White Paper
Comparison of the Prototype HTG Transcriptome Panel to RNA-Seq

Abstract: RNA-Seq is considered the gold standard for transcriptome analysis, but it requires complex sample preparation, substantial quantities of extracted RNA, and longer processing time to generate a quality sequencing library. A significant limitation of RNA-Seq is that it does not perform well on partially degraded samples having relatively low-quality RNA. This becomes an issue as archival 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.

The purpose of this White Paper is to present data generated using the prototype HTG transcriptome panel across multiple cancer indications, including melanoma, breast, colorectal, lung and prostate cancer. The panel presented here was able to profile the expression of approximately 20,000 RNA targets, using significantly less tissue than RNA-Seq and in less time. The prototype HTG panel successfully generated data for all samples tested using one or two sections, whereas 25% of samples tested failed to generate sufficient input for RNA-Seq, even with eight FFPE sections.

Collectively, the data presented here demonstrate that the prototype HTG panel is a competitive alternative to RNA-Seq for gene expression profiling while still maintaining the advantages of the targeted HTG chemistry, especially when using archival or small FFPE samples.

Introduction

Gene expression profiling is a powerful tool used to understand disease processes and to identify biomarkers that can be used for diagnosis, prediction of treatment efficacy and disease prognosis. Next Generation Sequencing (NGS) has revolutionized gene expression analysis by allowing researchers to interrogate tens of thousands of genes in a single experiment. Analysis of transcriptome-wide differential expression has provided insights into biological pathways and molecular mechanisms that regulate disease progression.[1]

Currently, RNA Sequencing (RNA-Seq) is considered the gold standard for quantification of transcriptome gene expression. RNA-Seq can quantify the abundance of RNA molecules in the entire transcriptome, detect novel transcripts, and enables discovery of complex genomic features, such as alternative splicing without prior knowledge of the genome sequence. Despite its advantages, RNA-Seq can be challenging in some situations, including in cases of minimal sample availability or partially degraded RNA. One of the most common sample types used in oncology is formalin-fixed paraffin-embedded (FFPE) tissue sections. RNA from FFPE tissues, however, is prone to degradation over time, making it increasingly difficult to extract RNA of sufficient quality and quantity for RNA-Seq analysis. These FFPE samples, and ultimately the extracted RNA, are often available in very limited amounts, especially with the increased testing demands on precious clinical samples. This is particularly true for cancer indications that routinely utilize core needle biopsies and fine needle aspiration biopsies for disease monitoring. Therefore, small FFPE biopsies are often disqualified from RNA-Seq technologies due to the lack of available material, regardless of age.

This is the second White Paper in a series intended to demonstrate the proof-of-concept for a human transcriptome panel that uses the core HTG EdgeSeq™ chemistry (referred to as HTG panel) for gene expression profiling of FFPE tissues. This HTG panel can profile the expression of approximately 20,000 mRNA targets and includes over 200 control probes that help ensure consistent and reproducible panel performance. The first White Paper demonstrated that the prototype HTG panel can accurately measure gene expression levels in breast cancer FFPE samples and showed (1) good directional alignment with RNA-Seq; (2) accurate measurement of differential gene expression as compared to RNA-Seq; and (3) that the response generated by the panel has a linear relationship to the concentration of the analyte present in the sample.[2] Since these initial proof-of-concept studies, the panel design has matured, and several improvements have been made to improve sensitivity and reproducibility.

The purpose of this White Paper is to present new data using the next iteration of the HTG panel across five different cancer indications. First, expression evaluation, carried out using FFPE tissue samples from five cancer indications, showed good separation of indications and a high degree of repeatability between replicates. Second, expression analysis from the HTG panel was compared to RNA-Seq across multiple indications. Third, the linearity and accuracy of the panel was interrogated

using spiked-in reference material in a FFPE tissue background of multiple cancer indications. Lastly, the accuracy of the differential expression analysis was evaluated using a tissue mixture study, as described in Holik et al.,[3] and the results were compared to RNA-Seq. In addition, the HTG panel had a significantly lower sample drop-out rate, particularly for samples sectioned from FFPE tissue blocks older than ten years, and always used fewer sections to meet sample input requirements. Overall, the data presented here expand the utility of the HTG panel from breast cancer to additional cancer indications and shows good directional alignment to RNA-Seq.

