White Paper

Gene Expression Profiling and Sample Type Comparison between the HTG Transcriptome Panel and RNA-Seq

Introduction

RNA Sequencing (RNA-Seq) is considered the gold standard for quantification of transcriptome-wide gene expression. RNA-Seq allows for nucleotide-level identification of RNA, in certain cases without prior knowledge of the genome or transcriptional targets. This is because RNA-Seq can detect most transcripts in a sample using non-targeted technology.[1]

However, this comes with several distinct drawbacks. First, RNA-Seq relies on availability of significant amounts of good quality material from a given sample. One reason for this is that RNA-Seq necessitates RNA extraction before samples can be analyzed. RNA extraction removes degraded RNA and impurities in a sample; however, it has the potential to introduce an extraction bias due to removal of small, partially degraded, and low abundance RNAs.[2] Second, as RNA-Seq is not targeted, sequencing reads may be "wasted" on RNAs of little to no interest. Third, the RNA-Seq workflow can be time consuming, taking up to two weeks including sample preparation and sequencing time. Lastly, the bioinformatics pipelines for RNA-Seq analysis tend to be complicated and not standardized, posing a significant obstacle for researchers and clinicians.

The reality of the clinical research environment is that patient samples are precious, often available in limited quantities, and in a state of preservation where partially degraded tissue may contain critical insights. The recently introduced HTG Transcriptome Panel (HTP), based on the HTG EdgeSeq™ technology, represents a novel breakthrough and addresses the needs of gene expression research by overcoming many limitations of RNA-Seq. First, the HTP uses a proprietary extraction-free chemistry for sample preparation. This is an important improvement over conventional RNA-Seq as it reduces the amount of material required, and eliminates the risk of RNA extraction bias, which may result from the removal of small or partly degraded RNA species during the extraction process. Second, the HTP is a targeted gene expression panel that generates mRNA expression data for over 19,000 human protein-coding genes, meaning fewer reads are used up on sequencing other RNA species. Third, the HTG EdgeSeq workflow can be completed in as little as 36 hours, including sample preparation and sequencing. Lastly, the HTG EdgeSeq platform employs a fully integrated web-based data analysis package, called HTG EdgeSeq Reveal. The Reveal software allows researchers to generate robust analyses as well as high quality, publication ready graphics giving researchers the advantage of a fully automated and standardized bioinformatics pipeline.

This is the fourth in a series of white papers generated by HTG describing the performance of the HTP. The purpose of this white paper is to demonstrate accuracy of the HTP relative to RNA-Seq, and provide evidence of comparable performance between samples processed using HTG's extraction-free chemistry and RNA isolated and purified from the same formalin-fixed paraffin-embedded (FFPE) tissue specimens (eRNA). The data presented here show a high level of accuracy in differential expression applications as well as a high degree of sample-level correlation, suggesting the HTP can be used as a competitive alternative to RNA-Seq for transcriptome profiling of either FFPE or eRNA, with distinct advantages.

Methods

Samples

Table 1 summarizes the samples used in the study and associated experimental details.

Table 1. *Summary of samples, sample inputs and experimental replication*

Sample (Number)	Sample Type	Number of Lysis/ Extraction Events	Sample Input per Replicate and Well	Replicates per Lysate or eRNA
Breast Cancer (1)	FFPE	2	HTP: 3.4 mm^2 tissue* HTP: 35 ng eRNA RNA-Seq: 100-200 ng eRNA	6
Melanoma (1)	FFPE	2	HTP: 3.4 mm^2 tissue* HTP: 35 ng eRNA RNA-Seq: 100-200 ng eRNA	6

** Sample input for directly lysed tissue, defined as mm^2/well, is not the same as tissue area required for lysis due to the minimum required lysis volume and dilution requirement for the sample preparation workflow. For example, a sample input of 3.4 mm^2/well would require 11 mm^2 FFPE tissue lysed in a minimum of 50 µL Lysis Buffer A.*

