

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

应用简报

作者

Jennifer Carter Jones、Alex Siebold
和 Anne Bergstrom Lucas

摘要

从福尔马林固定石蜡包埋 (FFPE) 组织中提取 RNA 并制备 RNA 测序 (RNA-Seq) 文库，可帮助研究人员鉴定和验证疾病发生、发展和治疗耐药性过程中的新生物标志物。然而，之前从 FFPE 样品中得到的 RNA 质量通常较差，限制了组织样品用作转录组学分析和二代测序 (NGS) 研究的来源¹。

近来，在进行 cDNA 文库的靶向序列捕获时，我们发现通过安捷伦链特异性 RNA 文库制备试剂盒改进的方案可获得核糖体污染极小的 RNA 测序数据以及良好的测序覆盖率。在本研究中，我们使用从肿瘤/正常组织的新鲜冷冻 (FF) 和 FFPE 样品中获得的 RNA 证实，方案无需进行预先核糖体清除或 poly(A) 选择，即可从 FFPE 样品中获得高质量的数据。当使用 SureSelect^{XT} RNA 直接捕获方案和 SureSelect^{XT} RNA 直接捕获试剂盒中的试剂时，我们发现在 FF 和 FFPE 样品中，转录本的上调和下调程度以及置信水平近似，说明 FFPE 组织块中储存的 RNA 在重要基因表达研究中的实用性。



Agilent Technologies

前言

在大多数情况下，福尔马林固定与石蜡包埋是很多基于显微镜检查的传统分析（如组织染色）中的标准样品保存方案。但是，通过固定引发的交联虽然对保持组织结构有利，但对保存组织内的核酸却并不理想。由 FFPE 获得的 RNA 质量参差不齐，具体取决于所保存的组织类型、保存方式、保存时间或保存良好程度以及 RNA 的提取方法。由 FFPE 获得的总 RNA 中通常会丢失 poly(A) 尾端，而 RNA 测序实验中用于分离 mRNA 的方法则通常会引入严重偏差。

为测试一种替代的 RNA 转录分析工作流程，分别从一对匹配的肿瘤/正常 FF 和 FFPE 样品中提取四份 RNA，并通过改进的 SureSelect RNA 富集方案对其进行处理。根据 FFPE 所得 RNA 的完整性对方案进行了适当改进，以创建用于靶向 RNA 测序的高质量文库。与标准工作流程相比，从 FFPE 中提取 RNA 的流程改变包括所使用的起始样品量、样品裂解所需的时间，以及调整 PCR 条件和 AMPure XP (Beckman Coulter Genomics) 磁珠清除步骤的方式，以实现整个工作流程中获得样品量的最大化。结果表明 FFPE 和 FF 样品的整体校准统计学结果十分相似，

另外通过比较肿瘤和正常的基因表达数据得出 FF 和 FFPE 外显子组靶向样品具有良好的相关性和一致性。

材料与方法

RNA 样品

分别从相应的正常和恶性结肠癌组织（来自 Asterand Technologies）中得到四份样本，对应新鲜冷冻 (FF) 和 FFPE 样品，并从中提取出总 RNA。我们针对四个不同的对应 RNA 样品类型，分别准备三个独立的技术重复样品：FF-正常、FF-肿瘤、FFPE-正常和 FFPE-肿瘤，共创建 12 个文库。

总 RNA 样品质量分析

多个 RNA 测序方案均推荐使用安捷伦生物分析仪或 Agilent TapeStation 生成 RNA 样品的 RNA 完整值 (RIN 或 RIN[®])，其中多数方案需要不小于 8.0 的高质量 RIN 值。但是，由 FFPE 组织提取的总 RNA 得到的 RIN 值非常低，

我们发现如此低的 RIN 值无法作为该新方案成功与否的最佳指标。我们发现，200 nt (DV_{200}) 以上的 RNA 片段百分比的计算值是 FFPE 样品质量更为可靠的指标²。当 FFPE 弥散条带中至少 20% 的片段大于 200 nt 时，推荐使用 200 ng 作为每次反应的总 RNA 起始量。使用由非 FFPE 获得的总 RNA 时，建议总 RNA 的起始量为 100 ng。下文的表 1 中列出了这些推荐值。

本研究采用了配备真核生物总 RNA Pico 芯片 (5067-1513) 的 Agilent 2100 生物分析仪，并根据生物分析仪曲线计算 RIN 和 DV_{200} 值。同时，还使用 NanoDrop 分光光度计分析 RNA 样品的质量，以获得 260/280 和 260/230 的吸光度比。理想情况下，这两个比值均应尽可能接近 2.0，并用 NanoDrop 的浓度值确定加入反应的总 RNA 的纳克数。

级别	%RNA > 200 nt	建议的起始量
高质量 FFPE RNA	> 50%	200 ng
低质量 FFPE RNA	20%-50%	200 ng
不适用的 FFPE RNA	< 20%	不建议进一步处理
完整 (非 FFPE) RNA	> 70%	100 ng

