TruSeq™ RNA Exome

A reproducible, economical solution for analyzing RNA isolated from formalin-fixed, paraffin-embedded (FFPE) tissues and other low-quality samples.

Highlights

- **High-Quality Data from Difficult Samples**
  Assesses degraded samples, including FFPE tissues, for RNA sequencing
- **Exceptional Coverage with Focused Content**
  Maximizes discovery power at reduced sequencing depth by targeting transcriptome coding regions
- **Low Sample Input**
  Maintains high data quality from as little as 10 ng total RNA
- **Integrated, Flexible Workflow Solution**
  Comprehensive workflow streamlines RNA exome capture sequencing and supports single-plexing or multiplexing up to 4-plex

Introduction

Millions of FFPE archival tissue samples provide an enormous and invaluable repository of information for disease research, especially cancer. Typically, these samples are associated with long-term phenotypic data that can yield insight into gene expression changes that occur during various disease states. Unfortunately, the fixation process and storage of FFPE samples can lead to high RNA degradation, making it difficult to perform reliable, reproducible gene expression profiling studies with RNA sequencing (RNA-Seq).[1,2] It is possible to extract usable RNA from FFPE samples, but current analysis methods produce highly variable results or require expensive deep sequencing. This generates significantly different views of the transcriptome, reducing the reliability of the data and increasing budget requirements.

High Quality Data from Difficult Samples

To overcome these challenges and make it easier to access the valuable information in FFPE and other low-quality samples, Illumina offers TruSeq RNA Exome, previously sold as the TruSeq RNA Access Library Prep Kit (Figure 1). This workflow enables researchers to apply the power of next-generation sequencing (NGS) technology to gene expression studies involving RNA isolated from low-quality samples. By focusing on the coding regions of RNA, TruSeq RNA Exome requires less input RNA and fewer reads, increasing the number of samples per run for more cost-effective transcriptome analysis.

Exceptional Coverage

TruSeq RNA Exome features a highly optimized probe set that delivers comprehensive coverage of coding RNA sequences. TruSeq RNA Exome includes > 425,000 probes, each constructed against the NCBI37/hg19 reference genome, covering 98.3% of the RefSeq exome. The probe set was designed to capture > 210,000 targets, spanning 21,415 genes of interest (Table 1).

Table 1: TruSeq RNA Exome Coverage Details

Coverage Specification	TruSeq RNA Exome
No. of Target Genes	21,415
No. of Target Exonic Regions	214,126
No. of Probes	425,437
RefSeq Exome Percent Covered	98.3%

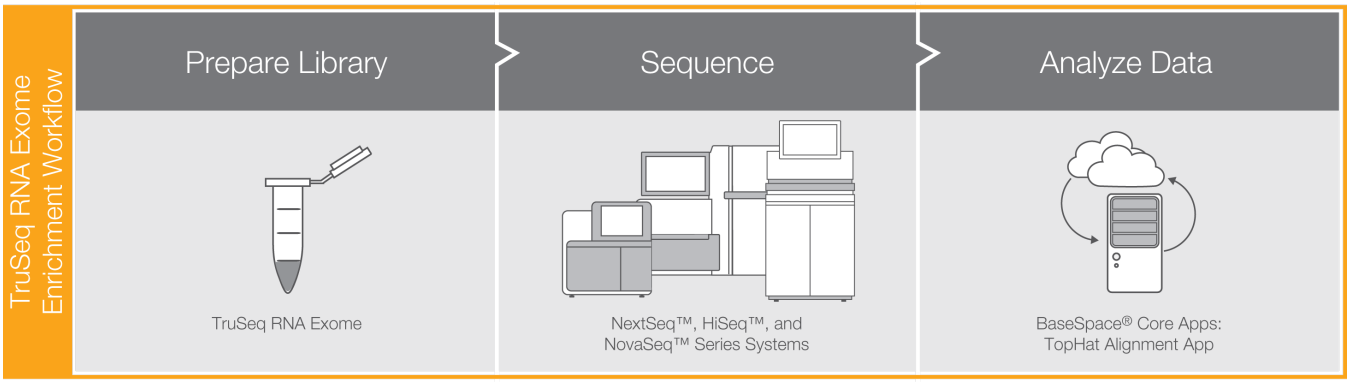

Figure 1: TruSeq RNA Exome Workflow—TruSeq RNA Exome is part of an integrated NGS solution that includes simplified library preparation and coding transcriptome capture, sequencing, and user-friendly data analysis.

