illumina®

RNA-Seq Offers Significant Advantages in Immuno-Oncology Research

TruSeq® RNA Exome performs with high analytical sensitivity and concordance compared to NanoString, while providing base-level resolution for a hypothesis-free approach to transcriptome analysis.

Introduction

Next-generation sequencing (NGS) has emerged as a powerful tool for facilitating new discoveries in immuno-oncology research. DNA sequencing has been successfully used to assess tumor mutational burden, identify neoantigens, and assess mutations that result in resistance to immunotherapy.[1-2] RNA analysis has also been important for identifying expressed neoantigens, characterizing the tumor microenvironment, and measuring expression levels of specific biomarkers that can be predictors of therapeutic response.[2-4]

Transcriptome analysis is helping to advance immuno-oncology research and guide rational biologically relevant therapeutic strategies. With high analytical sensitivity and a wide dynamic range, RNA sequencing (RNA-Seq) is a powerful quantitative method that can detect small variations in gene expression.[5-6] The first widely used methods for genome-wide profiling of RNA expression involved hybridization-based techniques, such as microarrays. However, as costs decreased, RNA-Seq became an increasingly attractive option for transcriptome profiling. By generating sequence information with nucleotide-level resolution, RNA-Seq is a hypothesis-free method that provides information that many other techniques do not, such as distinction of RNA isoforms and splice variants, detection of small genomic variants, and gene fusions (Table 1).[6-8] These data sets can also serve as retrospective samples that can be mined for content as new discoveries and hypotheses are made.

To harness these advantages, Illumina offers TruSeq RNA Exome (previously known as the TruSeq RNA Access Library Prep Kit), an RNA-Seq solution that targets coding RNAs and accurately measures gene expression for 21,415 genes. Starting with as little as 10 ng total RNA and optimized for formalin-fixed, paraffin-embedded (FFPE) tissues, TruSeq RNA Exome provides comprehensive coverage, including all exons of targeted genes. This white paper demonstrates the performance of TruSeq RNA Exome in comparison to a hybridization-based digital counting method for quantitative analysis (NanoString Technologies) of 57 tumor tissue samples.

Methods

Sample processing and sequencing were performed at Avera Precision Oncology Clinical Laboratory. RNA was extracted from 57 formalin-fixed, paraffin-embedded (FFPE) tumor tissue samples, then 100 ng RNA was analyzed with both TruSeq RNA Exome and the combination of two NanoString panels (nCounter PanCancer Immune Profiling Panel and the nCounter PanCancer Pathways Panel). Although TruSeq RNA Exome assesses 21,415 genes, concordance analysis was limited to the 1374 genes that are in common with the combined NanoString panels. For TruSeq RNA Exome, all samples yielded > 30 million reads, as is recommended for quantitative analysis. Gene expression was measured with RNA-Seq by calculating the value for fragment per kilobase per million (FPKM) reads. For NanoString, gene expression was measured by normalized counts. Analysis was performed using DESeq2 for TruSeq RNA Exome data and a Bioconductor normalization package for the hybridization technology[9,10] Differential expression analysis and variant analysis were performed for RNA-Seq data in BaseSpace® Sequence Hub with the Cufflinks Assembly & DE App and the RNA-Seq Alignment App.

Table 1: Analysis Options With RNA-Seq Data

Feature	Recommended Software Applications[a]
Differential Gene Expression Analysis	Cufflinks Assembly & DE, TopHat Alignment, RNA Express
Detection of Gene Fusions	Cufflinks Assembly & DE, RNA-Seq Alignment, TopHat Alignment
Detection of RNA Isoforms and Splice Variants	Cufflinks Assembly & DE, RNA-Seq Alignment, TopHat Alignment
Detection of SNVs and small indels	TopHat Alignment, RNA-Seq Alignment, BaseSpace Variant Interpreter

a. Available in BaseSpace Sequence Hub (www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps.html)

* Avera Precision Oncology Clinical Laboratory is an impartial third party.

† DESeq2 is a biostatistical tool used to estimates expression values and calculate differential expression between two groups. DESeq2 is a component of many RNA-Seq analytical workflows, including BaseSpace apps listed in Table 1.