HTG EdgeSeq Workflow

Specific details of the HTP workflow have been described previously.[3] Briefly, a single 5µm-thick section per sample was lysed using HTG's proprietary Lysis Buffer A (LBA) and treated with Turbo DNase to remove genomic DNA following the instructions for use. Alternatively, RNA extracted using the protocol below was diluted in lysis buffer to achieve the concentration of 1 ng/µL. Samples were then placed on an HTG EdgeSeq processor where gene-specific nuclease protection probes (NPP) were added. After allowing the NPPs to hybridize to their target RNAs, S1 nuclease was added to remove unhybridized NPPs and RNAs, leaving behind only NPPs hybridized to their target RNAs, resulting in a 1:1 ratio of probes and mRNA targets. Subsequently, sequencing adapters and molecular barcode tags were added by PCR; the tagged samples were cleaned, quantified, pooled, and sequenced using Illumina's NextSeq 500/550 system v2.5 (75 cycles). Data from the sequencer were processed and reported by the HTG EdgeSeq Parser software.

RNA-Seq Workflow

Prior to RNA-Seq testing, total RNA was extracted from four to eight 5µm-thick sections of FFPE tissue using the Qiagen RNeasy FFPE Kit, and DNase treated. One hundred nanograms of the isolated RNA were depleted of ribosomal RNA (rRNA) using KAPA RiboErase HMR Kit. The rRNA-depleted RNA was

fragmented before carrying out first-strand cDNA synthesis. The libraries were prepared using a KAPA RNA HyperPrep Kit. Adapter-ligated libraries were amplified with 12 cycles of PCR. Library concentration and fragment size distribution were determined using Agilent D1000 Screen Tape Assay on the 4200 TapeStation System. Libraries were sequenced using Illumina's NextSeq 500/550 High Output Kit v2 (300 Cycles).

Statistical Analysis

Differential gene expression analysis was performed between 12 replicates of the breast cancer and melanoma samples. To directly assess accuracy between the two platforms, technical replicates of the same two samples took the place of biological replication. RNA-Seq data were prefiltered to include the 19,398 genes represented in the HTP. The same differential expression model was fit on all sets of data following a typical workflow in which low-expression genes were filtered prior to model fitting and log fold-changes were shrunk using a Student's t prior.[4] The Pearson correlation coefficient between log-fold changes calculated for the two platforms represents accuracy of the differential expression analysis. Additionally, the median $\log_2$ counts per million (CPM)-transformed counts for directly lysed FFPE samples and RNA extracted from the same specimens were plotted in scatter plots, and the Pearson correlation coefficients were calculated.

Results

Accuracy of Differential Expression Analysis

The ability of the HTP to identify differentially expressed genes using HTG's extraction-free chemistry was compared to RNA-Seq utilizing eRNA. For this study, two individual FFPE samples, breast cancer and melanoma *(Table 1)*, were selected based on the measurable log-fold changes in gene expression between the two samples as determined by RNA-Seq. To ensure robustness of the analysis, two independent RNA extraction events (both HTP and RNA-Seq) and FFPE lysing events (HTP only) were carried out and six replicates of each sample were run at the recommended sample input for each platform for a total of 12 replicates per platform. A similar comparison had been carried out with FFPE lysates and published in a previous white paper, which showed good correlation to RNA-Seq with a Pearson correlation coefficient of 0.83.[3] Here, the cross-platform comparison was expanded to include eRNA as a sample type processed using the HTP *(Figure 1)*.

Similar to a previous study, the comparison of directly lysed FFPE tissue processed using the HTP compared to eRNA processed using RNA-Seq resulted in a Pearson correlation coefficient of 0.84 *(Figure 1, left)*. Interestingly, eRNA processed using the HTP produced a very similar correlation coefficient (0.83) suggesting that the HTP is equally capable of generating accurate differential expression data from eRNA as from fixed tissue samples. When directly lysed FFPE tissue and eRNA were tested using the HTP, they resulted in a Pearson correlation coefficient of 0.91 *(Figure 1, right)*. While a few outliers were observed in this comparison, they were all low-expressing genes (represented by a light shade of blue). The high degree of correlation suggests that directly lysed FFPE tissue and RNA extracted from FFPE produce comparable results of differential gene expression.

