSelect XT HS RNA (RNA XT HS)。比较这些工作流程的概述如图 1 所示。RNA XT HS 工作流程相对于 RNA 直接捕获工作流程具有明显的优势, 包括省去了 UDG 处理, 并将传统的过夜杂交替换为 90 分钟快速杂交, 大大缩短了周转时间。我们还希望确定能否在第一链和第二链合成之间去除 SPRI 纯化步骤, 并在 RNA XT HS 工作流程中省去冻干 RNA 起始材料 (有关详细信息, 请参见附录)。

比较工作流程性能

为比较两种文库制备过程的性能, 我们首先使用不同量的通用的人参比 RNA (UHRR) 和 SeraCare FFPE 肿瘤融合 RNA 参比物质 v2 (SeraCare) 作为分别代表完整样品和理想化 FFPE 样品的起始材料生成文库。用 SureSelect 人全外显子 V7 富集 RNA XT HS 和 RNA 直接捕获文库。最后, 使用 Illumina 测序仪对这些富集的文库进行测序, 并使用定制数据分析方案对数据进行分析 (有关详细信息, 请参见“实验”部分)。

表 1 汇总了通过分析 RNA XT 和 RNA 直接捕获测序数据得到的全局测序统计数据。我们发现, 测序的 RNA XT HS 文库在多个指标上与 RNA 直接捕获文库难分上下,

其中最明显的是高链特异性 (> 98%) 和低 rRNA 污染 (约 0.1%)。RNA XT HS 文库的高链特异性展示了 RNA XT HS 工作流程中省去 UDG 处理的方法的有效性, 便于利用不扩增含尿嘧啶模板的 PCR 酶。与 FFPE 样品相比, 完整样品的链特异性略高一些, 这是由于 FFPE 起始材料的质量有所下降。无论如何, 采用任何一种方案在 FFPE 起始材料中观察到的链特异性仍然很高 (对于 FFPE 材料而言, 高于 98%)。两种文库制备方法中 rRNA 污染的百分比始终较低, 这已在之前的靶向序列捕获方法中得到证明^[9]。

比较 RNA XT HS 和 RNA 直接捕获文库的外显子匹配率, 我们再次发现两者性能相当 (图 2)。缩短新型 RNA XT HS 工作流程的一个关键步骤是将传统的过夜杂交替换为快速杂交步骤。如图 2 所示, 不论样本起始量多还是少, 加速捕获不影响匹配率, 始终观察到 90% 的外显子读出率, 很少序列会被比对到基因间区。无论起始量或样品类型如何, V7 捕获文库靶向的转录本均占观察到的所有表达的 92.9%–94.1%。剩余 7% 左右归因于污染的高表达转录本, 例如线粒体基因或标注错误。