Results

High Concordance With Hybridization-Based Methods

Despite differences in analytical sensitivity, the accuracy of a given method for relative measurements of gene expression is usually validated by confirming similar results with a secondary method. Concordance between TruSeq RNA Exome and NanoString was demonstrated by looking at the correlation of the genes in a sample evaluated using both methods. Initial correlation was performed using a Pearson Correlation with application of variance stabilizing transformation and batch effect normalization.[11] A second correlation using Lin's Correlation was performed on the same data.[12] Two representative samples are shown (Figure 1), although similar results were found with all 57 samples. Mean value for Lin's Concordance Correlation Coefficient (CCC) was 0.98 with standard deviation of 0.03.

High concordance was also demonstrated using a heat map with spatially matched samples from the TruSeq RNA Exome analysis and NanoString analysis next to each other. Using a color scale to indicate relative gene expression levels for individual genes, expression patterns are visible with samples showing the same patterns with both methods. A closer view of representative samples shows high concordance for each gene in both methods (Figure 2).

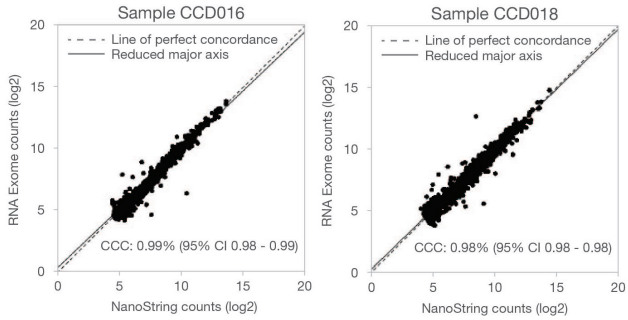


Figure 1: Concordance Between TruSeq RNA Exome and NanoString Displayed by Scatter Plots—Analysis of 57 tumor tissue samples using both methods was performed for 1374 genes commonly found in both methods. The scatter plots show expression for each gene from two representative ovarian tumor samples. Variance-stabilizing transformation was performed to allow for the measurements to appear on the same scale, and batch normalization was performed with the batch assumption based on the two distinct technologies. Lin's CCC and 95% confidence interval (CI) are shown. Mean value and standard deviation for Lin's CCC over all 57 samples was 97.7% and 3% respectively.

Sensitive Detection of Expressed Genes

To test assay sensitivity, the percentages of genes detected at low-level threshold cutoffs were compared for the two methods. When the detection threshold was set at ≥ 1 counts, TruSeq RNA Exome detected a higher proportion of the 1374 assessed genes (Figure 3A). For differential expression analysis, the detection of small changes in expression is more accurate when counts are increased. With the detection threshold raised to ≥ 50 counts, numerous genes were removed from detection by both methods, but TruSeq RNA Exome detected > 13% more genes than NanoString (Figure 3B).

