TruSeq® RNA Access Library Prep Kit

A reproducible, economical solution for analyzing RNA isolated from FFPE tissues and other low-quality samples.

Highlights

- **Focused for Affordability**
  Isolating human transcriptome coding regions maximizes discovery power at a fraction of the sequencing depth
- **High-Quality Data from Difficult Samples**
  Optimized for sequencing RNA from degraded samples, including formalin-fixed, paraffin-embedded tissues
- **Ideal for Samples with Limited Starting Material**
  Greatly reduced sample input requirements (as little as 10 ng total RNA) while maintaining high sensitivity

Introduction

Millions of formalin-fixed paraffin-embedded (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 may yield insight in gene expression changes that occur during various disease states. Unfortunately, the fixation process and storage of FFPE samples frequently leads to high RNA degradation, making it difficult to perform reliable, reproducible gene expression profiling studies. 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.

To overcome these challenges and make it easier to take advantage of the wealth of information in FFPE and other low-quality samples, Illumina offers the TruSeq RNA Access Library Prep Kit. This kit 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 region of RNA, the TruSeq RNA Access Kit requires less input RNA and fewer reads, increasing the number of samples per run for more cost-effective transcriptome analysis.

Focused Content

High capture efficiencies focus sequencing efforts on the content with the highest value. Sequencing of paired FFPE lung tumor and normal samples prepared using TruSeq RNA Access revealed that > 85% of the bases covered were within the coding and UTR regions of RNA (Figure 1).

Simple, Scalable Workflow

The TruSeq RNA Access Kit is designed and fully optimized for multiplexing capabilities, providing a simple, scalable workflow that is part of the Illumina integrated library preparation sequencing solution (Figure 2). Master-mixed reagents provide a quick start and make the process automation-friendly.

Starting from as little as 10 ng total RNA, 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 3A). This multi-sample 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 and intergenic regions. Pooled libraries are hybridized to biotin-labeled probes specific for coding RNA regions (Figure 3B). Specific targets within the pool are then captured by adding streptavidin beads that bind to the biotinylated probes (Figure 3C). Magnets pull the bound RNA fragments from the solution (Figure 3D). 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.

Superior Coverage

TruSeq RNA Access features a highly optimized probe set that delivers comprehensive coverage of coding RNA sequences. The kit 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).

