Journal of Immunological Methods xxx (xxxx) 113041

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Cross-platform comparison of immune-related gene expression to assess intratumor immune responses following cancer immunotherapy

Li Zhang[a, b, c], Jason Cham[d], James Cooley[e], Tao He[f], Katsunobu Hagihara[a, c], Hai Yang[b, c], Frances Fan[a], Alex Cheung[a], Debrah Thompson[e], B.J. Kerns[e], Lawrence Fong[a, c, *]

[a] Department of Medicine, University of California San Francisco, San Francisco, USA
[b] Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, USA
[c] Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA
[d] Department of Internal Medicine, Scripps Green Hospital, La Jolla, USA
[e] HTG Molecular Diagnostics, Inc., Tucson, AZ, USA
[f] Department of Mathematics, San Francisco State University, San Francisco, USA

ARTICLE INFO

Keywords:
Gene expression profiling
HTG EdgeSeq
Nanostring platform
Neoadjuvant immunotherapy
Classification performance
Prostate cancer

ABSTRACT

Neoadjuvant immunotherapy can induce immune responses within the tumor microenvironment. Gene expression can be used to assess responses with limited amounts of conventionally-fixed patient-derived samples. We aim to assess the cross-platform concordance of immune-related gene expression data. We performed comparisons across three panels in two platforms: Nanostring nCounter® PanCancer Immune Profiling Panel (nS), HTG EdgeSeq Oncology Biomarker Panel (HTG OBP) and Precision Immuno-Oncology Panel (HTG PIP). All tissue samples of 14 neoadjuvant GM-CSF treated, 14 neoadjuvant Provenge treated, and 12 untreated prostate cancer patients were radical prostatectomy (RP) tissues, while 6 prostatitis patients and 6 non-prostatitis subjects were biopsies. For all 52 patients, more than 90% of the common genes were significantly correlated ($p < 0.05$) and more than 76% of the common genes were highly correlated ($r > 0.5$) between any two panels. Co-inertia analysis also demonstrated high overall dataset structure similarity (correlation > 0.84). Although both dimensionality reduction visualization analysis and unsupervised hierarchical cluster analysis for highly correlated common genes ($r > 0.9$) suggested a high-level of consistency across the panels, there were subsets of genes that were differentially expressed across the panels. In addition, while the effect size of the differential testing for neoadjuvant treated vs. untreated localized prostate cancer patients across the panels were significantly correlated, some genes were only differentially expressed in the HTG panels. Finally, the HTG PIP panel had the best classification performance among the 3 panels. These differences detected may be a result of the different panels or platforms due to their technical setting and focus. Thus, researchers should be aware of those potential differences when deciding which platform and panel to use.

1. Introduction

Gene expression profiling (GEP) is a powerful tool with a wide range of purposes, including identifying relevant immune resistance mechanisms in the tumor microenvironment (Gajewski et al., 2011) and developing predictive biomarkers for active immunotherapy (Melero et al., 2014; Beard et al., 2013). There is a current need to perform highly multiplexed molecular testing to assess RNA expression from a variety of clinically relevant sample types, including: formalin-fixed, paraffin embedded (FFPE) tissue, biofluids, cell lysates, and extracted RNA. One limitation of gene expression profiling for molecular researchers is the difficulty of nucleic acid extraction from small and sometimes highly degraded samples and the introduction of bias by cDNA synthesis. In the case of biofluids, large amounts of this sample type are required to produce enough isolated RNA to test by traditional methods.

Two platforms commonly used for targeted GEP analysis are the HTG Molecular Diagnostics (HTG) EdgeSeq Platform and the nanoString nCounter Platform. HTG has developed the HTG EdgeSeq assays, a coupling of RNA extraction-free, nuclease protection assay with

* Corresponding author at: University of California San Francisco, 513 Parnassus Avenue, Room HSE301A, Box 0519, San Francisco, USA.
E-mail address: lawrence.fong@ucsf.edu (L. Fong).

https://doi.org/10.1016/j.jim.2021.113041
Received 31 August 2020; Received in revised form 14 March 2021; Accepted 16 March 2021
0022-1759/

next-generation sequencing (NGS) quantification of up to 2600 genes with a partially automated workflow. NanoString has developed the nCounter gene expression assays, which utilizes fluorescent-based optical-barcode targeted against genes of interest for a non-amplified measurement of gene expression. The nanoString technology is able to detect gene expression of up to 800 genes with an automated workflow. While both assays utilize short DNA based probes to hybridize to targets, HTG is able to boost both the number of probes available in each panel as well as to utilize the high sensitivity and wide dynamic range available with NGS detection. NanoString on the other hand is able to avoid amplification bias and shorten hands on workflow time by using fluorescent-based barcodes.