Methods

Samples

Twenty-four FFPE tissue blocks from five different cancer indications—melanoma, breast, prostate, lung, and colorectal—were tested on both HTG and RNA-Seq platforms, shown in *Table 1*. For each indication, four to six blocks were included: thirteen FFPE blocks were less than five years old, and eleven blocks were archival FFPE blocks that were at least ten years old. Fresh serial sections were cut at a thickness of 5µm from each block to generate the data described in this paper. For analysis on the HTG panel, 23/24 samples required a single 5-µm-thick section and a single sample required two sections to generate sufficient material following HTG's proprietary extraction-free lysis process whereas four to eight FFPE sections were lysed for RNA-Seq. The minimum sample requirement for RNA-Seq was dictated by the RNA extraction kit manufacturer. While all samples met the sample input requirement for the HTG workflow, six out of the 24 samples failed to generate sufficient extracted RNA to complete the RNA-Seq workflow, even when utilizing eight FFPE tissue sections.

Table 1. *Sample information.*

Cancer Indication	Age of FFPE Block	Number of Blocks	Met RNA-Seq Input Requirement
Breast	< 5 years	2	2/2
	>10 years	2	2/2
Prostate	< 5 years	2	2/2
	>10 years	2	0/2
Lung	< 5 years	3	2/3
	>10 years	2	0/2
Colorectal	< 5 years	3	3/3
	>10 years	3	3/3
Melanoma	< 5 years	3	2/3
	>10 years	2	2/2

HTG EdgeSeq Workflow

Twenty-three out of 24 samples required a single 5-µm-thick section of each FFPE tumor tissue and a single sample required two sections. Samples were lysed using HTG's proprietary lysis buffer and treated with DNase to remove genomic DNA. Samples were then placed on an HTG EdgeSeq processor where gene-specific nuclease protection probes (NPPs) 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, 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 sample testing, RNA-Seq requires total RNA be extracted from FFPE samples. Total RNA was prepared from four to eight 5-µm-thick sections of FFPE tumor tissue using the Qiagen AllPrep DNA/RNA FFPE Kit and DNase treated. While RNA was extracted for all 24 samples, six samples, four of them from blocks older than ten years, failed to generate sufficient material for RNA-Seq analysis. Depending on RNA yield, 100-200 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 twelve cycles of PCR. Library concentration and fragment size distribution were determined using Agilent D1000 Screen Tape Assay on the 4200 TapeStation System. Eighteen of the 24 samples yielded libraries with sufficient library concentrations and were sequenced using Illumina's NextSeq 500/550 High Output Kit v2.5 (150 cycles).

Update to Prototype HTG Panel

This section describes the improvements made to the prototype HTG transcriptome panel, highlighting the differences from White Paper 1 (WP1) to White Paper 2 (WP2). First, WP2 utilized the proprietary HTG Assay Architect, whereas WP1 did not. Assay Architect was designed specifically for use with HTG technology and is a custom software for designing probes based on tightly controlled parameters such as G/C content and melting temperature optimization. Together, this ensures low assay background and consistent performance in the qNPA reaction. Secondly, removal of a large number of non-coding genes, consuming up to 70% of the total reads, focused the targeted panel for WP2 and freed up vital reads necessary for low expressors while reducing the read depth requirement for samples. Lastly, the workflow for WP2 utilizes a more mature version of the HTG extraction-free sample lysis, that includes a new lysis buffer optimized for use with the DNase treatment. This new lysis method is faster than the standard extraction-free method and has the added benefit of removing genomic DNA and reducing assay background.

