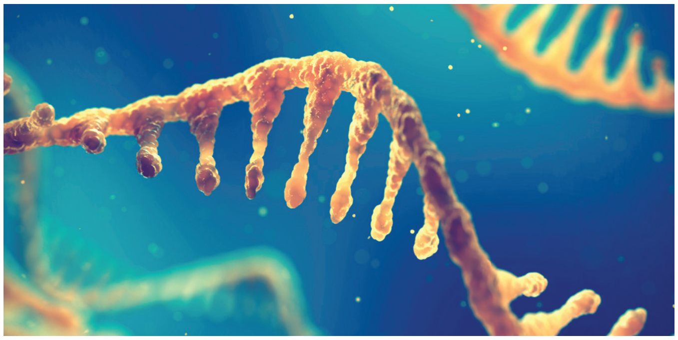
GenapSys
Sequencing without compromise

APPLICATION NOTE

GenapSys™ Sequencing Platform:

Targeted RNA Sequencing

- **Excellent gene expression concordance relative to established sequencing technologies**
- **Leverage hybrid capture based targeted RNA library prep methods**
- **Access to a personal, cost-effective sequencing platform for targeted RNA sequencing**
- **10-13 M highly accurate reads per sequencing run**
- **Low price per run and low price per sample**



Introduction

RNA sequencing (RNA-Seq) helps analyze the continuously changing cellular transcriptome, providing valuable information about the internal state of cells and transcriptional networks. In particular, differential gene expression analysis has played a significant role in the identification of pathway interactions and disease-state biomarkers.

Targeted RNA-Seq can overcome several challenges inherent in whole transcriptome RNA-Seq experiments, improving sequence coverage and sensitivity of detection of transcripts of interest that may be present in low amounts. Targeted RNA-Seq approaches are also advantageous in saving costs and simplifying analysis. Enrichment of transcripts of interest enables the capture of information about transcripts that could otherwise be missed or require a much greater number of sequencing reads to be detected.

The GenapSys Targeted RNA Sequencing workflow offers an affordable and simple Next-Generation Sequencing (NGS) solution for targeted RNA-Seq studies. Targeted RNA-Seq allows measurements of differential expression and allele specific gene expression across a large dynamic range. With targeted RNA-seq, one can obtain data across a large dynamic range, providing higher sensitivity and accuracy compared to conventional expression analysis methods.

Targeted RNA-Seq has considerable advantages over conventional expression analysis approaches such as microarrays or RT-PCR. The ability to perform deep sequencing provides higher sensitivity and a virtually unlimited dynamic range. Higher sensitivity allows for the detection of gene expression in low-abundance, low quality, or degraded FFPE samples. Targeted RNA-Seq provides a comprehensive base-by-base coverage of the area of interest and is better suited for discovery applications or for the measurement of rare transcripts compared to traditional methods.

Technology

The GenapSys™ Sequencer employs a novel electrical detection method that is capable of generating highly accurate DNA sequence information. With a CMOS based detector, simple fluidics, and low computational requirements, the GenapSys instrument is small, affordable, and accessible even to novice genomic scientists. Inside the sequencing chip are millions of individual sensors, each loaded with a single

Fig. 1 Simplified GenapSys workflow for Targeted RNA Sequencing.

bead coated in thousands of clonal copies of a particular DNA sequence. Individual nucleotides are flowed across the chip in succession and successful incorporation is detected by changes in impedance as the complementary DNA strand grows.

Methods

Whole transcriptome and targeted gene expression studies were performed using the following reference RNA Samples: Universal Human Reference RNA (UHRR, Agilent, Part #: 740000) and the Human Brain Total RNA (HBRR, Thermo Fisher Scientific, Ambion Part #: AM7962). Full-transcriptome libraries of the UHRR and HBRR samples were generated using the Lexogen SENSE mRNA-Seq Library Prep Kit v2, using the manufacturer's recommendations with approximately 500 ng of total RNA input. Final amplified libraries were size selected using the Sage Science Pippin Prep.

