# White Paper
## HTG EdgeSeq™ Reveal Immunophenotyping Signatures

*Abstract:* Tumors represent a complex microenvironment comprised of a variety of cell types, including proliferating tumor cells, blood vessels, tumor stroma, and infiltrating immune cells. To help researchers identify the relative abundance of immune and stoma cells in the tumor microenvironment, twenty-three new HTG EdgeSeq Reveal RUO Oncology Signatures were developed.

This paper describes the development, verification and performance of twenty-three HTG EdgeSeq Reveal signatures that can be applied to HTG EdgeSeq Precision Immuno-Oncology Panel data to identify the relative abundance of 19 immune and 4 stroma cell types. All signatures were evaluated for accuracy against the reference algorithm (xCell). In addition, spike-in of several immune and stromal cells demonstrated biological accuracy. Lastly, evaluation of the levels of the corresponding protein biomarkers demonstrated strong correlation of the signature outputs to immunohistochemistry results.

The data presented here show that the HTG EdgeSeq Reveal Immunophenotyping Signatures are accurate and robust, allowing researchers to quickly determine the relative abundance of twenty-three immune and stroma cell types using tumor FFPE samples tested with the HTG EdgeSeq Precision Immuno-Oncology Panel.

### Introduction

Tumors are composed of malignant cells embedded in a complex microenvironment comprised of a variety of cell types, including proliferating tumor cells, blood vessels, tumor stroma, and infiltrating immune cells. The composition of the tumor microenvironment (TME), especially the density of immune and stromal cells, has been shown to influence tumor progression and treatment outcomes. Knowledge of the cellular composition of the immune infiltrate of tumors, for example, has elucidated some of the mechanisms used by tumor cells to evade the immune response[1].

The development of gene expression profiling (GEP), which is based on the analysis of multiple genes, has increased our understanding of the role of immune and stromal cells in the TME. These methods are powerful tools to help gather information about the immune landscape of the tumor; however, alone they may not provide sufficient information about the tumor's specific cellular composition (i.e. abundance of specific immune cells). To overcome this limitation, computational methods have been developed to estimate the immune and stromal cellular content of tissue samples[2].

One of the most common computational tools to analyze relative cell abundance is gene enrichment. This approach utilizes the expression of genes that are characteristic for a cell type and computes an enrichment score that is high when the genes specific for a certain cell type are highly expressed in the sample (i.e., the cell type is enriched in the sample) and low otherwise. xCell is a recently published gene enrichment algorithm based on 489 gene sets extracted from large scale expression data from different projects and studies[3]. Similar to other gene enrichment algorithms (CIBERSORT, xCell, MCP, EPIC, TIMER and quanTIseq) the xCell algorithm estimates the relative abundance of several immune and stromal cell types.

The previously released HTG EdgeSeq Reveal Immune, Stroma, and TME signatures were built upon gene expression data from the HTG EdgeSeq Precision Immuno-Oncology Panel (PIP) using the xCell algorithm as a reference. PIP is a Research Use Only (RUO) gene expression profiling assay that measures the expression of 1,392 mRNAs involved in oncology and immune pathways and was designed to measure tumor immune response. To further elucidate the role of the TME in tumor immunobiology, 23 new HTG EdgeSeq Reveal Signatures were developed to build upon the previously released Immune, Stroma and TME signatures. These signatures provide an accurate and robust tool for measuring the relative abundance of 23 specific immune and stromal cell types. This white paper describes the development and validation of these signatures, available through the HTG EdgeSeq Reveal software and will help characterize the immune and stromal response in the TME further.

The 23 HTG EdgeSeq Reveal Immunophenotyping Signatures (*Table 1*) were developed to measure the relative cell abundance of immune or stromal cell types in formalin-fixed, paraffin-embedded tissue (FFPE) from solid tumors using gene expression data from the HTG EdgeSeq Precision Immuno-Oncology Panel. Four of the 23 signatures measure stromal cells, including adipocytes, endothelial cells, epithelial cells, and cancer-associated fibroblasts. The remaining 19 signatures measure the relative abundance of different immune cells, including different subtypes of T-lymphocytes, B-lymphocytes, phagocytes, and granulocytes.

***Table 1.*** *HTG EdgeSeq Reveal Immunophenotyping Signatures.*

| Immune Cells | | | Stromal Cells |
|---|---|---|---|
| *T-lymphocytes* | *B-lymphocytes* | *Mononuclear Phagocytes* | |
| CD4 T-cells | B-cells | Monocytes | Adipocytes |
| CD4 memory T-cells | Naive B-cells | Macrophages | Endothelial Cells |
| CD4 effector memory T-cells | Plasma Cells | Macrophage M1 | Epithelial Cells |
| CD8 T-cells | | Macrophage M2 | Fibroblasts |
| CD8 central memory T-cells | | Dendritic Cells | |
| CD8 effector memory T-cells | | | |
| Regulatory T-cells | | *Granulocytes* | |
| T-helper type 1 cells | | Mast Cells | |
| Natural Killer T-cells | | Neutrophils | |

## Development and Verification of the HTG EdgeSeq Reveal Immunophenotyping Signatures

The steps taken to develop and validate these signatures are illustrated below (*Figure 1*).