Focused Content

TruSeq RNA Exome provides high capture efficiencies that focus sequencing efforts on the high value content of RNA coding regions. To demonstrate this, libraries were prepared from FFPE lung tumor and normal samples using TruSeq Stranded Total RNA and TruSeq RNA Exome. Sequencing and analysis with the BaseSpace TopHat Alignment App revealed that TruSeq RNA Exome resulted in > 85% of the bases covered aligning to coding sequence and untranslated regions (UTR) of RNA, compared to < 40% with TruSeq Stranded Total RNA (Figure 3). By working with more focused content, TruSeq RNA Exome produces smaller data sets that enable faster data analysis and easier data handling.

Low Sample Input

High capture efficiency and coverage uniformity minimize the required sequencing depth to determine expression levels accurately and without bias. Starting with as little as 10 ng total RNA, it is possible to achieve the sequencing depth needed for accurate quantitation and detection of transcripts and gene fusions. This low input requirement makes TruSeq RNA Exome an ideal solution for precious samples with limited starting material.

Simple, Scalable Workflow

TruSeq RNA Exome is designed and fully optimized for flexibility in multiplexing needs, providing a simple, scalable solution that is part of the Illumina integrated NGS workflow that includes library preparation, sequencing, and data analysis (Figure 1).

Streamlined Library Prep

Stranded RNA-Seq libraries are prepared using accurate, proven TruSeq chemistry. This method adds unique oligonucleotides to each library, tagging them for downstream pooling into one lane (Figure 2A). This multisample pooling step allows more samples to be loaded in a single sequencing run, making high-throughput studies feasible. After libraries are pooled, they undergo a capture step that produces a targeted library, depleted of ribosomal RNA and intronic or intergenic regions. Pooled libraries are hybridized to biotin-labeled probes specific for coding RNA regions (Figure 2B). Specific targets within the pool are then captured by adding streptavidin beads that bind to the biotinylated probes (Figure 2C). Magnets pull the bound RNA fragments from the solution (Figure 2D). Captured RNA fragments are eluted from the beads and hybridized for a second enrichment reaction. After amplification, a targeted library is ready for cluster generation and subsequent sequencing. Reagents are supplied in quantities sufficient to provide flexibility in sample plexing from single-plex sequencing to 4-plex multiplexing. Master-mixed reagents provide a quick start and make the process automation-friendly.

