CASE STUDY

Uncovering Cell Type-Specific Expression Profiles in the Tumor Microenvironment with Ultra-Low Input RNA-Seq

Introduction

Advances in next generation sequencing (NGS) technologies have furthered our understanding of transcriptome dynamics and gene expression patterns. Since faithful characterization of the transcriptome depends largely on the quality and quantity of the input RNA, standard RNA-Seq approaches call for an ample amount (>500 ng) of intact RNA. Samples producing lower yields and degraded RNA typically require additional amplification steps as well as higher depths of sequencing to boost data output. These samples are prone to transcriptional bias and poor read mapping to exons[1,2].

Unfortunately, specimens in biomedical research are often restricted to minute quantities, such as formalin-fixed paraffin embedded (FFPE) tissue, sorted cells, biobank samples and microdissected tissue[3,4]. A team of researchers faced these limitations when studying the expression profiles of cancer and immune cells in the tumor microenvironment. Using fluorescence activated cell sorting (FACS), they isolated approximately 50 cells of each cell type from a glioblastoma. Since standard RNA-Seq techniques could not accommodate such limited starting material, the researchers approached GENEWIZ for a solution.

Using an optimized extraction-to-sequencing pipeline in combination with our Ultra-Low Input RNA-Seq service, GENEWIZ generated high quality transcriptomic data from approximately 50 cells. Here we show preparation of cDNA libraries from these sorted cells using a method that selectively amplifies full-length transcripts with minimal bias. Sequencing reads with strong quality scores and impressive mapping rates were achieved with detection of over 17,000 genes per sample. The quality and sensitivity of these results are commensurate with standard RNA-Seq experiments using micrograms of input RNA.

GENEWIZ®

(877) GENEWIZ (436-3949) ext 1
+1 (908) 222-0711 ext. 1

NGS@genewiz.com

genewiz.com/NGS

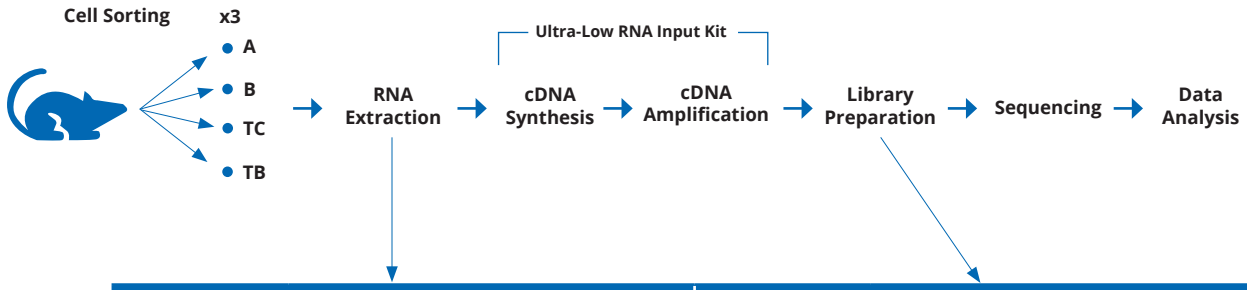
CASE STUDY

Viable Sequencing Libraries Prepared From Unquantifiable RNA

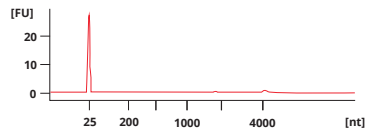
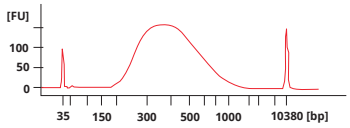
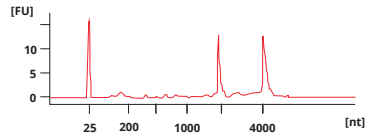
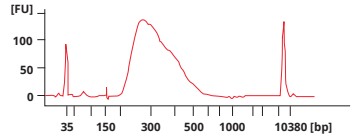
Total RNA was extracted from cell suspensions of astrocytes, B cells, cytotoxic T cells and bulk (CD3+ CD8-) T cells, containing approximately 50 cells each, using TRIzol reagent (Thermo Fisher Scientific). As expected, RNA concentration was below the detectable limit of a Qubit fluorometer or NanoDrop spectrophotometer. Using the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech), we performed reverse transcription on polyadenylated transcripts, followed by second-strand synthesis of full-length cDNA and amplification of the product (Figure 1A). Illumina-compatible sequencing libraries were constructed and then validated for sufficient concentration and appropriate fragment sizes (Figure 1B). Molecular indexes were added to each library, allowing samples to be pooled and sequenced on the Illumina HiSeq 2500 with a 1x100 bp single-end configuration. More than 30 million reads (>3 Gb total) were generated per sample, and on average 87% bases had quality scores above Q30.

Figure 1. Library Preparation

A.