Here, we compare the immune-related gene expression data across three panels: the HTG EdgeSeq Oncology Biomarker Panel (HTG OBP) and Precision Immuno-Oncology Panel (HTG PIP), and the nanoString nCounter® PanCancer Immune Profiling Panel (nS Immune) to assess the cross-platform concordance. Both HTG OBP and HTG PIP are research use only (RUO) gene expression profiling assays that measure expression of mRNAs involved in but not limited to oncology, immune pathway mapping and immune landscape assessment. HTG OBP detects 2560 transcripts associated with tumor biology, highlighting 24 groups and pathways. The probe set for HTG PIP comprises 1392 nuclease protection probes, including 4 negative control probes (ANT), 4 positive control probes (POS), and 10 process control probes. Gene groupings include current drug targets, lymphocyte lineage markers, B, T and NK cell activity markers, DNA repair genes, Toll-like receptors, interleukin and chemokine markers, and proteasome-associated genes. Both HTG panels are compatible with clinically relevant sample types such as FFPE, isolated cells, PAXgene and extracted RNA. The nS Immune panel is also an RUO multiplexed gene expression panel designed to quantify 770 genes. It is also fully compatible with clinically relevant sample types such as fresh-frozen tissue, FFPE tumor sections, isolated immune cell populations such as PBMCs, and cell lysates. NanoString has included 109 genes that are associated with 24 immune cell types and populations. The nS Immune panel also includes a unique cell profiling feature that uses gene expression data from 46 human genes to quantify the relative abundance of 14 immune cell types: B and T cells, Th1, T-regs, CD45 and CD8+ T cells, exhausted CD8+ T cells, cytotoxic and dendritic cells, macrophages, mast cells, neutrophils, natural killer (NK) cells, and NK CD56dim cells.

2. Materials and methods

2.1. Patients

Neoadjuvant GM-CSF (NeoGM) was a single-center phase 1 study of neoadjuvant GM-CSF prior to planned radical prostatectomy (RP) in patients with localized prostate cancer (NCT00305669) (Wei et al., 2016). All patients underwent RP within 5 days after the last dose of GM-CSF. Tissue samples of neoadjuvant Provenge (NeoProv) were derived from a single-arm, multicenter phase II trial of neoadjuvant sipuleucel-T treatment in men with localized prostate cancer administered prior to planned treatment (NCT00715104) (Fong et al., 2014; Hagihara et al., 2019). Subjects received sipuleucel-T at the standard two-week intervals for three planned doses prior to RP. Other patients were a group of patients who were at risk of harboring prostate cancer, based on digital rectal examination and/or serum prostatic-specific antigen (PSA), assessed at the University of California San Francisco (UCSF) Medical Center, or UCSF Comprehensive Cancer Center. All the tissue samples of 14 NeoGM treated, 14 NeoProv treated, and 12 untreated prostate cancer patients (Untreated) were RP tissues, while the tissues of 6 prostatitis patients and 6 non-prostatitis subjects were biopsies. All subjects gave written informed consent to participate in the protocol approved by the Institutional Review Boards (IRBs) of each participating institution.

2.2. Gene expression

RNA was extracted from FFPE RP sections using PureLinkTM FFPE Total RNA Isolation Kit (Invitrogen) and was evaluated for gene expression using the nCounter® PanCancer Immune Profiling Panel (NanoString Technologies, Inc.); and using the HTG EdgeSeq Oncology Biomarker Panel (HTG OBP) and Precision Immuno-Oncology Panel (HTG PIP) (HTG Molecular Diagnostics, Inc.).

2.3. Data standardization and normalization

For simple sample correlations (e.g., replicate comparisons) and quality control metrics, we use a simple per sample standardization called the $\log_2$(CPM):

$$\log_2\left(\frac{r_{gi}+0.5}{R_i+1}\times 10^6\right)$$

Where,

r_{gi} is the number of sequence reads for probe (g) and sample (i).

R_i is the number of mapped reads (library count) for sample (i).

The read to library count ratio is $\log_2$ transformed, to stabilize the variance, and the resulting measures are interpreted as "counts per million" (CPM). This standardization allows us to evaluate the probe level expression as a proportion of total counts on a sample level.

Normalization is a construct that is most influential in gene expression differential testing. Here we considered the median normalization suggested originally by Anders and Huber (Anders and Huber, 2010). A scaling factor s_i for each sample i is obtained for each gene g (over G genes) and samples, m, which is calculated as the median gene level expression value for each sample-gene count adjusted by the geometric mean over all genes. Any genes without expression over all samples are necessarily excluded from this scaling calculation. The formula for the scaling factor for the ith sample is:

$$s_i = median_g \frac{r_{gi}}{\left(\prod_{v=1}^{m} r_{gv}\right)^{1/m}}$$

where, r_{gi} is the raw count for the ith sample and gth gene.