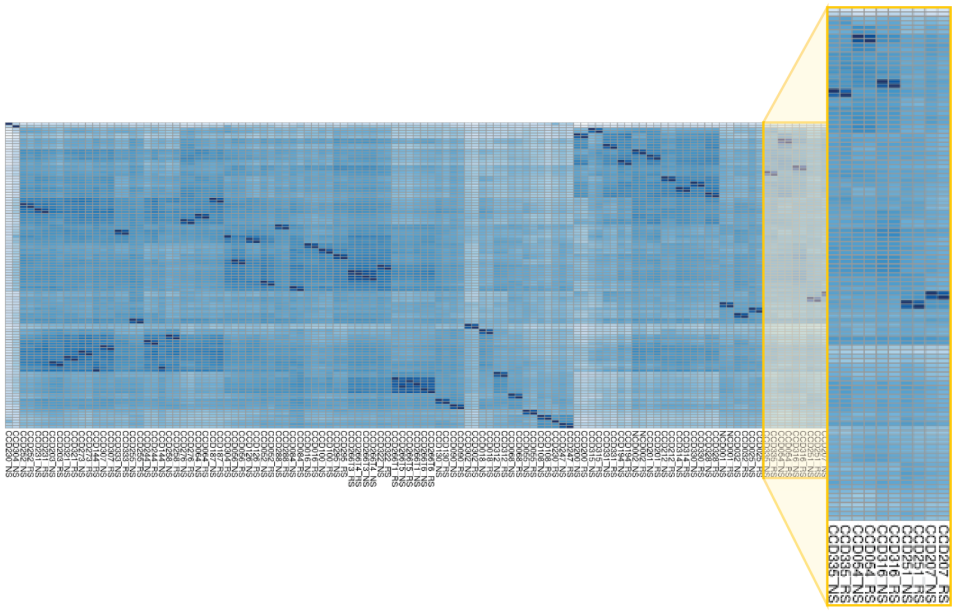


Figure 2: Concordance Between TruSeq RNA Exome and NanoString Displayed by Heat Maps—Heat maps show the gene expression profile of all 57 samples. Each column represents one experiment. The name of each experiment contains the sample ID and the experiment type (RS=TruSeq RNA Exome; NS=NanoString). Each row represents one gene, with color scale indicating expression level.

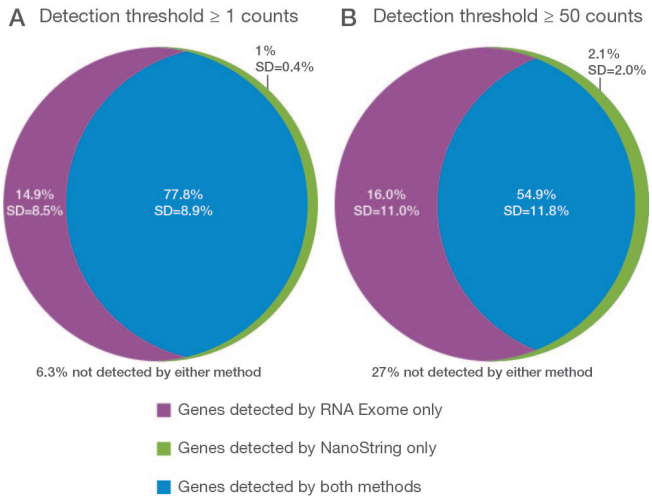


Figure 3: Genes Detected by TruSeq RNA Exome Panels and NanoString Panels—The average percent of detected genes from all 57 tumor samples, with standard deviations (SD) are shown. A total of 1374 unique genes were analyzed based on content of two NanoString panels, which excluded > 20,000 additional genes covered by TruSeq RNA Exome. Threshold cutoffs were applied to remove genes with counts lower than the stated cutoff.

Differential Gene Expression Analysis

Assessment of differentially expressed genes in related tissues can reveal important cellular changes that are undetected or overlooked by DNA analysis, such as epigenetic modifications, altered signal transduction pathways, or genetic variations in noncoding regions of the genome. A benefit of RNA-Seq is the ability to compare sequencing results between samples from different runs and experiments. This option may be useful for comparing populations of tumors from similar tissues. Because TruSeq RNA Exome targets >21,000 genes, detectable RNAs are not limited to tissue type, and the number of expression variants should be higher than with hybridization-based panels that have a smaller number of covered genes.

By comparing two representative ovarian tumor samples (CCD016, CCD084), 16,480 genes were identified and assessed by TruSeq RNA Exome, only 1123 of which were on the NanoString gene list. For differential expression, default cutoff value for the Cufflinks algorithm (q value < 0.05) was used to filter insignificant differences.

Among the filtered results, 75 genes were differentially expressed, only 10 of which were on the NanoString gene list. Results of expression analysis can be visualized with an interactive interface featuring adjustable filters and identification of outlier genes at the click of a button (Figure 4).