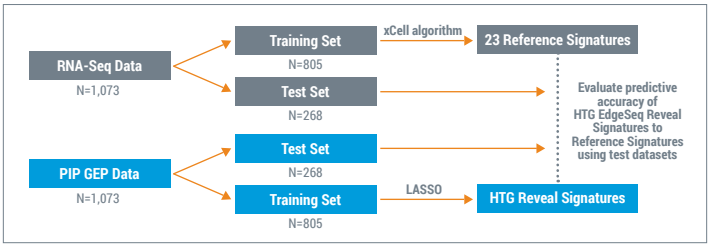



***Figure 1.*** *Overall signature development and verification workflow.*

The signatures were developed using 1,073 commercially procured FFPE cancer samples, including 364 colorectal, 325 gastric, and 384 ovarian cancer FFPE tissues. RNA-Seq and PIP gene expression data were generated at Q2 Solutions in Morrisville, NC and at the HTG VERI/O™ commercial laboratory in Tucson, AZ, respectively. The RNA-Seq and PIP gene expression data were randomly divided into training and test sets. Reference signatures for the 23 cell types were generated by applying the xCell algorithm to the RNA-Seq data from all the FFPE samples. The training set, which consisted of 805 randomly selected samples (75%), was used to develop the 23 HTG EdgeSeq Reveal Immunophenotyping Signatures. The test set, which included the remaining 25% of the samples, was used to assess the performance of the HTG Immunophenotyping Signatures, including determining how well they agreed with the xCell reference signatures. The scores were square root transformed to meet the normality assumption before they were used in the development of the HTG Signature Model.

### Signature Model Development

The HTG EdgeSeq Reveal Immunophenotyping Signatures were developed using a Least Absolute Shrinkage and Selection Operator (LASSO) model, which was applied to the training dataset. The LASSO model is a regularized regression method that incorporates variable selection through a shrinkage step of the parameter estimation wherein the coefficients for unimportant predictors preferentially are shrunk to zero. The LASSO model selected PIP target genes that would accurately predict each of the reference Immunophenotyping Signatures. Each HTG EdgeSeq Reveal Immunophenotyping Signature was computed as a linear combination of the standardized gene expression of selected PIP non-control probes weighted by the corresponding estimated LASSO coefficients.

### Verification

The ability of each HTG Immunophenotyping Signature to accurately predict the corresponding reference signature was assessed across the test dataset and was analyzed using Lin's concordance correlation coefficient (Lc) and Passing-Bablok (PB) regression. The correlation between the reference and predicted scores for the stromal cell signatures across the test dataset were plotted, and the Lc and estimated 95% confidence intervals were determined (*Figure 2*). The Lc values for adipocytes, endothelial cells, epithelial cells, and fibroblasts, were 0.76, 0.82, 0.89 and 0.86, respectively. The estimated PB intercepts were close to 0 and PB slopes were near 1 for all 23 signatures (data not shown), suggesting that there are no constant or proportional biases between the predicted and reference signature scores.

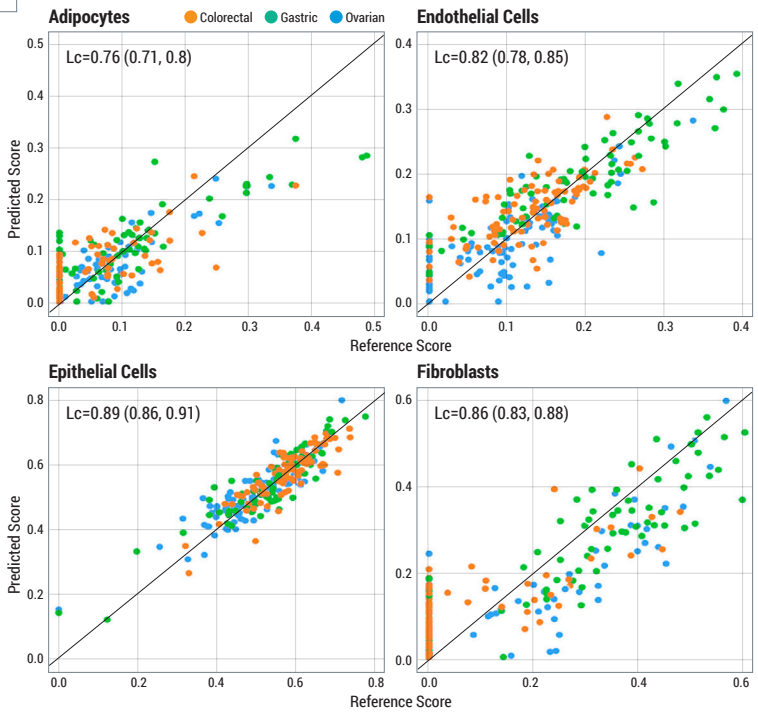



***Figure 2. Agreement between predicted and reference stromal cell signature scores.*** *Comparison of the reference xCell signatures to the corresponding HTG signatures using the test dataset. The y-axis represents the square root of the scores predicted by the HTG signatures and the x-axis represents the square root of the reference xCell signatures. The samples are colored according to cancer type (orange=colorectal cancer; green=gastric cancer; and blue=ovarian cancer). The black line is the unity line. Lin's concordance correlation coefficient (Lc) and the estimated 95% confidence intervals are shown in the upper left corner.*