表 1. 根据大于 200 nt 的 RNA 片段比例建议的总 RNA 起始量。理想情况下，FFPE 样品应至少有 20% ($DV_{200} > 20\%$) 的片段大于 200 nt，才可进行 RNA 直接捕获工作流程

工作流程概述

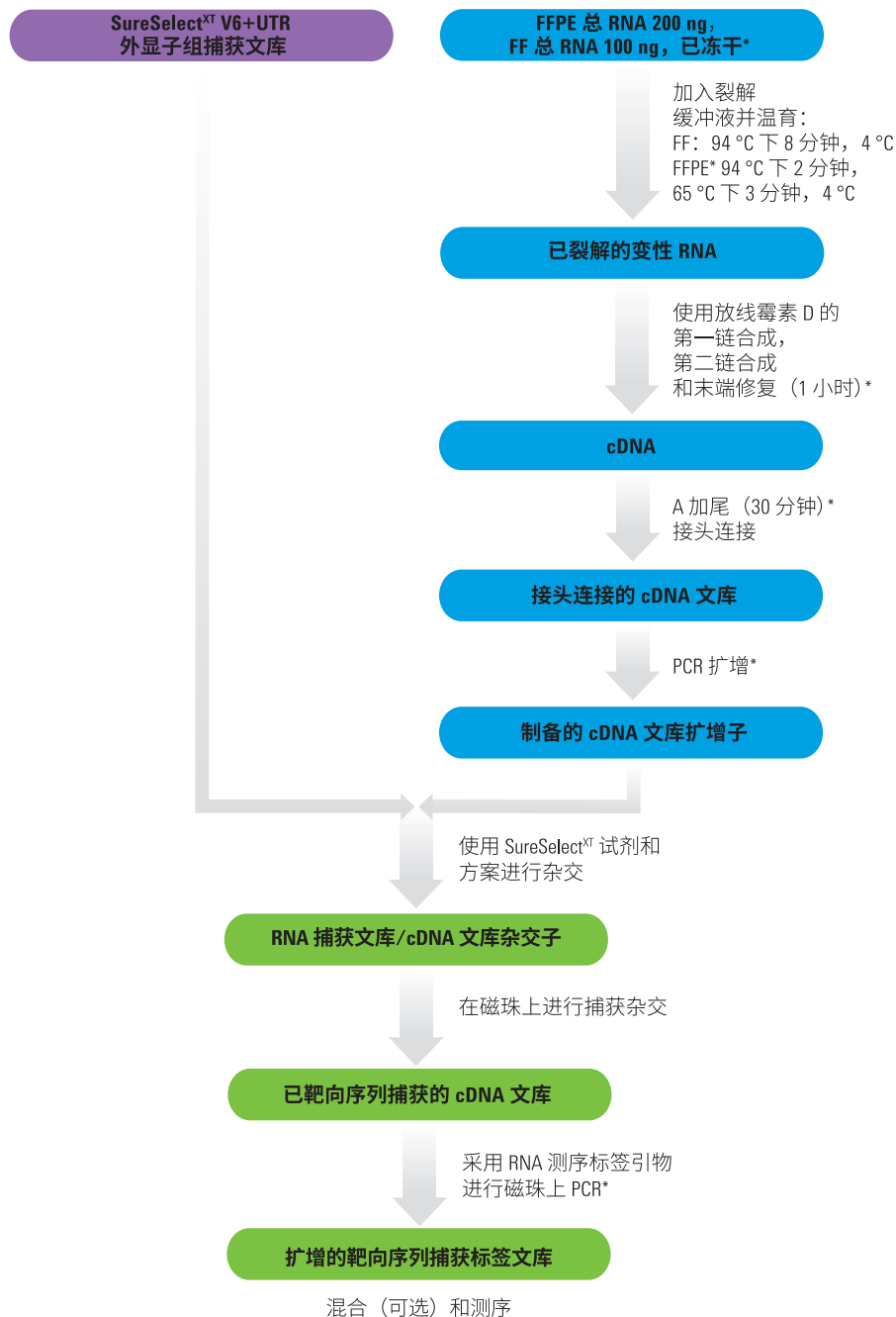


图 1. 新型 RNA 直接捕获方案省略了 Agilent SureSelect RNA 测序文库制备试剂盒中的初始 poly(A) 纯化步骤, 直接以总 RNA 开始。孵育和其他步骤的改进以星号标出

RNA 片段化

通常, 核糖体 RNA 清除或 poly(A) 分离是 RNA 测序文库制备的第一步。RNA 直接捕获方案中省略了这一步骤。在 RNA 直接捕获文库制备工作流程中, RNA 打断/引物退火是第一步。大多数情况下, FFPE 所得的 RNA 已经发生了一定程度的降解, 因此无需进行片段化, 或应运行一个简化的打断方案。下文的表 2 概述了可用于高质量 ($DV_{200} > 50\%$) 或低质量 ($DV_{200} > 20\%$) FFPE RNA 样品以及完整非 FFPE 样品的不同片段化/退火策略。