Results

Sample Qualification and Turn-Around Time

The key sample qualification rate and turn-around time advantages of the HTG panel relative to RNA-Seq are summarized in *Table 2.* First, a significantly higher amount of tissue was needed to generate sufficient extracted RNA for RNA-Seq analysis. Four to eight 5µm FFPE tissue sections were needed to generate sufficient libraries for RNA-Seq, whereas only one 5µm FFPE section was used for 23 of the 24 samples to generate library for the HTG panel. Additionally, six of the 24 samples failed to generate sufficient extracted RNA material for the RNA-Seq workflow and were therefore counted as sample failures. As described in *Table 1*, approximately half of the samples used in this study were older than ten years. Four of the five samples that failed to generate sufficient RNA for RNA-Seq were samples older than ten years. Conversely, all samples tested, regardless of age, successfully generated data using the HTG panel. Finally, samples processed on the HTG platform generated data in less than half the time required to generate data using the RNA-Seq workflow, including sample preparation and sequencing. Collectively, the data presented here highlight that the HTG panel uses significantly less sample overall and has a much lower failure rate for FFPE samples older than ten years.

Table 2. *Comparison of HTG and RNA-Seq platforms*

	HTG Panel	RNA-Seq
Number of FFPE Sections Used	1-2*	4-8
Sample Type Used	FFPE (extraction-free)	FFPE (extracted RNA)
Overall Pass Rate	100% (24/24)	75% (18/24)**
Pass Rate for Samples Older than 10 Years	100% (11/11)	63% (7/11)
Turn-Around Time	3 days	7 days

*Only one sample required two sections.
**RNA-Seq samples failed to generate sufficient extracted RNA to process samples.

Differential Expression and Repeatability Using the HTG Panel

To demonstrate the ability of the HTG panel to generate unique gene expression profiles for multiple cancer indications and show the potential repeatability of the panel, breast, prostate, melanoma and colorectal FFPE samples from *Table 1* were run in triplicate. The panel's ability to separate samples based on their gene expression profiles and cluster indications was assessed using t-Distributed Stochastic Neighbor Embedding (tSNE). The tSNE, a non-linear dimensionality reduction method, was used to visualize the indication clustering in the panel, shown in *Figure 1A* and samples from the same indications were run in duplicate using RNA-Seq and analyzed using tSNE plots, shown in *Figure 1B*. Analysis of the cancer indications tested show all indications exhibit good clustering patterns on the panel. Together, these data show equivalent sample clustering between the two methods tested, based on the gene expression profiles.

Figure 1A tSNE HTG panel all samples

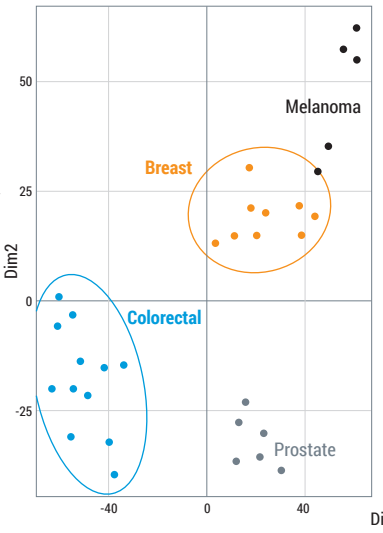

Figure 1B tSNE RNA-Seq all samples

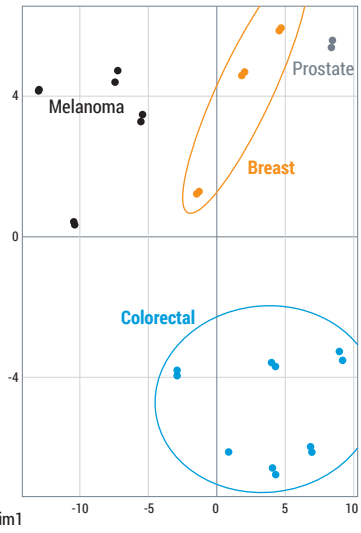

Figure 1. Differentiation of samples based on their gene expression visualized using tSNE. *tSNE plot of replicates from breast, prostate, colorectal and melanoma cancers for both the HTG panel (A) and RNA-Seq (B). Each dot represents a replicate, and each color represents the cancer indication. Circles indicate samples selected for use in Figure 3.*