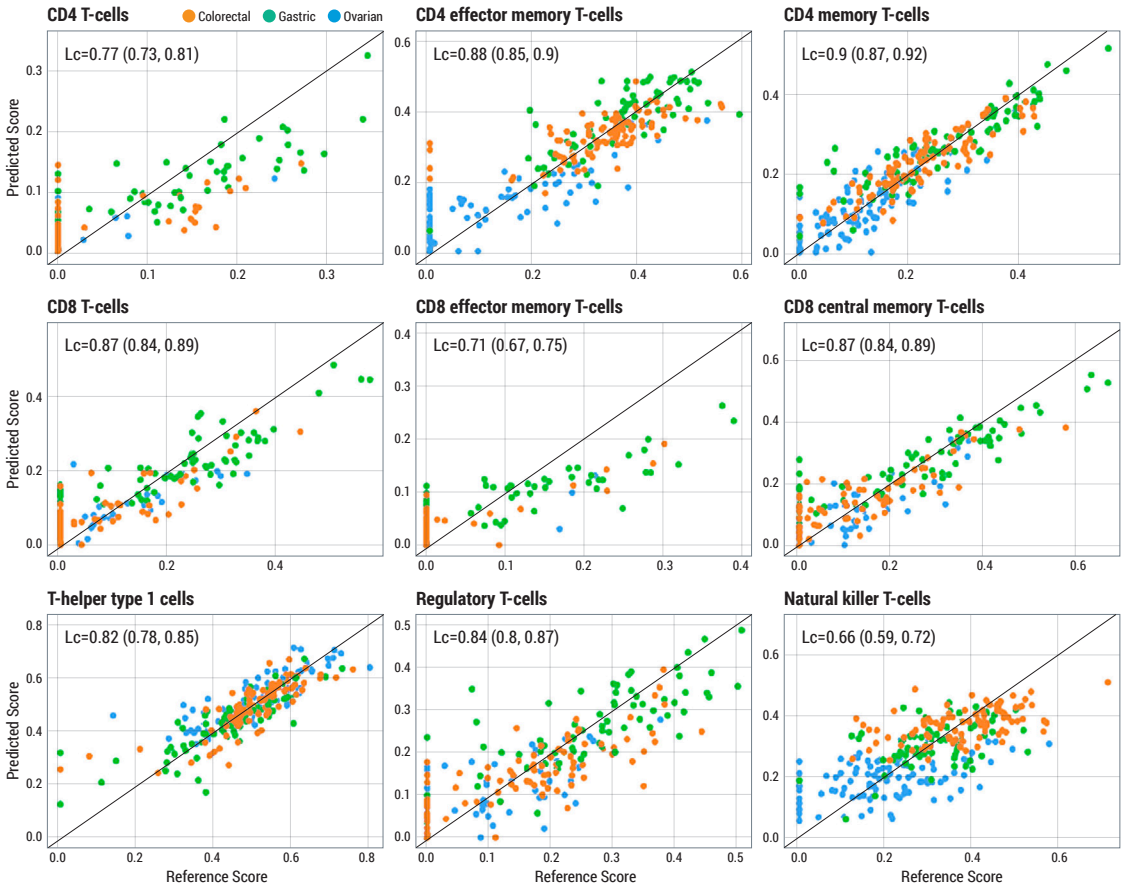



***Figure 3. Agreement between predicted and reference cell signature scores for T-lymphocytes.*** *Comparison of the reference xCell signatures to the corresponding HTG signatures using the test dataset. The y-axis represents the square root of the scores predicted by the HTG signatures and the x-axis represents the square root of the reference xCell signatures. The samples are colored according to cancer type (orange=colorectal cancer; green=gastric cancer; and blue=ovarian cancer). The black line is the unity line. Lin's concordance correlation coefficient (Lc) and the estimated 95% confidence intervals are shown in the upper left corner.*

The correlation between the reference and predicted scores were evaluated for the signatures that are specific to T-lymphocytes, including CD4 T-cells, CD4 memory T-cells, CD4 effector memory T-cells, CD8 T-cells, CD8 effector memory T-cells, CD8 central memory T-cells, T-helper type 1 cells, Regulatory T-cells, and Natural Killer T-cells (*Figure 3*). The Lc values comparing the signature scores from the test dataset and the reference signatures ranged between 0.66-0.9.