捕获前文库制备

我们使用 SureSelect^{XT} RNA 直接捕获试剂盒 (部件号 G7564A、G7564B) 中的试剂, 并在图 1 中以星号标注出该工作流程的变化。

步骤	温度	时间
高质量 FFPE RNA 样品*		
第 1 步	94 °C	3 分钟
第 2 步	65 °C	2 分钟
第 3 步	4 °C	保温
低质量 FFPE RNA 样品*		
第 1 步	65 °C	5 分钟
第 2 步	4 °C	保温
完整 (非 FFPE) RNA 样品*		
第 1 步	94 °C	8 分钟
第 2 步	4 °C	保温

表 2. 针对不同类型 FFPE 样品以及完整非 FFPE 所得总 RNA 的建议热循环仪温度。由于 FFPE 样品的差异较大, 每个温度下的孵育时间可能需要优化以获得最佳性能

对 SureSelect^{XT} RNA 靶向序列捕获方案中捕获前部分的改变，包括将第二链合成步骤的孵育时间延长至 60 分钟，以及将 A-加尾酶促反应的时间延长至 30 分钟。我们还将净化步骤中 AMPure XP 磁珠的比例从 1.2 倍体积增大至 1.8 倍体积。对接头连接的 cDNA 文库进行 PCR 扩增反应时，采用了 14 个循环。对于不同的 FFPE 总 RNA 样品类型，PCR 循环的次数可能需要增加或减少。

捕获和捕获后文库制备

RNA 直接捕获方案每次杂交反应都需要 200 ng 的 cDNA。每次反应中使用 5 μ L 的 91 Mb SureSelect 人全外显子组 V6+UTR (5190-8881) 捕获探针，并在 65 °C 下杂交 24 小时。RNA 直接捕获方案中另一个显著变化是使用 M-270 链霉亲和素磁珠（赛默飞世尔科技公司）替代 T1 链霉亲和素磁珠用于捕获杂交文库。在使用 M-270 磁珠捕获生物素化的文库时，我们可省略最终的洗脱步骤，并在仍与 M-270 磁珠（用不含核酸酶的水重悬）结合的情况下直接进行捕获文库的 PCR 扩增。采用 19 μ L（约为重悬 M-270 磁珠的一半）进行 12 个“磁珠上”PCR 反应循环，以对捕获文库进行扩增，并加上独特的索引标签。对采用另一半 M-270 链霉亲和素磁珠以得到测序步骤需要的 10 nM 左右浓

度的最终文库而言，不同 FFPE 的总 RNA 样品类型的 PCR 循环次数可能需要提高。

二代测序和数据处理

每个重复样品在捕获后经过独特的标记和混合，并通过 Illumina HiSeq 2500 High Output 测序仪（版本 4，采用 100 bp 配对末端化学测序法）在两个泳道中进行测序。采用 bcl2fastq v1.8.2 对 HiSeq 运行的输出进行多重性分解，生成的 fastq 文件通过 TopHat 2.0 修剪并与 hg19 对齐。我们通过使用博德研究所科学家开发的 RNA-SeQC 工具分析 TopHat BAM 文件，以此得到 RNA 测序的质量控制指标⁴。将每个样品的 RNA 测序数据归一化为两千万读出序列，用于下游基因表达的比较。

差异基因表达分析

RNA-SeQC 工具可用于生成一个含有基因表达指标的文本文件，其中表达指标由映射到 Ensembl 转录本的每百万个映射读出序列的转录本每千个碱基的读出序列 (RPKM) 表示。RPKM 指标可用于以基因长度和样品中映射读出序列的总数量对某个转录本的读出序列计数进行归一化，RPKM 测量结果还可说明样品和基因间差异的来源。我们采用 Strand NGS 软件⁵ 对 RPKM 基因表达数据进行分析与可视化。

结果

总 RNA 质量分析

我们使用 RNA 直接捕获方案对从恶性结肠癌和相邻正常组织“匹配的四份”对应的新鲜冷冻 (FF) 和 FFPE 样品中分离出的总 RNA 制备 RNA 测序文库。下文的图 2 中展示了生物分析仪真核生物总 RNA Pico 芯片得到的电泳图叠加迹线。针对这四种样品，分别计算了超过 200 nt 或 DV₂₀₀ 值的片段比例以及 RNA 完整值 (RIN)。FF 样品中可见 18s 和 28s 峰，而 FFPE 样品峰更为平滑，片段化更高，与预期相同。另外，使用 NanoDrop 分光光度计对 RNA 样品的质量进行了进一步测定，获得 260/280 和 260/230 的吸光度比。表 3 中总结了 NanoDrop 和生物分析仪得到的结果。