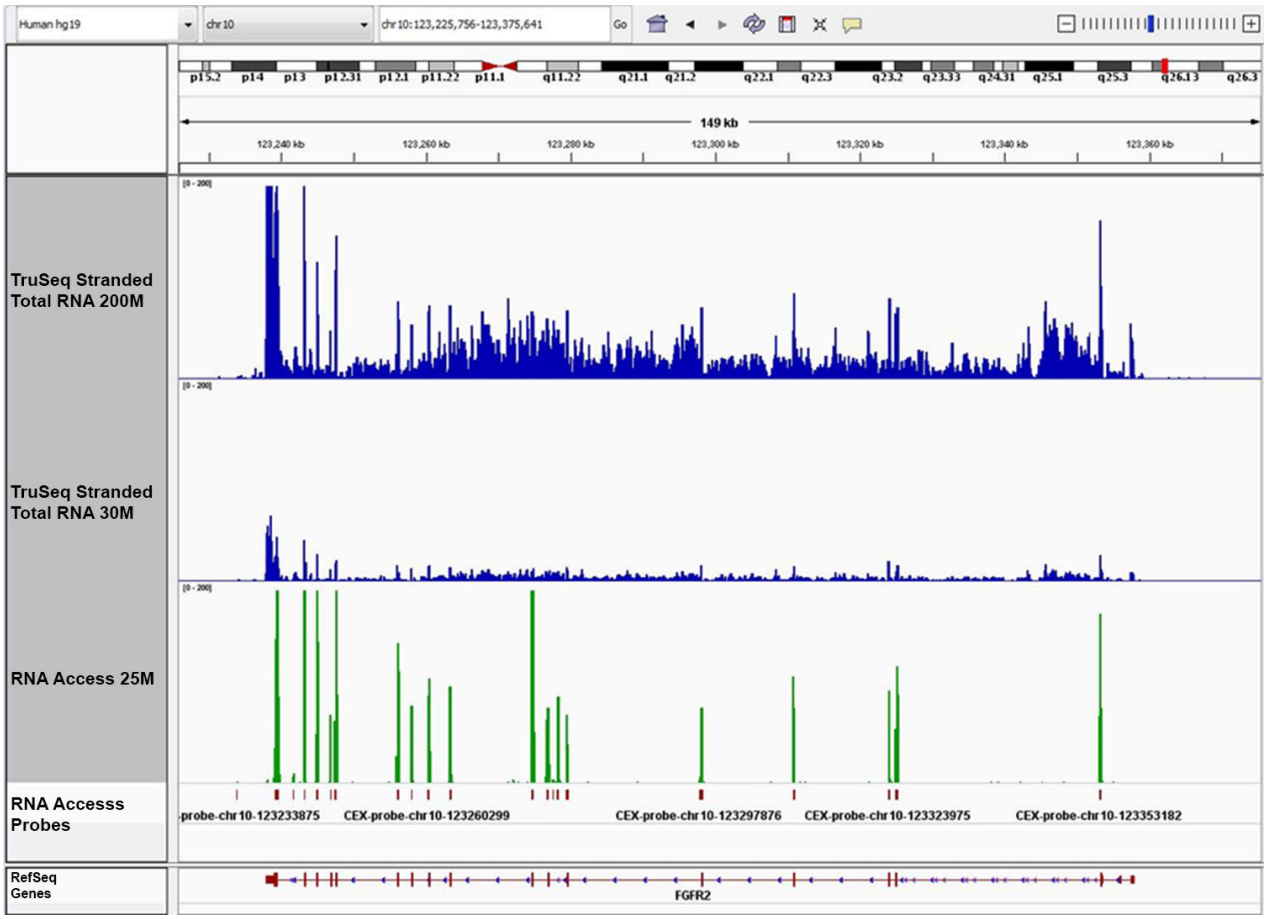
Table 1: Coverage Details

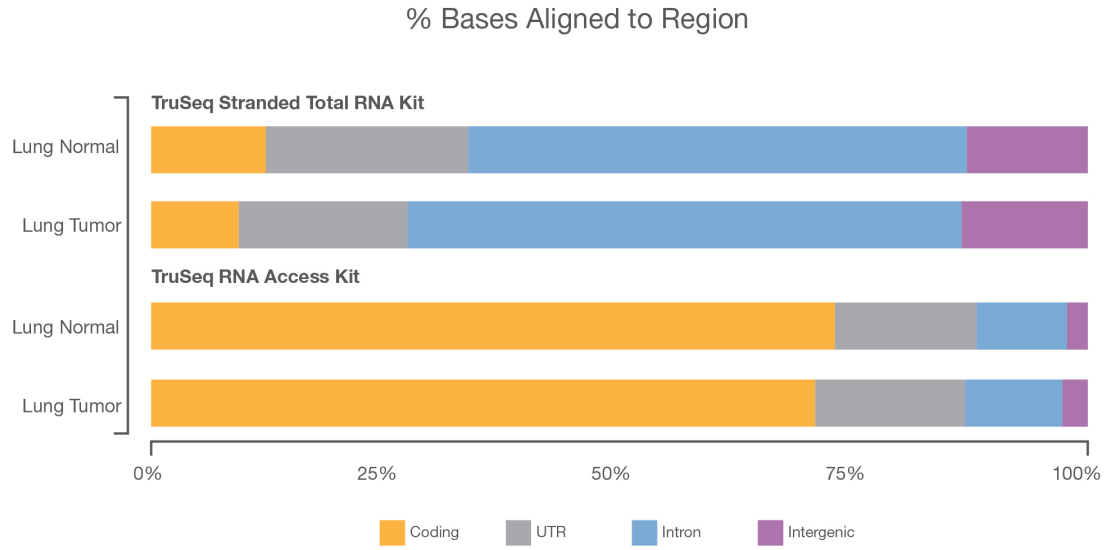
Number of target genes	21,415
Targeted exonic regions	214,126
RefSeq exome covered	98.3%
Number of probes	425,437

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 the TruSeq RNA Access Kit the ideal solution for small and/or precious samples.

Figure 1: Focus on RNA Coding Regions with the TruSeq RNA Access Kit





FFPE lung tumor samples were prepared with both the TruSeq Stranded Total RNA Kit (top two tracks) and the TruSeq RNA Access Kit (bottom track) and sequenced at 200 M and 25 M reads, respectively. Samples prepared with TruSeq RNA Access Kit show much deeper coverage of the exons, even at 1/8 the number of reads. The TruSeq Stranded Total RNA Kit data were down-sampled to 30 M reads (middle track) for comparison. Using the BaseSpace TopHat Alignment App, more than 85% of the data generated using the TruSeq RNA Access Kit aligned to transcripts (coding and UTR regions).

