White Paper

Overview of the Design and Performance of the HTG Transcriptome Panel

Abstract: While RNA-Seq is currently considered the gold standard for transcriptome analysis, 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. To address these issues, HTG Molecular Diagnostics has developed a targeted transcriptome panel that provides fast, accurate, and repeatable quantitative gene expression data.

The purpose of this White Paper is to summarize the data generated using the HTG Transcriptome Panel (HTP), from Feasibility through Verification. This panel was evaluated across multiple cancer indications, including melanoma, breast, colorectal, lung and prostate cancer tissue samples. The data presented here demonstrate that the HTG Transcriptome Panel can profile the expression of approximately 20,000 human mRNA targets, using significantly less tissue than RNA-Seq and in less time.

Collectively, the data presented here demonstrate that the HTP is a competitive alternative to RNA-Seq for gene expression profiling while still maintaining the advantages of the HTG EdgeSeq™ platform, including fast turnaround time, high sample success rate and low sample input, especially when using archival or small FFPE samples. At the release of this White Paper the HTG Transcriptome Panel has been formally launched and is available for customers to process in their own laboratory or through HTG's VERI/O laboratory.

Introduction

Gene expression profiling (GEP) 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, detect novel transcripts, and facilitate the discovery of complex genomic features such as alternative splicing all without prior knowledge of the genome sequence. Despite its advantages, RNA-Seq can be challenging in some situations, including in cases of limited sample availability or partially degraded RNA. The HTP addresses several of the limitations related to RNA-Seq. First, HTG EdgeSeq panels 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 efficiency 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 highly abundant non-coding and ribosomal 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 (HTG EdgeSeq Reveal).

This is the third paper in a series, intended to summarize the proof-of-concept, Feasibility and Verification data for a human transcriptome panel that uses the core HTG EdgeSeq chemistry for gene expression profiling of FFPE tissues. The HTP can profile the expression of 19,398 mRNA targets and includes over 200 control probes that help ensure consistent and reproducible panel performance. The first paper, titled "Proof-of-Concept for a Whole Transcriptome Panel Using HTG EdgeSeq Technology", showed that the HTP had

the potential to be used as a competitive alternative to RNA-Seq for gene expression profiling by demonstrating that it could 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] The data generated for the proof-of-concept used breast and kidney FFPE tissue with a prototype workflow.

The second white paper, titled "Comparison of the Prototype HTG Transcriptome Panel to RNA-Seq", presented Feasibility data using the next iteration of the HTP across five different cancer indications, expanding the utility of the HTP from breast to melanoma, prostate, lung and colorectal cancers. During Feasibility, the HTP demonstrated (1) the ability to differentiate indications 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.[3] In addition, the HTP showed a lower sample failure rate compared to RNA-Seq when using archival FFPE samples over ten years old. Overall, the data presented in that paper expand the utility of the HTP from breast cancer to additional cancer indications and shows good directional alignment to RNA-Seq.

The purpose of this white paper, the third in the series, is to briefly review the performance highlights from the previous two white papers and to demonstrate the performance of the design locked HTP workflow using data generated during Verification. New data presented here include (1) sample input robustness, (2) a multi-plate repeatability study, (3) sequencing configuration equivalency study and (4) a comparison to RNA-Seq with the final design locked workflow. Together, this series of three White Papers demonstrates that the HTP is accurate, reproducible and robust with faster turnaround time and higher sample pass rates across a variety of cancer indications. This provides researchers with a competitive alternative to RNA-Seq for identifying differentially expressed genes, allowing researchers to leverage the HTP for biomarker discovery and, potentially, development of clinical solutions.

Methods

Samples

Table 1 shows a list of different sample types used for Verification work with the design locked HTP workflow. See "Proof-of-Concept for a Whole Transcriptome Panel Using HTG EdgeSeq Technology" and "Comparison of the Protype HTG Transcriptome Panel to RNA-Seq" for specific samples used in those papers.

