Workflows for RNA Sequencing

A guide to Illumina solutions for next-generation RNA sequencing applications

illumina®

Introduction to RNA sequencing

The more researchers learn about the genome, the more they see how factors beyond the genetic sequence influence development, disease states, and more. To truly understand what is happening, scientists are turning towards a multiomics approach, looking at the transcriptome, epigenome, proteome, and metabolome, in addition to the genome to find comprehensive answers. In this guide, we focus on next-generation sequencing (NGS) workflows for studying the transcriptome.

A highly sensitive and accurate method for gene expression analysis across the transcriptome, NGS-based RNA sequencing (RNA-Seq) provides visibility into previously undetectable changes in gene expression, as well as enables the characterization of multiple forms of coding and noncoding RNA.[1,2] With RNA-Seq, researchers can detect the fine architecture of the transcriptome, such as transcript isoforms, gene fusions, single nucleotide variants (SNVs), and other features—without the limitation of prior knowledge,[2,3] in a single experiment.

RNA-Seq offers several advantages over other RNA analysis methods, such as qPCR and gene expression arrays. RNA-Seq provides a wider dynamic range than gene expression arrays, resulting in greater sensitivity and accuracy.[4] And unlike both qPCR and gene expression arrays, RNA-Seq can capture both known and novel features, enabling analysis of the transcriptome without a reference genome.[2] Furthermore, RNA-Seq is a popular choice for both model and nonmodel species, even when genetic tools and resources are limited.

Learn more about RNA-Seq at illumina.com/techniques/sequencing/rna-sequencing.html.

Integrated solutions for RNA-Seq workflows

Illumina RNA sequencing workflows seamlessly integrate library prep, sequencing, and data analysis to support transcriptome research. Illumina RNA library preparation solutions are available for a broad range of applications and sample types, including low-quality samples, such as formalin-fixed paraffin-embedded (FFPE) tissue, and for a wide range of input amounts. For large-scale studies, researchers can use high-throughput instruments like the NovaSeq™ 6000 and NextSeq™ 2000 Sequencing Systems, and multiplex up to 384 samples with unique dual indexes.

Learn more about Illumina RNA library prep solutions at illumina.com/techniques/sequencing/NGS-library-prep/rna.html.

RNA-Seq data can be quickly and securely transferred, stored, and analyzed in Illumina Connected Analytics or BaseSpace™ Sequence Hub, the Illumina multiomics cloud computing platforms. Both platforms offer in-cloud access to the DRAGEN™ Bio-IT Platform for accurate, ultra-rapid secondary analysis of RNA-Seq and other NGS data.

Learn more about RNA-Seq data analysis at illumina.com/informatics/sequencing-data-analysis/rna.html.

Illumina RNA-Seq data analysis solutions

Once the domain of bioinformatics experts, RNA-Seq data analysis is now more accessible than ever. Illumina offers push-button RNA-Seq software solutions packaged in intuitive user interfaces designed for biologists. These user-friendly tools support a broad range of NGS studies, from gene expression analysis to total RNA expression profiling and more.

DRAGEN RNA Pipeline

Perform ultra-rapid secondary analysis of RNA transcript data generated using NGS. Analysis includes spliced mapping and aligning, fusion detection, and gene expression quantification. Output gene expression data is compatible with the DRAGEN Differential Expression application.

DRAGEN Differential Expression App

Perform differential expression analysis on gene expression data generated using NGS. It runs the DESeq2 algorithm on Salmon quantification files from the DRAGEN RNA Pipeline to output genes and transcripts that are differentially expressed between two sample groups.

BaseSpace Correlation Engine

Compare expression profiles from RNA-Seq, qPCR, and gene expression arrays. Perform integrated analysis between DNA, RNA, and methylation studies. Compare molecular profiles from your own studies with results from curated public data repositories.

BaseSpace Sequence Hub

Illumina genomics cloud computing environment offering cost-effective, secure data storage and management, and easy-to-use tools for data analysis and collaboration.

Illumina Connected Analytics

Comprehensive cloud-based data management and analysis software platform empowering researchers to manage, analyze, and interpret large volumes of multiomics data in a secure, scalable, and flexible environment.

Considerations for RNA-Seq

This RNA-Seq workflow guide provides suggested values for read depth and read length for each of the listed applications and example workflows. Illumina recommends consulting the primary literature for your field and organism for the most up-to-date guidance on experiment design.

Read depth
For RNA-Seq, read depth (number of reads per sample) is typically used instead of coverage. Detecting low-abundance genes can require an increase in read depth. The ENCODE consortium (Encylopedia of DNA Elements) has published a set of ENCODE Guidelines and Best Practices for RNA-Seq to support your study aims.

Learn more about read depth for RNA-Seq at support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html.