The scaling factor is then used to modify the original gene count to obtain the normalized count value, r_{gi}^{norm}:

$$r_{gi}^{norm} = \frac{r_{gi}}{s_i}$$

The resulting data, r_{gi}^{norm}, will be used for differential testing analysis.

2.4. Statistical analysis

For each of the common genes, the pairwise relationships between each two panels were assessed by applying Pearson correlation coefficient (r) on both raw and CPM data. The RV coefficient was used to measure the overall relationship across all common genes for each two panels. The disadvantage of only examining common genes present across the platforms is that data from biologically significant genes may be lost if a gene is not represented on the platforms examined. Coinertia analysis (CIA) is a multivariate method that identifies correlationships in multiple datasets which contain the same samples (Culhane et al., 2003). CIA is not limited to the analysis of datasets containing the same number of genes, thus uses all the genes in each platform. CIA simultaneously finds ordinations (dimension reduction diagrams) from the datasets that are most similar by finding successive axes from the two datasets with maximum covariance. UMAP (McInnes and Healy, 1802), a dimensionality reduction technique, was applied to visualize the comparisons across the platforms. Unsupervised

hierarchical cluster analysis with Euclidean distance was utilized on the data from each platform.

We performed differential testing between treated patients (NeoProv or NeoGM) vs. untreated patients within each panel by Limma (Ritchie et al., 2015). We compared the effect size, $\log_2$ of fold change (FC) across the panels by Pearson correlation coefficient, where FC was calculated as the ratio of the normalized expression. In addition, we assessed the classification performance across all 3 panels, where five-fold crossed validation was used. Data were randomly split into 5 subsets of similar sizes. For each run, by using 4 folds of the data (training data), genes were selected by SVM-RFE (Guyon et al., 2002). Selected genes were classified by SVM (Furey et al., 2000). The 5th fold of the data (validation data) was used to calculate the classification accuracy and area under the curve (AUC) of receiver operating characteristic (ROC) curve.

We predicted the cell fractions for each sample within each panel by applying the cell composition deconvolution method (Chiu et al., 2019), which uses advanced statistical methods and reference samples from NCBI GEO to build the gene expression signature matrix. We assessed the correlation of the predicted cell fractions between the panels by Pearson correlation coefficient and then compared the predicted cell fractions between treated patients (NeoProv or NeoGM) and untreated patients within each panel by Wilcoxon rank sum test.

All the analyses were performed by R (https://www.r-project.org). Statistical significance was declared based on $p < 0.05$. No multiple testing adjustment was done.

3. Results

3.1. Overall consistency across 3 panels

The number of "matched" or common genes across each of the three panels is shown in Fig. 1A. Among the 2560 genes in HTG OBP, 1392 genes in HTG PIP and 770 genes in the nS Immune panel, 731 genes were matched between HTG OBP and HTG PIP, 430 genes were matched between HTG OBP and nS Immune panel, and 617 genes were matched between HTG PIP and nS Immune panel (Table 1). Based on CPM data, which considers the gene counts and sequencing read depth (HTG) or digital count depth (nS) of each sample, more than 90% of the common genes were significantly correlated ($p < 0.05$) and more than 76% of the common genes were highly correlated ($r > 0.5$) between any two panels. The RV coefficients range from 0.71 to 0.81