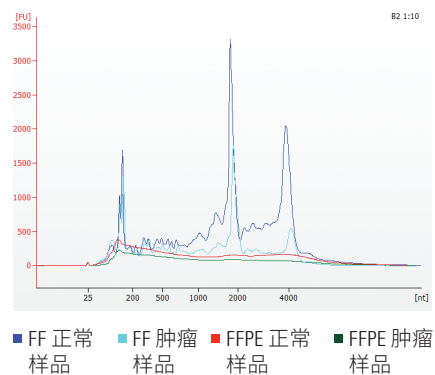


图 2. 安捷伦生物分析仪 RNA Pico 电泳图，展示了 FF 和 FFPE 所得总 RNA 样品的显著差别。不同颜色的线代表不同的 RNA 样品类型。200 nt 标记右侧的峰/弥散条带代表 > 200 nt 的片段

捕获前样品前处理结果

由于 FFPE 和 FF 总 RNA 样品的 DV₂₀₀ 值均大于 70%，因此我们冻干了 200 ng 的 FFPE 总 RNA 和 100 ng 的完整 FF 总 RNA。将冻干样品加入 19 μL 裂解缓冲液中进行重悬。根据表 1 中描述的裂解/退火反应条件，将 FFPE RNA 鉴定为“高质量 FFPE RNA”，将 FF RNA 鉴定为“完整（非 FFPE）”，并分别采用表 1 中推荐的孵育条件。根据 RNA 直接捕获方案的改进（包括更长的第二链合成和 A-加尾步骤），使用片段化/变性的 RNA 以及退火引物制备文库。清除步骤中加入的 AMPure XP 磁珠比例增大至 1.8 倍的反应体积。捕获前扩增步骤进行 14 次循环，然后在 Agilent TapeStation 4200 的 D1000 胶条上运行不同量的捕获前 cDNA 文库。图 3 展示了一个 FFPE 样品所得的电泳谱图迹线示例。采用 RNA 直接捕获方案，12 个样品中均可获得用于下一步靶向捕获的足够 cDNA，如图 4 所示。

样品	A260/A280	RIN	DV ₂₀₀
FF 正常样品	1.96	6.7	85%
FF 肿瘤样品	2.10	6.1	74%
FFPE 正常样品	1.93	2.3	75%
FFPE 肿瘤样品	1.93	2.3	74%

表 3. FF 和 FFPE 总 RNA 样品的 RNA 质量指标

捕获后样品前处理结果

与 SureSelect 人全外显子组 V6+UTR 捕获探针在 65 °C 下杂交 24 小时后，采用 M-270 链霉亲和素磁珠捕获杂交文库。标准的靶向 RNA 测序方案中的最终洗脱步骤可省略，并将 M-270 磁珠在 40 μL 不含核酸酶的水中重悬。

采用与 M-270 磁珠结合的 19 μL 文库对文库进行扩增与分别标记。这一最终扩增步骤中，采用 12 次 PCR 循环。使用 AMPure XP 磁珠对标记和扩增后的文库进行纯化，并采用 HSD1000 胶条在 Agilent TapeStation 4200 上对最终的 cDNA 捕获后文库进行测量。图 3 展示了一个 FFPE 样品所得的电泳谱图迹线示例。

采用 FFPE 所得样品获得的代表性 Agilent TapeStation 4200 电泳图：

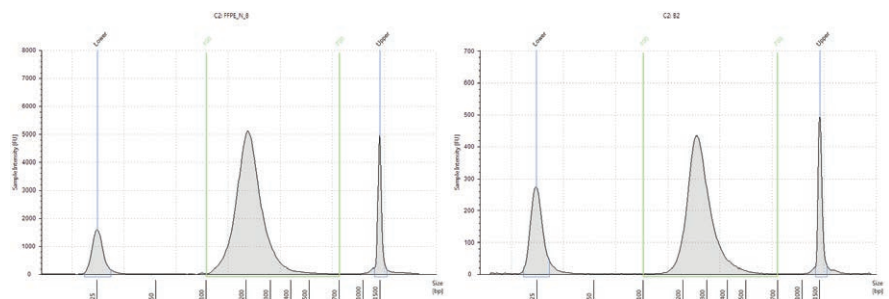


图 3. 分别采用 D1000 预制胶条和 HSD1000 预制胶条进行捕获前 PCR 分析（左侧）和捕获后 PCR 样品分析（右侧）的 Agilent 4200 TapeStation 迹线。两条迹线均来自 FFPE_N_B02 技术重复样品，但可代表所有样品