Read length: paired vs single reads
Paired-end RNA-Seq enables discovery applications such as detecting gene fusions in cancer and characterizing novel splice isoforms. For gene expression profiling, single-end sequencing may be sufficient, however paired-end reads can enable more accurate read alignment and the ability to detect insertion-deletion variants, which is not possible with single-read data.

Learn more about paired-end vs single-read sequencing at illumina.com/science/technology/next-generation-sequencing/plan-experiments/paired-end-vs-single-read.html.

RNA-Seq applications and example workflows

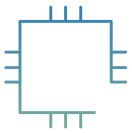
While RNA-Seq can be used for virtually any study that requires insight into the transcriptome, here we highlight three primary applications: gene expression profiling, whole-transcriptome sequencing, and single-cell RNA-Seq.

Gene expression profiling

To understand normal cell development and disease mechanisms, researchers frequently investigate differential expression during development, in specific tissues, or in response to varying conditions. RNA-Seq shows exceptional performance in profiling genes with low expression levels.[5] RNA-Seq is being used to assess gene expression profiles for the study of complex diseases and laying the groundwork for advances in precision medicine by identifying potentially therapeutic biomarkers.[5]

Workflow example: mRNA-Seq with high-quality RNA input

- Focus on the coding transcriptome
- Quantify gene expression, identify known and novel transcript isoforms, and measure allele-specific expression

Library prep	Sequencing	Data analysis
		
Illumina Stranded mRNA Prep Delivers accurate, unbiased detection of the protein-coding transcriptome with precise measurement of strand information[a] *RNA input* *25-1000 ng high-quality RNA*	*Estimated samples per flow cell*[b] *25M reads per sample* *2 × 75 bp read length* **NextSeq 550 System** Mid-output: 5 High-output: 16 **NextSeq 2000 System**[c] P2: 16 P3: 44 **NovaSeq 6000 System** SP: 32 S1: 64 S2: 164 S4: 384	**DRAGEN RNA Pipeline** **DRAGEN Differential Expression App** **BaseSpace Correlation Engine**

a. Strand-specific RNA-Seq allows researchers to predict sense and antisense transcript structures, identify overlapping regions of transcription, and estimate expression levels of sense and antisense genes.[6]

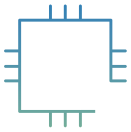
b. Sequencing parameters may vary; current details provided for illustrative purposes. Samples per flow cell depends on platform and flow cell configuration. For more information on instrument, flow cell, and reagent choice, visit illumina.com/systems/sequencing-platforms.html or illumina.com/library-prep-array-kit-selector.html.

c. P2 flow cells with the same sample throughput also available on the NextSeq 1000 System.

Learn more about mRNA-Seq at illumina.com/techniques/sequencing/rna-sequencing/mrna-seq.html.

Workflow example: RNA enrichment with low-input or degraded/FFPE samples

- Focus on the human RNA exome
- Quantify gene expression and identify novel transcript isoforms, SNVs, gene fusions, and allele-specific expression

Library prep	Sequencing	Data analysis
		
Illumina RNA Prep with Enrichment **Illumina Exome Panel** Enables cost-effective RNA exome analysis using sequence-specific capture of the coding regions of the transcriptome *RNA input* *10 ng minimum high-quality RNA* *20 ng minimum degraded/FFPE samples*	*Estimated samples per flow cell*[a] *25M reads per sample* *2 × 100 bp read length* **NextSeq 550 System** Mid-output: 5 High-output: 16 **NextSeq 2000 System**[b] P2: 16 P3: 44 **NovaSeq 6000 System** SP: 32 S1: 64 S2: 164 S4: 384	**DRAGEN RNA Pipeline** **DRAGEN Differential Expression App** **BaseSpace Correlation Engine**

a. Sequencing parameters may vary; current details provided for illustrative purposes. Samples per flow cell depends on platform and flow cell configuration. For more information on instrument, flow cell, and reagent choice, visit illumina.com/systems/sequencing-platforms.html or illumina.com/library-prep-array-kit-selector.html.

b. P2 flow cells with the same sample throughput also available on the NextSeq 1000 System.

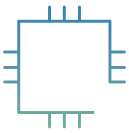
Learn more about targeted RNA-Seq at illumina.com/techniques/sequencing/rna-sequencing/targeted-rna-seq.html.

Whole-transcriptome sequencing

In addition to protein, genes encode a vast array of nonprotein-coding elements that play an instrumental role in orchestrating how a cell is organized from a transcriptional regulation perspective.[1] A major strength of RNA-Seq lies in its ability to identify novel features of the transcriptome. NGS-based RNA-Seq can sequence the whole transcriptome (eg, genes, gene variants, and noncoding transcripts) without the limitation of probe design, delivering a high-resolution, base-by-base view of coding and multiple forms of noncoding RNA activity. This provides a comprehensive picture of gene expression across the full transcriptome at a specific moment in time. With total RNA-Seq, the whole transcriptome—including both known and unknown regions—is captured.