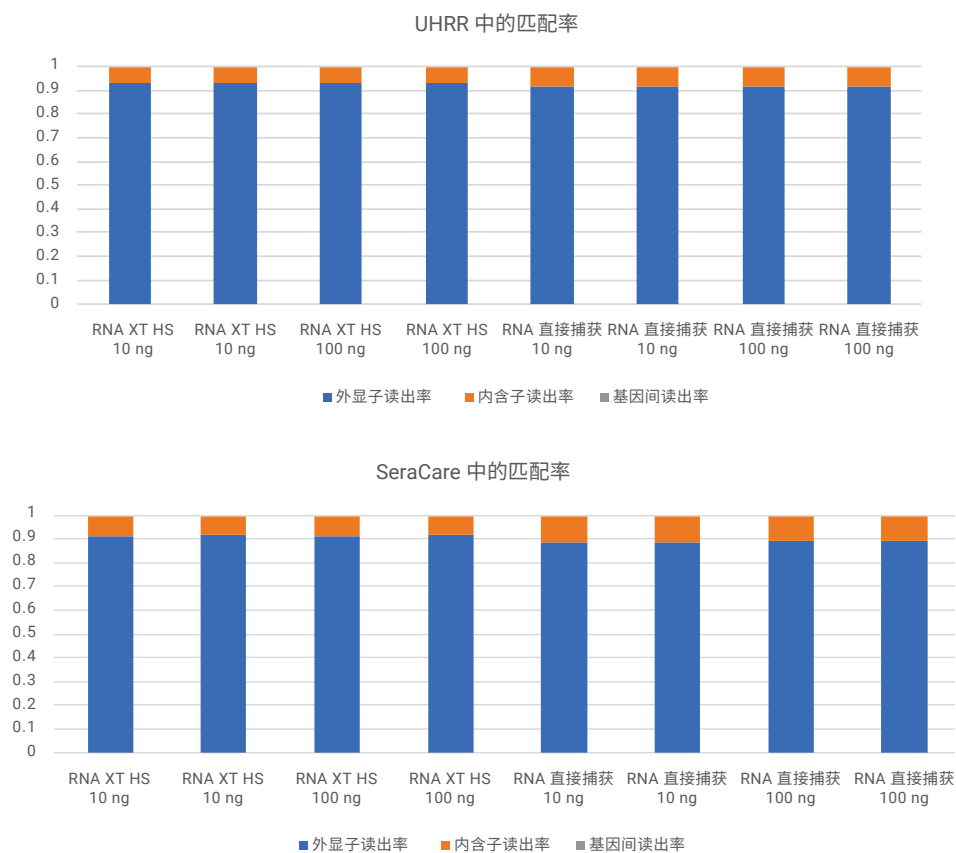


图 2. RNA XT HS 和 RNA 直接捕获测序文库的匹配率比较。使用 RNA XT HS 或 RNA 直接捕获工作流程由 UHRR (A) 和 SeraCare 样品 (B) 制备 RNA-seq 文库。显示了 Agilent SureSelect 人全外显子 V7 外显子组富集的测序文库与外显子、内含子和基因间序列的匹配率。怀疑内含子读出率反映了正在测序的未经处理的 mRNA

文库制备过程总体效率的一项关键指标是估计的起始文库复杂度。文库复杂度越高越好，这类文库相比复杂度低文库来讲可以提供更高的定量可信度。文库复杂度是基于测序深度（基于配对序列）和观察到的独特变异的数量（由重复率得出）计算得出。我们使用 Picard 方案（假定所有重复均为 PCR 重复）来计算估计的文库复杂度。这种方法低估了真实的文库复杂度，对此将在结果部分予以详细讨论。预期得到的文库复杂度列于表 1 中，并如图 3 所示。比较 RNA XT HS 和 RNA 直接捕获工作流程，结果发现，由高起始量 (100 ng) 新鲜样品和 FFPE RNA 起始材料构建的文库之间无显著差异。但是，当 UHRR 或 SeraCare 的起始材料减少至 10 ng 时，发现文库复杂度明显不同。首先，这些起始量较低的文库比起始量更高的文库复杂度更小、更简单。这点符合预期，因为人们尝试转换为测序文库的 RNA 起始材料的量越少，最终的文库复杂度就应越小。我们还发现，在起始量较低的情况下，简化的 RNA XT HS 工作流程所产生的文库的效率相比 RNA 直接捕获工作流程提高了约 1.5–2 倍。我们正在对 RNA XT HS 工作流程中导致效率提升的多种因素进行研究（数据未显示）。

估计的测序文库复杂度 (Picard 方案)

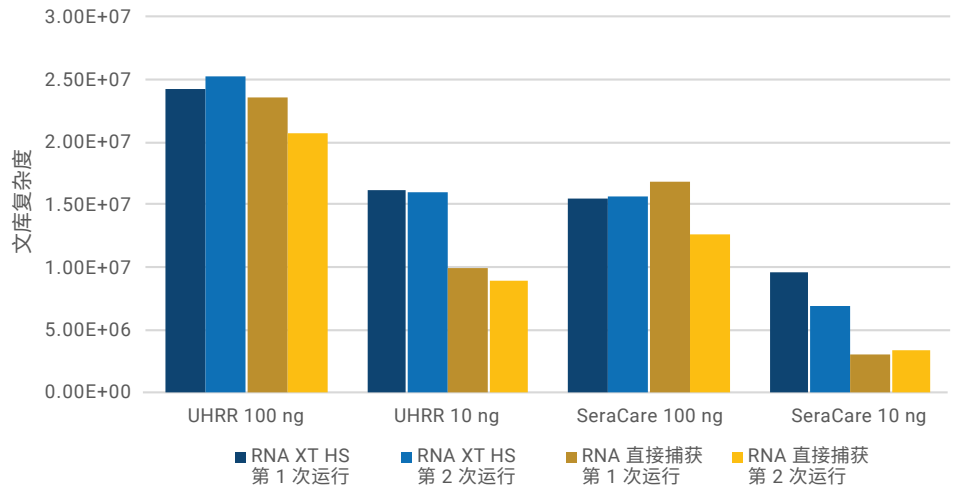


图 3. 在低样品起始量下，RNA XT HS 与 RNA 直接捕获文库之间的文库复杂度差异。在 Illumina HiSeq 4000 平台上，通过配对末端测序，对使用指示的来源和起始量制备的测序文库进行测序。根据观察到的重复率以及由 Picard RNA 分析工具确定的相应读出序列对的数量，确定估计的文库复杂度。该方案不区分 PCR 和片段重复