The correlation between the reference and predicted scores were evaluated using the test dataset for the B-lymphocyte-specific signatures, including B-cells, naive B-cells, and plasma cells (*Figure 4*). The high Lc values for all three comparisons suggest a high level of agreement between the reference and predicted signature scores for these cell types.

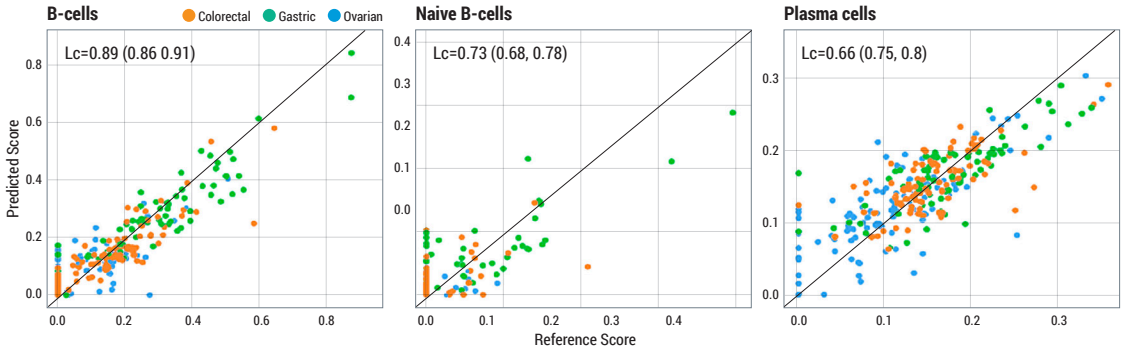



***Figure 4. Agreement between predicted and reference signature scores for B-lymphocytes.*** *Comparison of the reference xCell signatures to the corresponding HTG signatures using the test dataset. The y-axis represents the square root of the scores predicted by the HTG signatures and the x-axis represents the square root of the reference xCell signatures. The samples are colored according to cancer type (orange=colorectal cancer; green=gastric cancer; and blue=ovarian cancer). The black line is the unity line. Lin's concordance correlation coefficient (Lc) and the estimated 95% confidence intervals are shown in the upper left corner.*

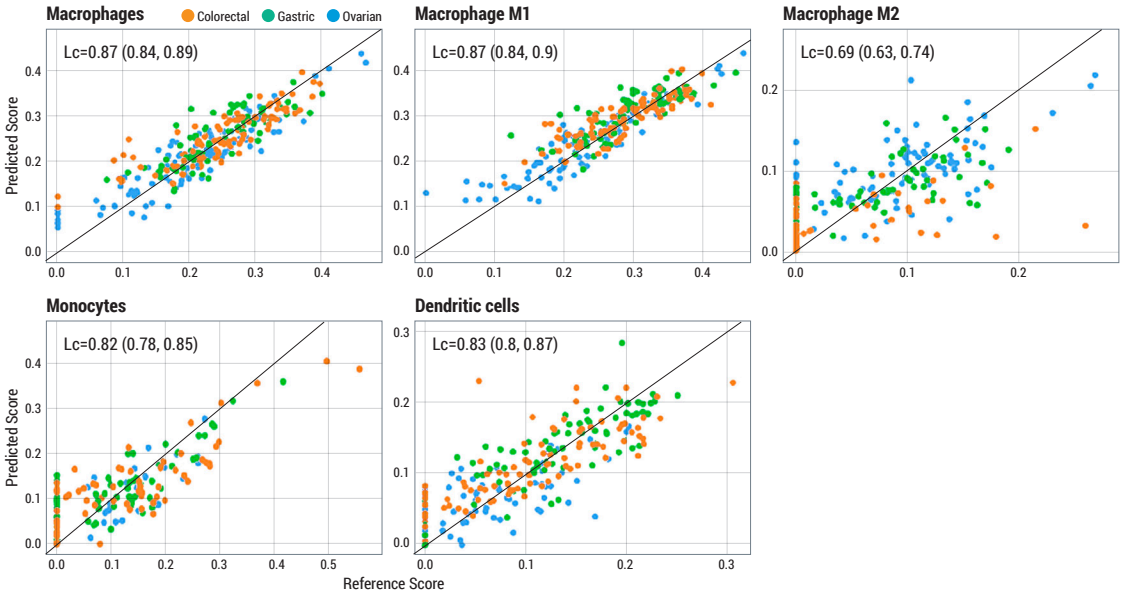



***Figure 5. Agreement between predicted and reference signature scores for mononuclear phagocytes.*** *Comparison of the reference xCell signatures to the corresponding HTG signatures using the test dataset. The y-axis represents the square root of the scores predicted by the HTG signatures and the x-axis represents the square root of the reference xCell signatures. The samples are colored according to cancer type (orange=colorectal cancer; green=gastric cancer; and blue=ovarian cancer). The black line is the unity line. Lin's concordance correlation coefficient (Lc) and the estimated 95% confidence intervals are shown in the upper left corner.*

The correlation between the reference and predicted scores were evaluated for the signatures that are specific to mononuclear phagocytes (*Figure 5*). The Lc values were 0.87, 0.87, 0.69, 0.82, and 0.83 for the Macrophages, Macrophage M1, Macrophage M2, Monocytes, and Dendritic cell signatures, respectively.