B.

	Extracted RNA		Sequencing Library	
	Qubit Conc. (ng/µL)	Bioanalyzer Electropherogram	Qubit Conc. (ng/µL)	Bioanalyzer Electropherogram
Sample A1	<1		2.9	
Control	9.3		2.0	

A. Workflow overview. Astrocytes (A), B cells (B), cytotoxic T cells (TC) and bulk T cells (TB) were sorted from glioblastomas in mice. RNA was extracted from 12 cell suspensions (3 biological replicates per cell type) and used for cDNA synthesis and subsequent library preparation.

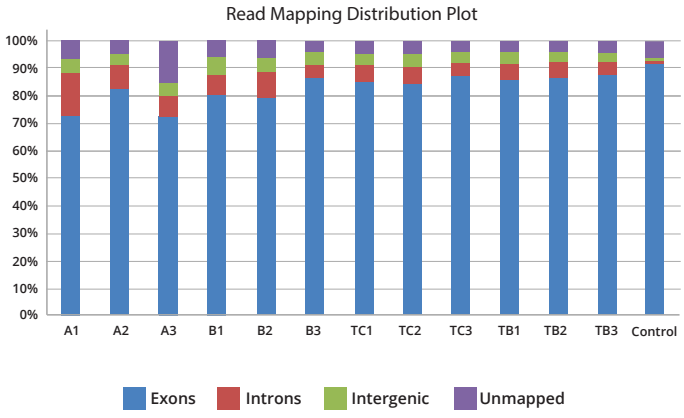
B. Evaluation of sample quality prior to sequencing. Extracted RNA and libraries were analyzed for concentration and distribution of fragment sizes. Results from a representative sample are shown. An RNA sample of high quality and measurable concentration was included as a control in the study.

Sequencing Reads Map Well to Exons, Representing High Gene Count

Following adapter trimming and read quality assessment, sequence reads were aligned to the mouse GRCm38 (mm10) reference genome. On average, 94% of reads were successfully mapped, exceeding the expected mapping rate of 70-90% for standard RNA-Seq projects[5] and that of the control sample (91%). Furthermore, 82% of reads aligned to previously annotated exons, representing an average of 17,324 genes (Figure 2). These results are highly comparable to a comprehensive RNA-Seq study, which detected a range of 15,578 to 22,295 expressed genes using micrograms of input RNA from 17 mouse tissues[6]. The high exon mapping rates indicate effective library preparation and allow downstream analysis such as quantification of differential gene expression.

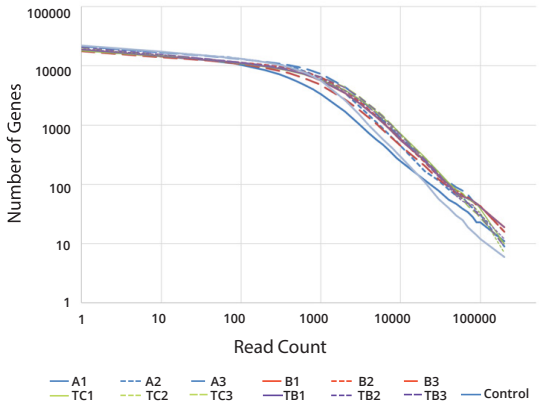
Figure 2. Read Alignment to the Reference Genome

A.



A. Percentage of mapped (exons, introns or intergenic regions) and unmapped reads. Astrocytes (A), B cells (B), cytotoxic T cells (TC) and bulk T cells (TB) from three replicates are shown.

B.