Table 1. *Sample information.*

	Cancer Indication	Sample Input (mm^2/well)	Sample Number	Replicates
Sample Input*	Breast, Colorectal, Lung, Melanoma, Prostate	0.106 to 13.68	2	3 (per input)
Repeatability	Breast, Colorectal, Lung, Melanoma, Prostate	3.42	1	6 (per plate)
RNA-Seq Comparison	Breast, Melanoma	3.42	1	6

** Sample input, 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 genomic DNase treatment. For example, a sample input of 3.42 mm^2/well would require 11 mm^2 FFPE tissue lysed in a minimum of 50 µL Lysis Buffer A.*

HTG EdgeSeq Workflow from Proof-of-Concept to Verification

This section will highlight the similarities and differences between the Proof-of-Concept, Feasibility and Verification workflows. Major changes include probe design, which originally used commercially available probe sequences and evolved to using the proprietary HTG Assay Architect software *(Table 2)*. In addition, any probes that did not meet performance standards during the development of the panel were re-designed and tested to ensure proper function for all 19,398 gene probes. Second, the sample preparation workflow evolved to include an optimized, robust workflow that includes shorter incubation times and a new genomic DNA removal step. Lastly, four unique Quality Control (QC) metrics were established and locked during Design and Development phase to assess RNA quality, sequencing read depth, background and presence of undigested genomic DNA (gDNA). These metrics were applied to all data generated for HTP Verification.

Table 2. *Proof-of-Concept, Feasibility and Verification workflow comparison.*

	Proof-of-Concept	Feasibility	Verification
Probe Design	Commercially available probe design	HTG's Assay Architect software used for probe design	HTG's Assay Architect software used for probe design
	Contained probes for non-coding genes	Removed probes for non-coding RNA	Redesigned, tested any poor performing probes
Sample Preparation	Commercially available reagents used for sample preparation	Samples lysed using new lysis buffer, Lysis buffer A	Optimized and guard banded sample preparation workflow
		Turbo DNase treated to remove gDNA	Determined appropriate equipment for efficient lysis
QC Metrics	None	None	QC0-QC3 applied to all data

Note: Each column in the table above represents work completed before each phase (Proof-of-Concept, Feasibility and Verification) was completed.

HTG EdgeSeq Workflow

Specific details for the HTP workflow for proof-of-concept and Feasibility were captured previously.[2,3] For this paper, samples were lysed using HTG's proprietary Lysis Buffer A and treated with Turbo DNase to remove gDNA. 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, 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 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 v2.5 (300 Cycles).

Results

Sample Success and Turnaround Time

The key sample qualification rate and turnaround time advantages of the HTP during Feasibility are summarized in *Table 3*. First, a significantly lower amount of tissue was needed to generate sufficient sample for HTP compared to the tissue required to generate extracted RNA for RNA-Seq analysis. Additionally, all samples tested, regardless of age, successfully generated data using the HTP. Conversely, four of the five samples that failed to generate sufficient sample for RNA-Seq were samples cut from blocks that were older than ten years. 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 prep and sequencing. Data from Feasibility (data not shown here) showed the HTP platform can achieve an equally high degree of repeatability. Pearson correlation coefficient (Pearson Cor.) of 0.9 to 0.98 were generated from samples less than five years old and greater than ten years old, highlighting the utility of this assay for archival blocks that may be too degraded to use on other GEP platforms. Collectively, the data presented here and in previous HTP White Papers[2,3] highlight that the HTP uses significantly less sample overall and has a much lower failure rate for FFPE samples older than ten years that RNA-Seq.

Table 3. *Comparison of HTG and RNA-Seq platforms.*

Subgroup	HTP	RNA-Seq
Number of FFPE Slides 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% (13/13)	63% (7/11)
Turnaround Time	3 days	7 days

** Only a single sample required two sections.*
*** Samples failed to generate sufficient extracted RNA to process using RNA-Seq.*

Sample Input

To determine the recommended FFPE sample input to use with HTP, eight sample inputs, ranging from 13.68 down to 0.106 (mm^2/well) were tested across five different cancer indications (*Table 1, Figure 1*). Samples were lysed at the highest sample input amount and serially diluted to the final concentrations.