表 1. SureSelect XT HS RNA 和 SureSelect XT RNA 直接捕获文库的全局测序统计数据：在 Illumina HiSeq 4000 平台上，对使用 RNA XT HS 或 RNA 直接捕获方案由相同的起始材料中生成的 RNA 测序文库进行测序。下采样至 2×10^7 读出序列后，使用 Picard RNA 分析工具生成文库统计数据。使用 UmiAwareMarkDuplicatesWithMateCigar 标记 MBC 后，利用相同的方案生成经 MBC 校正的统计数据。(A) 新鲜冷冻 (“完整”) RNA 起始材料；(B) FFPE 起始材料

A. 通用参比 RNA (UHRR)

方案	SureSelect XT HS RNA				SureSelect XT RNA 直接捕获方案			
	100	100	10	10	100	100	10	10
样品起始量 (ng)	100	100	10	10	100	100	10	10
分析的读出序列 (百万)	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9
rRNA 读出率 (%)	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
链特异性 (%)	99.1%	99.1%	99.0%	99.0%	99.1%	99.2%	99.1%	99.1%
重复率 (%)	17.3%	16.7%	24.3%	24.7%	17.9%	20.0%	35.8%	38.7%
估计的文库大小 ($\times 10^6$)	24.2	25.2	16.2	15.9	27.5	23.6	10.3	9.2
经 MBC 校正的重复率 (%)	7.1%	6.4%	17.1%	17.8%	不适用	不适用	不适用	不适用
经 MBC 校正的估计文库大小 ($\times 10^6$)	65.9	72.8	24.6	23.6	不适用	不适用	不适用	不适用

B. Seraseq FFPE 肿瘤融合 RNA 参比物质 v2 (SeraCare)

方案	SureSelect XT HS RNA				SureSelect XT RNA 直接捕获方案			
	100	100	10	10	100	100	10	10
样品起始量 (ng)	100	100	10	10	100	100	10	10
分析的读出序列 (百万)	18.6	18.6	18.7	18.6	19	18.8	18.9	18.9
rRNA 读出率 (%)	0.1%	0.2%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%
链特异性 (%)	98.6%	98.6%	98.6%	98.7%	98.9%	98.9%	98.9%	98.9%
重复率 (%)	24.8%	24.7%	36.0%	45.0%	23.5%	29.4%	69.1%	67.2%
估计的文库大小 ($\times 10^6$)	15.5	15.5	9.6	6.9	18.6	13.3	3	3.3
经 MBC 校正的重复率 (%)	13.2%	13.7%	29.3%	38.8%	不适用	不适用	不适用	不适用
经 MBC 校正的估计文库大小 ($\times 10^6$)	32.5	31.2	12.7	8.66	不适用	不适用	不适用	不适用

唯一分子标识符 (MBC) 对 RNA 测序文库的影响

如上文所述, 准确测量文库的复杂度对于评估文库制备过程的效率至关重要。上述文库复杂度的估计基于以下假设: 具有相同起始和终止的读出序列对是来源于相同的原始文库分子的 PCR 重复。但是, 随机打断也可能引起重复, 导致两个独立的片段具有相同的末端。由于片段重复是真正的独立测序文库组成部分, 因此文库复杂度的确定只能基于真实的 PCR 重复。

与 DNA 测序相比, RNA 测序可能导致更多的片段重复, 这是由于某些基因的高表达增加了一些随机片段重复的机会。这导致文库复杂度估计更加不准确。RNA XT HS 工作流程的一个潜在优势是可以使用包含单个 10 bp MBC 的 XT HS 接头构建文库。假设该 MBC 可用于区分 PCR 和片段重复。我们使用改良的数据分析方案 (利用 MBC 数据) 对上述 RNA XT HS 测序数据进行了重新分析 (注: RNA 直接捕获文库不包括 MBC, 因此未纳入该重新分析中)。该分析的结果报告于表 1 和图 4 中。