Lastly, the correlation between the reference and predicted scores were evaluated across the test dataset for the signatures that are specific to granulocytes, including neutrophils and mast cells (*Figure 6*). The Lc values were 0.68 for mast cells and 0.84 for neutrophils.

Overall, the Lc values comparing the reference scores generated using the xCell algorithm and the HTG EdgeSeq Reveal Immunophenotyping Signature models ranged from 0.66-0.90. For some signatures, the agreement was impacted by the presence of excess zeros or a narrow dynamic range of scores. The high amount of zero scores increased the variability and underestimated the predicted values compared to the reference. Consequently, some of these signatures were tested in the accuracy studies described below to further assess the performance of the signatures. Taken together, the results demonstrate a good degree of correlation between the reference scores generated using the xCell algorithm and the HTG EdgeSeq Reveal Immunophenotyping Signature model.

## Spike-in of Immune and Stromal Cells Demonstrates Biological Accuracy of the Immunophenotyping Signatures

The next step was to confirm that the signatures accurately measured the abundance of immune and stromal cells in tumor tissue. The biological accuracy of nine signatures, including signatures for B-cells, naive B-cells, CD4 T-cells, CD8 T-cells, regulatory T-cells, neutrophils, macrophages, dendritic cells, and epithelial cells, was evaluated by spiking these cell types into a gastric cancer FFPE tissue lysate and testing whether the signature outputs correlated with the number of cells in the

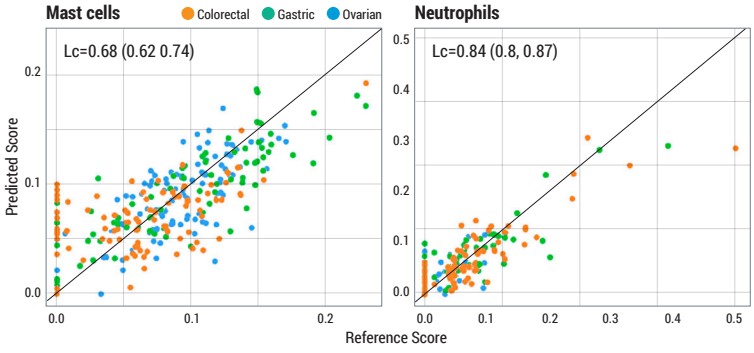



***Figure 6. Agreement between predicted and reference signature scores for granulocytes.*** *Comparison of the reference xCell signatures to the corresponding HTG signatures using the test dataset. The y-axis represents the square root of the scores predicted by the HTG signatures and the x-axis represents the square root of the reference xCell signatures. The samples are colored according to cancer type (orange=colorectal cancer; green=gastric cancer; and blue=ovarian cancer). The black line is the unity line. Lin's concordance correlation coefficient (Lc) and the estimated 95% confidence intervals are shown in the upper right corner.*

lysate. These signatures were chosen based on the importance of the cell type in tumor immunobiology and to test signatures that contained excessive zero scores, which influenced the signature's predictive performance in the development dataset. The gastric cancer FFPE sample had been run on PIP and was screened using the immune and stroma signatures. The sample had low immune and stroma scores.

The cells used were commercially procured human primary immune or stromal cells that were spiked into an FFPE lysate at concentrations ranging from 500 to 4,000 cells/well. PIP gene expression data were used to calculate the HTG EdgeSeq Reveal Immunophenotyping Signature scores at each concentration and for every cell type tested. The correlation between the scores and the cell number was plotted and analyzed by calculating the Pearson correlation coefficient, which expresses the linear covariation between two sets of scores, for each cell type. As expected, the signature scores increased as more cells were spiked into the FFPE sample, and the Pearson correlation coefficients ranged from 0.71-0.98 (*Figure 7*). Collectively, these findings demonstrate that there is a strong correlation between the number of immune and stromal cells present in the tissue and the Immunophenotyping Signature outputs; thereby, confirming that the HTG EdgeSeq Immunophenotyping Signatures tested can be used to estimate immune and stromal cell infiltrates in the tumor robustly and with good accuracy.