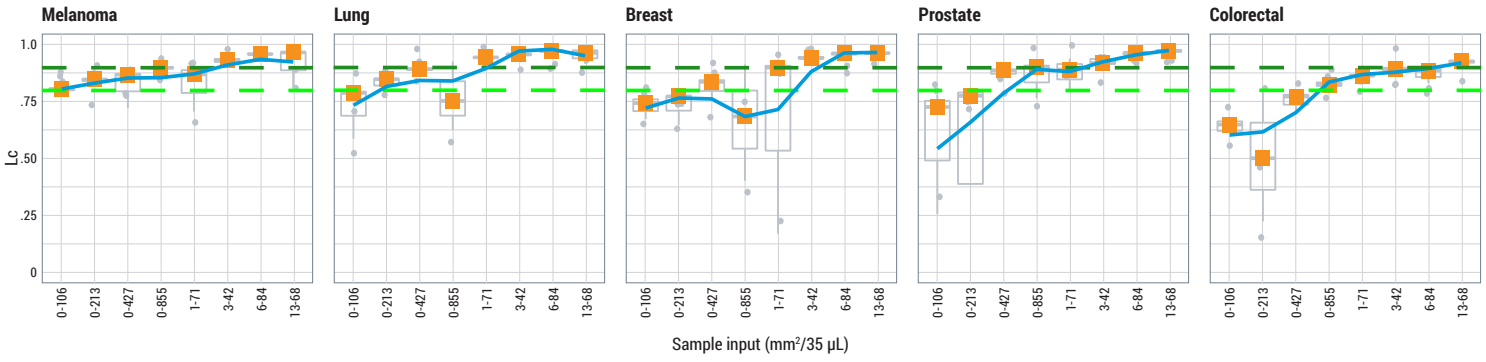


Figure 1. Sample input robustness demonstrated across eight input amounts ranging from 0.106 to 13.68 mm^2/well. *Five cancer indications were run at eight sample inputs (X-axis). The Lc values (Y-axis) at each sample input were determined and plotted. The orange squares represent the median Lc of samples at the given sample input. The blue line represents the smoothing line for the average median Lc values. The dashed lines denotes Lc values of 0.80 (light green) and 0.90 (dark green).*

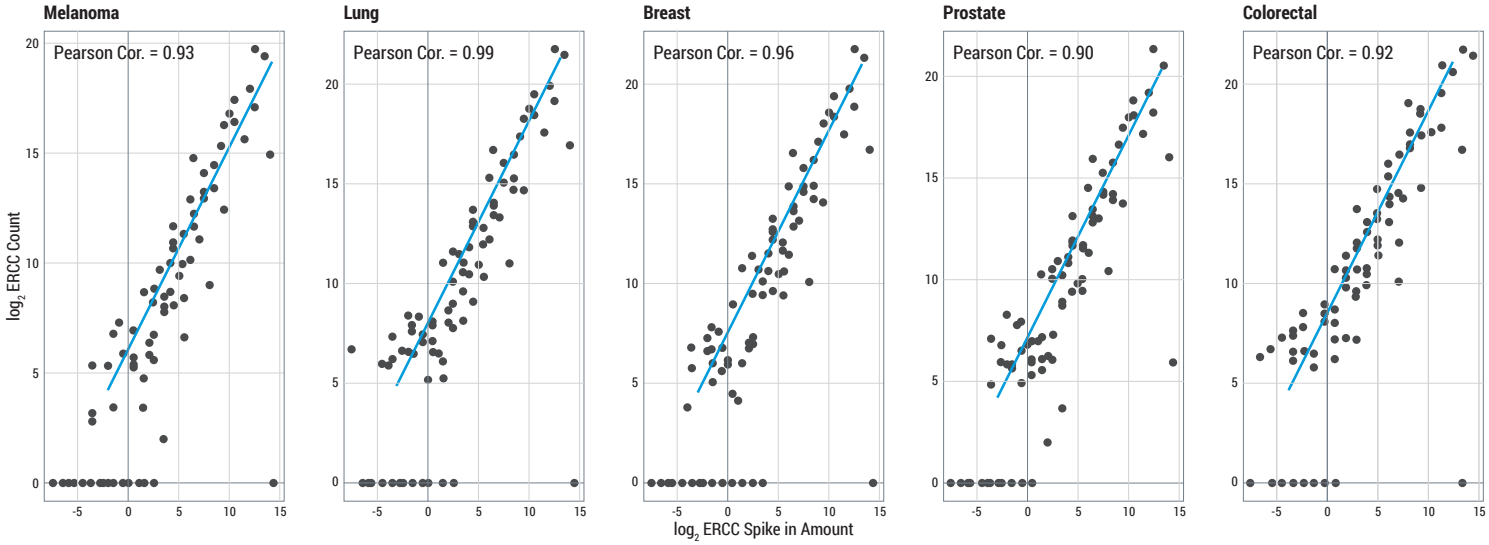


Figure 2. Assay linearity and dynamic range using spiked-in ERCC transcripts. *The predicted amount of log_2 ERCC spike-in (X-axis) were correlated with the actual counts of log^2 ERCC (Y-axis). Pearson Cor. are in the upper left-hand corner of each plot.*