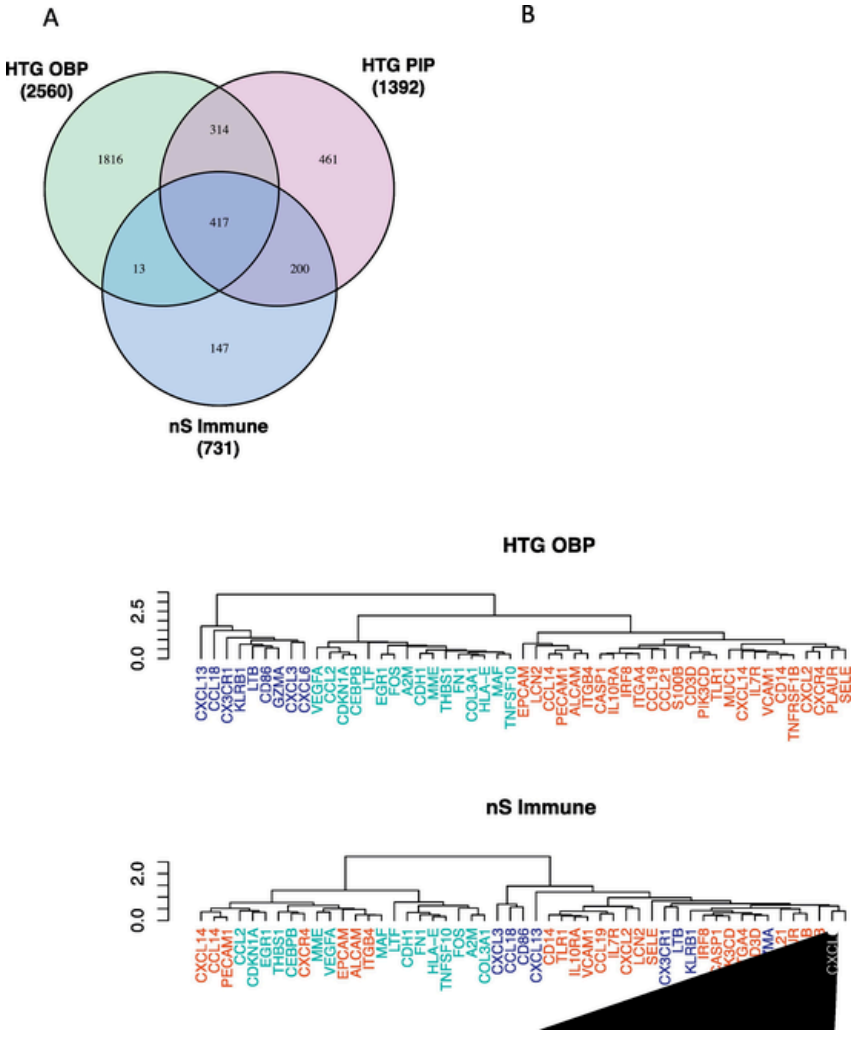

Fig. 1. Gene expression comparison across HTG platform (HTG OBP and HTG PIP) and nanostring platform (nS Immune). (A) A Venn diagram show the number of common and unique genes across the three panels. (B) UMAP plots the reduction to two dimensions. Each dot represents a gene from each panel. (C) Hierarchical clustering of HTG OBP and nS Immune panel of top correlated genes ($r > 0.9$). Top panel plots the dendrogram of 3 clusters based on HTG OBP dataset with randomly assigned colors. Bottom panel plots the dendrogram of the clusters based on nS Immune panel but using the same color code in HTG OBP data of the top panel. (D) Hierarchical clustering of HTG PIP and ns Immune panel of top correlated genes ($r > 0.9$) Top panel plots the dendrogram of 3 clusters based on HTG PIP dataset with randomly assigned colors. Bottom panel plots the dendrogram of the clusters based on nS Immune panel but using the same color code in HTG PIP data of the top panel.

Table 1
Gene consistency comparison across platforms.

		HTG OBP vs. HTG PIP	nS Immune vs. HTG OBP	nS Immune vs. HTG PIP
Raw Data	# of genes matched (% of matched genes)	731 (39.9%)	430 (14.8%)	617 (22.7%)
	# of genes significantly correlated ($p < 0.05$)	607 (83.0%)	317 (73.7%)	528 (85.6%)
	# of high correlated genes ($r > 0.5$)	477 (65.3%)	219 (50.9%)	415 (67.3%)
	RV coefficient	0.81	0.71	0.81
CPM Data	# of genes significantly correlated (p < 0.05)	662(90.6%)	403 (93.7%)	560 (90.8%)
	# of high correlated genes (r > 0.5)	557(76.2%)	362 (84.2%)	475 (77.0%)
	RV coefficient	0.79	0.79	0.80

based on the raw data and 0.79–0.80 based on CPM data. Targeted sequencing data is not normalized following the standard RNA-Seq workflow. However, the data is typically transformed to CPM, which removes variability between samples receiving a different number of reads. Thus, the CPM data is more concordant than the raw data (for example, when comparing HTG OBP with nS Immune, the percentage of significantly correlated genes increased from lower than 73.7% to 93.7%). Together, these results show that there is a high degree of correlation that is even higher for CPM normalized data.

Next, we applied CIA to compare gene expression profiles across the three platform. The relationships between these datasets are described by the correlations between the pairs of ordinations along the first (F1) and second pair of axes (F2). The results in Table 2 show that 99.96–99.97% of the total variance are represented by the F1 in each analysis. There is a high correlation between pairs of ordinations on each axis. This, therefore, presents a good summary of the co-structure between the any two panels studied. For all 3 comparisons, the correlation between the first axes (F1) of the two ordinations was 0.89, and the correlation between second axes (F2) of the two ordinations was 0.84–0.87. These high values partly result from the maximization of covariance or the product of the correlation and the squared variances projected onto the co-inertia axes. These results show that the expression data of all genes across all three panels are highly concordant.