## Method Comparison Demonstrates Strong Correlation Between Signature Outputs and IHC

Currently, immunohistochemistry (IHC) is one of the most common methods used to estimate the immune and stromal cell content within a sample; therefore, the outputs of some of the signatures were compared to multiple commonly used immune cell IHC markers. Sixty-one non-small cell lung cancer (NSCLC) FFPE samples were run on PIP and stained for CD4 and CD8. IHC analyses were performed by a pathologist to assure that areas with viable tumor cells and non-necrotic areas were analyzed. Briefly, the number of cells staining positive for each immune cell marker were counted across five areas. The tumor area immune cell count (TAIC), or the average number of positively stained immune cells per $mm^2$ for each marker, was compared to the outputs of several Immunophenotyping Signatures. The correlations between the TAIC for CD8 and CD4 and the CD8 and CD4 T-cell Immunophenotyping Signature scores across 61 NSCLC FFPE samples were plotted and the Pearson correlation coefficient

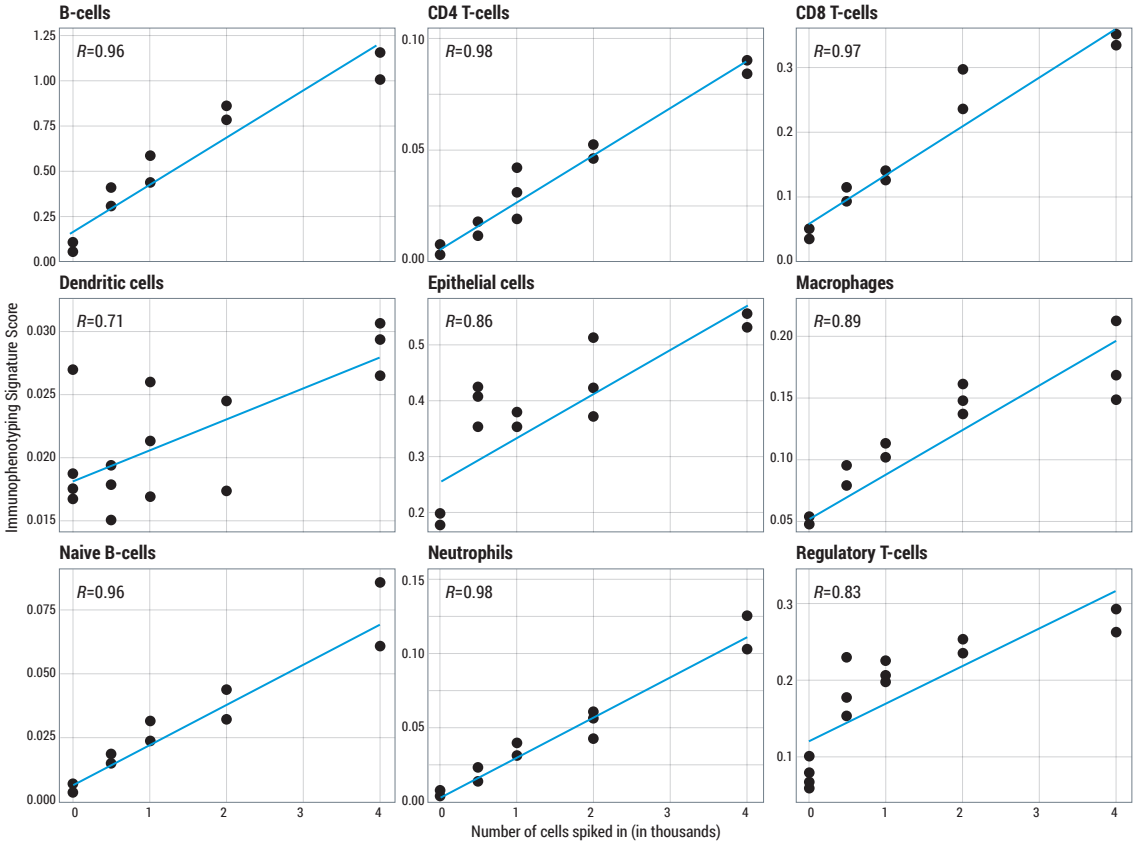



***Figure 7. Correlation between signature outputs and the amount of immune or stromal cells spiked in.*** *The y-axis represents the scores predicted by each of the HTG signatures and the x-axis is the number of immune or stromal cells spiked into the FFPE sample. The Pearson correlation coefficient is shown in the upper left corner for each cell type. The blue line represents the estimated linear regression line.*