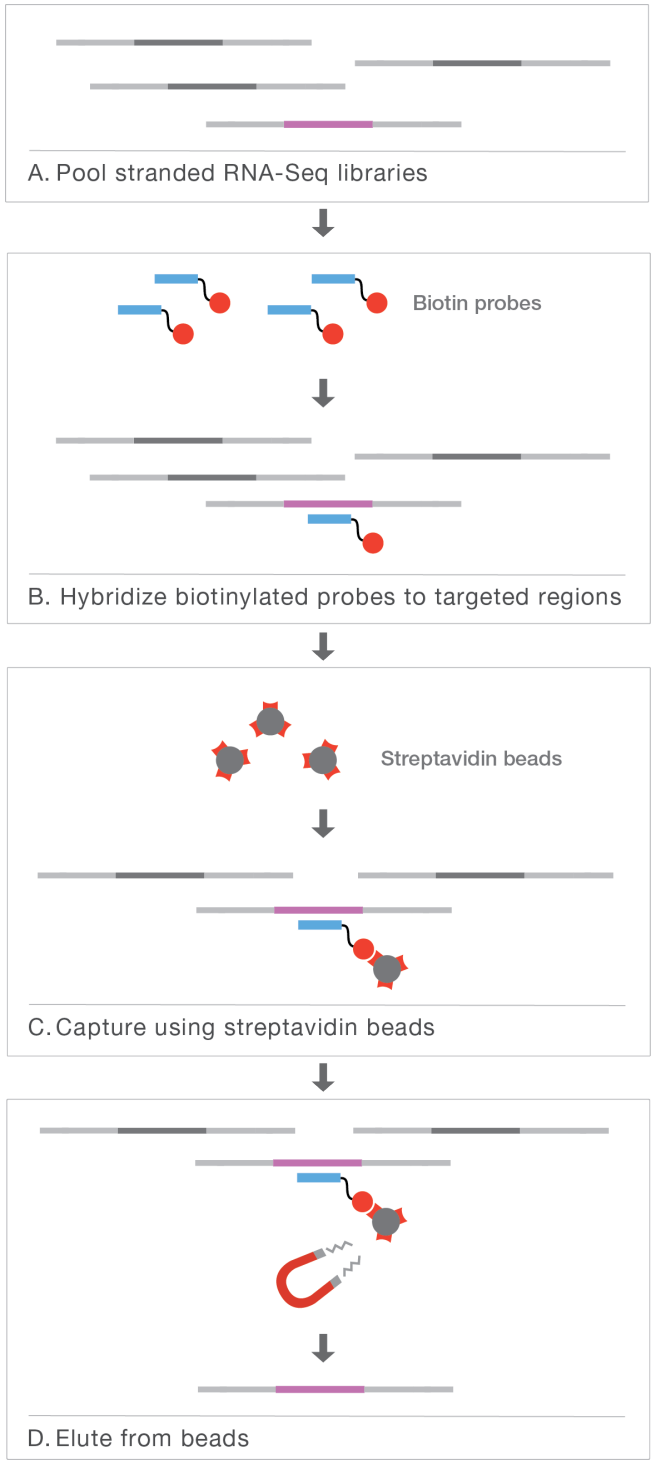

Figure 2: TruSeq RNA Exome Capture Chemistry—TruSeq RNA Exome provides a simple and streamlined protocol for isolating targeted regions of interest from samples.

Figure 3: Focus on RNA Coding Regions with TruSeq RNA Exome—FFPE lung tumor and normal samples were prepared with both TruSeq Stranded Total RNA and TruSeq RNA Exome. Libraries were sequenced at 200 M and 25 M reads, respectively. (A) Samples prepared with TruSeq RNA Exome show much deeper coverage of the exons, even at 1/8 the number of reads. The TruSeq Stranded Total RNA data were down-sampled to 30 M reads for comparison. (B) Using the BaseSpace TopHat Alignment App, more than 85% of the data generated using TruSeq RNA Exome aligned to transcripts (coding and UTRs).

Cost-Effective Sequencing

By focusing on the coding regions of RNA and combining TruSeq RNA Exome with high-throughput instruments like the NextSeq, HiSeq, and NovaSeq Series of systems, laboratories can sequence 5× more samples per run without sacrificing data quality (Figures 3 and 4). TruSeq RNA Exome produces highly accurate information that increases the percentage of usable exonic reads in the assembly of the coding regions of highly fragmented RNA. Reads are focused on the regions of interest, effectively extending read budgets (Table 2) without sacrificing gene fusion discovery power (Figure 5).

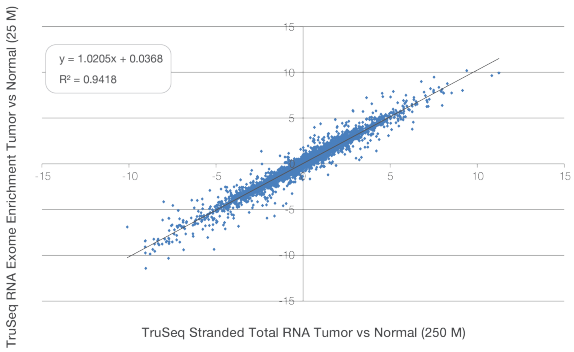

Figure 4: Accurate Data with a Fraction of the Reads—Libraries were prepared with TruSeq RNA Exome (25 M reads) and TruSeq Stranded Total RNA (250 M reads) from lung tumor and normal FFPE samples. Differential expression analysis reveals that $\log_2$ fold-change values correlate highly across the full dynamic range ($R^2 = 0.9418$).