Additionally, replicates of all five indications run on the HTG panel were evaluated to assess the potential repeatability of the HTG panel. First, a pairwise comparison was performed, and Pearson correlation coefficient (Pearson Cor.) was calculated for replicates of each sample and cancer indication. This analysis was separated for samples sectioned from blocks less than five years old and blocks greater than ten years old to evaluate any potential impact on the data quality caused by FFPE block age. Representative correlation plots for samples less than five years old are shown in *Figure 2, top row*. These example plots are consistent with the observed correlation coefficients across all samples and replicates less than five years old, which ranged from 0.9 to 0.98. Next, Pearson Cor. were calculated for samples sectioned from blocks greater than ten years old to assess any negative impact on replicate correlation using older archival samples. Representative correlation plots for samples greater than ten years old are shown in *Figure 2, bottom row*. These example plots are consistent with the observed correlation coefficients across all samples and replicates greater than ten years old, which ranged from 0.9 to 0.98. These data demonstrate the HTG panel can achieve an equally high degree of repeatability with samples that are less than five years old and with samples that are more than ten years old. This highlights the utility of this assay for archival blocks that may be too degraded to use on other GEP platforms.

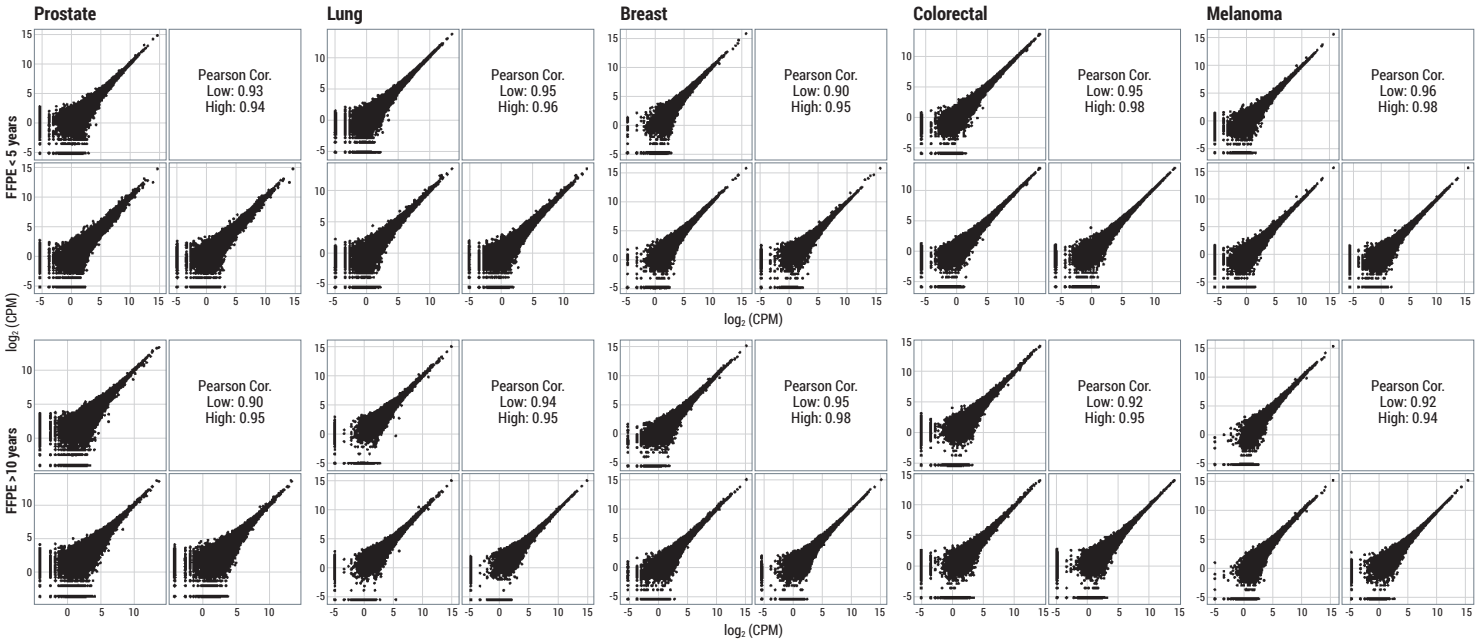

Figure 2. Correlation plots for pairwise comparisons of technical replicates of prostate, lung, breast, colorectal and melanoma cancer tissue samples tested using the HTG panel. *The highest and lowest pair-wise Pearson Cor. are reported in the upper right corner for each sample. Plots on the top row were generated from samples less than five years old and plots on the bottom row were generated from samples greater than ten years old.*