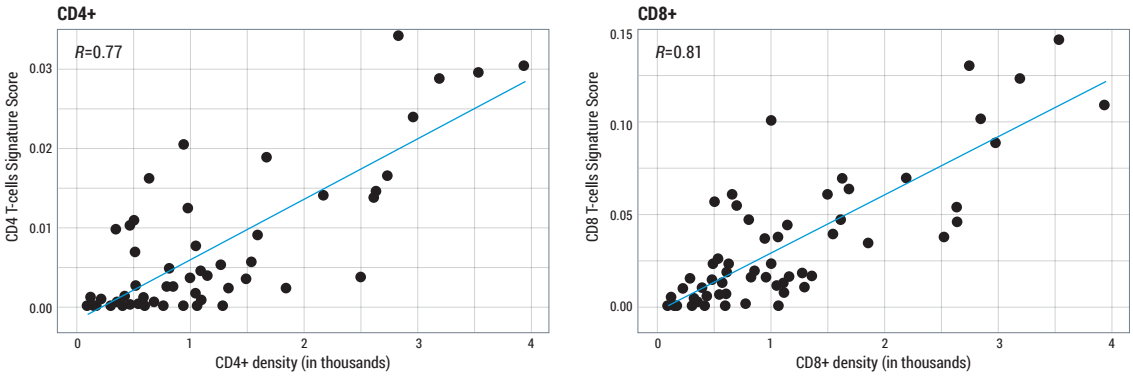



***Figure 8. Strong correlation between signature outputs and IHC stains for CD4 and CD8 T-cells.*** *The y-axis represents the signature output for CD4 T-cells (left panel) and CD8 T-cells (right panel). The x-axis displays the density of cells staining positive for CD4 or CD8. The Pearson correlation coefficient is shown in the upper left corner. The blue solid line represents the estimated linear regression line.*

were calculated (*Figure 8*). The Pearson correlation coefficient between the CD4 T-cell signature scores and the CD4 IHC staining was 0.77, showing good correlation. Similarly, the Pearson correlation coefficient between CD8 IHC and the signature scores for CD8 T-cells was 0.81, demonstrating a strong, positive correlation. Overall, these results demonstrate a high degree of correlation between the HTG EdgeSeq Reveal Immunophenotyping Signatures tested and the levels of corresponding protein biomarkers by IHC.

This study also provides further evidence that the signatures are tissue-agnostic since they can accurately predict the CD4 and CD8 T-cell content in a separate cohort and cancer type.

## Precision of HTG EdgeSeq Reveal Immunophenotyping Signatures

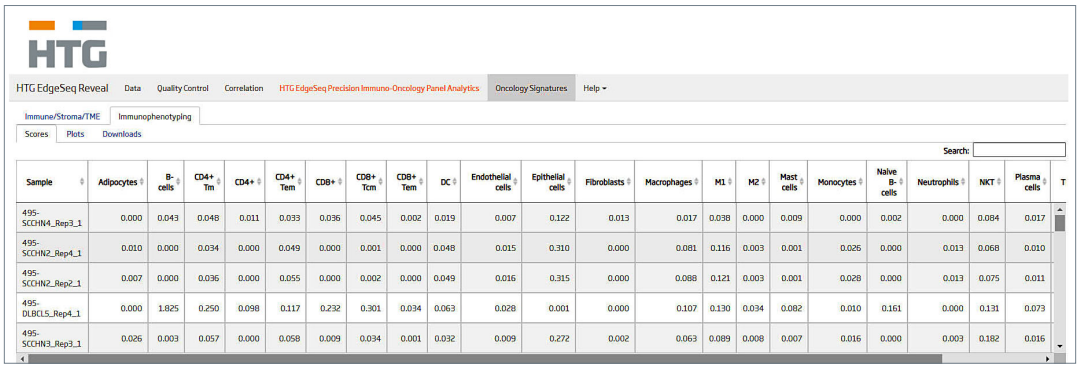
The precision of the Immunophenotyping Signatures was evaluated across intra-plate (well to well), and inter-experimental factors (including manufacturing lot, operator, and processor). Ten FFPE samples, covering melanoma, ovarian, colorectal, and gastric cancers, were run at the recommended sample input for PIP. The precision of each signature was assessed by calculating an interclass correlation coefficient and a variance component analysis using a mixed-effect model. These analyses revealed high repeatability between experimental factors and technical replicates. The imprecision of each sample, expressed as the standard deviation (SD), ranged from 0.003 to 0.03 for all the signatures.

## Signature Implementation and Data Interpretation

The 23 Immunophenotyping Signatures are applied to HTG EdgeSeq PIP data using the HTG EdgeSeq Reveal software. The HTG EdgeSeq Reveal software is a fully integrated web-based data analysis software suite that can analyze data quality and generate publication quality figures. The available HTG EdgeSeq Reveal Oncology Signatures include the previously released Immune, Stroma and TME signatures and now the 23 Immunophenotyping Signatures. Any HTG EdgeSeq PIP datasets, including new and previously generated datasets, loaded into the Reveal software will cause a new tab to appear titled 'Oncology Signatures'. Clicking on this tab will generate two new tabs, labeled Immune/Stroma/TME and Immunophenotyping. Clicking on "Immunophenotyping" will open the Immunophenotyping Signature tab. Within this tab the user will be able to interrogate samples for their immunophenotyping score using several bioinformatic tools including Score, Plots and Downloads (*Figures 9-11*). The "Plots" tab allows researchers to generate heatmaps and radar charts, both with the ability to customize the cell types that are included.