3.2. Divergences among high correlated genes

In addition to direct comparison of gene expression data, we used UMAP to compare the global structure of the expression data of all genes across 3 panels for the 52 patients simultaneously. In Fig. 1B, there were no clear clusters to distinguish the expression data from the 3 panels suggesting that there is a high degree of consistency. However, there was a small divergence at the top right tail. Most of the divergent genes were in the long tail of HTG OBP and HTG PIP panel and were within the chemokine ligand (CCL) family (CCL3, CCL7, etc.) and the interleukin family (IL3, IL5, etc.). We also performed an independent unsupervised hierarchical cluster analysis for highly correlated common genes ($r > 0.9$). The highly correlated genes were first clustered based on the HTG OBP dataset into three clusters (Fig. 1C, top panel). These genes were then clustered based on the nS Immune panel but the same color code as in the HTG OBP data was used (Fig. 1C, bottom panel). Although the majority of genes in the red and light blue coded clusters stay the same, some of the blue and red coded genes changed their cluster allocations in nS Immune clusters, such as CXCL14, CCL14, CXCR4, etc. Similarly, we compared the clustering results between HTG PIP and nS Immune panel (Fig. 1D) and found that CCL18, HLA – DQB1, CXCL6, FCGR2B, LTB, CX3CR1, CXCL10, ITGAL, TLR1 were the genes shuffled most.

Table 2
Results of Co-inertia analysis (CIA) analysis across platforms (based on CPM data).

	HTG OBP vs. HTG PIP	nS Immune vs. HTG OBP	nS Immune vs. HTG PIP
% Inertia (F1)	99.96	99.97	99.97
% Inertia (F2)	0.02	0.02	0.02
Correlation of ordinations (F1)	0.89	0.89	0.89
Correlation of ordinations (F2)	0.84	0.87	0.87

3.3. Variation in correlation with clinical characteristic

We also assessed the gene expression data within a clinical context. First, we performed differential testing between the neoadjuvant treated patients (NeoProv or NeoGM) and untreated patients within each panel by Limma (Ritchie et al., 2015). The volcano plots (Fig. 2A-C) present -$\log_{10}$(pvalue) vs. $\log_2$(FC) based on the data from each panel, where FC (fold change) was directly calculated by Limma and stood for the difference of expression of the common genes between two cohorts. Supplementary Table 1 presents the list of the genes identified as significantly differentially expressed between samples from treated and untreated patients based on the 3 panels. We compared the effect size, $\log_2$(FC) across the panels (Fig. 2). Overall, the $\log_2$(FC) across the panels were significantly correlated ($p < 0.001$). The correlation varied with the highest correlation between HTG PIP and nS Immune ($r = 0.6$, $p < 0.001$), while the correlation was 0.4 and 0.45 between HTG PIP and HTG OBP, and between nS Immune and HTG OBP, respectively (Fig. 2D, 2E and 2F). Due to the small study cohort size, each panel could only identify a small amount of significantly differentiated genes ($p < 0.05$). Some genes were only identified as significant in one panel (blue or green dots) while some genes were identified as significant across multiple panels (red dots). For example, 8 genes were significant in all 3 panels, while 16 genes detected as significantly expressed in both HTG OBP and PIP but not in nS Immune (Fig. 2G). Only one gene (PLK3) and 5 genes (MR1, S100A8, THBD, TNFAIP3, TREM1) was overlapped between nS Immune and HTG OBP or PIP respectively. 28 genes were detected by the nS Immune panel itself, 82 gene by HTG PIP only and 232 genes by HTG OBP only. Although the number of genes within each panel differ, the percentage of detection of significant genes were similar between the two HTG panels (PIP = 8% and OBP = 10%), while nS Immune was slightly lower at 5%. In general, the effect size, $\log_2$(FC) across the panels were significantly but moderately correlated. There were common significant genes detected by different platforms but some that were only detected by the HTG platform.

In addition, we evaluated the classification performance across the 3 panels by classification accuracy and AUC. Due to the small sample size, we utilized a 5-fold crossed validation and selected the top 30 genes within each run for each panel. Table 3 shows that for almost each run, HTG PIP had the highest accuracy and AUC, while nS Immune always had the lowest classification accuracy (Table 3). The average taken over 5 runs showed that the HTG PIP panel performed best. The heatmap is a plot of the selection frequency within each panel, where only the genes that were selected at least 3 times among the 5 runs across 3 panels were retained (Fig. 2H).

Based on cell composition deconvolution analysis, the gene expression data from HTG OBP and HTG PIP had few samples with M1, Natural Killer cells, and T helper cells (the corresponding predicted cell proportion was 0). nS Immune, however, did have a few samples with NK cells, Treg, M1, and T helper cells. In general, the predicted immune cell proportions between HTG OBP and PIP were consistent for most cell types, except CD8 T cells. However, the HTG PIP panel had 5 predicted cell proportions significantly correlated with those of the nS Immune panel but HTG OBP did not have any predicted cell proportions signifi-

