JCI The Journal of Clinical Investigation

Differential immune profiles distinguish the mutational subtypes of gastrointestinal stromal tumor

Gerardo A. Vitiello, … , Shan Zeng, Ronald P. DeMatteo

J Clin Invest. 2019. https://doi.org/10.1172/JCI124108.

Research In-Press Preview **Immunology** **Oncology**

Gastrointestinal stromal tumor (GIST) is the most common human sarcoma, frequently characterized by an oncogenic mutation in the *KIT* or *platelet-derived growth factor receptor alpha (PDGFRA)* genes. We performed RNA sequencing of 75 human GIST tumors from 75 patients, comprising the largest cohort of GISTs sequenced to date, in order to discover differences in the immune infiltrates of *KIT* and *PDGFRA*-mutant GIST. Through bioinformatics, immunohistochemistry, and flow cytometry, we found that *PDGFRA*-mutant GISTs harbored more immune cells with increased cytolytic activity when compared to *KIT*-mutant GISTs. *PDGFRA*-mutant GISTs expressed many chemokines, such as CXCL14, at a significantly higher level when compared to *KIT*-mutant GISTs and exhibited more diverse driver-derived neoepitope:HLA binding, both of which may contribute to *PDGFRA*-mutant GIST immunogenicity. Through machine learning, we generated gene expression-based immune profiles capable of differentiating *KIT* and *PDGFRA*-mutant GISTs, and also identified additional immune features of high PD-1 and PD-L1 expressing tumors across all GIST mutational subtypes, which may provide insight into immunotherapeutic opportunities and limitations in GIST.

Find the latest version:

https://jci.me/124108/pdf

Differential immune profiles distinguish the mutational subtypes of gastrointestinal stromal tumor

Gerardo A. Vitiello[# 1], Timothy G. Bowler[# 2], Mengyuan Liu[1, 3], Benjamin D. Medina[1], Jennifer Q. Zhang[1], Nesteene J. Param[1, 3], Jennifer K. Loo[1], Rachel L. Goldfeder[4], Frederic Chibon[5], Ferdinand Rossi[1, 3], Shan Zeng[1], and Ronald P. DeMatteo[1, 3]

[#] These two authors contributed to this work equally.

[1]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[2]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[3]Department of Surgery, The University of Pennsylvania, Philadelphia, PA, USA
[4]Genome Technologies, The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
[5]INSERM U1037, Cancer Research Center of Toulouse, 31037 Toulouse, France

Disclosures: The authors have declared that no conflicts of interest exist.

Word count: 6,129

Figures: 7

Supplementary Figures: 5

Supplementary Tables: 7

Corresponding author:
Ronald P. DeMatteo, MD
3400 Spruce Street
Philadelphia, PA 19104
Phone: (215) 662-7539
Fax: (215) 614-0363
Email: Ronald.dematteo@uphs.upenn.edu

Abstract

Gastrointestinal stromal tumor (GIST) is the most common human sarcoma, frequently characterized by an oncogenic mutation in the *KIT* or *platelet-derived growth factor receptor alpha* (*PDGFRA*) genes. We performed RNA sequencing of 75 human GIST tumors from 75 patients, comprising the largest cohort of GISTs sequenced to date, in order to discover differences in the immune infiltrates of *KIT* and *PDGFRA*-mutant GIST. Through bioinformatics, immunohistochemistry, and flow cytometry, we found that *PDGFRA*-mutant GISTs harbored more immune cells with increased cytolytic activity when compared to *KIT*-mutant GISTs. *PDGFRA*-mutant GISTs expressed many chemokines, such as CXCL14, at a significantly higher level when compared to *KIT*-mutant GISTs and exhibited more diverse driver-derived neoepitope:HLA binding, both of which may contribute to *PDGFRA*-mutant GIST immunogenicity. Through machine learning, we generated gene expression-based immune profiles capable of differentiating *KIT* and *PDGFRA*-mutant GISTs, and also identified additional immune features of high PD-1 and PD-L1 expressing tumors across all GIST mutational subtypes, which may provide insight into immunotherapeutic opportunities and limitations in GIST.

Introduction

Gastrointestinal stromal tumor (GIST) is the most common sarcoma (1), often originating in the stomach or the small intestine. GIST is a paradigm for targeted molecular therapy, since the majority of GISTs contain an activating mutation in *exon 11* of the *KIT* proto-oncogene, for which there are effective and well-tolerated tyrosine kinase inhibitors (2, 3). Imatinib mesylate, which inhibits the KIT oncoprotein, improved the median survival in patients with advanced and metastatic GIST from one to five years (3), while adjuvant imatinib improved recurrence-free survival in patients with resectable disease (4).

Platelet derived growth factor receptor alpha (*PDGFRA*)-mutant GIST is the second most common form of GIST. Approximately half of *PDGFRA*-mutant GISTs contain a *D842V* substitution, which is inherently resistant to imatinib therapy (5). The majority of *PDGFRA*-mutant GISTs develop in the stomach (6, 7), while *KIT*-mutant GISTs can arise in the stomach, small intestine, or rectum (8), suggesting an innate and important biologic difference between these driver mutations. Other GIST mutation types also associate with predictable clinicopathologic features. For example, *KIT exon 9*-mutant GISTs nearly always develop in the small intestine (9), n*eurofibromin 1*-mutant (*NF1*-mutant) GISTs have a predilection for the duodenal-jejunal flexure (10), and succinate dehydrogenase deficient (*SDH*-deficient) GISTs are indolent, multifocal, and one of the few GISTs, along with those driven by kinase fusions, which metastasize via the lymphatic system (11, 12). The underlying biologic mechanisms linking mutation type and clinicopathologic features are not well understood.

Although *PDGFRA D842V* (*D842V)*-mutant GISTs do not respond to imatinib, natural history shows that recurrence-free survival is more favorable in patients with any *PDGFRA* mutation when compared to a *KIT* mutation (4). This finding, along with the predictable behavior of GIST mutation types, has us hypothesize that the mutational driver may impact other aspects of tumor biology, specifically the tumor microenvironment and host immune response. Our group has extensively characterized the immune infiltrate in *Kit*-mutant GIST using a genetically engineered mouse model and human specimens (13-17), providing clear rationale for multiple immunotherapy trials in patients with GIST, but the immune response to *PDGFRA*-mutant GIST is currently unknown.