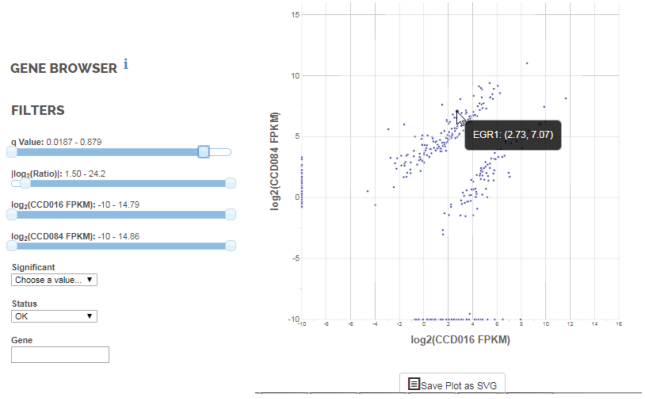


Figure 4: Interactive Scatter Plot Showing Differential Expression Analysis—Two ovarian tumor samples (CCD016, CCD084) were analyzed with TruSeq RNA Exome, then results analyzed with the Cufflinks Assembly & DE App. Adjustable filters enable the inclusion or exclusion of genes according to expression level, fold difference between the two samples, and significance (q value). Blue dots above and below the filtered zone represent individual genes, for which more information can be accessed through an interactive user interface in BaseSpace Sequence Hub.

Detection of Variants

Although DNA sequencing is often used to identify small variants in tumors, RNA-Seq is often performed on the same sample to confirm that the variants are expressed. However, RNA-Seq can be used alone to detect expressed variants. With numerous analysis options available in BaseSpace Informatics Suite, original sequence data can be accessed from archived samples and assessed for numerous types of variants, such as identification and interpretation of small variants (Figure 5). Because RNA-Seq provides nucleotide-level resolution for each read, it can also be used to identify gene fusions (Figure 6).

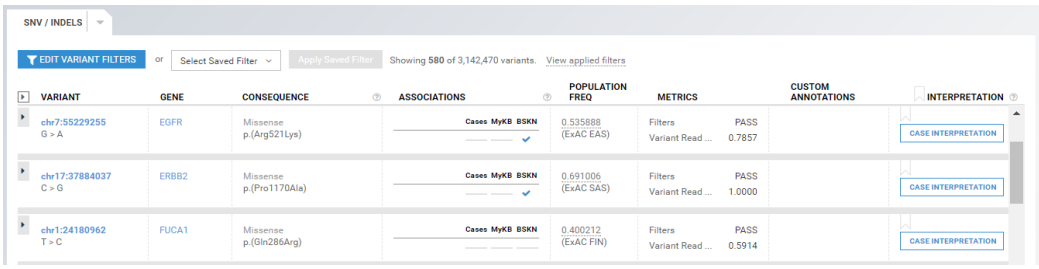
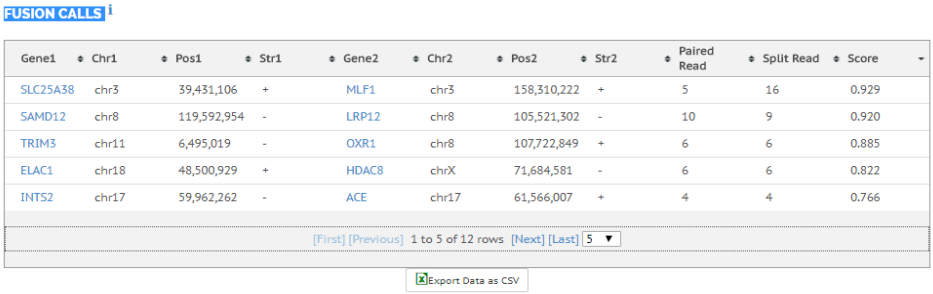


Figure 5: Detection of SNVs and Indels—A representative ovarian tumor sample (CCD016) was sequenced using TruSeq RNA Exome. The resulting data was analyzed in BaseSpace Sequence Hub with the RNA-Seq Alignment App to generate a variant calling file (VCF). The VCF file was then analyzed with BaseSpace Variant Interpreter to generate a list of detected variants with known associations with cancer.