如图 4 所示, 校正片段重复会导致对相应的文库复杂度的过高估计, 尤其是对于较高的起始量, 其中如果不考虑片段重复, 文库复杂度会被低估三倍左右。预期在文库更大且相应起始量更高的情况下, 将产生更大的偏差, 因为如果将来源于相同编码序列的更多分子处理到文库中, 则观察到片段重复的机会更高。

接下来, 我们希望在—组“实际”样品中探讨 MBC 校正的影响。因此, 我们使用—组匹配的乳腺肿瘤和邻近的正常组织 (两者均作为新鲜冷冻样品以及 FFPE

样品进行储存和处理), 生成 RNA XT HS 文库。使用如上所述的有关 UHRR 和 SeraCare RNA XT HS 文库的步骤, 对这些文库进行富集、测序和分析。这些文库的汇总统计数据可参见表 2 和图 4。

结果发现, 与 UHRR 和 SeraCare 文库相比, 在肿瘤/正常新鲜冷冻样品和 FFPE 样品测序数据中观察到优异的链特异性和 rRNA 读出率。我们还观察到 FFPE 文库明显小于新鲜冷冻文库,

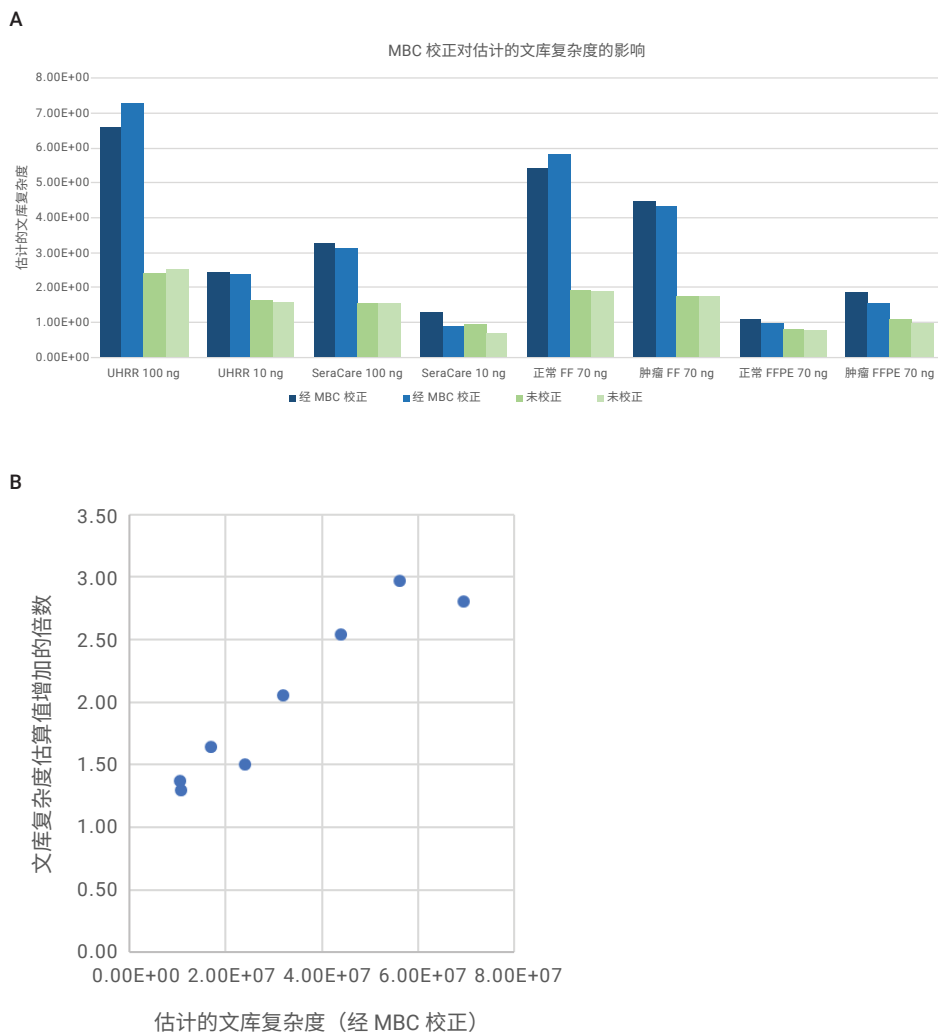


图 4. 唯一分子标识符 (MBC) 对文库大小估计值的影响。RNA XT HS 文库的大小通过 Picard 去冗余 (未校正) 或使用 MBC (经 MBC 校正) 后计算得出。Picard 去冗余假定所有重复均为 PCR 重复, 而 MBC 校正则区分 PCR 重复与随机片段重复。A) 由一系列不同样品和起始材料得出的几种 RNA XT HS 测序文库的估计文库复杂度。B) 经 MBC 校正的文库与未校正文库之间的差异在较大文库 (因此起始量也相应较高) 中更明显

