illumina®

Demystifying RNA Sequencing Data Analysis

Answers to some of the most common RNA-Seq data analysis FAQs.

Introduction

Next-generation sequencing (NGS) has revolutionized many areas of genomics research including the study of the transcriptome. RNA sequencing (RNA-Seq) is a highly sensitive and accurate method of measuring expression across the transcriptome. It provides visibility to previously undetected changes occurring in various disease states, in response to therapeutics, under different environmental conditions, and across a broad range of other study designs—all without the limitation of prior knowledge.

RNA-Seq also allows researchers to detect both known and novel features in a single assay, enabling the detection of transcript isoforms, gene fusions, single nucleotide variants, allele-specific gene expression, and other features. Although the advantages of RNA-Seq compared to previous methods have led to an increase in the adoption of RNA-Seq, many researchers have questions regarding RNA-Seq data analysis. Here we address the most common questions and concerns about RNA sequencing data analysis methods.

1. "I'm not an informatics expert."

RNA-Seq data analysis is now easier and more accessible than ever before; bioinformatics experience is not required. The user-friendly BaseSpace® genomics computing environment* includes a broad range of RNA-Seq data analysis software apps for transcript abundance measurement, detection of novel transcripts, coding single nucleotide polymorphisms (cSNPs), gene fusions, and more.

To learn about BaseSpace Core Apps for RNA-Seq analysis, visit the RNA-Seq data analysis page (www.illumina.com/applications/sequencing/rna/rna-seq-data-analysis.html). Not sure how to use BaseSpace? For a step-by-step run-through of a typical RNA-Seq data analysis workflow, download the BaseSpace RNA-Seq Technical Note (www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote-basespace-rna-seq.pdf).

2. "I'm only interested in using well-established, peer-reviewed data analysis methods."

BaseSpace provides click-and-go access to the most widely adopted RNA analysis pipelines in the research community. These include TopHat2, Cufflinks, TopHat Fusion, STAR, and DESeq2. Users can leverage the growing Illumina NGS ecosystem in BaseSpace, one of the largest collections of commercial and open source data analysis software tools currently available.

3. "I don't have the data storage and computer infrastructure in place."

As the cost of NGS decreases, data volumes are constantly increasing. Historically, the quantity of NGS data posed significant challenges for many laboratories. Assistance from bioinformatics and IT professionals was required to store and maintain growing data repositories. BaseSpace addresses such challenges with big data management by providing a virtually unlimited NGS data storage solution. Whether storing gigabytes, terabytes, or even petabytes of sequencing data, BaseSpace can accommodate your needs. Furthermore, all data in BaseSpace is automatically backed-up when it's written. Because of this service, the traditional NGS data management headaches of making and maintaining backup copies of data are relieved.

For more information on BaseSpace storage and security, download the BaseSpace Data Security Technical Note (www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_basespace_security.pdf).

* BaseSpace is the Illumina genomics computing environment for NGS data analysis and management. This flexible solution is available as both a cloud-based and onsite platform.

4. "I am not comfortable with or allowed to use a cloud-based system."

For labs that have special storage, network bandwidth, security, or compliance requirements, we created BaseSpace Onsite—a local storage and analysis solution. The BaseSpace Onsite System is a local version of the BaseSpace cloud that enables the secure storing, sharing, and analysis of NGS data from the NextSeq® 500 and HiSeq® 2500 Systems.

The BaseSpace Onsite System features an intuitive web interface that guides you through sequencing experiments, from sample management and run preparation through variant calling. You can track thousands of samples and experiments without complicated spreadsheets. It guides you through importing sample information, planning library preparation setup for multiplexing, and setting up runs. Syntax errors and index compatibility are checked at each step of the process.

To learn more about the BaseSpace Onsite System, download the BaseSpace Onsite System Specification Sheet (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet-basespace-onsite.pdf).

5. "Can I plug data output files generated by BaseSpace Apps into downstream analysis software that is not available through BaseSpace?"

The RNA-Seq apps in BaseSpace provide output files that can be directly input into a broad range of available secondary analysis solutions. Illumina sequencing systems are also designed to support multiple data analysis pipelines. Output file formats include *.bcl, FASTQ, BAM, *.vcf, *.csv, and *.txt and can be exported for analysis by third-party software. As a complement to the BaseSpace Core Apps for RNA Analysis, the BaseSpace environment also includes a growing body of software solutions for visualization, analysis, and sharing.

6. "I have legacy data from qPCR and GEX arrays that I want to retain and build upon."

NextBio® Research is a technology and platform agnostic tool that performs analysis across large collections of genomic and proteomic data to reveal novel discoveries and/or evidence to support existing hypotheses. With NextBio Research, data sets from microarrays, RNA-Seq, and other genomics/proteomics platforms can be compared through a combination of rank-based enrichment statistics, metaanalyses, and biomedical ontologies. NextBio Research also provides a continuously curated collection of public high-throughput data to enable researchers to stay up-to-date on information regarding any set of genes or proteins. Together with private data uploaded and shared across a research team or institution, NextBio Research provides comprehensive tools for extracting value from all data available, both public and private.

For a more in-depth discussion, read the Data Correlation Details: Enrichment Analysis Technical Note (www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote-data-correlation-enrichment.pdf).

7. "How do I compare my genomic data to public data and identify possible biomarkers or disease mechanisms?"

With NextBio Research applications, you can identify mechanisms of disease, drug targets, and prognostic or predictive biomarkers. NextBio Research supports multiple data interpretation functions including:

- Comparison of profiles from RNA-Seq, qPCR, and GEX arrays
- Storage of all current and/or past private data for improved security and convenience
- Integrated analysis across DNA, RNA, and methylation studies
- Comparison of molecular profiles from your own studies with results from curated public data repositories

Explore the growing library of curated, public genomic data with our easy-to-use, web-based tools that mine public data and create billions of novel correlations. NextBio Research applications support downstream analysis for gene function studies, drug and disease mechanisms, and cross-species analysis. Illumina offers a free version of NextBio Research to users at academic, government, or non-profit institutions.†

To learn more about NextBio solutions, visit the NextBio page (www.illumina.com/informatics/research/biological-data-interpretation/nextbio.html).

† Researchers interested in more advanced functionality can subscribe to the full version of NextBio Research, which offers a complete solution for data storage, analysis and integration.

8. "What if I get into trouble and need help during a run or an analysis?"

Whether you have basic data analysis questions that require immediate attention or you have advanced questions requiring in-depth consultations, Illumina can help. Beyond immediate phone and email support, Illumina customer service and support teams provide a full suite of expedient solutions from initial trainings, to instrument support, personalized consultation, and ongoing NGS education. Illumina customer support offerings include:

Illumina Technical Support

- Global, 24/5 phone and email support in the Americas, Europe, and Asia-Pacific.
- Illumina TS specialists can perform desktop sharing with GoToAssist—a powerful tool for quick identification and diagnosis of issues over the phone.
- For faster case handling, enter your case number at the main phone menu to be routed directly to the TS specialist handling your case.

Illumina University Training

- Instructor-Led Training at Your Chosen Facility
- Instructor-Led Training at an Illumina Training Center
- On-Line Courses
- Webinars

Illumina Consulting

- Proof-of-Concept Services for instrument and library preparation testing
- Concierge Services for design assistance and product optimization

For more on these and other service and support offerings, visit: www.illumina.com/services/instrument-services-training.html

Illumina • 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.

 Pub. No. 470-2015-004 Current as of 06 November 2015

illumina®