In this study, we performed RNA sequencing on 75 surgical specimens from 75 human GIST patients to characterize the immune infiltrates of different GIST mutations. Surprisingly, we discovered that *PDGFRA*-mutant GIST harbors more immune cells than *KIT*-mutant GIST, which may be related to oncogene-specific cytokine production or more HLA-diverse neoepitope recognition, suggesting that patients with *PDGFRA*-mutant GIST might benefit from therapeutic immunomodulation. We trained a random forest machine learning algorithm on RNAseq data from all of our *KIT* and *PDGFRA*-mutant GISTs, as well as on a subset of untreated, primary, gastric (UPG) *KIT* and *PDGFRA*-mutant GISTs, in order to characterize a *PDGFRA*-specific immune landscape. Finally, we identified the top immune features correlating with high PD-1 and PD-L1 expression across all GIST specimens to identify potential barriers to and synergistic opportunities for successful immune checkpoint blockade in GIST.

Results

ssGSEA identifies immune cell pathway enrichment in PDGFRA-mutant GIST.

We performed RNA sequencing and principal component analysis (PCA) of 75 human GIST specimens comprising various mutation types (n=37 *KIT*-mutant, n=24 *PDGFRA*-mutant, n=7 *SDH*-deficient, n=4 multiple drivers, n=2 wild type, and n=1 *NF1*-mutant; **Supplemental Table 1**). The combined PCA of all mutational drivers demonstrated clustering by mutational subtype (**Supplemental Figure 1, top**), which was most apparent in *PDGFRA*-mutant and *SDH*-deficient GISTs (**Supplemental Figure 1, bottom**). Notably, *KIT*-mutant GISTs appeared to cluster into three distinct groups, which upon closer inspection of clinicopathologic features is related to treatment status and tumor location.

Given what our group has discovered regarding the immune infiltrate of GIST (13-15, 17), the metabolic characteristics of imatinib-treated GIST (18), and the association of cell cycle activity with GIST aggressiveness (19, 20), we used our sequencing data to perform single-sample gene set enrichment analyses (ssGSEA) focused on immune, metabolic, and cell cycle pathways (**Figure 1**). Using published gene sets (21), along with assessment of previously published metabolic, cell cycle, and immune pathways (22), we identified 136 gene sets for inclusion in ssGSEA (**Supplemental Table 2**).

ESTIMATE and Cyt scoring, which infer the quantity and cytolytic activity of the immune infiltrates respectively (23, 24), revealed that *KIT*-mutant GIST had a range of immune cell infiltration and cytolytic activity (**Figure 1**). Meanwhile, *SDH*-deficient, *NF1*-mutant*, PDGFRA D842V* (*D842V*)-mutant GISTs with a concurrent *CDKN2A* (p16)

deletion, and wild type GISTs (defined as non-*KIT*, non-*PDGFRA*, non-RAS activated, and non-*SDH* deficient as previously described (25)) exhibited generally low ESTIMATE and Cyt scores and a lack of immune cell pathway enrichment.

Conversely, ESTIMATE, Cyt, and ssGSEA identified increased immune cell infiltration, greater immune cell activity, and a significant enrichment of immune-related gene sets in *PDGFRA*-mutant GISTs, an unexpected finding which suggested that *PDGFRA*-mutant GISTs were more immunogenic than other GIST mutations (**Figure 1**). Cell cycle activity pathways did not clearly correlate with immune cell infiltration across all GIST driver mutation types, while metabolic activity appeared to inversely correlate with the quantity of immune infiltrate, specifically in *KIT exon 11*-mutant GISTs. Overall, these data show that while tumors driven by a particular GIST mutation can have variable immune profiles, *PDGFRA*-mutant GIST contains the strongest gene expression-based immune signature when compared to other GIST mutations.

PDGFRA mutant GIST is more immunologically active than KIT mutant GIST.

In order to validate our ssGSEA finding that *PDGFRA*-mutant GIST was highly enriched in immune cell quantity and cytolytic activity, we first compared the clinicopathologic characteristics of all *KIT* and *PDGFRA*-mutant GISTs (**Supplemental Table 3; n=61**). Though not associated with immune cell infiltration in GIST, tumor size, mitotic rate, and tumor location are predictive of GIST aggressiveness and recurrence free survival after surgical resection (26). *PDGFRA*-mutant GISTs were more commonly found in stomach when compared to *KIT*-mutant GISTs, and there were significantly more males in our *PDGFRA*-mutant GIST cohort. There was also a trend toward

decreased mitotic activity in *PDGFRA*-mutant when compared to *KIT*-mutant GISTs (**Supplemental Table 3**). ESTIMATE and Cyt scores, however, did not correlate with mitotic rate across all *KIT* and *PDGFRA*-mutant GISTs (**Supplemental Figure 2A**). Cyt score and the overall number of CD45 mRNA transcripts were significantly higher in *PDGFRA*-mutant when compared to *KIT*-mutant GISTs (**Figure 2A**), while there was also a trend towards increased ESTIMATE score and CD8 mRNA expression, suggesting that *PDGFRA*-mutant GISTs may exhibit greater immune cell infiltration and activation than *KIT*-mutant GISTs.

Since we observed significant differences in tumor location between *KIT* and *PDGFRA*-mutant GISTs, and metastatic lesions and treatment status have also been shown to alter the immune infiltrate in GIST (14, 27), we then compared only untreated, primary, gastric (UPG) *PDGFRA* and *KIT*-mutant GISTs to minimize confounders and maximize the potential to observe important biologic differences that may be related to the oncogenic driver (**Supplemental Table 4; n=22**). UPG *PDGFRA* and *KIT*-mutant GISTs with a low mitotic rate exhibited significantly higher ESTIMATE scores when compared to UPG GISTs with a high mitotic rate (**Supplemental Figure 2B),** suggesting that immune cell infiltration may be partly related to GIST aggressiveness when controlling for tumor location and treatment status**.** However, when also controlling for the mutational driver in addition to tumor location and treatment status, mitotic rate did not inversely correlate with immune cell infiltration (**Supplemental Figure 2C**), implying that the oncogenic driver may also be contributing to differences in immune response. In fact, ESTIMATE score, Cyt score, and the mRNA expression of CD45 were significantly greater in UPG *PDGFRA*-mutant GISTs when compared to

UPG *KIT*-mutant GISTs, while the expression of CD8 mRNA was also increased (**Figure 2B**). Furthermore, gene set enrichment analyses showed that UPG *PDGFRA*-mutant GISTs were more significantly enriched in the overall immune response, the adaptive immune response, antigen binding, and alpha/beta T cell activation pathways among others (**Figure 2C**), including lymphocyte activation, lymphocyte differentiation, and B cell activity (not shown), suggesting that the difference in immune infiltration and activity may be related to the oncogenic driver.