这是一个得到广泛报道的现象。使用 MBC 校正法分析 FFPE 测序数据时，我们发现文库估计复杂度改善了 0.3–3.5 倍。因此，在区分 PCR 重复和片段重复的情况下，将恢复假定所有重复都是 PCR 重复时丢失的数据。这表明，为最大程度获取 FFPE RNA 样品所提供的测序信息，准确鉴别这些样品中的 PCR 重复和片段重复至关重要。

基因融合的检测

靶向 RNA-seq 已被证明在检测基因融合事件方面具有实用性，对于具有挑战性的样品 (FFPE) 而言尤其如此。我们想进一步了解 RNA XT HS 文库构建流程通过外显子探针捕获（相比于仅针对融合基因捕获）是否可以用于融合基因的检测。如上文所述，我们使用 SureSelect 人全外显子 V7 对来自所有样品 (UHRR、SeraCare、新鲜冷冻和 FFPE) 的 RNA XT HS 文库进行靶向序列捕获，并在 2×150 读取长度下对作为配对末端文库的富集文库进行测序。使用 STAR-Fusion 分析肿瘤/正常新鲜冷冻和 FFPE 测序数据，并通过 FusionInspector 进行可视化。

FusionInspector 结果检测到一系列推断的融合，包括已知的假阳性，例如 VDJ 重组事件，表明某些样品中存在免疫细胞。表 3 中移除了 VDJ 重组。数据分析显示，样品中存在 15 种潜在的基因融合，以前报道的 15 个融合伙伴基因中的 12 个参与了与肿瘤相关的基因融合，尽管后者不存在于观察到的组合中。融合蛋白之一 FCHSD2-FAM168A（在表 3 中以灰色突出显示）先前

表 2. 新鲜冷冻和 FFPE Agilent SureSelect RNA XT HS 文库的全局测序统计数据。使用 RNA XT HS 方案，由一组匹配的肿瘤和正常邻近组织（作为新鲜冷冻组织和 FFPE）生成的 RNA 测序文库。在 Illumina HiSeq 4000 平台上对这些文库进行测序。下采样至 2×10^7 读出序列后，使用带有或不带 MBC 标签的 Picard RNA 分析工具生成文库统计数据

样品类型	新鲜冷冻				FFPE			
	正常	正常	肿瘤	肿瘤	正常	正常	肿瘤	肿瘤
来源								
样品起始量 (ng)	70	70	70	70	70	70	70	70
分析的读出序列 (百万)	18.8	18.8	18.7	18.8	18.6	18.6	18.7	18.7
rRNA 读出率 (%)	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%
链特异性 (%)	99.0%	99.0%	99.1%	99.2%	98.4%	98.4%	98.8%	98.8%
重复率 (%)	21.0%	21.0%	23.0%	23.0%	41.0%	43.0%	33.0%	35.0%
估计的文库大小 ($\times 10^6$)	19	18.8	17.3	17.3	7.9	7.4	10.9	9.8
经 MBC 校正的重复率 (%)	8.0%	8.0%	10.0%	10.0%	33.0%	35.0%	22.0%	25.0%
经 MBC 校正的估计文库大小 ($\times 10^6$)	53.9	58.2	44.7	43.1	11	9.93	18.6	15.4

表 3. 新鲜冷冻和 FFPE 正常/肿瘤样品中基因融合的检测。对 RNA XT HS 测序文库进行测序，然后使用 STAR-Fusion 方案分析基因融合。在每一行中列出了从一个或多个样品中检测到的推断融合以及支持该融合检测的证据。连接读出序列和支持读出序列表示为每百万个唯一文库读出序列对中的读出序列。每一列代表两次技术重复的平均值。在 TGCA 数据库中，以灰色突出显示的融合被鉴别为与乳腺癌相关的融合，并可能是该肿瘤的驱动性突变。从分析中除去观察到的指示样品中存在免疫细胞的 VDJ 重组事件。检测基于对每个样品的 $(0.7-1.4) \times 10^7$ 唯一读出序列对的分析结果