Comparison of HTG Panel Data to RNA-Seq

The ability of the HTG panel to identify differentially expressed genes was compared to RNA-Seq. A similar comparison was carried out in the previous White Paper and showed good correlation to RNA-Seq. Here we are expanding that comparison to include the additional indications of colorectal and breast cancer to demonstrate that these two platforms correlate well across multiple indications. Samples for this comparison were chosen from the tSNE plot, shown in *Figure 1 (circled samples)*. Samples and indications for this analysis were selected based on the indication with the greatest number of overlaps between the two platforms to ensure a more robust data set for analysis and included samples from blocks less than five years old and blocks greater than ten years old. A comparison of the fold changes between the cancer indications on each of the two platforms was used to evaluate the HTG panel's ability to detect differentially expressed genes *(Figure 3)* in comparison to RNA-Seq. Pearson Cor. was calculated to determine the agreement of the log fold changes between the two platforms. Pearson Cor. of 0.82 was calculated for the comparison of these indications, between the two platforms, closely matching the correlation presented in the previous White Paper. Taken together, these data show a high degree of concordance between the two platforms suggesting that the HTG panel is comparable to RNA-Seq for gene expression analysis across multiple indications.

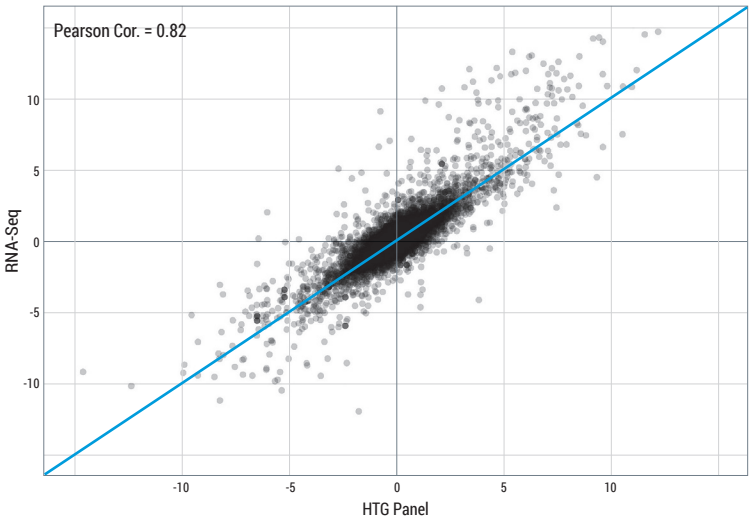

Figure 3. Comparison of differential expression analysis results between the HTG panel and RNA-Seq for a representative set of breast and colorectal cancer FFPE samples. *Log fold change values for RNA-Seq are plotted on the y-axis and log fold change values for the HTG panel are plotted on the x-axis. The blue line represents the unity line.*

Accuracy of Differential Expression Analysis, Linearity and Dynamic Range of the HTG Panel

To assess the accuracy and dynamic range of the HTG panel, differential expression was measured using exogenous RNA controls developed by the External RNA Controls Consortium (ERCC). The ERCCs are spiked in at various known ratios, which then serve as a truth set which is used to benchmark the accuracy of transcript measurement, as described in Munro et al[4]. The control probes used here are a commercially available mixture of 92 synthetic ERCC standards (Ambion, Life Technologies). These standards are divided into two mixes, Mix 1 and Mix 2. Each mix is composed of four subpools of 23 ERCC, each with defined abundance ratios between the two mixes, shown in *Figure 4.* So, for example, a 4:1 ratio means that the 23 ERCC probes from subpool A are at a 4:1 ratio between Mix 1 and Mix 2. ERCC Mix 1 and Mix 2 were spiked into different FFPE lysate preparations (lung, breast, colorectal and prostate). Samples were then processed using the HTG panel and the observed log fold change values of the ERCC spike-ins were then compared to the expected values, listed in *Table 3*, to assess accuracy of differential expression detection. It is important to note that the expected log 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 FFPE matrix.