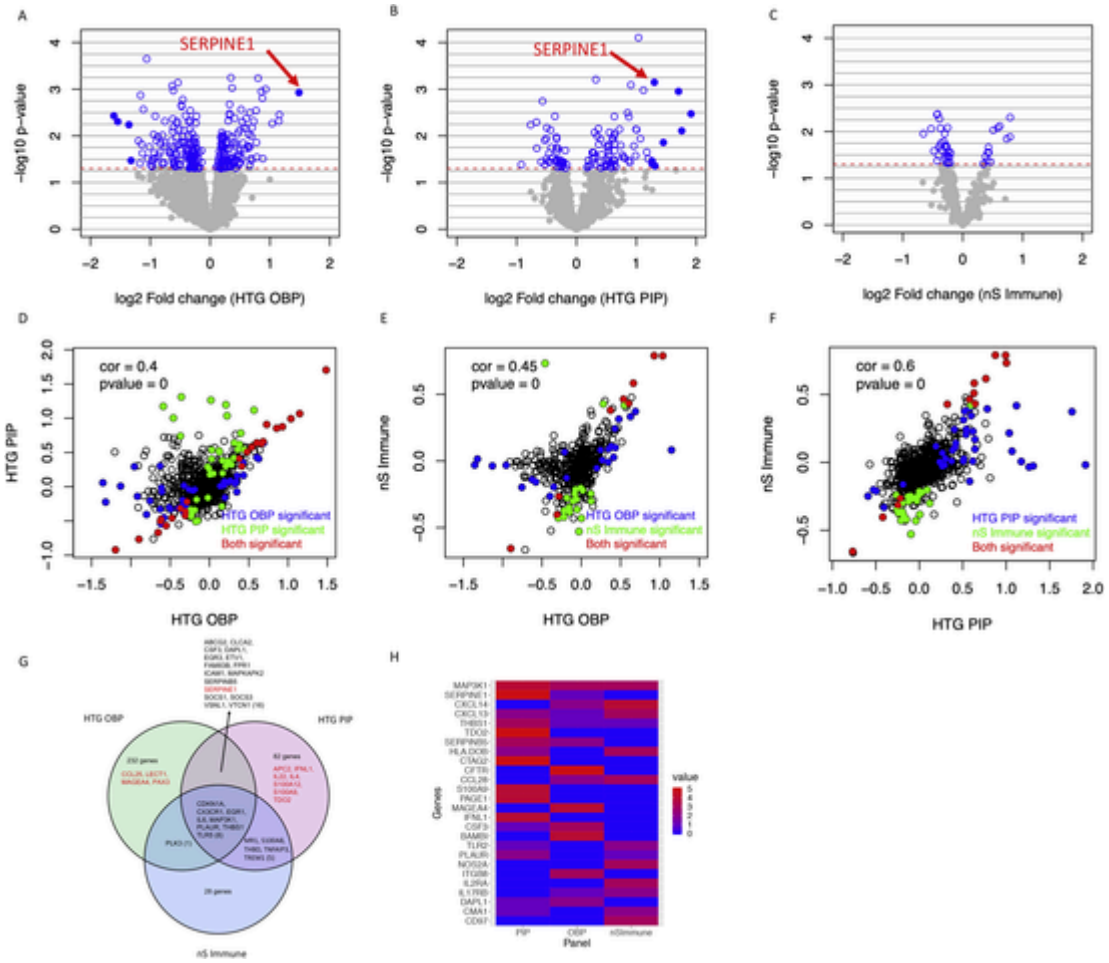

Fig. 2. Comparison of the relationships of the gene expression with clinical characteristics across panels. Effect size is defined as $\log_2$(FC), where FC is folder change obtained by differential testing of patients receiving neoadjuvant treatment (NeoProv or NeoGM) vs. untreated patients using Limma. Pearson's correlation coefficient was calculated for each fig. (A-C) Volcano plot based on each panel. Each dot represents a single gene, genes in grey have $p \geq 0.05$, and genes in blue have $p < 0.05$ while solid blues stand for $|\log2(FC)| > 1.2$. The red dash line stands for $p = 0.05$. (A) Volcano plot based on HTG OBP panel. (B) Volcano plot based on HTG PIP panel. (C) Volcano plot based on nS Immune panel. (D—F) Correlation of $\log_2$(FC) between panels. (D) Correlation of $\log_2$(FC) between HTG PIP and HTG OBP. Blue and green dots presented HTG OBP and HTG PIP significant genes ($p < 0.05$), respectively. Red dots presented the genes significant in both HTG OBP and HTG PIP genes. (E) Correlation of $\log_2$(FC) between HTG OBP and nS Immune. Blue and green dots presented HTG OBP and nS Immune significant genes ($p < 0.05$), respectively. Red dots presented the genes significant in both HTG OBP and nS Immune genes. (F) Correlation of $\log_2$(FC) between HTG PIP and nS Immune. Blue and green dots presented HTG PIP and nS Immune significant genes ($p < 0.05$), respectively. Red dots presented the genes significant in both HTG PIP and nS Immune. (G) A Venn diagram show the number or the list of the genes differentially expressed ($p < 0.05$) across the three panels. The genes highlighted in red are the ones also have $|\log_2(FC)| > 1.2$. (H) A heatmap presents the top genes selected to classify the patients who receiving neoadjuvant treatment from untreated by each panel. For each panel, we performed 5 runs, within each run, top 30 genes were selected. We only kept the genes that were selected more than 3 times among 5 runs across 3 panels. The heatmap plotted the selection frequency within each panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Comparison of classification performance across platforms (based on CPM data).