	新鲜冷冻				FFPE			
	正常		肿瘤		正常		肿瘤	
推断的融合	连接	支持	连接	支持	连接	支持	连接	支持
MYLK--LPAR6	0.00	0.00	8.86	0.86	0.00	0.00	1.68	0.00
FCHSD2--FAM168A	0.00	0.04	5.60	0.21	0.00	0.00	2.79	0.04
FAM157A--RB1	0.00	0.00	2.59	0.04	0.00	0.00	0.91	0.00
RP4-565E6.1--HYDIN	0.00	0.00	1.47	0.00	0.00	0.00	1.33	0.00
CDR2--FRG1	0.00	0.00	0.42	0.00	0.00	0.00	0.48	0.00
SLC7A5--RP11-645C24.2	0.00	0.04	0.33	0.20	0.00	0.00	0.09	0.04
PHRF1--TXNDC5	0.00	0.00	2.78	0.17	0.00	0.00	0.00	0.00
NUDT1--AC004840.8	0.00	0.00	1.71	0.00	0.00	0.00	0.00	0.00
MPZL1--RCSD1	0.00	0.00	1.34	0.06	0.00	0.00	0.00	0.00
CCDC66--SLMAP	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00
UBE2Q2--C15orf27	0.00	0.00	0.77	0.06	0.00	0.00	0.00	0.00
POLR2J--AC004980.9	0.00	0.00	0.00	0.00	11.53	0.00	0.00	0.00
POLR2J--UPK3B	0.00	0.00	0.00	0.00	3.96	0.00	0.00	0.00
RP11-634B7.4--TRIM58	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RB1--MYLK-AS1	0.00	0.00	0.16	0.03	0.00	0.00	0.00	0.00

已被确定为与乳腺癌相关，并可能是该肿瘤的驱动性突变^[5]。对于所有其他肿瘤特异性融合，TCGA 数据库中列出至少一个融合伙伴基因（通常是两个），并与肿瘤形成相关。我们发现融合转录本 A:B、A 和 B 两者均存在于肿瘤数据库中。但是，不存在 A:B 的列表。尽管存在 A:C 和 D:B 融合转录本的示例。这些数据表明，富含外显子组的 RNA XT HS 文库使研究人员能够鉴别整个转录组水平下的融合，即使具有挑战性的样品也适用。

结论

为了使靶向 RNA-seq 文库的构建效率更高，我们通过结合使用 SureSelect XT RNA 直接捕获文库制备试剂盒与 SureSelect XT HS 靶向序列捕获试剂盒来改进当前的 SureSelect XT RNA 直接捕获工作流程。新型 SureSelect XT HS RNA 工作流程对工作流程进行了多项重大改进：

- 直接将裂解缓冲液加入 RNA 样品中，以替代 RNA 起始材料冻干
- 省去第一链 cDNA 合成后的 SPRI 磁珠纯化步骤
- 利用另一种 PCR 酶，该酶不扩增含尿嘧啶的模板，无需 UDG 处理即可保持链特异性
- 将探针（“诱饵”）捕获程序由 24 小时杂交更改为 1.5 小时快速杂交方案
- 将 XT 测序接头替代为 XT HS 接头，该接头使 MBC 能够用于更准确的文库复杂度评估和测序读出序列的“恢复”

我们发现 RNA XT HS 工作流程将文库和靶向序列捕获周转时间从 2-3 天缩短至 1-2 天。新型简化工作流程产生的 RNA-seq 文库与 SureSelect XT RNA 直接捕获方案生成的文库在链特异性、rRNA 读出率和高起始量下的匹配率方面无明显差异。在较低的起始量 (10 ng) 下，我们发现缩短的工作流程不影响总体性能，而且似乎更高效。我们还发现，改进的接头 (XT HS) 设计中包含 MBC，现在能够鉴别片段重复。这一改进的片段重复鉴别方法改善了测序结果，可避免使用标准的开始-停止复制方法时遇到的读出序列丢失。