Targeted gene expression libraries were generated by hybrid capture based enrichment of whole transcriptome libraries. In particular, we use the IDT Pan Cancer Panel (v1.5) to enrich whole transcriptome RNA-Seq libraries of the UHRR and HBRR reference standards. Enriched and amplified libraries were size selected using the Sage Science Pippin Prep instrument. The whole transcriptome and targeted gene expression libraries were each run on the GenapSys Sequencer and the Illumina NextSeq platform. FASTQ files obtained from the sequencing runs were analyzed using the same analysis workflow, as described below.

The human transcriptome FASTA file (hg38, release 33) was downloaded from GenCode. The index files were generated using Salmon (v1.1.0) based on the human transcriptome FASTA file. Adapter sequences in the FASTQ files were removed using cutadapt. Quantification of RNA was performed using Salmon (v1.1.0) based on the adapter-removed fastq files. The correlation between samples was calculated using Pearson's coefficient method for genes with a minimum five read count or more.

Results

Whole transcriptome RNA-Seq libraries for the UHRR (Fig 2A) and HBRR (Fig 2B) reference standards were run on the GenapSys and Illumina NextSeq platforms. Each sample had > 30M reads, and the gene expression (transcripts per million, TPM) for both runs showed high correlation with an R^2 value of 0.963 for the UHRR sample and 0.958 for the HBRR sample. We also plotted the relative change in gene expression between the UHRR and HBRR samples, to model the typical transcriptomic experiment, which compares an experimental condition against another experimental condition or the baseline condition. The fold change (in $\log_2$ units) for the UHRR/ HBRR gene expression data is plotted for both GenapSys and Illumina data (Fig 3). There is a strong