	Accuracy			AUC[a]		
	HTG OBP	HTG PIP	nS Immune	HTG OBP	HTG PIP	nS Immune
Fold 1	87.5%	100.0%	62.5%	100.0%	100.0%	100.0%
Fold 2	75.0%	87.5%	62.5%	91.7%	100.0%	66.7%
Fold 3	77.8%	88.9%	33.3%	77.8%	94.4%	83.3%
Fold 4	85.7%	85.7%	85.7%	100.0%	100.0%	100.0%
Fold 5	50.0%	62.5%	37.5%	60.0%	73.3%	73.3%
Average	75.2%	84.9%	56.3%	85.9%	93.6%	84.7%

[a] Area under the curve (AUC) of receiver operating characteristic (ROC) curve.

cantly correlated with the nS Immune panel (Supplementary Table 2). Interestingly, the predicted proportions of Dendritic cells (DCs) (7% vs. 5%, $p = 0.032$), Treg (4% vs. 2%, $p = 0.002$) and memory CD8 T cell (33% vs. 36%, $p = 0.018$) from HTG OBP were found significantly different between treated and untreated patients (Supplementary Table 3), which were not observed in the other two panels.

4. Discussion

Clinical and research sample sizes are limited while the demand for additional molecular information continues to increase. Understanding one's options for sample conservation and profiling has become a critical part of effective experimental design while maximizing sample utilization. This escalating issue has led to the current assessment of the cross-platform concordance of immune-related gene expression data, where we compared the immune-related gene expression data across

three unique panels of two technology platforms: HTG OBP, HTG PIP and the nS Immune. There was a high level of cross platform concordance even when using extracted RNA from relatively degraded FFPE samples.

We used multiple different statistical and bioinformatical methods to compare the gene expression dataset of 52 patients across 3 panels. More than 76% of the common genes across the panels were significantly and highly correlated ($p < 0.05$ and $r > 0.5$). In addition, the CIA results show that the expression data of all genes across all three panels are highly concordant. Furthermore, the effect size of the common genes across the panels was significantly correlated and there were common significant genes detected by different platforms.

Despite the overall consistency among HTG OBP, HTG PIP, and nS Immune, there are some distinct results when comparing across platforms. When investigating the global structure of the expression data, there was a small divergence in UMAP. Most of the genes that make up the divergence identified between the HTG OBP and HTG PIP panels are from the CCL and IL families. The genes making up the divergence on UMAP across the panels have a wide range of functions. These functions range from inducing differentiation of lymphoid and myeloid progenitor cells (IL-3), to activation of B cells, T cells or eosinophils (IL-5, IL-9, IL-12) to pro or anti-inflammatory effects (IL-13, IL-17, IL-19, IL-22, IL-23, IL-24, IL-27). The CCL genes identified on UMAP are also consistent with the findings from the unsupervised hierarchical clustering. Genes encoding for chemokines and chemokine receptors (CXCL14, CCL14, CXCR4, CXCL4, CCL18, CXCL13, CX3CR1, CXCL6) that clustered in the HTG OBP or PIP panels did not cluster together in the nS immune platform. Chemokines and chemokine receptors vary in expression and are produced by a wide range of cells including cancer cells, stromal cells, endothelial cells, and tumor infiltrating cells. (Vindieux et al., 2009) Thus, the divergence in UMAP and the differences in clustering across platforms particularly among genes encoding for interleukins, chemokines and chemokine receptors are more likely a result of either tissue heterogeneity as opposed to probe design or assay chemistry.

There were some genes detected that were significant in both HTG OBP and PIP but were not detected via the nS platform (Fig. 2D). Of the 16 differentially expressed genes detected by HTG OBP and PIP but not nS, four of which (CSF3, DAPL1, SERPINB5, SERPINE1) were also among the genes with highest selection frequency (Fig. 2H). These genes are all involved in tumor progression either through mediating tumorigenesis, proliferation, metastasis, avoiding the immune system, or conferring resistance. For example, loss of SERPINB5, a tumor suppressor gene, has been associated with increasing malignancy in breast, prostate, thyroid and skin cancer. (Yang et al., 2016) Additionally, SERPINE1 has been found to be highly expressed in a wide range of cancers including prostate cancer and induces tumor vascularization to facilitate tumor metastasis. (Li et al., 2018) The differential expression detected by the HTG platform as opposed to the nS platform may be a result of either a larger panel or amplification in HTG OBP and HTG PIP compared to the nS platform. Interestingly, HTG PIP consistently had the best classification accuracy and classification performance. HTG EdgeSeq chemistry is able to increase sensitivity with an increased number of probes and amplification. Researchers should be aware of the potential differences in results introduced due to the two distinct platforms. Our findings suggest that the data produced from each assay was consistent globally but a subset of genes were divergent and their expression were differentially expressed between treated vs. untreated patients.