To further validate the observed differences in immune cell infiltrate between *PDGFRA* and *KIT*-mutant GISTs, we performed immunohistochemical (IHC) staining for CD45 and CD8 on *KIT* and *PDGFRA*-mutant GISTs, including additional tumors not included in the sequencing cohort (**Figure 2D**). On gross microscopic examination, it appeared that *PDGFRA*-mutant GISTs contained more $CD45^+$ and $CD8^+$ cells, with a proportion of immune cells clustered around perivascular structures, a finding that we have found to be associated with adaptive immunity ([Medina BD and DeMatteo RP], unpublished observations). Quantification of CD45 and CD8 IHC staining, including staining of the most infiltrated and cytolytically active *KIT*-mutant GIST in the *KIT*-mutant cohort (Human GIST 203), showed that *PDGFRA*-mutant GISTs had significantly more $CD45^+$ and $CD8^+$ cells per 20X high powered field (HPF) than *KIT*-mutant GISTs (**Figure 2E**). Moreover, flow cytometric analysis of *KIT* and *PDGFRA*-mutant GIST specimens confirmed that *PDGFRA*-mutant GISTs harbor significantly more $CD45^+$ immune cells than *KIT*-mutant GISTs (**Figure 2E**). Together, these findings confirm that *PDGFRA*-mutant GISTs contain more immune cells with increased cytolytic activity when compared to *KIT*-mutant GISTs with similar clinicopathologic features.

CIBERSORT and DGE analysis identify unique immune signatures in GIST.

To discover additional differences in the immune landscape between UPG *PDGFRA* and UPG *KIT*-mutant GISTs, we profiled a previously defined set of 117 relevant immune features (**Supplemental Table 5**) (28, 29). In addition to Estimate and Cyt scores, this profile included CIBERSORT scores, which infer various immune cell frequencies through RNA-seq deconvolution, as well as 93 relevant immune-related genes. UPG *PDGFRA*-mutant GISTs appeared to have higher proportions of intratumoral CD4 T cell and macrophage expression (**Figure 3, middle**), while UPG *KIT*-mutant GISTs primarily expressed CD4 and macrophage signatures only in the most immune-cell infiltrated tumors. Furthermore, B cell and CD8 T cells signatures were also expressed in GISTs with high ESTIMATE and Cyt scores, confirming what we and others have shown regarding the immune cell prevalence in GIST (14, 27).

Differential gene expression (DGE) analysis revealed that 20/93 (21.5%) immune-related genes were significantly differentially regulated between UPG *KIT* and UPG *PDGFRA*-mutant GISTs (bold p-values <0.05, **Figure 3, bottom**). Specifically, UPG *PDGFRA*-mutant GISTs expressed significantly more CCR5, BTLA, CD96, CD48, TNFRSF9, TNFSF8, CCR4, CXCL11, CXCR4, KDR, IL6R, TNFRSF8, TNFSF14, TIGIT, TNFRSF17, HLA-DQA2, CXCL14, and CXCL12. Meanwhile, *KIT*-mutant GISTs expressed significantly more TNFSF18 and MICA. Together, these data suggest that *PDGFRA* and *KIT*-mutant GISTs elicit a different quality of immune infiltrate, which may ultimately provide driver-specific strategies for immunotherapy.

PDGFRA and KIT mutant GISTs have distinct signaling and cytokine signatures.

To determine if intracellular signaling or cytokine differences could be contributing to the observed differences in immune infiltration between UPG *PDGFRA* and UPG *KIT*-mutant GISTs, we utilized gene set enrichment analysis (GSEA) to detect differences in these pathways. On GSEA, we first observed enrichment of a known *PDGFRA*-mutant GIST signaling pathway gene set from a prior report analyzing 3 *PDGFRA*-mutant GISTs (**Figure 4A, top)** (30). Compared to UPG *KIT*-mutant GISTs, UPG *PDGFRA*-mutant GISTs also demonstrated significant transcriptional activation of the PI3K signaling pathway (**Figure 4A, bottom)**, which may be driven by expression changes in *PDGFRA*, *RELN*, *KDR*, or *ERK*, four genes that have been found to be significantly differentially expressed between *PDGFRA* and *KIT*-mutant GISTs not only in our cohort but also previously (30). Furthermore, many cytokine and chemokine pathways were significantly upregulated in UPG *PDGFRA*-mutant GISTs (**Figure 4B**), which we found to be related to increased expression of *CXCL14, CCL7, CCL19, VEGFA*, and *KITLG* (**Figure 4C**).

The differences observed in *CXCL14* expression were particularly intriguing. First, *CXCL14* mRNA was expressed at significantly higher levels in nearly all UPG *PDGFRA*-mutant GISTs. RT-PCR of bulk, human *PDGFRA* and *KIT*-mutant tumors also revealed that *PDGFRA*-mutant GISTs produced approximately 10 times more *CXCL14* mRNA than *KIT*-mutant tumors, and 21 times more *CXCL14* mRNA than the *KIT*-mutant GIST-T1 (**Figure 4D, top**) and GIST-882 cell lines (not shown). Finally, *CXCL14* has been shown to contribute to NK, CD4, CD8, and T regulatory cell chemotaxis and tumor regression in other tumor models (31, 32). UPG *PDGFRA*-mutant GISTs expressed

CXCL14 mRNA at higher levels than what has been shown to lead to immune-mediated tumor regression in HPV^+ head, neck and cervical cancers, while *KIT*-mutant tumors expressed CXCL14 mRNA at levels below this threshold (**Figure 4D, bottom**). Together, these findings suggest that CXCL14 expression may contribute to the observed differences in immune infiltration between *PDGFRA* and *KIT*-mutant GISTs (31, 32).

PDGFRA mutation produces multiple, HLA-diverse, strong binding neoepitopes.

Given the association of neoantigen presentation with inflamed tumor/immune microenvironments in other malignancies, we hypothesized that *PDGFRA*-mutant GISTs may produce more potent neoepitopes when compared to *KIT*-mutant GISTs. To explore this, we first mapped out the 8-, 9-, and 10-mer neoepitopes produced by all driver mutations in our cohort, generating neoepitopes with the mutation placed at each amino acid position in the mutant peptide. Then, we tested the binding affinity of each neoepitope to the matched, patient-specific HLA type in which that mutation was found.

First, we wanted to determine if total neoepitope burden and the number of high affinity neoepitopes produced by each GIST specimen correlated with immune infiltration or cytolytic activity in GIST, since neoepitope burden has been shown to correlate with increased response to immunotherapy in a variety of cancers (33, 34). However, total neoepitope burden and the number of high-affinity (<500nM binding) or very high-affinity (<50nM binding) neoepitopes did not correlate with immune infiltration or cytolytic activity across all mutations in the GIST cohort (**Figure 5A**) or across *KIT* and *PDGFRA*-mutant GISTs specifically (not shown).