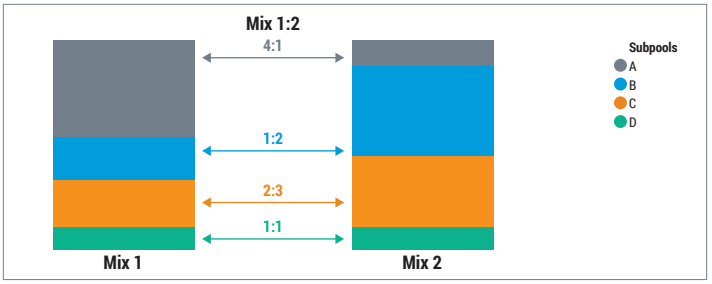

Figure 4. Design of ERCC RNA control ratio mixtures. *Two mixtures of the same 92 ERCC RNA transcripts are prepared such that four subpools with 23 transcripts per subpool are in four defined abundance ratios between the two mixtures. The abundance ratios of the four subpools in Mix 1 and Mix 2 were 4:1, 1:2, 2:3, and 1:1.*

Table 3. *Expected and observed log fold-changes of ERCC subpools.*

Subpool	Ratio of ERCCs Mix 1:Mix 2	Expected log_2 Fold-Change	Observed log_2 Fold-Change
A	4:1	2.00	2.18
B	1:2	-1.00	-1.12
C	2:3	-0.58	-0.58
D	1:1	0	0.09

The HTG panel's linearity and dynamic range were then assessed using the same 92 ERCC standards, which span a million-fold concentration range. Multiple samples and cancer indications were tested because the ERCC standard signal can fluctuate based on the RNA input from the sample. Pearson Cor. were calculated for all five samples and are shown in *Figure 4.* The results show Pearson Cor. between 0.90 - 0.99 for all samples. These data demonstrate that, as expected, the counts generated by the panel for the ERCC transcripts