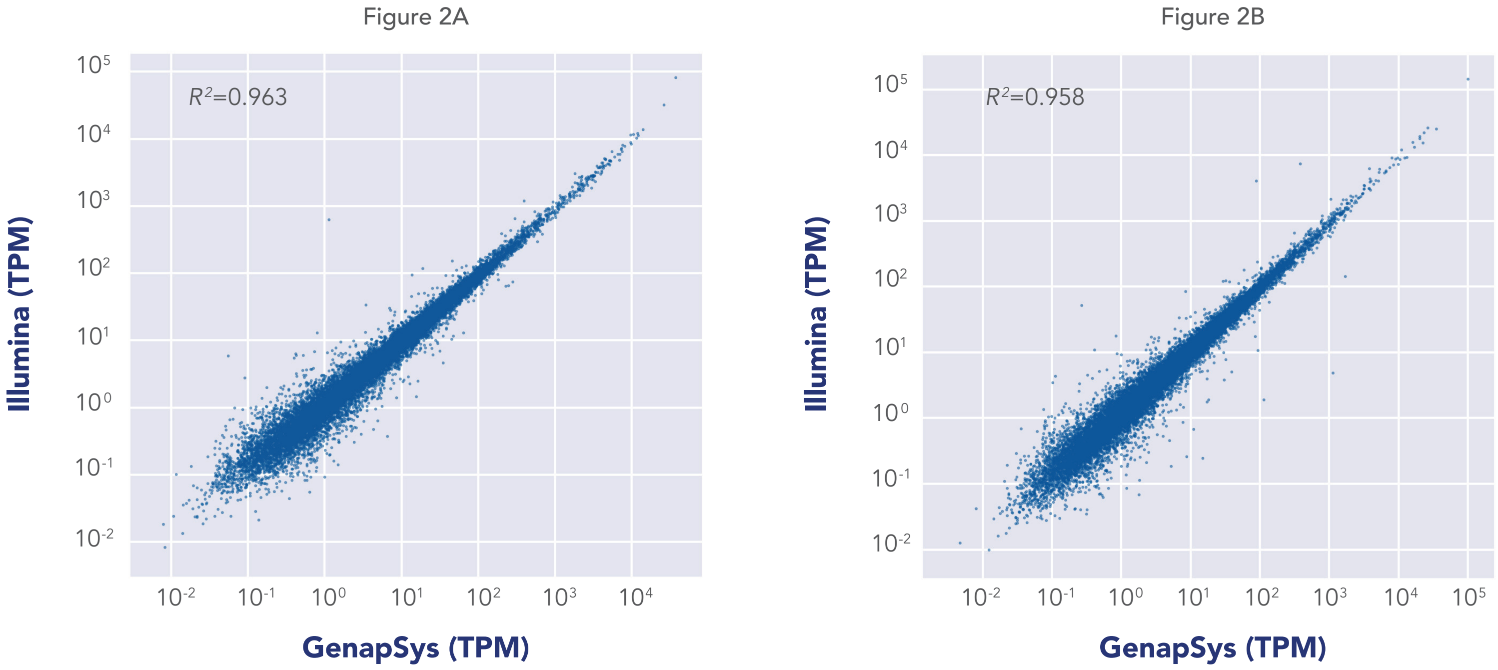


Figure 2A

Figure 2B

Fig. 2 Whole transcriptome RNA-Seq data for the (A) UHRR and (B) HBRR samples is plotted to compare gene expression between GenapSys and Illumina, as measured by transcripts per million (TPM).

To learn more about the GenapSys Sequencing Platform, visit GenapSys.com

Figure 3

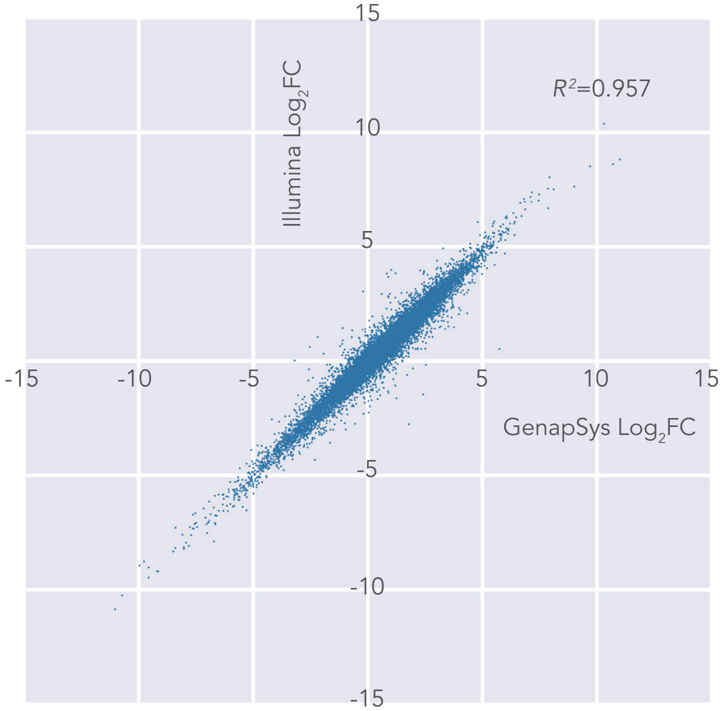


Fig. 3 Fold change ($\log_2$ units) in gene expression for the UHRR and HBRR whole transcriptome RNA-Seq libraries was compared between the GenapSys and Illumina platforms. There is a strong correlation in the fold change measurements between the platforms, with an R^2 value of 0.957.