Comparison of Directly Lysed FFPE and eRNA

To further investigate the impact of RNA extraction on HTP gene expression data, eRNA and directly lysed FFPE samples from

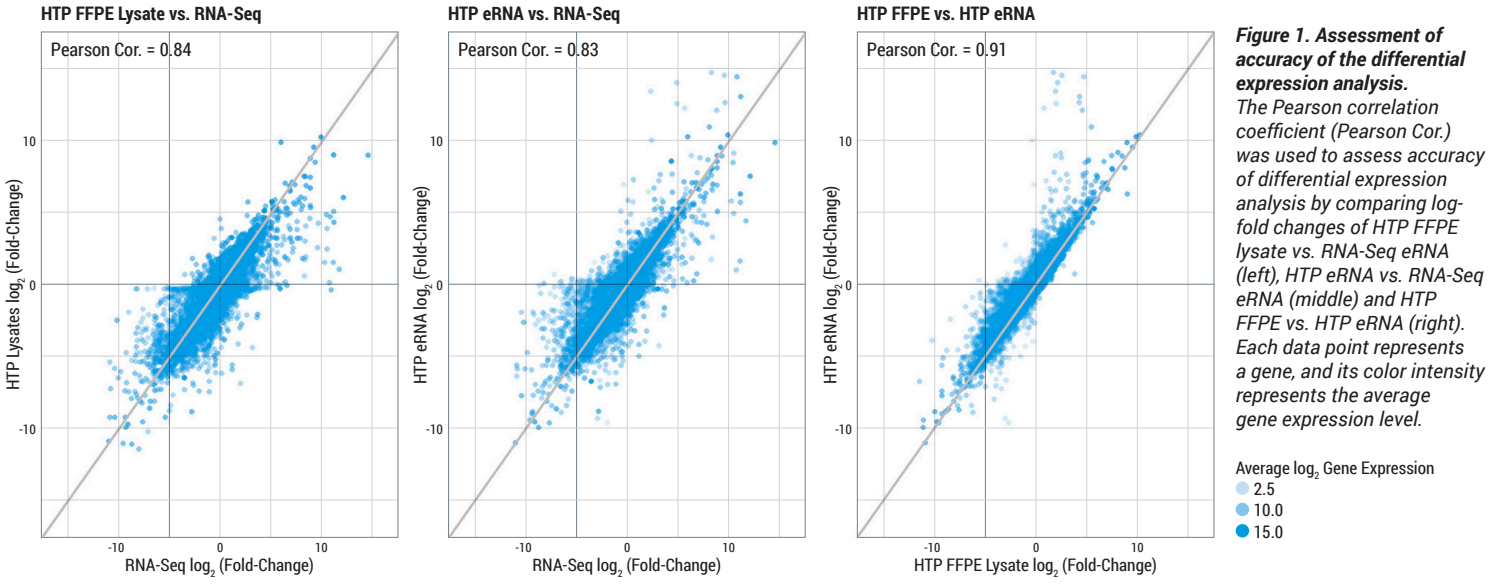

Figure 1. Assessment of accuracy of the differential expression analysis. *The Pearson correlation coefficient (Pearson Cor.) was used to assess accuracy of differential expression analysis by comparing log-fold changes of HTP FFPE lysate vs. RNA-Seq eRNA (left), HTP eRNA vs. RNA-Seq eRNA (middle) and HTP FFPE vs. HTP eRNA (right). Each data point represents a gene, and its color intensity represents the average gene expression level.*