Workflow example: Total RNA-Seq

- Analyze both coding and noncoding transcripts
- Identify novel transcript isoforms, gene fusions, variants, and allele-specific expression
- Use as low as 1 ng input

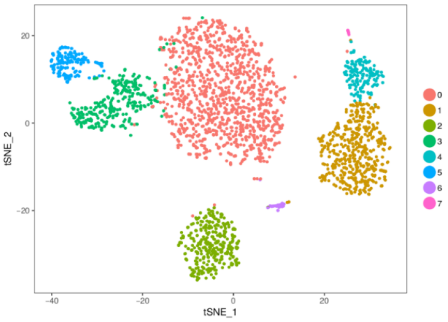
Library prep	Sequencing	Data analysis
		
Illumina Stranded Total RNA Prep with Ribo-Zero™ Plus Detects coding and noncoding transcripts for whole-transcriptome sequencing with integrated enzymatic depletion of ribosomal RNA *RNA input* *1-1000 ng high-quality RNA* *10 ng minimum for optimal performance and FFPE samples*	*Estimated samples per flow cell*[a] *50M reads per sample* *2 × 75 bp read length* **NextSeq 550 System** Mid-output: 2 High-output: 8 **NextSeq 2000 System**[b] P2: 8 P3: 22 **NovaSeq 6000 System** SP: 16 S1: 32 S2: 82 S4: 200	**DRAGEN RNA Pipeline**

a. Sequencing parameters may vary; current details provided for illustrative purposes. Samples per flow cell depends on platform and flow cell configuration. For more information on instrument, flow cell, and reagent choice, visit illumina.com/systems/sequencing-platforms.html or illumina.com/library-prep-array-kit-selector.html.

b. P2 flow cells with the same sample throughput also available on the NextSeq 1000 System.

Learn more about total RNA-Seq at illumina.com/techniques/sequencing/rna-sequencing/total-rna-seq.html.

Single-cell RNA-Seq



Looking at RNA at the single-cell level can reveal the cell types present and how individual cells are contributing to the function of complex biological systems. Single-cell RNA-Seq is an NGS method that examines the transcriptomes of individual cells, providing a high-resolution view of cell-to-cell variation and cellular subpopulations in complex tissues. Advances in microfluidic technologies have enabled high-throughput single-cell profiling where researchers can examine hundreds to tens of thousands of cells per experiment in a cost-effective manner.

Workflow example: Single-cell RNA-Seq

- Identify cell types and attribute transcriptional activities to specific cell types
- Discover new cell types that may serve novel functions in complex systems
- Detect transcriptional patterns in lower-frequency cell types
- Resolve transcriptional changes down to individual cell types to inform mechanistic and pathway models

Library prep	Sequencing	Data analysis
Chromium Single Cell Gene Expression Solution (10x Genomics) Encapsulates and barcodes 3’ end-tagged mRNA to measure gene expression for hundreds to tens of thousands of single cells	*Estimated samples per flow cell*[a] *5000 cells per sample* *20K read pairs per cell* **NextSeq 550 System** Mid-output: 1 High-output: 4 **NextSeq 2000 System**[b] P2: 4 P3: 11 **NovaSeq 6000 System** SP: 8 S1: 16 S2: 41 S4: 100	**DRAGEN Single-Cell RNA Pipeline** **Cell Ranger ARC** (10x Genomics)

a. Sequencing parameters may vary; current details provided for illustrative purposes. Samples per flow cell depends on platform and flow cell configuration. For more information on instrument, flow cell, and reagent choice, visit illumina.com/systems/sequencing-platforms.html or illumina.com/library-prep-array-kit-selector.html.

b. P2 flow cells with the same sample throughput also available on the NextSeq 1000 System.

Learn more about single-cell sequencing workflows and download our eBook at illumina.com/techniques/sequencing/rna-sequencing/ultra-low-input-single-cell-rna-seq.html.

References

1. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. 2011;12(2):87-98. doi:10.1038/nrg2934
2. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57-63. doi:10.1038/nrg2484
3. Wilhelm BT, Landry JR. RNA-Seq—quantitative measurement of expression through massively parallel RNA-Sequencing. *Methods*. 2009;48:249-57. doi:10.1016/j.ymeth.2009.03.016
4. Su Z, Labaj PP, Li S, et al. A comprehensive assessment of RNA-Seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotech*. 2014;32:903-914. doi:10.1038/nbt.2957
5. Xu J, Gong B, Wu L, et al. Comprehensive assessments of RNA-Seq by the SEQC consortium: FDA-led efforts advance precision medicine. *Pharmaceutics*. 2016;8(1). doi:10.3390/pharmaceutics8010008
6. Mills JD, Kawahara Y, Janitz M. Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. *Curr Genomics*. 2013;14(3):173-181. doi:10.2174/1389202911314030003

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-00034 v1.0