We observed that the majority of patients with an oncogenic mutation in *PDGFRA* or *KIT* produced at least one high-affinity neoepitope (<500 nM binding), regardless of mutation subtype (**Figure 5B**). We also found that the *D842V* mutation produced 34 unique, high-affinity neoepitope:HLA specific binding combinations across the predicted HLA-types in our cohort of 16 *D842V*-mutant patients (**Figure 5B**), while *KIT exon 11* and other *PDGFRA* mutations on average only produced 2-3 per mutation (**Supplemental Table 6**). The *D842V* mutation generated 6 different neoepitopes that bound to 12 different HLA types in 14 patients in our cohort (**Supplemental Table 7**), one of which is the most common HLA type in the United States and present in 95.7% of Caucasians (HLA*A02:01) (35, 36), suggesting that this mutation could produce an immune response in a wide variety of patients, whereas *KIT exon 11* neoepitopes bound to less prevalent HLA-types.

Since the number of patients with *D842V* mutations (n=16) exceeded the number of patients with any individual *KIT* mutation (n=1 to 4, **Supplemental Table 7**) and this may obscure our HLA observations, we explored mutation-specific neoepitope binding diversity more broadly against all validated NetMHCPan3.0 HLA-types (**Figure 5C**). Again, nearly all *PDGFRA* and *KIT* GIST mutations produced at least one high affinity neoepitope capable of binding to many different HLA-types. *D842V* neoepitopes bound with a very high affinity (<50 nM binding) to HLA*A02:01 and many other prevalent HLA types. Interestingly, 50% (4/8) of all *PDGFRA* mutations produced at least one very high-affinity neoepitope (<50nM binding) to HLA*A02:01, compared to only 14% of *KIT* mutations (2/14). Thus, while all *KIT* and *PDGFRA*-mutant GISTs appear to produce high-affinity neoepitopes, *PDGFRA*-mutant GISTs appear to produce more high-affinity

neoepitopes to more common HLA types. Given that GIST is often driven by a single oncogenic mutation, and GIST harbors a rich immune infiltrate, this raises the possibility that oncogenic GIST mutations produce neoepitopes that may classify as what has recently been provocatively described as high quality neoepitopes (37).

Machine learning identifies immune gene signatures for GIST.

Through bioinformatics prediction and biologic validation methods, it appears that *PDGFRA*-mutant GIST contains more immune cells with a higher cytolytic activity than *KIT*-mutant GIST, and both subsets contain a unique immune profile. This makes characterizing a global immune infiltrate in GIST challenging, and suggests that specific immunotherapy approaches may ultimately need to be targeted to the specific oncogenic mutation in GIST. We therefore sought to define the most important immune features differentiating *PDGFRA* and *KIT*-mutant GISTs through machine learning, which is an unbiased approach not dependent on previously defined gene set analysis, allowing for the discovery of novel and impartial observations. We specifically used all *KIT* and *PDGFRA*-mutant GISTs (**Supplemental Table 3**; **n=61**), as well as only UPG *KIT* and *PDGFRA*-mutant GISTs (**Supplemental Table 4; n=22**) to develop two random forest models, and assumed that these immune differences would most likely be related to the oncogenic driver.

The combination of salient immune features capable of most accurately profiling a *PDGFRA* or *KIT*-mutant GIST were identified based on each feature's "importance," which is calculated as part of caret for R's implementation of randomForest (38). The "feature importance" metric is a measure indicating how useful a feature is for

separating samples into their distinct classes. The features with the highest importance provide the classifier with the highest increase in performance.

We included all 117 immune features to initially create the model (**Supplemental Table 5**) and subsequently retrained the model using fewer features. We noted that including more features in the model increased the model's performance on the training set, but decreased the model's performance on the test set, a phenomenon known as overfitting (**Supplemental Figure 3A**). Therefore, to prevent overfitting, we retrained the model using only the 6 most important features, which we believe will broaden the applicability of our model to future GIST samples. When restricting the model to the six most relevant features (**Figure 6A**), 27/30 *KIT*-mutant GISTs were correctly classified as *KIT*-mutant in the training set, while 14/20 *PDGFRA*-mutant GISTs were correctly assigned as *PDGFRA*-mutant (Sensitivity = 70%, Specificity = 90%). The top immune features included CXCL14, TGFBR1, TNFSF9, MICA, TNFRSF25, and CD96 (**Figure 6B**). Notably, the *PDGFRA*-mutant tumors that were incorrectly classified as *KIT*-mutant by our model had lower ESTIMATE and Cyt scores than correctly classified *PDGFRA*-mutant tumors (**Supplemental Figure 4A**), while *KIT*-mutant tumors incorrectly classified as *PDGFRA*-mutant had more variable CXCL14 mRNA expression and higher TGFBR1 and CD96 mRNA expression than correctly classified *KIT*-mutant tumors **(Supplemental Figure 4B).** In our separate cohort of testing samples (n=11), our 6-feature model correctly identified *KIT* and *PDGFRA*-mutant GISTs with 91% accuracy (**Figure 6C**). To demonstrate the importance of these 6 immune features in classifying *KIT* and *PDGFRA*-mutant GISTs, we excluded these features and retrained our random

forest model, which reduced classifier performance and model accuracy from 91% to 72.7% (**Supplemental Figure 3B**).

We further validated our model on an external cohort of GISTs from the CINSARC study ((39), n=12). Application of our immune profiling model to the CINSARC cohort correctly differentiated *KIT* and *PDGFRA*-mutant GISTS with 83.3% accuracy (**Figure 6C**). Similar results were obtained when we created a model using only UPG *PDGFRA* and *KIT*-mutant GISTs (**Figure 6D, 6E, and 6F**), in which CXCL14, IDO, TNFSF14, KDR, MICA, and TNFRSF9 were the most important features defining the model.

We employed the same machine learning approach to identify gene expression-based immune features associated with high PD-1 and PD-L1 expression across all GISTs (n=75), in order to discover novel therapeutic targets and potential barriers to immune checkpoint blockade in GIST. Interestingly, our 6-feature model showed that increased CD27 expression was highly related to PD-1 upregulation (Sensitivity=90%, Specificity=87.1%, **Figure 7A**), along with PDCD1LG2, BTLA, CD40, CXCL14, and MICB, suggesting that these pathways may represent synergistic opportunities for PD-1 immunomodulation in GIST (**Figure 7B**). Using only these 6 features, our model was able to correctly predict high PD-1 expression in 92.9% of the testing cohort and 66.7% of the external CINSARC GIST cohort (**Figure 7C**).