捕获前 PCR 产率：肿瘤与正常，FF 与 FFPE（捕获需要的 200 ng）

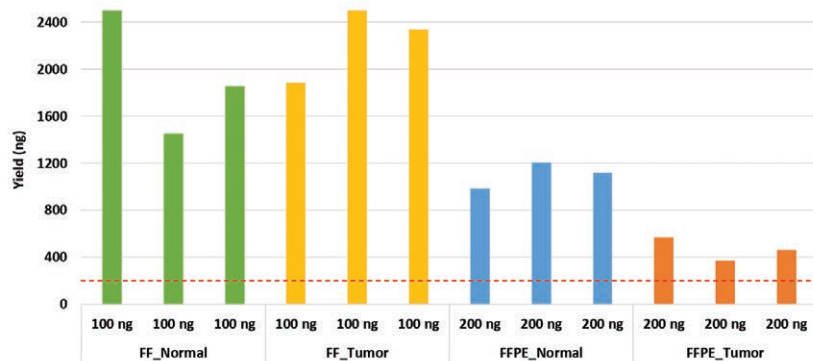


图 4. 对 FF 样品和 FFPE 样品分别使用 100 ng 和 200 ng 的总 RNA 起始量时，技术重复样品之间的捕获前文库产率对比。虽然不同技术重复样品之间产率有一定的差别，但 FF 样品比相应 FFPE 样品所得的产率更高，即使样品前处理时总 RNA 的起始量仅为后者的一半。图中的红线指示了继续进行杂交捕获步骤所需的最小产率 (200 ng)

图 5 中展示了 12 个样品的捕获后产率比较。捕获前和捕获后产率以及实验中关键条件的总结均列于表 4 中。

二代测序、数据处理和 RNA 测序 QC 指标

12 个技术样品在捕获后经过分别的标记和混合，并通过 Illumina HiSeq 2500 (100 bp 配对末端测序) 在两个泳道中进行测序，使每个样品约产生八千万个配对末端读出序列。将所有 12 个文库的读出序列数目归一化为每个文库两千万个读出序列，以便进行比较。一般而言，与 FF 样品相比，在 FFPE 样品的预期链中观察到的读出序列更少，如图 6 所示。所有样品在 SureSelect 人全外显子组 V6+UTR 捕获探针靶向区域中的读出序列数量十分相似 (黄色柱状, 约 87%)，平均有 95% 的读出序列处于靶向区域的 100 bp 内 (数据未显示)。FF 样品的重复读出序列的平均比

例为 17%，FFPE 样品的这一比例为 24% (绿色柱状)。出现这一较高的重复结果比例并不意外，主要是由于 FFPE 总 RNA 样品的裂解量较高。本实验中的核糖体 RNA 读出序列比例极低，占到总读出序列的 0.3%-1.3%，图 6 中将这一比例 rRNA 的匹配读出序列值乘以 10 倍 (红色柱状)，以实现更好的可视化。

匹配的读出率

我们比较了 FFPE 和 FF 样品间人转录组学内容的平均匹配统计结果。基因间读出率指的是匹配到基因间的读出序列比例，外显子读出率指的是匹配到外显子的读出序列比例，内含子读出率指的是匹配到内含子的读出序列比例。图 7 展示了所有文库的非常相似的外显子 (约 90%)、内含子 (约 9%) 和基因间 (约 1%) 匹配率。内含子读出序列数目反映了 SureSelect 人全外显子组 V6+UTR 文库中存在的内含子探针。

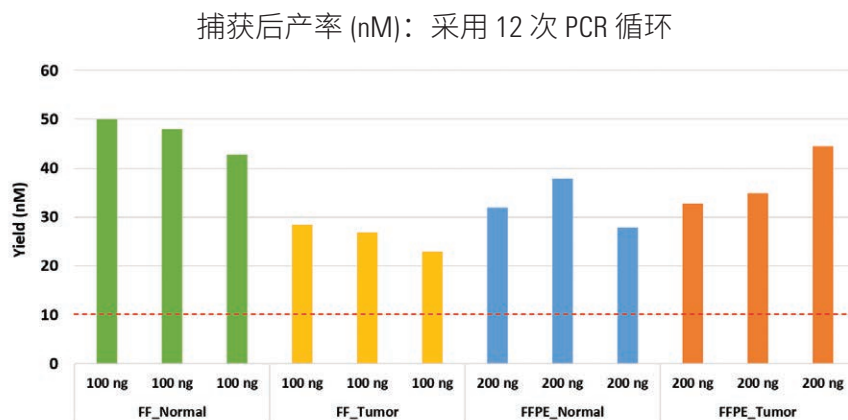


图 5. 使用 200 ng cDNA 起始量进行杂交反应时，不同技术重复样品间的捕获后文库产率比较。图中的红线指示了合并标记样品以进行测序所需的最小浓度 (10 nM)

方案改进和 PCR 产率的快速指南总结：

样品	RIN	DV ₂₀₀	总 RNA 起始量 (ng)	片段化时间/温度	捕获前 PCR 循环次数	重复样品捕获前产率 (µg)	捕获后 PCR 循环次数	重复样品捕获后产率 (nM)
FF 正常样品	6.7	85%	100	94 °C 下 8 分钟	14	2.8、1.5、1.9	12	50、48、43
FF 肿瘤样品	6.1	74%	100	94 °C 下 8 分钟	14	1.9、2.5、2.3	12	29、27、23
FFPE 正常样品	2.3	75%	200	94 °C 下 3 分钟 65 °C 下 2 分钟	14	1.0、1.2、1.1	12	32、38、28
FFPE 肿瘤样品	2.3	74%	200	94 °C 下 3 分钟 65 °C 下 2 分钟	14	0.6、0.4、0.5	12	33、35、45