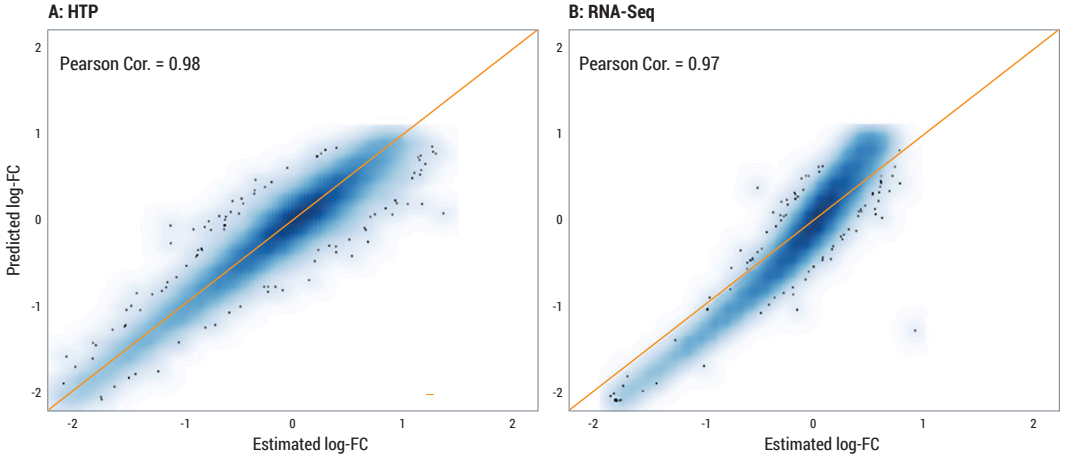
The recommended sample input for the FFPE samples was identified by evaluating technical replicate repeatability by Lin's Concordance Correlation Coefficient (Lc). Lc values above 0.9 suggest high repeatability and are considered robust. *Figure 1* illustrates the Lc values for one sample from each indication across the range of concentrations tested. As expected, the Lc values decreased as the sample input decreased. While Lc values were above 0.9 for most samples run at 1.71 mm^2 per well, Lc values were above 0.9 for all samples when run above 3.42 mm^2. To make sure that the sample input is robust, the recommended sample input for FFPE in the HTP is 3.42 mm^2 per well.

Dynamic Range and Accuracy of the HTG Transcriptome Panel

To assess the dynamic range, differential expression was measured during Feasibility, using exogenous RNA controls developed by the External RNA Controls Consortium (ERCC). To do this, the ERCCs were spiked in at various known ratios, which then serves as a truth set 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) which span a million-fold concentration range. Multiple samples and cancer indications were tested with these spike-ins as the ERCC standard signal can fluctuate based on the RNA input from the sample. The observed linear response spans the 106-fold concentration range covered by the ERCC controls, indicating that the HTP exhibits strong linearity across a large dynamic range *(Figure 2)*.

The accuracy of the HTP was demonstrated during Feasibility with the use of a tissue mixture study, which was executed by generating mixtures of previously characterized samples at defined ratios. The basic premise of the study relies on two samples with distinct gene expression profiles that are combined

Figure 3: 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 HTP 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 orange line is the unity line.*



at known ratios. The accuracy of expression can be assessed by building a prediction model based on each individual sample and comparing the expected and observed log_2 fold-changes in different tissue mixtures.[5] The correlation measured by Pearson Cor., between the observed and predicted log-fold changes were 0.98 for the HTP and 0.97 for RNA-Seq, shown in *Figure 3A* and *Figure 3B*, respectively. These findings demonstrate that the differential expression analysis generated by the HTP is comparable to the accuracy of RNA-Seq.