PD-L1 mRNA expression correlated with PD-L1 protein expression in our cohort (**Supplemental Figure 5**). When training our PD-L1 random forest model, expression of key antigen presenting machinery, including B2M, HLA.DRA, and TAP2, and immune responsive cytokines CXCL10 and LTA were predictive of increased PD-L1 expression

across all GISTs (Sensitivity=76.7%, Specificity=87.1%, **Figure 7D**), suggesting a relationship between immune checkpoint ligand upregulation and immune recognition (**Figure 7E**). Applying the model of these antigen presenting machinery features and immune cytokines, along with PD-L2, which was also identified as a feature predictive of high PD-L1 expression, our model correctly identified high PD-L1 expression with 85.7% accuracy in the testing cohort and 91.7% of the CINSARC GIST cohort (**Figure 7F**).

Discussion

The dramatic success of immune checkpoint blockade in a variety of notoriously difficult-to-treat cancers has nearly standardized immunomodulation as an approach for cancer treatment (40-42). However, subsequent research has advanced our understanding of how cancer type and tumor-cell specific characteristics, such as genomic mutational burden can predict response to immunotherapy (33, 34). It has also recently been shown that the state of the immune microenvironment can prevent response to immunotherapy if not appropriately addressed (43). In fact, it has become clear that the complex relationship between cancer cell intrinsic factors and the subtleties of the cancer-type specific immune microenvironment may make a generalized strategy of immune checkpoint blockade ineffective (44). Therefore, it has become increasingly evident that cancer-specific immunotherapeutic strategies are needed.

In this study, we performed RNA sequencing on 75 human GIST specimens from 75 patients to characterize the immune landscape of different GIST mutational drivers, which we believe is the largest cohort of GISTs to be analyzed with next-generation RNA sequencing. After observing potential differences in immune infiltration, we focused on the two prevalent GIST mutational subtypes, namely *KIT* and *PDGFRA*, and found that *PDGFRA*-mutant GISTs contain more immune cells than *KIT*-mutant GISTs. This suggests that patients with *PDGFRA*-mutant GIST may ultimately have a greater potential to respond to immunotherapy.

Through differential gene expression analysis, we identified mutation specific immune landscapes that may be significant for directing GIST immunotherapy. First,

PDGFRA-mutant GISTs contained higher expression of CD96, suggesting that NK cells may be playing a more significant role in the immune environment of *PDGFRA*-mutant GIST when compared to *KIT*. In fact, NK cell infiltrate has been suggested to be a contributor to long-term survival in *KIT*-mutant GIST (45), which may help explain why *PDGFRA*-mutant GIST patients have a better overall prognosis (4). Similarly, *PDGFRA*-mutant GISTs had higher CD4, CCR4, and CCR5 expression (indicative of T regulatory cells), and higher expression of the immunomodulators BTLA, CD48, TNFRSF9, and TIGIT. *PDGFRA*-mutant GISTs, therefore, may be more receptive to a variety of immunotherapeutic approaches, such as targeting TIGIT, BTLA, CD48, or TNFRSF9 in addition to T regulatory cells, NK cells, and PD-1/PD-L1 checkpoint blockade.

Comparison of UPG GISTs of the *KIT* and *PDGFRA*-mutant phenotypes revealed that mitotic rate, and indicator of GIST aggressiveness, is inversely related to the quantity of the immune infiltrate. While this concept has been described in other cancers, tumor proliferative index has not yet been shown to affect the quantity of the immune infiltrate in GIST. *PDGFRA*-mutant GIST has been shown to have a lower proliferative index when compared to *KIT*-mutant GISTs (46). However, when controlling for mutational driver the relationship between mitotic rate and immune cell infiltration in UPG GISTs loses significance, suggesting that oncogenic driver may be more important for dictating the immune infiltrate. In fact, comparison of UPG *KIT* and *PDGFRA*-mutant GISTs showed distinct cytokine profiles, which we hypothesize may be contributing to immune cell chemotaxis. CXCL14 is a relatively novel chemokine with a clear role in immune cell attraction (31, 32, 47). CXCL14 was found to be significantly upregulated in *PDGFRA*-mutant GISTs compared to *KIT*, not only in our cohort but also

in a previously published cohort (30). Unfortunately, studying this correlation has proven challenging, as there is currently no *PDGFRA*-mutant GIST mouse model or human *PDGFRA*-mutant GIST cell line in which to explore what appears to be a clear biological relationship between PDGFRA tyrosine kinase activation and CXCL14 secretion. Thus, the implication of our finding at this time remains unclear.

It is clear, however, that in cancers with a high number of mutations, tumor mutational burden correlates with immunotherapy response (33). However, the role of neoepitopes in the immune response of cancers with a single, oncogenic mutation (i.e., GIST) has not been described. To our surprise, nearly every GIST mutation in our cohort produced a neoepitope capable of binding with high-affinity to patient-specific HLA Class I, suggesting that even cancers with one mutation can generate a neoepitope recognizable by the immune system. We discovered that the *D842V* mutation produced 6 different high affinity neoepitopes, which bound to 12 different HLA types of patients in this cohort alone, one of which is the most common HLA type in the United States and in Caucasians (35, 36). While we have not shown that these neoepitopes produced an immune response in these patients, it is intriguing that GISTs with a larger immune infiltrate appear to have a mutation that generates peptides with broad HLA binding specificity, suggesting that this mutation may be of higher quality (37). Moreover, the HLA-specific binding affinities of neoepitopes generated by mutation-specific GIST subtypes reported in this study may provide a road map for patient selection in future neoantigen vaccination trials.

We employed machine learning in order to identify key differences and to profile the expression of immune-related genes between *PDGFRA* and *KIT*-mutant GISTs.

Generating a gene-expression based immune-profile which accurately predicted the *KIT* or *PDGFRA*-mutant genotype not only supports our hypothesis that the genotype may impact the immune infiltrate, but also introduces another method by which profiling the immune landscapes of *KIT* and *PDGFRA*-mutant tumors can lead to more personalized cancer immunotherapy. Not surprisingly, CXCL14 appeared to be a top feature of *PDGFRA*-mutant GIST, again suggesting that CXCL14 expression is oncogene-specific and may be responsible for the observed differences in immune infiltration and activity. TGFBR1, TNFSF9, MICA, TNFRSF25, CD96, IDO, TNFSF14, KDR, and TNFRSF9 all contributed to the predictive capacity of the immune profiling models, which resulted in >90% diagnostic accuracy in our testing set and in 83% of the CINSARC validation set. We have previously shown the importance of IDO in dictating the immune phenotype of GIST (13). The role of NK cells and the additional immune modulatory receptors identified in this study should continue to be explored.