当我们检测“实际”样品时，发现即使对于 FFPE 样品，RNA XT HS 工作流程仍可继续生成高质量数据。初步的基因融合分析在新鲜冷冻和 FFPE 肿瘤样品中检测到潜在的基因融合，展示了使用 RNA XT HS 对 FFPE RNA 进行测序的潜在应用案例。尽管没有显示基因表达或剪接数据，但我们的初步分析表明，靶向 RNA seq 可用于全局基因表达分析、剪接变异检测、变异表达检测和等位基因特异性表达分析。最后，SureSelect RNA XT HS 工作流程的开发使其与 DNA SureSelect XT HS 靶向序列捕获试剂盒“对齐”，从而更轻松地对同一样品中的 DNA 和 RNA 进行平行测序。这在多种应用中可能有利，包括常规的多组学研究以及用于 TMB-MSI 分析和融合检测的样品平行处理。

缩写词

FFPE，福尔马林固定石蜡包埋；TPM，每千个碱基的百万转录本；nt，核苷酸；UHRR，通用人参比 RNA；MBC，分子条形码；UDG，尿嘧啶去糖基化酶；

参考文献

- 1) Cieslik, M., *et al.* The Use of Exome Capture RNA-Seq for Highly Degraded RNA with Application to Clinical Cancer Sequencing. *Genome Res.* **2015**, 25, 1372-1381
- 2) Mittempergher, L., *et al.*, MammaPrint and Blueprint Molecular Diagnostics Using Targeted RNA Next-Generation Sequencing Technology. *J. Mol. Diagn.* **2019**, 21, 808-823
- 3) Jones, J. C.; Alex Siebold, A.; Lucas, A. B. SureSelect XT RNA 直接捕获方案可提供同步的 FFPE RNA 样品的转录组富集和核糖体清除，安捷伦科技公司应用简报，出版号 5991-8119CHCN，**2017**
- 4) Haas, B., *et al.* STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv.120295* (**2017**)
- 5) Hu, X., *et al.* TumorFusions: An Integrative Resource for Cancer-Associated Transcript Fusions. *Nucleic Acids Res.* **2018**, 4, 46(D1), D1144-D1149

附录

1. RNA 测序文库构建的详细方案

RNA 样品前处理

根据制造商说明使用 Qiagen RNeasy FFPE 试剂盒从 FFPE 卷中分离总 RNA。使用 Qiagen RNeasy mini 试剂盒从冷冻组织中分离 RNA。在 Agilent 2100 生物分析仪系统上，使用 RNA 6000 Pico 试剂盒分析所有总 RNA 样品。在将样品加载到芯片上之前，将样品加热至 80 °C 并保持 2 分钟。RIN 和 DV200 值通过 Bioanalyzer 软件计算得出。所有检测的总 RNA 样品的这些样品质量指标列于表 S1 中。

使用 Agilent SureSelect XT RNA 直接捕获试剂盒中的试剂制备 cDNA

注：在 DMSO 中提前制备 4 µg/µL 的放线霉素-D 储备液，并以单次使用的等分试样 (3 µL) 冷冻保存于 -20 °C 下。

注：将 SureSelect XT RNA 直接捕获试剂盒中的裂解混合液、第一链反应预混液以及第二链酶和寡核苷酸混合物置于冰上解冻，并高速涡旋 5 秒，然后在使用前短暂离心。

1. 用 4 µL 不含核酸酶的水制备总 RNA 样品。实验之间的起始量有所不同，并显示于结果图示中

注：允许较小的 RNA 样品起始体积 (小于 4 µL)，但是不建议采用较大的 RNA 样品起始体积 (大于 4 µL)。

2. 将裂解混合液加入 RNA 样品中，使最终样品体积达到 20 µL

3. 根据 SureSelect XT RNA 直接捕获试剂盒方案 (G9691) 的建议，使用基于单个 RNA 样品质量指标的条件，在

表 S1. 样品质量指标和裂解条件

样品描述	RIN	DV200	裂解
通用人参比 RNA (UHRR)	9.2	94%	94 °C, 8 min
Seraseq FFPE 肿瘤融合 RNA 参比物质 v2	2	54%	94 °C, 3 min; 65 °C, 2 min
正常乳腺, 冷冻	6.3	94%	94 °C, 8 min
肿瘤乳腺, 冷冻	5	88%	94 °C, 8 min
正常乳腺, FFPE	2.1	48%	65 °C, 5 min
肿瘤乳腺, FFPE	2	47%	65 °C, 5 min