Performance Robustness

To evaluate the robustness of the final HTP workflow, two studies were executed. The first was a standard precision study utilizing five identical plates containing replicates of five cancer indications *(Table 1)* run across multiple lots, operators, days and instruments. Briefly, multiple replicates of FFPE samples from five cancer indications were run at 3.42 mm^2/well, the recommended sample input for this Panel. The overall precision, measured by Lc, across operators, instrument, day and formulation lots is summarized in *Table 4*. The mean and median Lc values for FFPE were 0.940 and 0.952 *(Table 4)*, respectively, suggesting that the performance of FFPE samples in the HTP is indeed repeatable.

Second, the HTP is available in two configurations for sequencing on an Illumina NextSeq 500/550, 8-sample and 24-sample. To ensure equivalent performance of the panel across the 8 and 24-sample sequencing configuration, the repeatability of gene expression was determined by measuring the Lc values between log_2(CPM) transformed count values of well-pairs between the two configurations. Strong agreement was observed between the 8 and 24-sample sequencing configuration with mean and median Lc values of 0.985 *(Table 4)*. Taken together, these data show good precision in gene expression across operator, instrument, day, and reagent lot as well as good agreement across the two available sequencing configurations.

Table 4. *Performance robustness.*

	Mean	Median
Overall Precision	0.940	0.952
8 vs 24-sample Sequencing configuration	0.985	0.985

Comparison to RNA-Seq

The ability of the HTP to identify differentially expressed genes was compared to RNA-Seq. A similar comparison was carried out during proof-of-concept and Feasibility, and both showed good correlation to RNA-Seq with a Pearson Cor. of 0.77 and 0.82, respectively. Here we are building on that comparison by using the final design locked workflow to accomplish two goals. First, to show that the HTP gene expression correlates well to RNA-Seq and second, to show that the HTP generates very similar results between the Feasibility and Verification workflows. For this study, two individual FFPE samples, a breast cancer and melanoma sample *(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 (RNA-Seq) and FFPE lysing (HTP) events were carried out and six replicates of each sample were run at the recommended sample input for each platform. A comparison of the fold changes between the cancer indications on each of the two platforms was used to evaluate the ability of the HTP to accurately detect differentially expressed genes. Pearson Cor. was calculated to determine the agreement of the log fold changes between the two platforms. *Figure 4* shows a Pearson Cor. of 0.83 for the comparison of these indications, between the two platforms, closely matching the correlation presented in the previous White Paper. Again, these data show a high degree of concordance between the two platforms suggesting that the HTP is comparable to RNA-Seq for gene expression analysis across multiple indications.

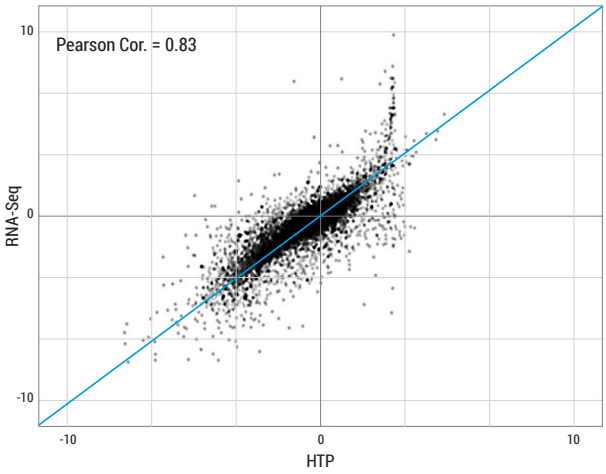


Figure 4: Comparison of differential expression analysis results between HTP and RNA-Seq for a representative set of melanoma and breast cancer FFPE samples. *Log fold change values for RNA-Seq are plotted on the Y-axis and log fold change values for the HTP are plotted on the X-axis. The blue line represents the unity line.*

Integration with HTG EdgeSeq Reveal software

The HTP is fully compatible with the HTG EdgeSeq Reveal software, a powerful, simple-to-use platform for interrogating and visualizing HTG EdgeSeq data. This allows users to gain insights into complex biology by applying QC metrics and using data analysis that enable researchers to gain actionable insights quickly and reliably. The QC metrics were designed to identify samples with insufficient sample input, insufficient sequencing read depth, high background and presence of genomic DNA. These metrics are designed to help researchers exclude low quality sample data and prevent them from affecting the analysis. In addition, there is a new user adjustable gene filtering step *(Figure 5)* that is designed to allow researchers to exclude genes