Since immunotherapy has not yet shown efficacy in advanced and metastatic GIST (48), we sought to explore the immune profiles of PD-1 high and PD-L1 high GISTs across our entire GIST cohort to identify both therapeutic opportunities and potential barriers to effective immunotherapy. CD27 was found to be a significant feature predictive of high PD-1 expression in our random forest model, which correctly identified high PD-1 expression in 93% of our testing GIST cohort. Agonistic CD27 antibodies already exist and have shown efficacy with anti-PD-1 in other models, suggesting that this may be an effective strategy for GIST (49). Similarly, the co-stimulatory and checkpoint receptors CD40 and BTLA, along with the immune checkpoint ligand PDCD1LG2 and the MHC I protein MICB were features predictive of

high PD-1 expression. We have previously shown that CD40 ligation enhances antitumor immunity in our *Kit exon 11* mouse model (17), and these other receptors should be considered in future trials.

Notably, applying our PD-1 model to the external CINSARC cohort only resulted in 67% accuracy, even though our PD-L1 model, in which the same methodology was used, was more accurate (92%). One reason may be related to the small sample size of the external cohort. However, these models identify high-priority targets for further experimental validation.

Finally, expression of key antigen presenting machinery including B2M, PDCD1LG2, HLA-DRA, and TAP was found to be highly predictive of PD-L1 expression not only in our cohort (86% accuracy) but also in the external CINSARC GIST cohort (92% accuracy), suggesting a tumor-cell specific mechanism of linking neoepitope presentation with immune suppression. A similar finding has been observed in HIV infected cells, where MHC class I machinery was upregulated concurrently with PD-L1 (50). This is an important concept to consider, as efforts to block PD-L1 may result in tumor cell changes that also downregulate MHC class I antigen presentation, which is essential for the CD8 T cell-mediated immune response. High PD-L1 expression has also recently been shown to independently correlate with a worse prognosis in soft tissue sarcomas, which was attributed to an increased degree of immune exhaustion, suppression, and negative regulation (51). In contrast, efforts to enhance antigen presentation using neoantigen peptide vaccination or GM-CSF-expressing tumor cell vaccines may require PD-L1 inhibition.

Ultimately, it is becoming increasingly evident that immunotherapy requires a specific approach targeted not only to the tumor type but also the tumor-specific immune microenvironment. Through RNA sequencing and four trained machine learning models, we have characterized important and consistent differences in the immune landscape of GISTs driven by different oncogenic mutations, which we believe may help guide future immunotherapy trials in GIST.

Materials and Methods

Human GIST specimens and clinicopathologic features. Fresh surgical specimens were collected immediately upon resection and flash frozen in liquid nitrogen. RNA was extracted using the RNeasy kit (Qiagen) as per the kit-specified protocol. Clinicopathologic features were obtained via chart review of patient and pathologic records.

***Bioinformatics*.** Poly-A selection and next-generation RNA sequencing was performed on an Illumina HiSeq 2500 platform by the institutional Integrated Genomics Operations core facility. A minimum of 40-50 million 50bp paired-end reads was obtained for each sample. RNA sequencing reads were processed using Trimmomatic v0.36 (52) using the following parameters recommended for paired end Illumina sequencing: PE -phred33, ILLUMINACLIP: TruSeq3-PE.fa:2:30:10, LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 4:15, and MINLEN: 36. Sequencing reads were then aligned to the human genome (version hg38_r88) and gene-level counts were calculated using STAR version 2.6.0a (53). Read counts were then normalized and differential gene expression (DGE) analyses were performed on indicated groups using the R software package DESeq2 (54, 55). Gene set enrichment analysis (GSEA) was performed using the java GSEA software package (version 3.0) and the molecular signature databases (MSigDB) version 6.2 from the Broad Institute (56, 57).

The activity of cytotoxic T-cells within each sample was estimated utilizing the CYT score, which is calculated from the geometric mean of the normalized read counts for Perforin (PRF1) and Granzyme A (GZMA) as previously described (23). The immune

cell compartment of these samples was further quantified from the normalized read counts using the R package ESTIMATE (23). The ESTIMATE algorithm is composed of a non-immune “stromal score” parameter and an “immune score” parameter. In this work, we refer to ESTIMATE score as the “immune score” subcomponent of the ESTIMATE algorithm, which infers the degree of immune cell infiltration within tumor tissue based on expression data of 141 immune-related genes previously shown to correlate with the presence of leukocytes (24). The absolute quantities of individual immune cell subtypes was inferred from the normalized read counts by using the CIBERSORT program (29), which employs an externally-validated leukocyte gene signature matrix (LM22) of 547 genes designed to distinguish 22 cell phenotypes including seven T cell types, B cells, macrophages, monocytes, and NK cells (29). Heatmaps of gene expression data with corresponding clinicopathologic characteristics were created with the R package ComplexHeatmap (58). Single sample gene set enrichment analyses (ssGSEA) were performed using the R software package GSVA (59).

Neoepitopes and HLA prediction. Patient HLA haplotypes were inferred using the program Seq2HLA (60). For neoepitope identification, the amino acid sequences of patient-specific mutations were mapped to combinations of 8-, 9-, and 10-mer peptides, with the mutation evaluated at each amino acid position. Each patient-specific peptide was then tested with the respective patient-specific HLA-type using NetMHCPan3.0 to identify potential high affinity neoepitopes (61, 62). Neoepitope burden was defined as any epitope with a predicted binding affinity within the top 2% when compared to 400

random natural peptides, as recommended by NetMHCPan3.0. High affinity neoepitopes were further filtered using a binding affinity threshold of 500nm, while very high affinity neoepitopes were defined as <150nM binding affinity. Finally, to explore mutation-specific neoepitope binding diversity, mutation-specific neoepitopes were tested more broadly against all validated NetMHCPan3.0 HLA-types and visualized using heatmap.2 in R.

Machine learning. Random forest modeling with 5-fold cross validation was performed on the training set using the software package caret for R to identify top predictors (38). One hundred and seventeen features, comprising normalized read counts of various immune genes and bioinformatically inferred CIBERSORT scores were used to train the model (**Supplemental Table 5**). For characterization of *PDGFRA*-specific immune features, *KIT* and *PDGFRA*-mutant GIST samples were first partitioned randomly into training (80% of samples) and testing sets (20% of samples). The training data set was used to train a random forest machine learning classifier using 5-fold cross validation. During each of the 5 folds, a random forest model was trained on 80% of the training data, and training accuracy, sensitivity, and specificity was calculated; the remaining 20% of the training data was then used to calculate testing accuracy, sensitivity and specificity. Finally, our trained, machine learning model was applied to the hold-out, 20% testing set and to the external validation CINSARC cohort (n=12, (39)) in order to assess model accuracy on a naïve sample set.