Cost-Effective Solution

The Illumina HiSeq® 2500 and NextSeq™ 500 Sequencing Systems offer the highest throughput capabilities combined with highest number of reads in the industry. By focusing on the coding regions of RNA, laboratories can sequence more samples per run—5× more than when sequencing total RNA—without sacrificing data quality (Figures 1 and 4). In fact, the TruSeq RNA Access Kit 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 relevant to your studies, effectively extending your read budget (Table 2).

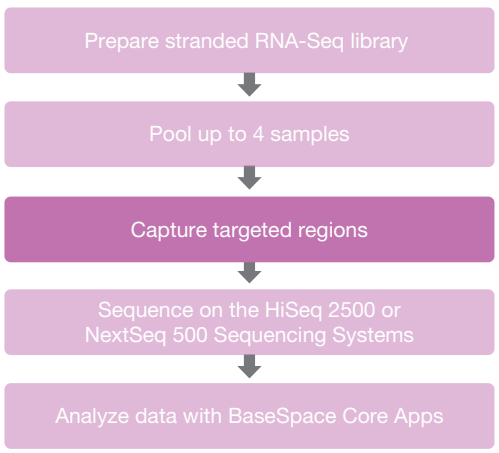
Efficient Discovery

Sequencing total RNA provides access to the entire transcriptome and is effective for FFPE samples; however, this may be more information than researchers need. Using TruSeq RNA Access, researchers can enrich for the coding regions and focus on the high-value content of interest without sacrificing gene fusion discovery power (Figure 5).

Table 2: The TruSeq RNA Access Kit Extends Read Budget

Application	Fresh/Frozen or FFPE RNA-Seq
Reads per sample	25 M (2 × 75 bp)
MiSeq System v3 chemistry	1 sample per run
NextSeq 500 System Mid-Output	5 samples per run
NextSeq 500 System High-Output	16 samples per run
HiSeq 2500 System Rapid-Run	24 samples per run
HiSeq 2500 System High-Output	160 samples per run

Figure 2: Overview of the TruSeq RNA Access Workflow

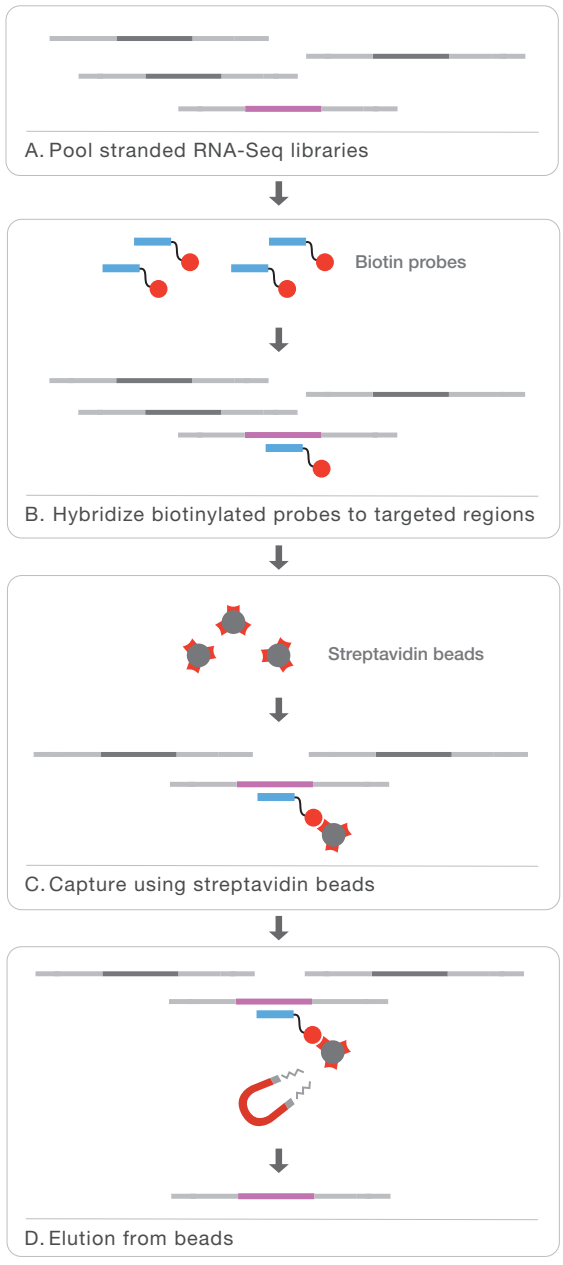


The TruSeq RNA Access Kit uses an integrated workflow to simplify library preparation and coding transcriptome capture for sequencing the coding regions of RNA. This increases the number of samples per run and the amount of usable data generated.