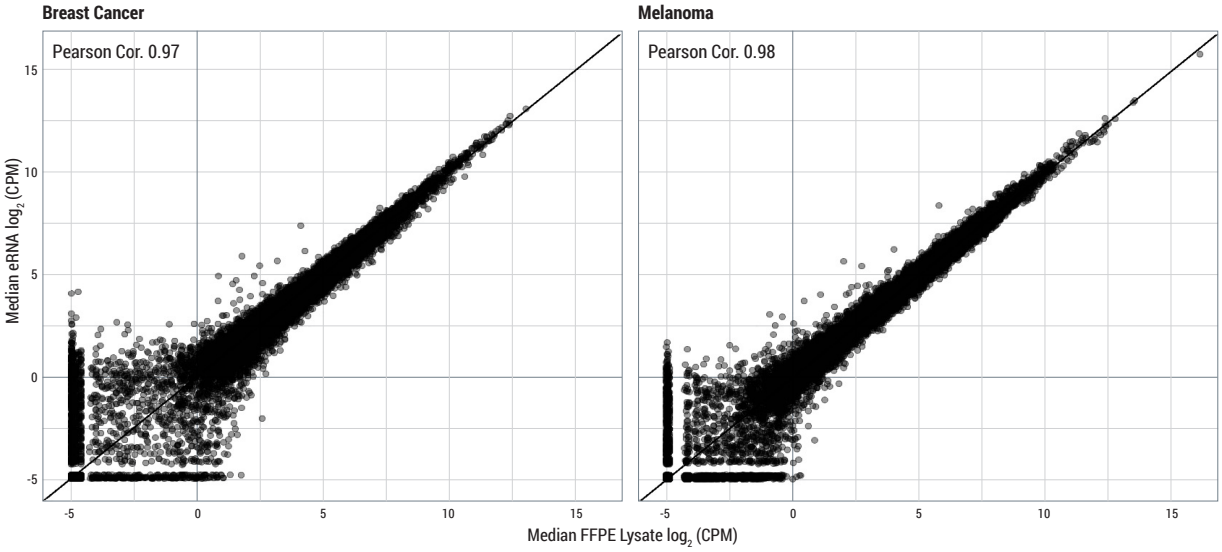

Figure 2. Comparison of FFPE lysates and eRNA using the HTP. *The median log_2(CPM)-transformed count values for directly lysed FFPE samples and eRNA from the breast cancer (left) and melanoma (right) samples were plotted. The Pearson Cor. are reported in the upper left corner; the black diagonal line is the unity line.*

breast cancer and melanoma specimens were tested using the HTP. The two sets of median log_2(CPM)-transformed expression values were compared for each of the two samples *(Figure 2).*

The Pearson correlation coefficients were calculated to determine the agreement of the expression data between the two sample types. Pearson correlation coefficients of 0.97 and 0.98 between the median log_2(CPM)-transformed count values obtained from the directly lysed FFPE tissue and eRNA from the same tissue for the breast cancer and melanoma specimens, respectively, were observed. These data demonstrate a high degree of correlation between the two sample types suggesting that the HTP generates highly comparable data from FFPE tissue lysates and RNA extracted from the same specimens. By extension, this also suggests that the HTP is not significantly impacted by extraction bias that may be present when using RNA-Seq methods, likely owing to the underlying HTG EdgeSeq technology which enables the detection of fragmented RNA using probes that only require a 50-nucleotide-long mRNA target for hybridization.

Conclusions

Although RNA-Seq is widely accepted as the gold standard for genome-wide transcriptomics analysis, it has significant requirements for sample quantity and RNA quality, as well as for highly skilled technical and bioinformatics personnel needed to generate and analyze the results, making it not well suited for many applications. The data presented in this report demonstrate that the HTP performs similarly to RNA-Seq in terms of the detection of differentially expressed genes for directly lysed FFPE tissue as well as eRNA. Additionally, the HTP eliminates the risk of extraction bias and generates data that are accurate and highly correlated for these two sample types, providing researchers a great deal of flexibility and confidence in their data. The combination of extraction free processing, excellent performance across various sample types and RNA quality, reduced sequencing demands, and simplified bioinformatics, yields an ultra-efficient gene expression profiling workflow ideal for mRNA research across many disease areas.

References

1: Yuan J, Hegde PS, Clynes R, *et al.* (2016). Novel technologies and emerging biomarkers for personalized cancer immunotherapy. *J Immunother Cancer.* 4:3.

2: Sultan M, Amstislavskiy V, Risch T, Schuette M, Dokel S, Ralser M, Balzereit D, Lehrach H, Yaspo ML. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics.* 15:675.

3: White Paper — Overview of the Design and Performance of the HTG Transcriptome Panel, Revision 1 released on 04-Aug-2021.

4: Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

HTG Molecular Diagnostics, Inc.
Call 1-877-289-2615
www.htgmolecular.com

C-115-3.22.R0