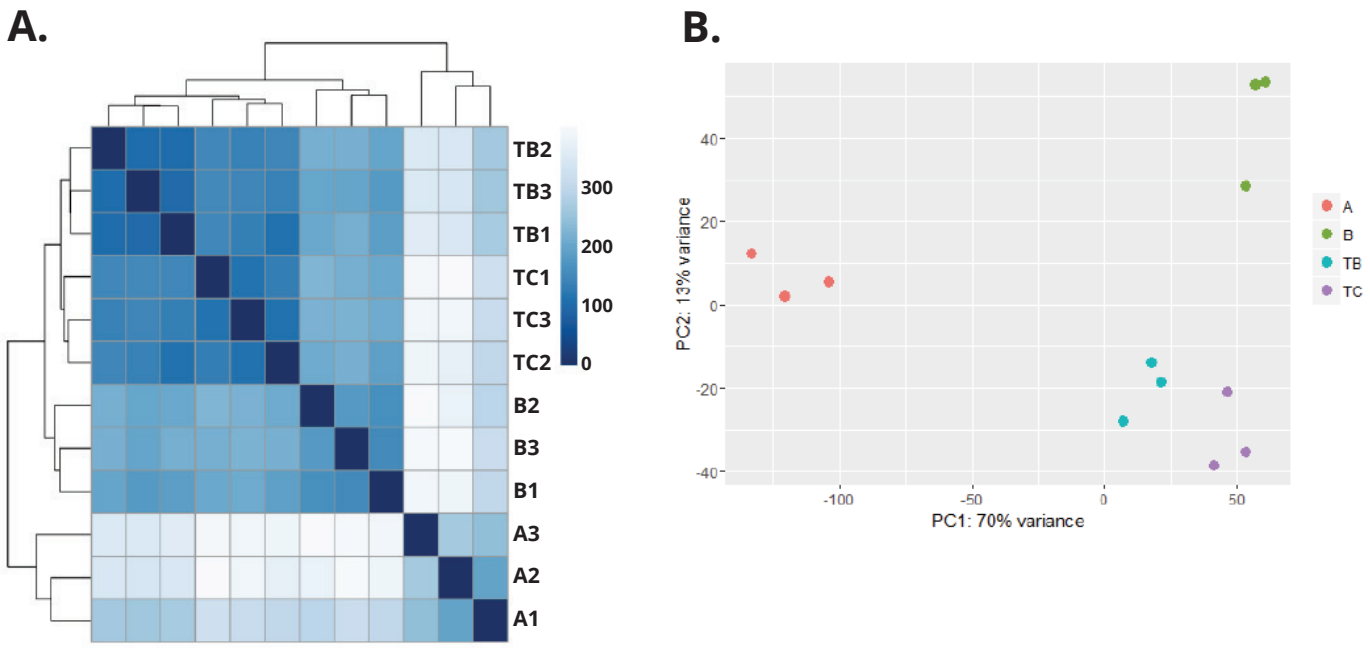


B. Plot of genes with a minimum number of read hits. An average of 17,324 genes have at least 10 hits per sample.

Cell Types Can Be Distinguished By Expression Data

To measure the relatedness of the samples according to their gene expression profiles, cluster analysis and principal component analysis were performed. High similarity was observed between biological replicates of the same cell type, and the largest variance was observed between astrocytes and the three groups of lymphocytes (Figure 3). These findings illustrate how our Ultra-Low Input RNA-Seq delivers reproducible transcriptomic data with sufficient complexity to differentiate cell types.

Figure 3. Comparison of Expression Profiles Between Samples



A. Heatmap of sample-to-sample distances. Astrocytes (A), B cells (B), cytotoxic T cells (TC) and bulk T cells (TB) from three replicates are shown.

B. Principal component analysis plot.

Conclusion

With <1 ng of RNA extracted from approximately 50 cells, we generated NGS data rivaling that of standard approaches using micrograms of RNA. Notably, these results were published in the journal *Science*[8]. Ultra-Low Input RNA-Seq empowers researchers to uncover gene expression patterns and assess transcriptomic heterogeneity in scant source material, a common limitation in oncology, clinical research and developmental biology.

Methods

Reverse transcription and cDNA amplification were performed with the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech). Libraries were constructed using the Illumina Nextera XT kit and analyzed for concentration (Qubit DNA assay and NanoDrop; Thermo Fisher Scientific), size distribution (Agilent Bioanalyzer) and quantification of viable sequencing templates via qPCR. Sequencing was performed on the Illumina HiSeq 2500 in Rapid Run Mode with 1x100 bp single-end configuration. Sequencing reads were assessed for overall quality, followed by adapter trimming and removal of low quality data. They were then mapped to the reference genome using CLC Genomics Workbench (QIAGEN) with total gene hit counts and RPKM values calculated. Differential expression analysis was performed using DESeq2 (Bioconductor). Sample distances were measured by log2, transforming the normalized counts data and calculating the Euclidean distance between samples[7]. Plots were generated using the R package 'pheatmap' for the sample distance plot and the DESeq2 function 'plotPCA' for the principal component analysis plot.

References

1. Adiconis, X. et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods* 10, 623-629 (2013).
2. Gallego Romero, I., Pai, A., Tung, J. & Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology* 12, 42 (2014).
3. Gilbert, M. et al. The isolation of nucleic acids from fixed, paraffin-embedded tissues–which methods are useful when?. *PLoS ONE 2*, e537 (2007).
4. von Ahlfen, S., Missel, A., Bendrat, K. & Schlumpberger, M. Determinants of RNA Quality from FFPE Samples. *PLoS ONE 2*, e1261 (2007).
5. Conesa, A. et al. A survey of best practices for RNA-seq data analysis. *Genome Biology* 17, (2016).
6. Li, B. et al. A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. *Scientific Reports 7*, (2017).
7. Team, R. C. R: A language and environment for statistical computing. Vienna, Austria: *R Foundation for Statistical Computing*; 2013. http://www.R-project.org/ (2013).
8. Quail, D. et al. The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas. *Science* 352, (2016).