表 4. 总 RNA 质量指标、方案改进以及捕获前和捕获后 PCR 产率的总结

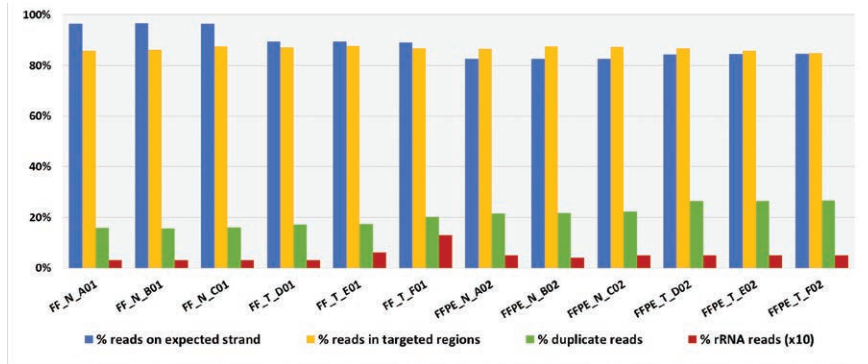


图 6. 所制备和测序的 FF 和 FFPE 文库的链特异性和靶向序列捕获效率

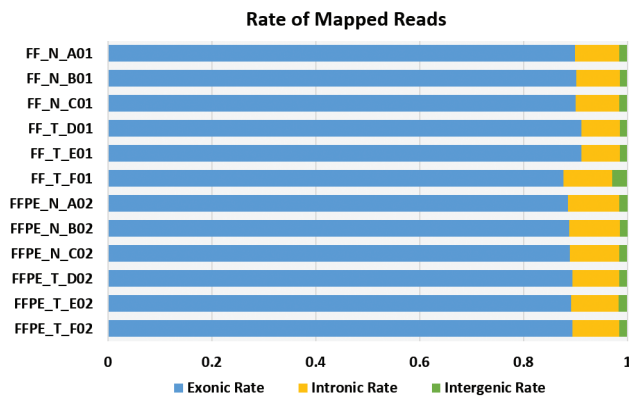


图 7. 外显子、内含子和基因间的匹配读出率

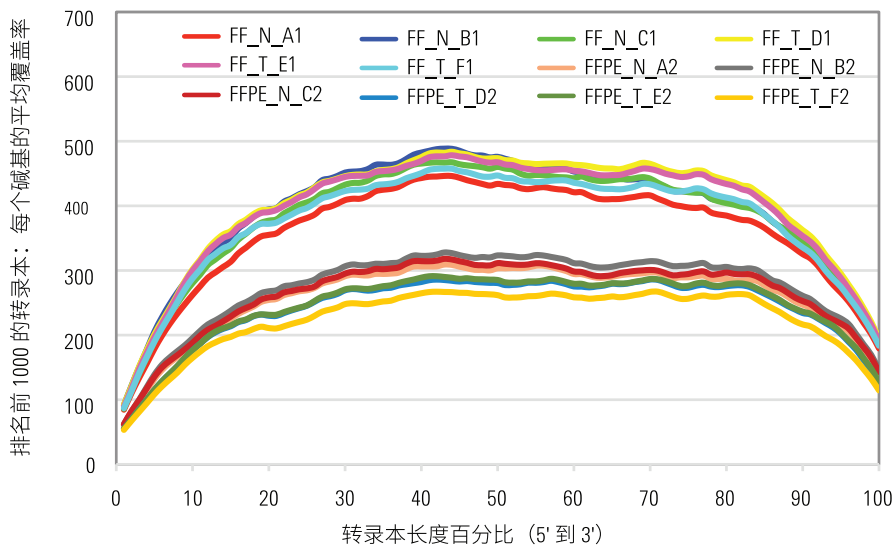


图 8. 前 1000 个表达基因 5' 到 3' 端转录本的覆盖率。线的颜色与图中样品名称的颜色相对应

转录本 5' 和 3' 端的覆盖率

基因 5' 和 3' 端的覆盖率对正确鉴定转录本十分重要。我们测定了已知转录本从 5' 到 3' 端每个长度百分比的平均覆盖率，以确定在 FF 和 FFPE 文库间是否存在偏差差异。FF 文库中前 1000 个表达基因的平均覆盖率显著高于 FFPE 文库的相应值，这是由于前者的文库复杂度更高（重复率更低）。尽管存在文库复杂度的差异，图 8 仍展示出 FF 和 FFPE 样品在转录本整个长度中的覆盖率十分接近。