Given the total sample size of our cohort and to prevent overfitting of our model, random forest modeling with 5-fold cross validation was re-performed on the training set

using only the top 6 identified features, and the model was then applied to the naïve testing set and CINSARC external data set to assess for model accuracy. Feature "Importance" is a variable calculated by the caret package, with a higher number indicating that the feature is more important for model accuracy (38). The confusion matrices shown in Figures 6 and 7 reflect the model with the highest accuracy after 5-fold cross validation and optimization by caret. The p-value used by caret is calculated using an exact binomial test in which a one-sided test is performed to see if the model accuracy rate is better than the "no information rate," which represents the proportion of data within the majority class. For generation of PD-1 and PD-L1-specific immune features, the same approach was used, with groups split into high and low expression based on median PD-1 (median=14.1) or PD-L1 (median=209.5) normalized read counts in the training set.

Quantitative real-time PCR. Total RNA was extracted from snap-frozen human tumors, reverse transcribed, and amplified with PCR TaqMan probe for *CXCL14* (Hs01557413_m1). Quantitative PCR was performed using a ViiATM7 real-time PCR system (Applied Biosystems). Data were calculated by the $2^{-\Delta\Delta Ct}$ method as described by the manufacturer's protocol and were expressed as fold increase over the GIST-T1 or GIST-882 cell line control (63, 64).

Flow cytometry. Flow cytometric analysis was immediately performed at the time of specimen collection on freshly obtained human GIST specimens after tumor dissociation as before (15). Cells were analyzed on a BD FACS Aria or LSR Fortessa

(BD Biosciences). A human-specific antibody for CD45 (2D1) was obtained from BD Biosciences.

Western Blot*.* Protein from flash frozen GIST tissue or cell lines was analyzed as before (18). Antibodies used against glyceraldehyde-3 phosphate dehydrogenase (GAPDH; Clone D16H11) and PD-L1 (Clone E1L3N) were purchased from Cell Signaling Technology.

Histology. Freshly collected tumors were fixed in 4% paraformaldehyde, embedded in paraffin, and cut into 5-μm sections. CD45 (PD7/26 + 2B11) and CD8 (SP57) staining was performed by the institutional Molecular Cytology Core. Slides were scanned with MIRAX scan (Zeiss) and photographed with Pannoramic Viewer. For quantification of CD45 and CD8 staining, cell counting was performed manually in a blinded fashion at 20X magnification (HPF). The number of $CD45^+$ and $CD8^+$ cells per HPF was calculated by examining 5 HPFs per tumor and plotting the average per tumor.

Statistical Analysis. Some data and graphs were created using Prism 7.0 (Figures 2A, 2B, 2E, 4C, 4D, 5A, 5B, 6B, 6E, 7B, 7E; GraphPad Software). Statistical tests were performed as described in the Figure legends, and when possible, adjusted p-values calculated by DESeq2 were used to assess for significance, with a threshold of adjusted $p<0.05$. For Figures 7B and 7E, a False Discovery Rate (FDR) approach was taken using the Benjamini, Krieger, and Yekutieli two-stage step-up method with a q-value threshold of <0.1 in order to assess for significance.

Data Availability and CINSARC External Validation GIST Cohort. RNA sequencing data presented in this work will be available through the Sequencing Read Archive under submission no. SUB5135588 upon publication of this work. RNA sequencing data of fresh-frozen GIST tumors from the publically available CINSARC cohort (39) were obtained from the Sequence Read Archive under accession no. SRP057793.

Study Approval. Tumor specimens were obtained from 2011 to 2017 from 75 GIST patients who underwent surgical resection at Memorial Sloan Kettering Cancer Center. Prior to resection, patients were consented under an IRB-approved protocol. Using a prospectively maintained institutional database, clinicopathologic characteristics were documented and encrypted, and all research was conducted in compliance with Health Insurance Portability and Accountability Act regulations.

Author Contributions

GAV, TGB, BDM, and RPD contributed to study conception and design. GAV, TGB, and RPD contributed to the development of research methodology. GAV, TGB, BDM, JQZ, JKL, NJP, SZ, ML, and FC contributed to the acquisition of data and conduction of experiments. GAV, TGB, ML, RG, NJP, FR, and RPD contributed to the analysis and interpretation of data. GAV, TGB, and RPD contributed to the writing, review, and revision of the manuscript. All authors reviewed and contributed comments that were incorporated into the final version of this manuscript. Study supervision was provided by RPD.

Acknowledgements

The investigators were supported by NIH grants R01 CA102613 and T32 CA009501, The David Foundation, Betsy Levine-Brown and Marc Brown, David and Monica Gorin, and the GIST Cancer Research Fund (RPD); and the Division of Loan Repayment NIH grant L30 TR002111 (GAV). The Molecular Cytology Core Facility at Memorial Sloan Kettering was supported by Cancer Center Support Grant P30 CA008748. We thank Russell Holmes for logistical and administrative support.

References

1. Ducimetiere, F., et al. 2011. Incidence of sarcoma histotypes and molecular subtypes in a prospective epidemiological study with central pathology review and molecular testing. *PLoS One* 6:e20294.
2. Hirota, S., et al. 1998. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* 279:577-580.
3. Joensuu, H., and DeMatteo, R.P. 2012. The management of gastrointestinal stromal tumors: a model for targeted and multidisciplinary therapy of malignancy. *Annu Rev Med* 63:247-258.
4. Corless, C.L., et al. 2014. Pathologic and molecular features correlate with long-term outcome after adjuvant therapy of resected primary GI stromal tumor: the ACOSOG Z9001 trial. *J Clin Oncol* 32:1563-1570.
5. Corless, C.L., et al. 2005. PDGFRA mutations in gastrointestinal stromal tumors: frequency, spectrum and in vitro sensitivity to imatinib. *J Clin Oncol* 23:5357-5364.
6. Lasota, J., Dansonka-Mieszkowska, A., Sobin, L.H., and Miettinen, M. 2004. A great majority of GISTs with PDGFRA mutations represent gastric tumors of low or no malignant potential. *Lab Invest* 84:874-883.
7. Wardelmann, E., et al. 2004. Association of platelet-derived growth factor receptor alpha mutations with gastric primary site and epithelioid or mixed cell morphology in gastrointestinal stromal tumors. *J Mol Diagn* 6:197-204.
8. Cavnar, M.J., et al. 2017. Rectal Gastrointestinal Stromal Tumor (GIST) in the Era of Imatinib: Organ Preservation and Improved Oncologic Outcome. *Ann Surg Oncol* 24:3972-3980.
9. Oppelt, P.J., Hirbe, A.C., and Van Tine, B.A. 2017. Gastrointestinal stromal tumors (GISTs): point mutations matter in management, a review. *J Gastrointest Oncol* 8:466-473.
10. Burgoyne, A.M., et al. 2017. Duodenal-Jejunal Flexure GI Stromal Tumor Frequently Heralds Somatic NF1 and Notch Pathway Mutations. *JCO Precision Oncology*:1-12.
11. Miettinen, M., et al. 2011. Succinate dehydrogenase-deficient GISTs: a clinicopathologic, immunohistochemical, and molecular genetic study of 66 gastric GISTs with predilection to young age. *Am J Surg Pathol* 35:1712-1721.
12. Shi, E., et al. 2016. FGFR1 and NTRK3 actionable alterations in "Wild-Type" gastrointestinal stromal tumors. *J Transl Med* 14:339.
13. Balachandran, V.P., et al. 2011. Imatinib potentiates antitumor T cell responses in gastrointestinal stromal tumor through the inhibition of Ido. *Nat Med* 17:1094-1100.
14. Cavnar, M.J., et al. 2013. KIT oncogene inhibition drives intratumoral macrophage M2 polarization. *J Exp Med* 210:2873-2886.
15. Seifert, A.M., et al. 2017. PD-1/PD-L1 Blockade Enhances T-cell Activity and Antitumor Efficacy of Imatinib in Gastrointestinal Stromal Tumors. *Clin Cancer Res* 23:454-465.