Deconvolution analysis allows researchers to infer cell type composition based on gene expression. Our analysis revealed that HTG OBP and HTG PIP were correlated in all subtypes except for CD8+ T cells. However, HTG OBP did not have any significantly correlated cell types compared to the nS platform. Thus, further investigation is necessary with flow cytometry to characterize the proportion of each cell type would be needed to validate the deconvolution analysis.

Overall, for studies where there is limited sample or where sample conservation is required, the newer version of HTG molecular RNA profiling technology (HTG PIP) is a comparable option.

Funding

LF is supported by NIH R01CA223484 and the Prostate Cancer Foundation. LZ receives the support of the UCSF Academic Senate Committee on Research.

Author contributions

L.Z. and L.F. concepted and designed the experiment, and acquired the data. L.Z., and H.Y. performed the data analysis. All authors participated in the interpretation of study results, and in the drafting and approval of the final version of the manuscript.

Data and materials availability

Datasets generated during and/or analyzed during the study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

LF has received research funding from Abbvie, Bavarian Nordic, BMS, Dendreon, Janssen, Merck and Roche/Genentech. JC and BK are employees of HTG. DT was an employee of HTG during manuscript preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jim.2021.113041.

References

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Beard, et al., 2013. Gene expression profiling using nanostring digital RNA counting to identify potential target antigens for melanoma immunotherapy. Clin. Cancer Res. 19 (18), 4941–4950.

Chiu, Y.J., Hsieh, Y.H., Huang, Y.H., 2019. Improved cell composition deconvolution method of bulk gene expression profiles to quantify subsets of immune cells. BMC Med. Genet. 12 (169). https://doi.org/10.1186/s12920-019-0613-5.

Culhane, A.C., Perrière, G., Higgins, D.G., 2003. Cross-platform comparison and visualisation of gene expression data using co-inertia analysis. BMC Bioinformat. 4, 59.

Fong, L., Carroll, P., Weinberg, V., Chan, S., Lewis, J., Corman, J., Amling, C.L., Stephenson, R.A., Simko, J., Sheikh, N.A., et al., 2014. Activated lymphocyte recruitment into the tumor microenvironment following preoperative sipuleucel-T for localized prostate cancer. J. Natl. Cancer Inst. 10611. https://doi.org/10.1093/jnci/dju268.

Furey, T.S., et al., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16 (10), 906–914.

Gajewski, T.F., et al., 2011. Molecular profiling to identify relevant immune resistance mechanisms in the tumor microenvironment. Curr. Opin. Immunol. 23, 286–292.

Guyon, I., et al., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1–3), 389–422.

Hagihara, K., Chan, S., Zhang, L., Oh, D.Y., Wei, X.X., Simko, J., Fong, L., 2019. Neoadjuvant sipuleucel-T induces both Th1 activation and immune regulation in localized prostate cancer. Oncoimmunology. 8 (1), e1486953.

Li, S., Wei, X., He, J., Tian, X., Yuan, S., Sun, L., 2018. Plasminogen activator inhibitor-1 in cancer research. Biomed. Pharmacother. 105, 83–94.

McInnes, L., Healy, J., 1802. UMAP: uniform manifold approximation and projection for dimension reduction. ArXiv e-prints 3426, 2018.

Melero, I., et al., 2014. Therapeutic vaccines for cancer: an overview of clinical trials. Nat. Rev. Clin. Oncol..

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and

microarray studies. Nucleic Acids Res. 43 (7), e47. https://doi.org/10.1093/nar/gkv007.

Vindieux, D., Escobar, P., Lazennec, G., 2009. Emerging roles of chemokines in prostate cancer. Endocr. Relat. Cancer 16 (3), 663–673. https://doi.org/10.1677/ERC-09-0109.

Wei, X.X., Chan, S., Kwek, S., Lewis, J., Dao, V., Zhang, L., Cooperberg, M.R., Ryan, C.J., Lin, A.M., Friedlander, T.W., Rini, B., Kane, C., Simko, J.P., Carroll, P.R., Small, E.J., Fong, L., 2016. Systemic GM-CSF recruits effector t cells into the tumor microenvironment in localized prostate cancer. Cancer Immunol Res 4 (11), 948–958 PMID: 27688020.

Yang, S., Yeh, C., Chou, Y., et al., 2016. Serpin peptidase inhibitor (SERPINB5) haplotypes are associated with susceptibility to hepatocellular carcinoma. Sci. Rep. 6, 26605. https://doi.org/10.1038/srep26605.