When analyzing the output from these signatures, it is important to keep in mind that the numbers generated are relative abundance and do not indicate a specific number of cells for each cell type. Most relative abundance scores in the data described in this paper fell between the values of 0 and 1.0, however there is no limit to the scale and numbers above

***Figure 9. Scores.*** *Clicking on this tab will generate scores for each of the 23 Immunophenotyping Signatures for each sample passing QC metrics. This easy to read table has samples in the left column and lists scores for each cell type in the rows adjacent to each sample. Scores can be downloaded under the "Downloads" tab. Note: there is an additional QC metric step applied to samples when generating immunophenotyping scores. While unlikely, it may result in the loss of additional samples.*



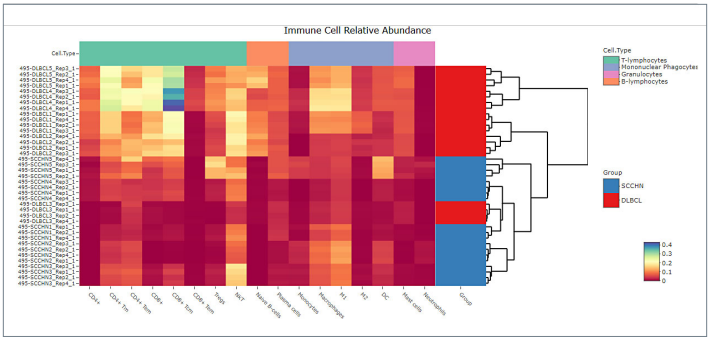



***Figure 10. Heatmap.*** *A data visualization technique that shows magnitude of a phenomenon as color in two dimensions and is a great way to identify trends in large datasets. Clicking on this tab will generate a heatmap which will allow the user to cluster datasets by samples or cell types. Individual cell types can be added and removed from the chart by selecting them on the left side of the screen.*

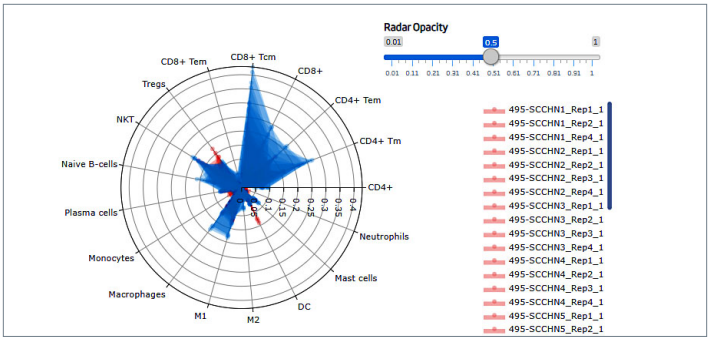



***Figure 11. Radar charts.*** *A graphical method of displaying multivariate data in the form of two-dimensional projections that tend to preserve separation of clusters as well as information on the relative contributions of various markers in differentiating phenotypes. The transparency of the radar chart can be adjusted using the "Radar Opacity" function and individual cell types can be added and removed from the chart by selecting them on the left side of the screen.*

1.0 were recorded for some cell types, although the accuracy may be impacted for scores significantly outside the dynamic range of scores observed in the training and test sample sets. Once data analysis is complete, use the "Download" tab to download your data. The Immunophenotyping scores can be downloaded in a single Excel file containing three tabs; raw data, QC summary and Immunophenotyping scores for additional biostatistical analysis. In addition, this tab allows researchers to customize the heatmap to meet publication requirements including figure size, font size and file format.

## Conclusion

This paper describes the development, verification and performance of twenty-three HTG EdgeSeq Reveal signatures, based on established, peer-reviewed Immunophenotyping Signatures (xCell). Primary development and verification of the HTG EdgeSeq Reveal Immunophenotyping Signatures used 1,073 FFPE tumor tissue samples, run on both RNA-Seq and HTG EdgeSeq Platforms, to generate training and test data. Accuracy in the prediction of all 23 signature scores was evaluated using Lin's concordance correlation coefficient, which generated Lc values between 0.66 and 0.9. In addition, Passing-Bablok regression analysis did not indicate the presence of any proportional bias. Additional evaluation of the signatures included testing spike-ins of cells and comparison to IHC markers. Series of nine purified stromal and immune cells were spiked into a FFPE background to confirm the HTG EdgeSeq Immunophenotyping Signatures can be used to accurately estimate immune and stromal cell infiltrates. Pearson correlation coefficients ranging from 0.71-0.98 demonstrated the biological accuracy of these signatures. Additional evaluation between the signature output for CD4 and CD8 T-cells as compared to IHC showed strong correlation with Pearson correlation coefficients of 0.77 and 0.81, respectively. The data presented above show that these signatures correlate well to three different methods of evaluation, RNA-Seq, purified cell spike-ins and IHC. Lastly, the precision of the Immunophenotyping Signatures was evaluated revealing high repeatability between experimental factors and technical replicates showing that these signatures are both accurate and repeatable. These signatures are available in the HTG EdgeSeq Reveal software for datasets generated using the HTG EdgeSeq Precision Immuno-oncology Panel.

## References

1: Fridman WH, et al. The immune contexture in human tumours: impact on clinical outcome. Nat. Rev. Cancer. 2012. 12:298-306.

2: Sturm G, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics. 2019. 35:436-445.

3: Aran D, et al. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017. 18:220.