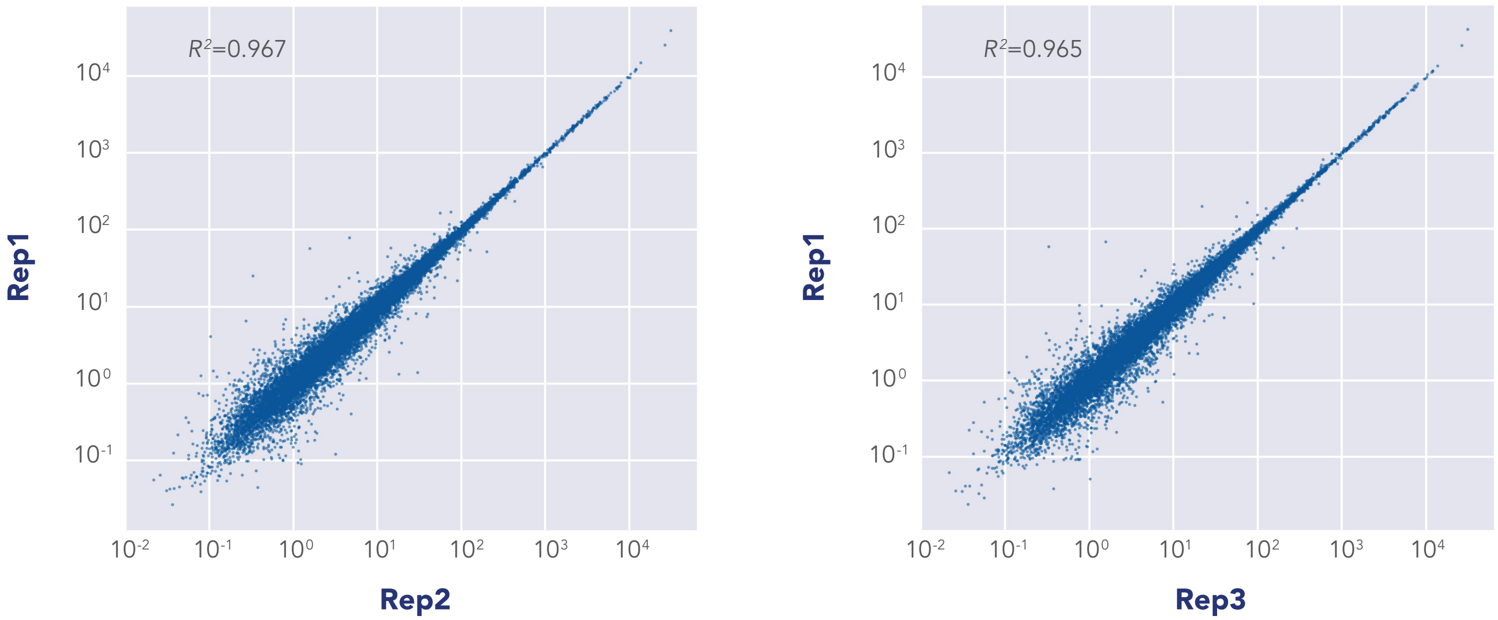


Figure 4A

Figure 4B

Fig. 4 RNA-Seq runs on the GenapSys platform show high reproducibility. Three technical replicates were performed on the whole transcriptome UHRR library. Correlation between (A) Replicate 1 and 2 and (B) Replicate 1 and 3, as measured by transcripts per million (TPM), show high concordance.

To learn more about the GenapSys Sequencing Platform, visit GenapSys.com

correlation in the fold change values, indicating that the gene expression data does not show any bias across a wide range of expression levels and genes. Finally, we demonstrated that gene expression is reproducible across runs. In Fig. 4, we show that 3 technical replicates of the UHRR gene expression libraries show strong reproducibility in the data, with R^2 values of 0.967 and 0.965 between gene expression levels of replicates.

Targeted gene expression libraries were generated by enriching a whole transcriptome library with the IDT xGen Pan Cancer panel ver 1.5. This panel consists of 7816 probes spanning 800 kb of the human genome, for enrichment of 127 significantly mutated genes implicated in 12 different types of tumor tissue. Fig. 5 shows the correlation for the targeted gene expression libraries for the UHRR (Fig. 5A) and HBRR (Fig. 5B) samples run on the GenapSys and Illumina NextSeq platforms. The gene expression (in transcripts per million, TPM) for both runs showed high correlation with an R^2 value of 0.99 for the UHRR sample and 0.954 for the HBRR sample. Thus, we demonstrate that we can obtain gene expression information of a targeted set of genes, thereby increasing the throughput of the number of samples and simplifying the analysis.