表 S2. 第一链反应混合物

试剂	一次反应的体积	八次反应的体积 + 过量 10%
RNA-seq 第一链反应预混液	8 µL	70.4 µL
放线霉素-D (120 ng/µL)	0.5 µL	4.4 µL
总计	8.5 µL	74.8 µL

表 S3. SPRI 清洗方案参数

AMPure 磁珠体积	105 µL (1.8X 体积)
磁珠孵育时间	5 min
70% 乙醇清洗 (执行两次)	200 µL
37 °C 下的干燥时间	1-2 min 或更短
洗脱体积	50 µL 不含核酸酶的水

- 将 28.5 µL 反应物置于预编程的 SureCycler 8800 (或等效的热循环仪) 中加热，以使 RNA 样品裂解。用于每个样品的裂解参数列于表 S1 中。裂解后，将样品保存在冰上，直至准备进行第一链合成步骤为止
- 在水中将 4 µg/µL 放线霉素-D 的 DMSO 储备液稀释至 120 ng/µL (3 µL 放线霉素-D + 97 µL 水)
- 制备八个样品大量反应混合物用于第一链合成 (表 S2)。对反应混合物进行涡旋混合，在使用前将其保存在冰上
- 将 8.5 µL 第一链反应混合物加入冰上的各 20 µL 裂解样品中。通过短暂涡旋并离心，将样品混合
- 将 28.5 µL 反应物置于预编程的 SureCycler 8800 中在 25 °C 下孵育 10 分钟，然后在 37 °C 下孵育 40 分钟，然后保存在 4 °C (或置于冰上) 下，直至进行第二链合成
- 在使用前，对第二链 + 末端修复酶混合物和 RNA-seq 第二链 + 末端修复寡核苷酸管进行涡旋混合
- 将 25 µL 第二链 + 末端修复酶混合物 (蓝色盖) 加入冰上的 28.5 µL 第一链反应物中
- 随后立即加入 5 µL 第二链 + 末端修复寡核苷酸混合物 (黄色盖)
- 将样品加盖，涡旋混合，短暂离心，然后放回冰中

12. 将 58.5 μ L 反应物置于预编程的 SureCycler 8800 中在 16 $^{\circ}$ C 下孵育 60 分钟，然后保存在 4 $^{\circ}$ C（或置于冰上）下，直至进行 SPRI 纯化
13. 经过第二链合成后，使用表 S3 中概述的方案对 SPRI 进行 cDNA 纯化

使用 SureSelect XT HS 文库制备试剂盒中的试剂制备 SureSelect XT HS cDNA 文库

14. 使用用于 Illumina 配对末端多重测序试剂盒和方案 (G9702) 的 SureSelect XT HS 靶向序列捕获系统中详述的试剂和说明，执行 cDNA 的末端修复、dA 加尾和 XT HS 接头连接

注：从第 27 页的第 3 步开始遵循

SureSelect XT HS 方案，但是例外和修改概述于本方案的第 15–20 步中。

15. SureSelect XT HS 方案第 34 页的第 6 步。根据该方案中的说明进行文库的捕获前 PCR 扩增，改良的 PCR 循环次数如下所示：
 - a. UHRR, Seraseq™ v2, 用于 12 次 PCR 循环（高质量 RNA）
 - b. 乳腺 quad 样品，用于 14 次 PCR 循环（低质量 RNA）
16. 利用生物分析仪或 TapeStation 对捕获前文库进行评估，以评估文库分子大小的产量和分布

SureSelect XT HS cDNA 文库的靶向序列捕获和测序

17. SureSelect XT HS 方案第 46 页的第 1 步。使用 200 ng 捕获前文库作为起始材料进行靶向序列捕获，其中采用 SureSelect XT HS 快速杂交试剂。

利用 5 μ L SureSelect 人全外显子 V7 外显子组探针进行杂交

18. 根据该方案执行链霉亲和素磁珠捕获和后续清洗
19. 根据该方案针对所有样品使用 12 个 PCR 循环执行捕获后 PCR 扩增。利用生物分析仪或 TapeStation 对捕获后文库进行评估，以评估文库分子大小的产量和分布
20. 所有文库均在 Illumina HiSeq 4000 平台上以 2 \times 150 的读取长度进行测序

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

www.agilent.com

仅限研究使用。不可用于诊断目的。

本文中的信息、说明和指标如有变更，恕不另行通知。

PR7000-2381

© 安捷伦科技（中国）有限公司，2020

2020年1月29日，中国出版

5994-1644ZHCN