Table 2: TruSeq RNA Exome Extends Read Budget

Sequencing System	Fresh/Frozen or FFPE RNA-Seq[a]
MiSeq™ System v3 Chemistry	1 sample per run
NextSeq 500 System Mid-Output Flow Cell	5 samples per run
NextSeq 500 System High-Output Flow Cell	16 samples per run
HiSeq 2500 System Rapid-Run Mode	24 samples per run
HiSeq 2500 System High-Output Mode	160 samples per run
NovaSeq 6000 System S2 Flow Cell	132 samples per run

a. Sequenced at 25 M reads per sample (2 × 75 bp).

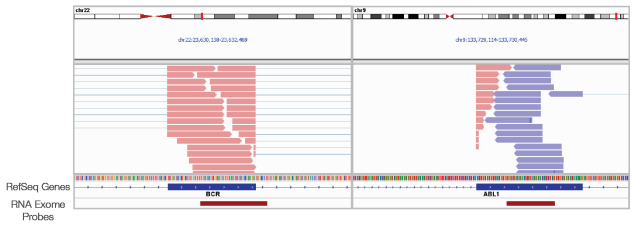

Figure 5: Efficient Gene Fusion Discovery—TruSeq RNA Exome enables detection of expressed fusion transcripts without the need to design probes specific for the fusion junction. The well-characterized BCR-ABL fusion is detected efficiently in the Universal Human Reference RNA (UHRR) sample at 25 M reads.

Convenient, Easy Data Analysis

TruSeq RNA Exome data sets can be analyzed using RNA-Seq Software Apps in BaseSpace Sequence Hub. These apps provide expert-preferred data analysis tools with intuitive, click-and-go user interfaces designed for informatics novices. TopHat 2 enables high-confidence alignment for abundance measurement as well as the detection of splice junctions, gene fusions, and SNPs. CuffDiff enables sensitive transcript discovery and differential expression analysis. TopHat Fusion delivers robust, high-confidence detection of gene fusions, while the Illumina Isaac™ pipeline delivers reliable variant calling.[3] Output files can be used in a broad range of secondary analysis solutions. RNA-Seq Apps enable easy aggregation of multi-sample reports, notification of job completion on mobile devices, and efficient file organization for collaboration and sharing.

Summary

FFPE samples offer a wealth of information that has been difficult to access historically. As part of an integrated Illumina sequencing solution, TruSeq RNA Exome offers a reproducible, economical method for sequencing RNA from FFPE and other low-quality samples.

Learn More

To learn more about RNA exome capture sequencing, visit www.illumina.com/techniques/sequencing/rna-sequencing/rna-exome-capture-sequencing.html.

Ordering Information

Product	Catalog No.
TruSeq RNA Library Prep for Enrichment (48 samples)	20020189
TruSeq RNA Enrichment (up to 48 samples at 4-plex, 12 enrichments)	20020490
TruSeq RNA Single Indexes Set A (12 indexes, 48 samples)	20020492
TruSeq RNA Single Indexes Set B (12 indexes, 48 samples)	20020493
Exome Panel (45 Mb)	20020183

References

1. von Ahlfen S, Missel A, Bendrat K, Schlimpberger M. Determinants of RNA quality from FFPE samples. *PLoS ONE*. 2007;2(12):e1261.
2. Penland SK, Keku TO, Torrice C, et al. RNA expression analysis of formalin-fixed paraffin-embedded tumors. *Lab Invest*. 2007;794:383–391.
3. Raczy C, Petrovski R, Saunders CT, et al. Isaac: ultrafast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*. 2013;29:2041–2043.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

illumina®