Convenient, Easy Data Analysis

Working with more focused content leads to a smaller data set. In turn, this leads to easier data analysis and data handling. TruSeq RNA Access data sets can be analyzed using RNA-Seq Software Apps in the Illumina BaseSpace™ cloud computing environment. These apps provide expert-preferred data analysis tools (TopHat/Cufflinks) packaged in an intuitive, click-and-go user interface designed for informatics novices and pre-configured workflows that support a range

Figure 3: TruSeq RNA Access Capture Chemistry

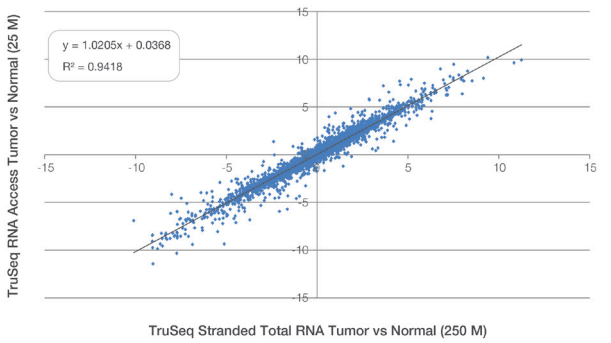


The TruSeq RNA Access Kit provides a simple and streamlined method for isolating targeted regions of interest from samples.

of common transcriptome data analysis needs. TopHat 2 enables high-confidence alignment for abundance measurement as well as the detection of splice junctions, gene fusions, and cSNPs. CuffDiff enables sensitive transcript discovery and differential expression analysis. TopHat Fusion delivers robust, high-confidence detection of gene fusions, while the Illumina Isaac[1] pipeline delivers reliable variant calling.

Simple-to-follow prompts guide users through the entire process, starting from selecting the files generated by the sequencer, to filtering and visualizing analyzed data and results. RNA-Seq Apps Software generates output files that may be directly input into a broad range of available secondary analysis solutions. Core Apps for RNA are available as part of the BaseSpace Core Apps, with the goal of making the most frequently used sequencing applications for secondary analysis available within a single environment. Core Apps enable easy aggregation of multi-sample reports, notification of job completion on mobile devices, and efficient file organization for collaboration and sharing.

Figure 4: Accurate Data with a Fraction of the Reads



Differential expression analysis was performed after sequencing total RNA from FFPE lung tumor and lung normal samples prepared with the TruSeq RNA Access Library Prep Kit (25 M reads) and the TruSeq Stranded Total RNA Sample Prep Kit (250 M reads). Data reveal that $\log_2$ fold-change values correlate highly across the entire dynamic range of expression regardless of the total number of reads collected.

Summary

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

Learn More

To learn more about the TruSeq RNA Access Library Prep Kit, visit www.illumina.com/Access.

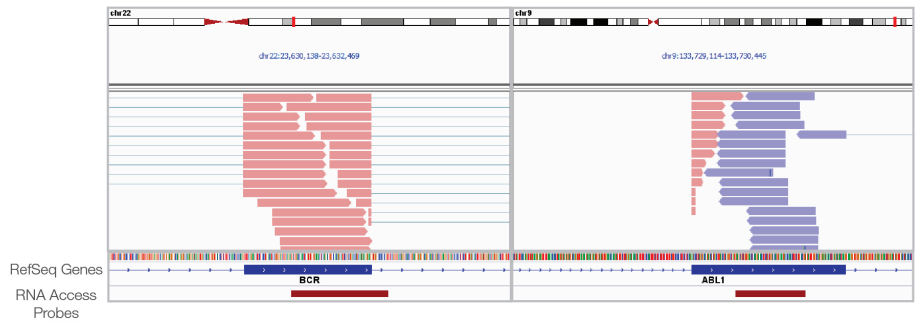
Reference

1. http://bioinformatics.oxfordjournals.org/content/early/2013/06/04/bioinformatics.btt314

Ordering Information

Product	Catalog No.
TruSeq RNA Access Library Prep Kit - Set A (48 samples, 12 indexes)	RS-301-2001
TruSeq RNA Access Library Prep Kit - Set B (48 samples, 12 indexes)	RS-301-2002

Figure 5: Efficient Gene Fusion Discovery



The TruSeq RNA Access Library Prep Kit 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.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY.

© 2014 Illumina, Inc. All rights reserved.
Illumina, Genetic Energy, HiSeq, NextSeq, TruSeq, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 470-2013-004 Current as of 31 March 2014

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA