White Paper
Proof-of-Concept for a Whole Transcriptome Panel Using HTG EdgeSeq™ Technology

Abstract: Next-Generation Sequencing (NGS) has emerged as a powerful tool with the ability to generate an unprecedented quantity and quality of data. Over time, RNA-Seq has established itself as the gold standard for the NGS-based quantification of gene expression but, it requires high sample input amounts, high sample quality, and has bias caused by RNA extraction. HTG EdgeSeq technology has been able to overcome many of the limitations related to RNA-Seq, but until now has only been tested with targeted panels of less than 3,000 genes.

The purpose of this White Paper is to introduce a new prototype HTG EdgeSeq panel that covers the entire human transcriptome, and still retains the advantages of smaller targeted panels. The data show proof-of-concept for a prototype targeted HTG EdgeSeq whole transcriptome panel that can profile the expression of more than 20,000 RNA targets in FFPE samples. A comparison of gene expression profiling between HTG EdgeSeq and RNA-Seq platforms showed that the HTG EdgeSeq panel has good directional alignment with RNA-Seq, accurately measures differential gene expression, and has a linear response to spike-in targets present in FFPE samples.

Overall, the data presented here show that the prototype HTG EdgeSeq whole transcriptome panel has the potential to be used as a competitive alternative to RNA-Seq for gene expression profiling, while maintaining the advantages of a smaller targeted panel. Specifically, the HTG EdgeSeq platform requires less sample input and is capable of successfully processing samples of low quality.

Introduction

RNA sequencing (RNA-Seq) is a popular method for the study of gene expression, as it enables the detection and quantification of all RNA molecules in biological samples without prior knowledge of the sequence of interest[1]. RNA-Seq, however, suffers from technical complexity in sample preparation, requiring substantial quantities of extracted RNA, and extensive processing to generate a sequencing library. Another significant limitation of RNA-Seq is that it does not perform well on samples having relatively low-quality RNA. This becomes an issue as archival formalin-fixed paraffin-embedded (FFPE) tissue sections are often the sole means of addressing specific clinical and biological questions and tend to be low quality due to RNA degradation. Additionally, both the sample, and the RNA that may be extracted from that sample, are often available in very limited amounts.

HTG EdgeSeq technology addresses several of the limitations related to RNA-Seq[2]. First, HTG EdgeSeq assays use an extraction-free method of sample preparation. This is an important improvement over RNA-Seq because it eliminates the risk of RNA extraction bias, which results from the removal of small or partly degraded RNA species during the extraction process. The extraction-free process also means RNA is not lost due to low yields of the extraction process, so less sample input is required to generate equivalent amounts of addressable RNA. Second, the HTG EdgeSeq technology uses quantitative nuclease protection chemistry to detect RNA species of 50 nucleotides or more; thus, short, or fragmented RNAs normally removed during RNA extraction can be measured. Third, the HTG EdgeSeq platform is a targeted gene expression profiling platform that only generates information for specific transcripts in a sample, meaning fewer reads are used up on sequencing RNAs of limited utility such as non-coding RNAs. Finally, the HTG EdgeSeq workflow can be completed in a much shorter time frame than RNA-Seq and employs a fully integrated web-based data analysis package, that allows for a standardized bioinformatics pipeline.

The purpose of this White Paper is to demonstrate proof-of-concept for a prototype whole transcriptome panel that uses the core HTG EdgeSeq chemistry for gene expression profiling of FFPE tissues. The prototype HTG EdgeSeq whole transcriptome panel presented here can profile the expression of more than 20,000 RNA targets in FFPE samples. To demonstrate that the prototype HTG EdgeSeq panel can accurately measure gene expression levels in FFPE tissue, a set of breast and kidney cancer FFPE samples were analyzed for gene expression on both HTG EdgeSeq and RNA-Seq platforms and compared. Overall, the data presented here show: (1) good directional alignment with RNA-Seq as demonstrated using three example genes; (2) accurate measurement of differential gene expression as confirmed by spike in of synthetic external RNA controls (ERCC) as well as comparison to RNA-Seq; and (3) that the response generated by the panel has a linear relationship to the concentration of the ERCC analyte present in the sample.

Methods

Samples

Sixteen breast cancer FFPE tissues were commercially procured and tested using both the prototype HTG EdgeSeq whole transcriptome panel and RNA-Seq. The 16 samples can be divided into four groups (*Table 1*) based on the immunohistochemistry (IHC) expression of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (Her2).

In addition to the breast cancer samples, five kidney carcinoma FFPE tissues were also tested. These data are not included in this paper because the libraries generated with these samples using RNA-Seq had extremely low diversity and failed to provide informative data. The kidney FFPE samples were ten years old and arguably contained highly degraded RNA. These samples, however, generated quality data with the prototype HTG EdgeSeq whole transcriptome panel, highlighting the advantages that HTG EdgeSeq technology offers when working with older tissue.

Table 1. *Sample information.*

Cancer	Sample Type	Number Tested
ER/PR+, Her2- Breast Cancer	FFPE	6
ER/PR-, Her2+ Breast Cancer	FFPE	5
ER/PR-, Her2- Breast Cancer	FFPE	4
Undetermined Breast Cancer	FFPE	1
Kidney Cancer	FFPE	5

HTG EdgeSeq Workflow

HTG EdgeSeq libraries were generated by following the HTG EdgeSeq workflow outlined in *Figure 1*. Briefly, one 5-µm-thick section of each FFPE tumor tissue was lysed in HTG's proprietary lysis buffer and includes a new DNase treatment step to remove genomic DNA. 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. Sequencing adapters and tags were added by PCR. The tagged samples were cleaned, quantified, and sequenced on Illumina's NextSeq 550 System. Data from the NGS instrument were processed and reported by the HTG EdgeSeq parser software.

RNA-Seq Workflow

Unlike samples processed using HTG EdgeSeq technology, samples for use in RNA-Seq must first go through RNA extraction. In this study, total RNA was prepared from three 5-µm-thick sections of FFPE tumor tissue using the Qiagen AllPrep DNA/RNA FFPE Kit. One hundred nanograms of the isolated RNA were depleted of ribosomal RNA (rRNA) using KAPA RiboErase HMR Kit. The rRNA-depleted RNA was fragmented using high temperature in the presence of magnesium before carrying out first strand cDNA synthesis. Sequencing libraries for whole transcriptome analysis were prepared using 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 4200 TapeStation System. Illumina's NextSeq 550 System was utilized for sequencing analysis.

Results

To test the performance of the prototype HTG EdgeSeq whole transcriptome panel, several experiments were done to evaluate the assay's ability to detect differentially expressed genes in FFPE tissues. The linearity and dynamic range of the assay, as well as the accuracy of the assay were also evaluated using ERCCs. Whenever possible, the performance of the panel was compared to RNA-Seq, which is considered the gold standard for whole transcriptome gene expression quantification.

Differential Expression Analysis using HTG EdgeSeq and RNA-Seq in Breast Cancer FFPE

The identification of differentially expressed genes is one of the primary uses of RNA-Seq and HTG EdgeSeq platforms. Therefore, the ability to detect differential gene expression on both platforms was evaluated using eight breast cancer FFPE samples, including three triple-negative (ER/PR- and Her2-) and five estrogen receptor positive, Her2-negative (ER+/Her2-) samples. These two breast cancer subtypes are known to have distinct gene expression profiles. The agreement of the log fold changes between the two cancer subtypes analyzed by RNA-Seq and HTG EdgeSeq was assessed using Lin's concordance correlation coefficient (Lc). Both platforms

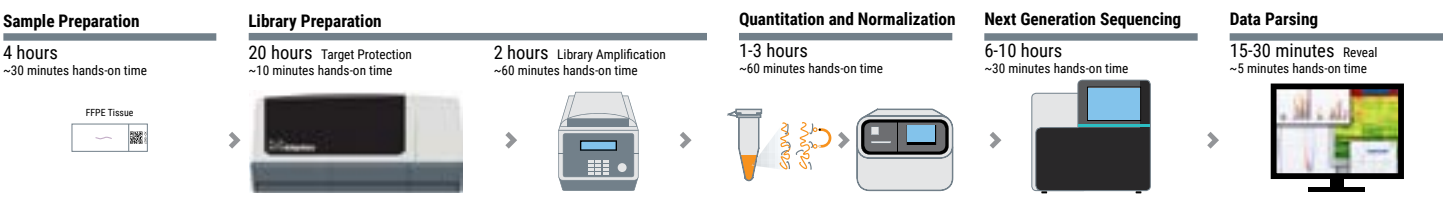

Figure 1. *HTG EdgeSeq workflow*

exhibited similar ability to detect differentially expressed genes, with an Lc value of 0.77 as illustrated in *Figure 2*. These data demonstrate that both platforms perform similarly in the detection of differentially expressed genes among two breast cancer subtypes.

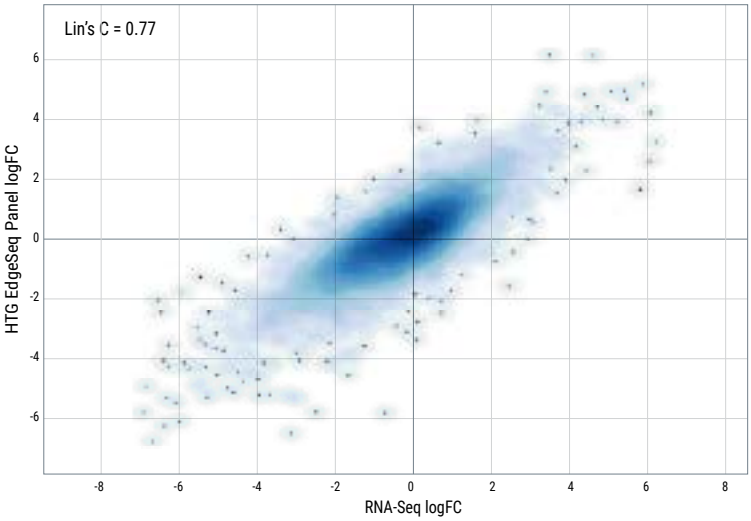

Figure 2. *Comparison of differential expression analysis results between the prototype HTG EdgeSeq whole transcriptome panel and RNA-Seq using two breast cancer subtypes. Each dot represents a gene, log fold change values for RNA-Seq are plotted on the x-axis, log fold change values for the prototype HTG EdgeSeq panel on the y-axis. The Lc value was determined to be 0.77.*

Next, the correlation between a subset of genes known to be differentially expressed between the two breast cancer subtypes was evaluated. *Figure 3* illustrates the correlations between the prototype HTG EdgeSeq panel and RNA-Seq for estrogen *(ESR1)*, progesterone *(PGR)*, and epidermal growth factor *(ERBB2)* receptors. Each dot in the panels represents one of the 16 breast cancer samples tested. The Pearson correlation coefficients comparing the two platforms were 0.96, 0.89 and 0.79 for *ESR1, PGR,* and *ERBB2,* respectively. Collectively, the data from *Figures 2* and *3* demonstrate that the prototype HTG EdgeSeq panel showed good concordance with RNA-Seq for gene expression profiling of tumor FFPE tissues. These FFPE samples had also been evaluated for estrogen and progesterone receptors and Her2 expression by immunohistochemistry (IHC). The gene expression results of ER, PR, and Her2 were compared to IHC across the 16 samples. In *Figure 3*, the prototype HTG EdgeSeq whole transcriptome panel accurately identified most of the samples based on their receptor status.

Accuracy of the Differential Expression Analysis

Next, the accuracy of the differential expression analyses was tested using exogenous RNA controls developed by the External RNA Controls Consortium (ERCC). The ERCC standards (Ambion, Life Technologies) are mixtures of 92 synthetic RNA control transcripts (Mix1 and Mix2). Each ERCC Mix contains four subgroups, consisting of 23 transcripts spanning a 10^6-fold concentration range. In these studies, a different ERCC mixture was spiked into each FFPE sample so that the abundance ratios of the mixtures in the four subpools were 1:1, 1:2, 2:3 and 4:1, as described in Munro et al[3]. By spiking the ERCC standards at defined ratios, the accuracy in quantifying gene expression differences could be determined. Specifically, the observed log fold change values of the ERCC spike-ins were compared to the expected values, which are listed in *Table 2*. The observed log fold-change values for the ERCC spike-ins are also listed in *Table 2* and shown in *Figure 4*. It is important to note that the expected fold-changes can be perturbed slightly by experimental factors that are not related to the assay, such as the amount of RNA in a sample. The results shown demonstrate that the assay can accurately detect the fold-change of ERCC transcripts at four different ratios in a complex whole transcriptome milieu.

Table 2. *Expected and observed fold-changes of ERCC spike-in mixes.*

Subgroup	Mix1:Mix2 Ratio	Expected log Fold-Change	Average Observed log Fold-Change
A	4.00	2	2.16
B	1.00	0	0.19
C	0.67	-0.58	-0.50
D	0.50	-1	-0.83

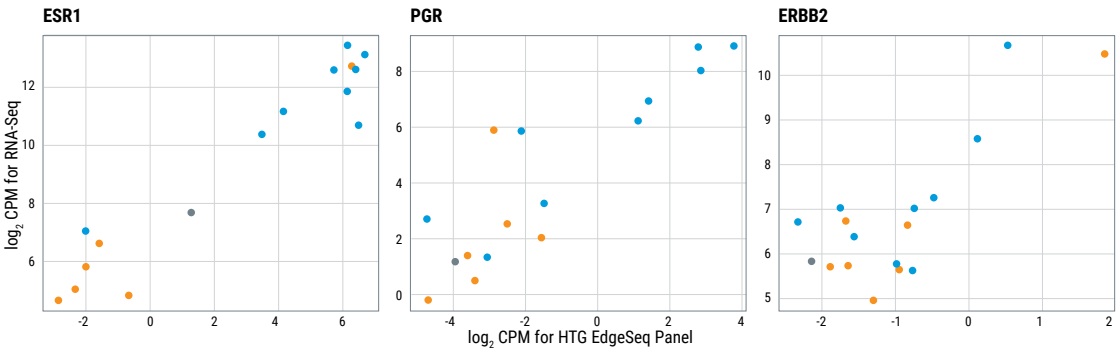

Figure 3. *Correlations between the prototype HTG EdgeSeq panel and RNA-Seq for estrogen (ESR1), progesterone (PGR), and epidermal growth factor (ERBB2) receptors. log_2 counts per million (CPM) values for ESR1, PGR, ERBB2 using the prototype HTG EdgeSeq whole transcriptome panel (x-axis) vs RNA-Seq (y-axis). Each point represents a sample. Point color indicates patient status as assessed by IHC: Blue = Positive, Orange = Negative, Grey = not determined.*

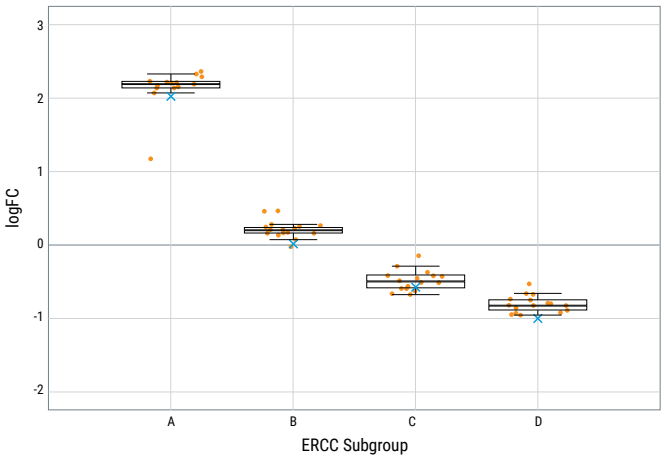

***Figure 4**. Prototype HTG EdgeSeq panel detects ERCC transcripts at predicted values. Breast and kidney FFPE sample lysates were mixed at either 50:50 or 75:25 ratio. ERCC controls were spiked into the samples at different ratios. Subgroup A was spiked into the tissue at a ratio of 4:1 and at 1:1, 2:3, and 1:2 for groups B, C, and D, respectively. log fold-change values are plotted in the y-axis and the four ERCC subgroups are listed on the x-axis. Each dot corresponds to a sample. The X for each subgroup represents the expected log fold change values for each subgroup.*