Summary

The GenapSys Targeted RNA Sequencing Solution offers a highly sensitive and accurate method for analyzing specific transcripts of interest. By harnessing the broad dynamic range of NGS sequencing, researchers can obtain more sensitive and accurate measurements for their targeted gene expression studies. The GenapSys Targeted RNA Sequencing Solution delivers high-quality NGS data in a compact, accessible and cost-effective platform.

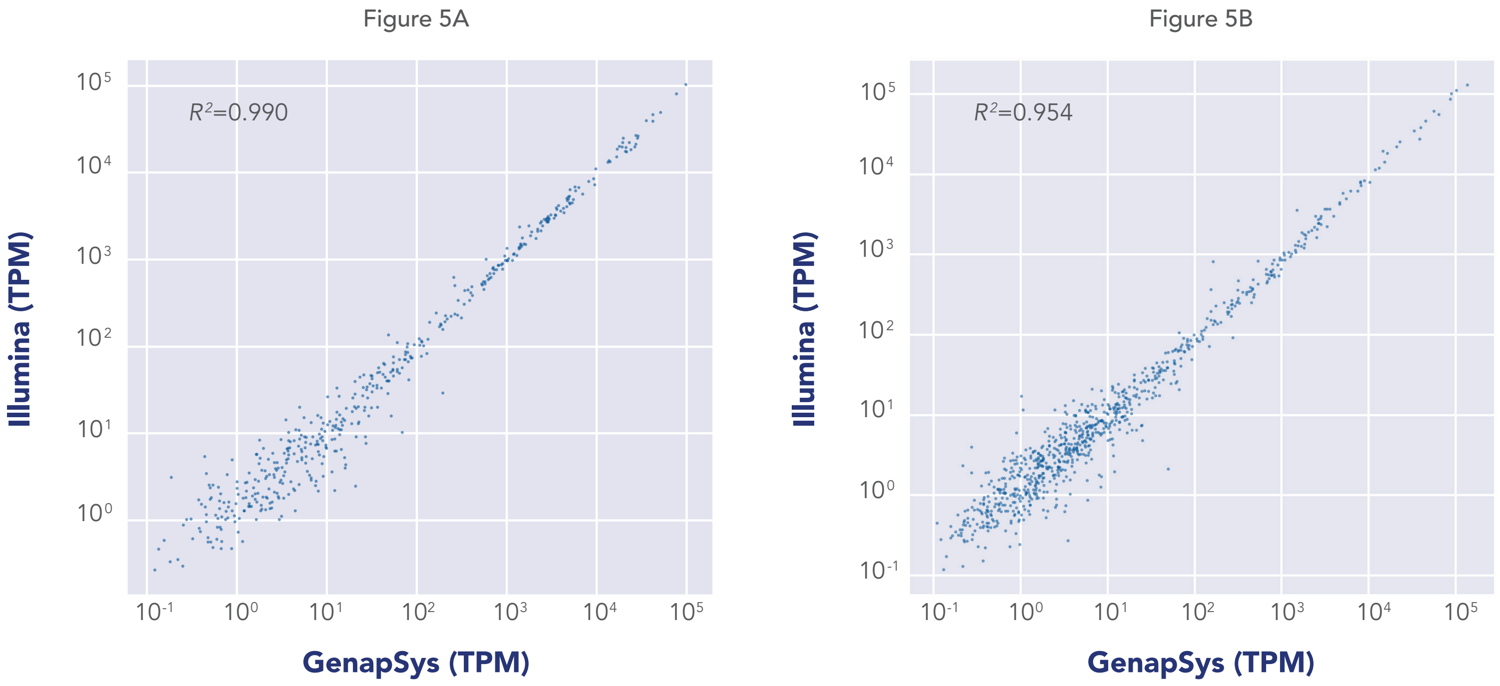


Fig. 5 Targeted gene expression data for the (A) UHRR and (B) HBRR samples is plotted, to compare gene expression between GenapSys and Illumina, as measured by transcripts per million (TPM).

GENAPSYS, INC.
200 Cardinal Way, 3rd Floor, Redwood City, CA 94063 • (800) 796-7710 • GenapSys.com

GenapSys
Sequencing without compromise