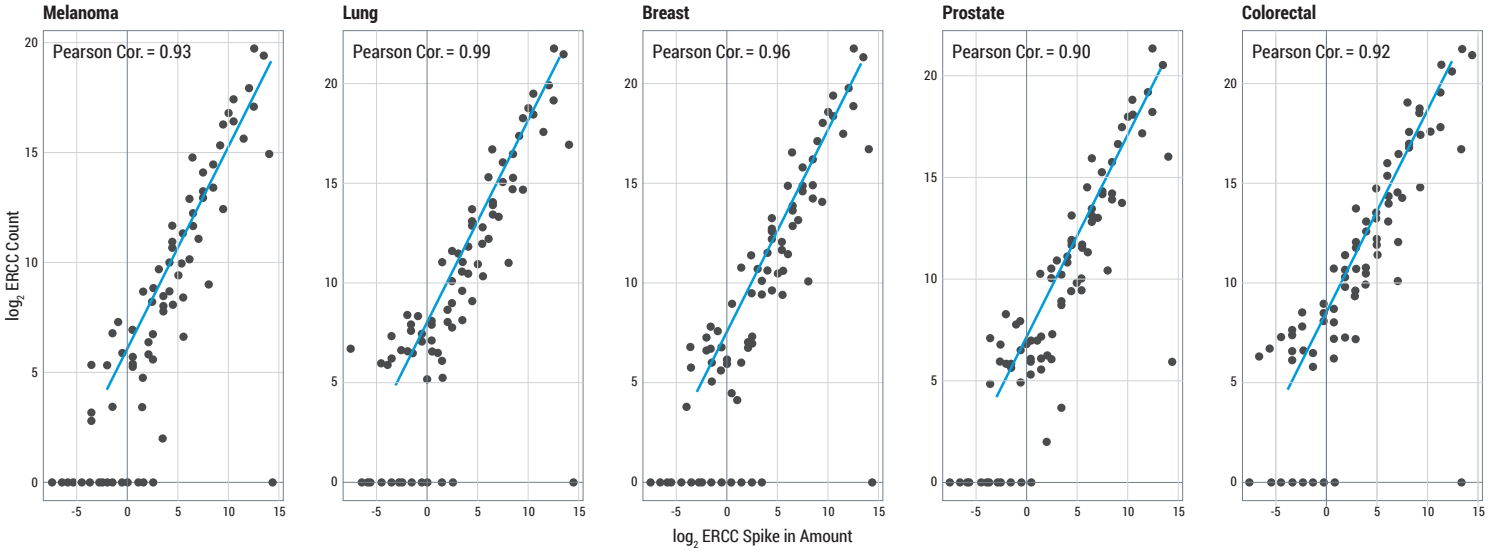

Figure 5. Assay linearity and dynamic range using spiked-in ERCC transcripts. *The predicted amount of log_2 ERCC spike-in (x-axis) were correlated with the actual counts of log_2 ERCC (y-axis). The blue line represents the unity line and Pearson Cor. coefficients are in the upper left-hand corner of each plot.*

are proportional to, and highly correlated with, the amount of ERCCs spiked into each sample. In addition, the data also illustrate that the linear response spans the million-fold concentration range covered by the ERCC controls, indicating a large dynamic range. Together, these data demonstrate that the HTG panel shows strong linearity across a wide dynamic range, further suggesting that this panel is suitable for interrogation of transcriptome-wide gene expression profiling.

Assessment of Accuracy of the Differential Expression Analysis Using a Tissue Mixture Study

A second approach to test the accuracy of the differential expression analysis utilized a tissue mixture study, which is executed by generating mixtures of previously characterized samples in defined ratios, described in Holik et al[3]. The basic premise of the study relies on two samples with distinct gene expression profiles—in this case, melanoma and colorectal cancer samples were used—resulting in many differentially expressed genes. When these two samples are combined at known ratios, the accuracy of differential expression can be assessed by comparing the expected and observed $\log_2$ fold-changes in different tissue mixtures. The gene expression results from each sample run at 100% were used to identify differentially expressed genes between the two samples and to create a prediction model that was then tested with the tissue mixtures. The predicted log-fold changes were correlated with the observed log-fold changes in the mixed samples. The results of the analysis for the HTG panel, in *Figure 6,* show the correlation measured by Pearson Cor. between the observed and predicted log-fold changes. Observed correlations were 0.98 for the HTG panel and 0.97 for RNA-Seq, shown in *Figure 6A* and *Figure 6B*, respectively. The comparison between expected and predicted fold-changes for RNA-Seq display a small bias towards zero for larger absolute fold-changes. Since this bias appears relatively constant it is most likely due to a slight error in the mixing proportions for the RNA-Seq experiment. These findings demonstrate that the differential expression analysis generated by the HTG panel demonstrates accuracy comparable to that of RNA-Seq.