技术重复样品间表现出的良好相关性

我们发现，除 FF 肿瘤样品之外，在使用 Strand NGS 程序中的主成分分析对 RPKM 读出序列进行分析时，每个样品的三个技术重复样品均可紧密聚类，如图 9 所示。此分析中鉴定的前两个主成分分别是组织生物学类型（肿瘤/正常，占 x 轴中 39% 的差异）以及组织保存方法（占 y 轴中 29% 的差异）。

使用 Strand NGS 程序对 RPKM 读出序列进行分层聚类分析（欧式相似度测量/Wards 关联规则）时，我们发现样品可根据样品类型得到主层聚类，其中正常样品与肿瘤样品分别聚类到不同的分支，而与 FFPE 或 FF 保存条件无关，如图 10 所示。样品类型后

的第二层聚类为保存条件，其中 FF 和 FFPE 样品处于不同的分支中。分层聚类图还表明，虽然许多转录本在不同样品类型和保存条件间有相似的表达水平，但仍存在与正常或肿瘤样品类型相关的显著差异。

FF 和 FFPE 样品相似的基因表达结果

选取所有重复样品中至少有一部分映射读出序列 (RPKM > 0) 的 75263 个 Ensembl 转录本，发现肿瘤和正常样品的 \log_2 比值有极高的相关性。倍数变化（肿瘤与正常）的显著性如图 11 中的火山图所示，其中 x 轴为倍数变化的 \log_2 值，y 轴为经负 \log_{10} 校正后的 p 值。显著表达（经 Benjamini Hochberg 多重测试校正后的 p 值 < 0.05，绿色水平线）且表达呈两倍变化（绿色垂直先）的转录本以红色标出，每种保存条件下显著表达的转录本数目也在图 11 中以红色标出。

为进一步展示 FF 和 FFPE 保存的总 RNA 样品在基因表达比值一致性方面的关联，将图 11 两张火山图中 x 轴表示的肿瘤/正常倍数变化的 \log_2 值重新在图 12 中作图并进行比较。比较两种条件下均表达的 75236 个 Ensembl 转录本的基因表达 \log_2 比值发现，两者之间有绝佳的一致性（斜率为 0.93）和相关性（拟合曲线的 $r^2 = 0.70$ ），如图 12 所示。

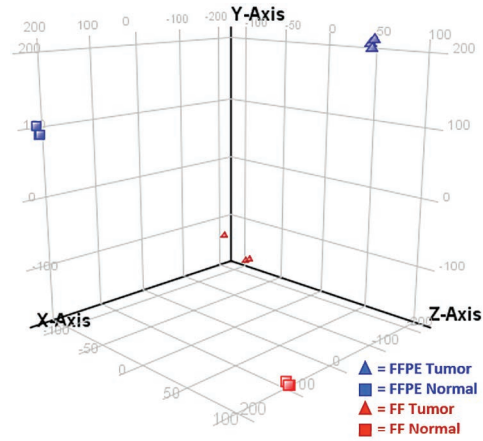


图 9. 主成分分析，说明技术重复样品间的相关性很强，可通过细胞生物学（肿瘤/正常状态）和组织保存方法（FFPE/FF）实现清晰的样品分离

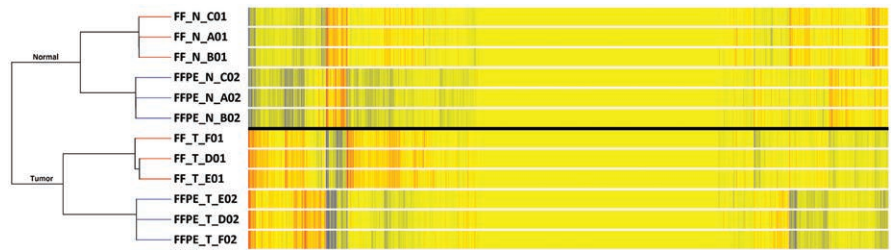


图 10. RPKM 表达的分层聚类分析结果，说明聚类的第一层与样品类型有关而非保存条件时，样品间的生物学差异（肿瘤与正常）大于 FFPE 和 FF 组织保存方法间的差异