FUSION CALLS

Gene1	Chr1	Pos1	Str1	Gene2	Chr2	Pos2	Str2	Paired Read	Split Read	Score
SLC25A38	chr3	39,431,106	+	MLF1	chr3	158,310,222	+	5	16	0.929
SAMD12	chr8	119,592,954	-	LRP12	chr8	105,521,302	-	10	9	0.920
TRIM3	chr11	6,495,019	-	OXR1	chr8	107,722,849	+	6	6	0.885
ELAC1	chr18	48,500,929	+	HDAC8	chrX	71,684,581	-	6	6	0.822
INTS2	chr17	59,962,262	-	ACE	chr17	61,566,007	+	4	4	0.766

[First] [Previous] 1 to 5 of 12 rows [Next] [Last] 5

Export Data as CSV

Figure 6: Detection of Fusions—A representative ovarian tumor sample (CCD016) was sequenced using TruSeq RNA Exome. The resulting data was analyzed in BaseSpace Sequence Hub with the RNA-Seq Alignment App using the fusion option.

Summary

Although hybridization-based methods have traditionally been used for genome-wide expression analysis, RNA-Seq has emerged as a method that demonstrates higher analytical sensitivity and wider dynamic range.[6-8] NanoString technologies also feature digital counting, which makes them attractive for quantitative measurements, although the coverage is limited to genes that are included in any given panel, and sensitivity may be affected by the lack of an amplification step.[6] High analytical sensitivity may be important for low-expressed genes, or moderately expressed genes in underrepresented cells.[5] In this study, overall concordance in measurement of relative expression levels was high between the two methods, but TruSeq RNA Exome demonstrated better performance in terms of analytical sensitivity when compared directly to NanoString. Furthermore, TruSeq RNA Exome can detect > 20,000 genes not covered by the two NanoString panels, offering a more comprehensive evaluation of the biology of a particular sample.

With a linear dynamic range similar to other RNA-Seq methods, TruSeq RNA Exome offers analytical sensitivity for low-expressed targets, while simultaneously offering a comprehensive view of the sample. TruSeq RNA Exome captures 214,126 exonic regions from 21,415 genes, enabling the distinction of related gene family members and splice variants. TruSeq RNA Exome also enables detection of single-nucleotide variants, small indels, and novel gene fusions within the transcriptome in FFPE tissues. By casting a broad net, researchers can test hypotheses, gather a multitude of data sets, and create archival comprehensive data sets on precious samples that can be revisited when new discoveries are made. When combined with powerful data analysis apps in BaseSpace Sequence Hub, TruSeq RNA Exome provides flexibility and high-quality data for immuno-oncology research.

Learn More

To learn more about the TruSeq RNA Exome, visit: www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-rna-access.html

References

1. van Rooij N, van Buuren MM, Philips D, et al. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol*. 2013;31(32):e439-442.
2. Lu YC, Yao X, Crystal JS, et al. Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Clin Cancer Res*. 2014;20(13):3401–3410.
3. Cimino-Mathews A, Thompson E, Taube JM, et al. PD-L1 (B7-H1) expression and the immune tumor microenvironment in primary and metastatic breast carcinomas. *Hum Pathol*. 2016;47(1):52–63.
4. Concha-Benavente F, Srivastava RM, Trivedi S, et al. Identification of the cell-intrinsic and extrinsic pathways downstream of EGFR and IFNγ that induce PD-L1 expression in head and neck cancer. *Cancer Res*. 2016;76 (5):1031–1043.
5. Yan H1, Dobbie Z, Gruber SB, et al. Small changes in expression affect predisposition to tumorigenesis. *Nat Genet*. 2002;30(1):25-26.
6. Yuan J, Hegde PS, Clynes R, et al. Novel technologies and emerging biomarkers for personalized cancer immunotherapy. *J Immunother Cancer*. 2016;4:3. doi:10.1186/s40425-016-0107-3.
7. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*. 2014;9(1):e78644. doi: 10.1371/journal.pone.0078644.
8. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.
9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
10. Bioconductor Open Source Software for Bioinformatics. bioconductor.org/help/workflows/ExpressionNormalizationWorkflow/. Accessed May 19, 2017.
11. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127.
12. Lin L. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255-268.

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