Figure 6A HTG Panel

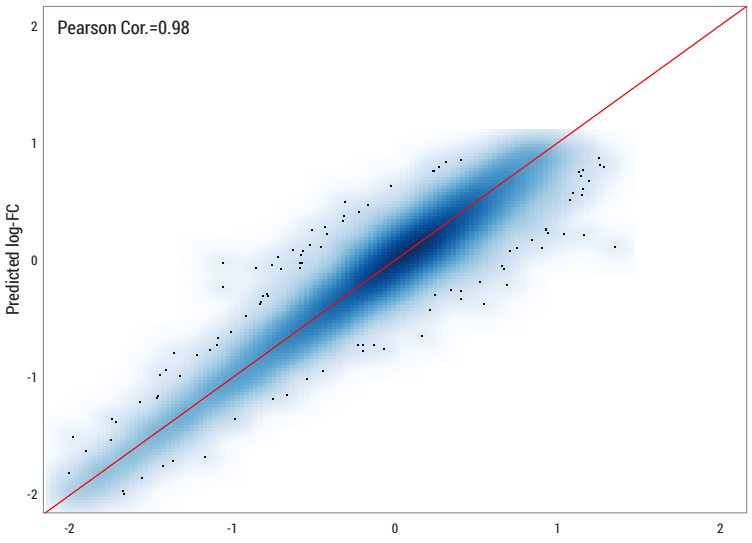

Figure 6B RNA-Seq

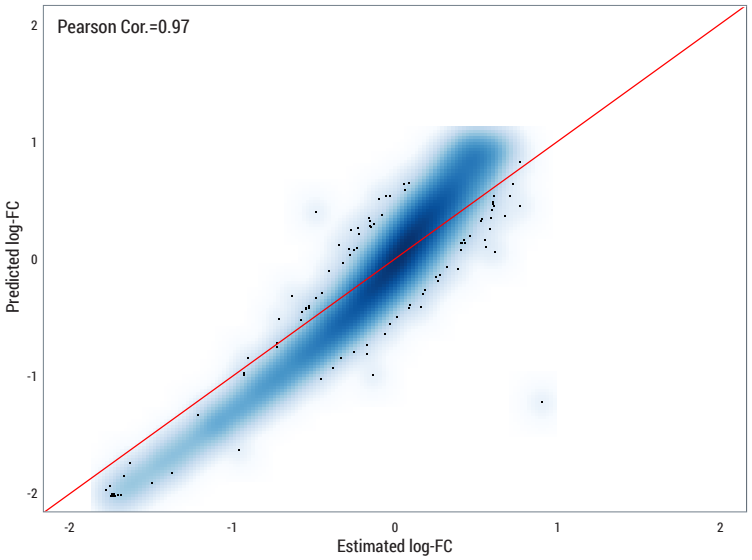

Figure 6. Assessment of accuracy of the differential expression analysis of a tissue mixture study. *The Pearson Cor. between the predicted and observed log fold-changes (log-FC) of gene expression in colorectal and melanoma samples for the HTG panel are 0.98 (A). Pearson Cor. between the predicted and observed log fold-changes in mixtures of colorectal and melanoma samples for RNA-Seq are 0.97 (B). The red line represents the unity line.*

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 data, making it difficult to use in some settings. The data presented in this report demonstrate that the prototype HTG transcriptome panel performs similarly to RNA-Seq in the detection of differentially expressed genes and addresses several of the limitations of RNA-Seq, namely the large amount of FFPE material used, and a complicated and time-consuming workflow.

The HTG panel enables profiling of approximately 20,000 unique mRNA targets in FFPE with a much simpler workflow, and with the ability to accommodate low quality samples or samples that are available in limited amounts. The data presented here show that the HTG panel requires less sample input (typically,

one section) due to the extraction-free sample preparation as compared to four to eight sections required for RNA extraction and subsequent RNA-Seq analysis, for all samples tested in this study. Second, the HTG panel workflow is significantly faster and less labor-intensive compared to the traditional RNA-Seq workflow; thus, providing critical answers in less time. Third, the HTG workflow provides a fully integrated web-based data analysis package that is simple-to-use and allows for a standardized bioinformatics pipeline.

The data presented in this report show that the HTG panel can be used as a competitive alternative to RNA-Seq for identifying differentially expressed genes. The panel demonstrated (1) the ability to differentiate samples based on their gene expression profiles; (2) high repeatability among replicates; (3) exceptional accuracy using spiked-in reference material and (4) equivalent accuracy of differential expression analysis in comparison to RNA-Seq in a tissue mixture study. Overall, the data presented here show that the HTG panel shows excellent robustness across a variety of cancer indications, 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 or FFPE samples with limited amounts. The HTG panel described here provides equivalent results for gene expression analysis, with a fraction of the sample input, high sample pass rate and excellent repeatability across a variety of cancer indications.

References

1: Yan H, Dobbie Z, Gruber SB, et al. (2002) Small changes in expression affect predisposition to tumorigenesis. Nat Genet. 30(1):25–26.

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

3: Holik AZ, Law CW, Liu P, Wang Z, Wang W, Ahn K, Asselin-Labat M, Smyth GK, and Ritchie ME. (2016). RNA-Seq mixology: designing realistic control experiments to compare protocols and analysis methods. Nucleic Acids Res. 45:1-18.

4: 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.