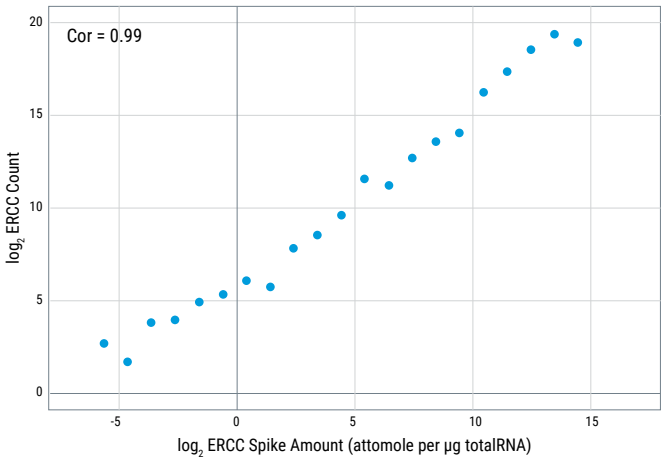

***Figure 5**. The relationship between signal and abundance for ERCC spike-in controls. The log_2 (CPM) for the ERCC transcripts are plotted on the y-axis vs ERCC spike-in amount on the x-axis. The Pearson correlation coefficient was determined to be 0.99.*

The ERCC transcripts were also used to assess the assay's linear range. *Figure 5* depicts the results of an ERCC experiment in a representative FFPE sample. The results demonstrate that the counts generated by the panel for the ERCC transcripts are proportional to, and highly correlated with, the amount of ERCCs spiked into the sample. The data also illustrates that the response is linear across the 10^6-fold concentration range covered by the ERCC controls. Collectively, the ERCC studies demonstrate the accurate measurement of differential gene expression by the prototype assay.

Conclusion

This report evaluates the performance of the prototype HTG EdgeSeq whole transcriptome panel. Collectively, these data show that the panel is comparable to RNA-Seq for gene expression analysis of FFPE tissue and may be superior to RNA-Seq for the analysis of archived FFPE samples in limited quantities. Interestingly, a set of five kidney FFPE samples could not be included in the analysis because RNA-Seq analysis failed to provide informative data due to extremely low diversity libraries prepared from these archival samples. The Lc value, which represents the correlation of log-fold changes determined by each method was 0.77, suggesting high concordance between the two platforms. The ERCC titration experiments show the ability of the HTG EdgeSeq prototype panel to detect transcripts at a predicted fold-change value in a complex whole transcriptome background, highlighting the panel's ability to measure differential expression accurately. Together, these data demonstrate the proof-of-concept for the prototype HTG EdgeSeq whole transcriptome panel. The whole transcriptome panel will provide researchers with an alternative to RNA-Seq for identifying differentially expressed genes, especially when using archival or small FFPE samples, and leveraging the HTG EdgeSeq platform for biomarker discovery and, potentially, development of clinical solutions.

References

1: Wang Z, Gerstein M, Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 10(1):57–63.

2: Ran D, Moharil J, Lu J, Gustafson H, Culm-Merdek K, Strand-Tibbitts K, Benjamin L, Navratil M. (2020). Platform comparison of HTG EdgeSeq and RNA-Seq for gene expression profiling of tumor tissue specimens. J of Clin Oncol. 38, 3566–3566.

3: Munro SA, Lund SP, Pine PS, Binder H, Clevert DA, Conesa A, Dopazo J, Fasold M, Hochreiter S, Hong H, et al. (2014). Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. Nat Commun. 5, 5125–5135.