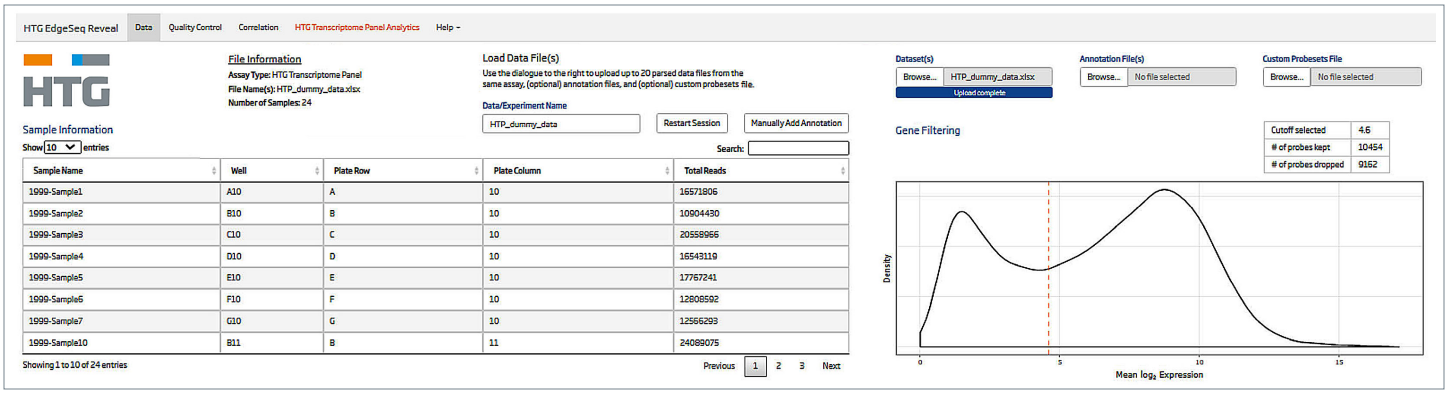


Figure 5: Example of visual output from HTP data the in HTG EdgeSeq Reveal software. *The HTG EdgeSeq Reveal software applies QC metrics, allows for gene filtering and myriad commonly used gene expression profiling tools.*

with low or no expression from further analysis. The HTG EdgeSeq Reveal contains an array of valuable biostatistical tools, such as volcano plots, heatmaps and differential gene expression analysis, that can be used to quickly generate meaningful data and reports without a complicated analysis pipeline.

Conclusions

Although RNA-Seq is widely accepted as the gold standard for genome-wide transcriptome 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 the performance of the HTP and shows that the launched 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 HTP enables profiling of 19,398 unique mRNA targets in FFPE tissue samples 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 HTP requires less sample input and is significantly faster compared to the traditional RNA-Seq. Second, the HTP platform showed accuracy similar to RNA-Seq and is highly reproducible with sample inputs of 3.42 mm^2 of tissue and greater. Third, over the course of these three white papers we have shown that the HTP consistently correlates well with RNA-Seq with correlations of 0.77, 0.82 and 0.83 from proof-of-concept, Feasibility and Verification, respectively. Keeping in mind that these two platforms use different sample input types (extraction-free lysate vs extracted RNA) and different chemistry (targeted vs non-targeted) we believe these correlations to be remarkable. Lastly, the HTP is compatible with the HTG EdgeSeq Reveal software, a web-based application that is simple to use and a powerful tool for interrogating complex data sets and generating publication quality outputs.

Together, the data presented here demonstrates that the HTP is accurate, reproducible and robust, with faster turnaround time and higher sample pass rates across a variety of cancer indications. The HTP can provide researchers with a competitive alternative to RNA-Seq for identifying differentially expressed genes, that may be superior to RNA-Seq for the analysis of archived FFPE samples or FFPE samples that exist in limited amounts, allowing researchers to leverage the HTP for biomarker discovery and, potentially, development of clinical solutions.

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

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.

5: 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.

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

Rev 1: 04-AUG-2021