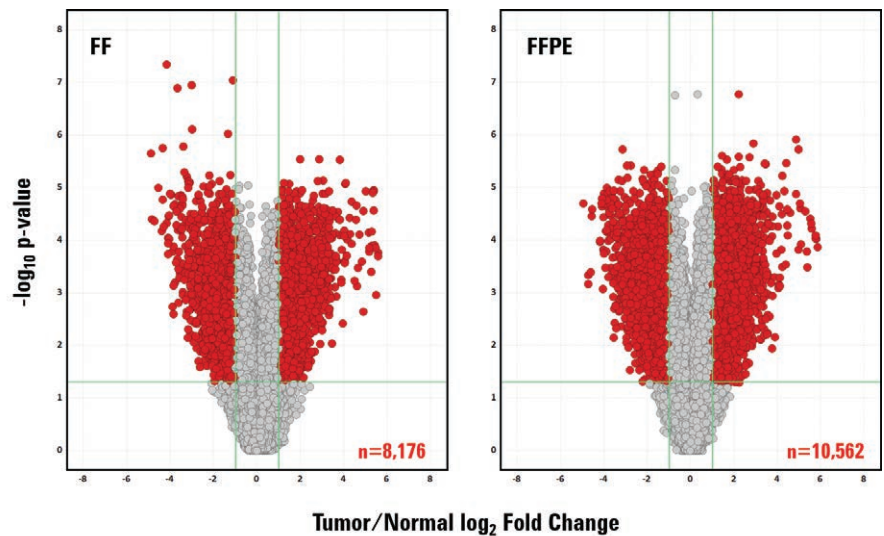


图 11. 显著倍数变化的火山图分析，其中 x 轴表示肿瘤与正常的 \log_2 倍数变化之比，y 轴表示倍数变化的显著性。红色的点表示表达 > 2 倍且校正 p 值 > 0.05 的基因。图中的绿色垂直线代表两倍的表达变化，绿色水平线则表示 p 值的 0.05 截止限

结论

本应用简报展示了 RNA 直接捕获测序方案是一种对 FFPE 样品所得的总 RNA 测序的优化方法，提供的靶向捕获基因表达数据与匹配的来自新鲜冷冻组织的 RNA 的数据具有高度一致性。此外，该工作流程借助杂交捕获进行转录组分析，无需在文库制备初始进行 poly(A) 选择或核糖体 RNA 清除。这一改进后的工作流程还可在样品仍与 M-270 链霉亲和素磁珠结合时直接进行捕获后扩增反应（“磁珠上” PCR），因而可减少额外的处理步骤，避免降低产率以及测序过程中产生非预期偏差的风险。

我们证实，从 FFPE 和 FF 样品中提取的总 RNA 除了可产生整个转录本相等的测序覆盖率之外，还具有相似的在靶百分比和读出序列匹配统计结果。虽然 FFPE 产生文库的复杂度比 FF 样品有所降低（前者 PCR 重复率更高），但是两者最终的上调和下调基因表达测量结果却十分相近。PCA 和分层聚类的结果说明了样品的主要区分因素依然是生物学差异，与保存方法（FF 与 FFPE）相比，样品来源（肿瘤与正常）对表达数据的影响更大。

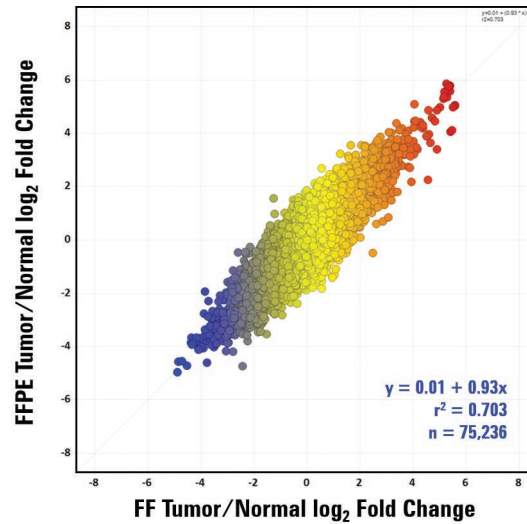


图 12. FF 和 FFPE 总 RNA 样品的基因表达倍数变化比值之间的相关性

综上所述，我们的数据证实了安捷伦的简化 RNA 直接捕获 RNA 测序工作流程的确可用于 FFPE 所得 RNA 的文库制备，从该文库中可获取具有生物学意义的测序结果。

参考文献

1. Cieslik M, Chugh R, Wu YM, Wu M, Brennan C, Lonigro R, Su F, Wang R, Siddiqui J, Mehra R, Cao X, Lucas D, Chinnaiyan AM, and Robinson D “The use of exome capture RNA-Seq for highly degraded RNA with application to clinical cancer sequencing” *Genome Research* **25**, 1372-1381 (2015)
2. “Evaluating RNA Quality from FFPE Samples” Illumina publication number 470-2014-001
3. “SureSelect^{XT} RNA Target Enrichment for Illumina Multiplexed Sequencing Strand Specific Library Prep and Target Enrichment Protocols” (用于 Illumina 多重测序链特异性文库制备和靶向序列捕获方案的 SureSelect^{XT} RNA 靶向序列捕获)，安捷伦出版号 G9691-90000
4. Deluca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G “RNA SeQC: RNA-seq metrics for quality control and process optimization” *Bioinformatics* **28**, 1530-1532 (2012)
5. Strand NGS software, Version 2.8, Build 230243. © Strand Life Sciences, Bangalore, India

可靠结果，完整方案

PR7000-0679
© 安捷伦科技（中国）有限公司，2017
2017年5月24日，中国出版
5991-8119CHCN

如需了解更多信息，请访问：

www.agilent.com/genomics

查找当地的安捷伦客户中心：

www.agilent.com/genomics/contactus

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

仅限研究使用。不可用于诊断目的。



Agilent Technologies