16. Sommer, G., et al. 2003. Gastrointestinal stromal tumors in a mouse model by targeted mutation of the Kit receptor tyrosine kinase. *Proc Natl Acad Sci U S A* 100:6706-6711.
17. Zhang, J.Q., et al. 2018. Macrophages and CD8(+) T Cells Mediate the Antitumor Efficacy of Combined CD40 Ligation and Imatinib Therapy in Gastrointestinal Stromal Tumors. *Cancer Immunol Res* 6:434-447.
18. Vitiello, G.A., et al. 2018. Mitochondrial Inhibition Augments the Efficacy of Imatinib by Resetting the Metabolic Phenotype of Gastrointestinal Stromal Tumor. *Clin Cancer Res* 24:972-984.
19. Zeng, S., et al. 2017. Wnt/beta-catenin Signaling Contributes to Tumor Malignancy and Is Targetable in Gastrointestinal Stromal Tumor. *Mol Cancer Ther* 16:1954-1966.
20. Zeng, S., et al. 2017. ETV4 collaborates with Wnt/beta-catenin signaling to alter cell cycle activity and promote tumor aggressiveness in gastrointestinal stromal tumor. *Oncotarget* 8:114195-114209.
21. Lawrence, M.S., et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499:214-218.
22. Senbabaoglu, Y., et al. 2016. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol* 17:231.
23. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. 2015. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160:48-61.
24. Yoshihara, K., et al. 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612.
25. Pantaleo, M.A., Nannini, M., Corless, C.L., and Heinrich, M.C. 2015. Quadruple wild-type (WT) GIST: defining the subset of GIST that lacks abnormalities of KIT, PDGFRA, SDH, or RAS signaling pathways. *Cancer Med* 4:101-103.
26. Gold, J.S., et al. 2009. Development and validation of a prognostic nomogram for recurrence-free survival after complete surgical resection of localised primary gastrointestinal stromal tumour: a retrospective analysis. *Lancet Oncol* 10:1045-1052.
27. Cameron, S., Gieselmann, M., Blaschke, M., Ramadori, G., and Fuzesi, L. 2014. Immune cells in primary and metastatic gastrointestinal stromal tumors (GIST). *Int J Clin Exp Pathol* 7:3563-3579.
28. Charoentong, P., et al. 2017. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* 18:248-262.
29. Newman, A.M., et al. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453-457.
30. Kang, H.J., et al. 2005. Correlation of KIT and platelet-derived growth factor receptor alpha mutations with gene activation and expression profiles in gastrointestinal stromal tumors. *Oncogene* 24:1066-1074.
31. Cicchini, L., et al. 2016. Suppression of Antitumor Immune Responses by Human Papillomavirus through Epigenetic Downregulation of CXCL14. *MBio* 7.

32. Lee, H.T., et al. 2017. A Crucial Role of CXCL14 for Promoting Regulatory T Cells Activation in Stroke. *Theranostics* 7:855-875.
33. Goodman, A.M., et al. 2017. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther* 16:2598-2608.
34. Hellmann, M.D., et al. 2018. Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* 33:843-852 e844.
35. Ellis, J.M., et al. 2000. Frequencies of HLA-A2 alleles in five U.S. population groups. Predominance Of A*02011 and identification of HLA-A*0231. *Hum Immunol* 61:334-340.
36. Maiers, M., Gragert, L., and Klitz, W. 2007. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol* 68:779-788.
37. Balachandran, V.P., et al. 2017. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* 551:512-516.
38. Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28:1-26.
39. Lesluyes, T., et al. 2016. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur J Cancer* 57:104-111.
40. Couzin-Frankel, J. 2013. Breakthrough of the year 2013. Cancer immunotherapy. *Science* 342:1432-1433.
41. Topalian, S.L., et al. 2012. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 366:2443-2454.
42. Wolchok, J.D., Rollin, L., and Larkin, J. 2017. Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* 377:2503-2504.
43. Messenheimer, D.J., et al. 2017. Timing of PD-1 Blockade Is Critical to Effective Combination Immunotherapy with Anti-OX40. *Clin Cancer Res* 23:6165-6177.
44. Topalian, S.L., Drake, C.G., and Pardoll, D.M. 2015. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 27:450-461.
45. Menard, C., et al. 2009. Natural killer cell IFN-gamma levels predict long-term survival with imatinib mesylate therapy in gastrointestinal stromal tumor-bearing patients. *Cancer Res* 69:3563-3569.
46. Campanella, N.C., et al. 2017. Lack of microsatellite instability in gastrointestinal stromal tumors. *Oncol Lett* 14:5221-5228.
47. Lu, J., Chatterjee, M., Schmid, H., Beck, S., and Gawaz, M. 2016. CXCL14 as an emerging immune and inflammatory modulator. *J Inflamm (Lond)* 13:1.
48. D'Angelo, S.P., et al. 2017. Combined KIT and CTLA-4 Blockade in Patients with Refractory GIST and Other Advanced Sarcomas: A Phase Ib Study of Dasatinib plus Ipilimumab. *Clin Cancer Res* 23:2972-2980.
49. Buchan, S.L., et al. 2018. PD-1 Blockade and CD27 Stimulation Activate Distinct Transcriptional Programs That Synergize for CD8(+) T-Cell-Driven Antitumor Immunity. *Clin Cancer Res* 24:2383-2394.
50. Pai, S.I., et al. 2018. HLA class I antigen processing machinery (APM) component expression and PD-1:PD-L1 pathway activation in HIV-infected head and neck cancers. *Oral Oncol* 77:92-97.

51. Bertucci, F., et al. 2017. PDL1 expression is a poor-prognosis factor in soft-tissue sarcomas. *Oncoimmunology* 6:e1278100.
52. Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
53. Dobin, A., et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21.
54. Love, M.I., Huber, W., and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
55. R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
56. Mootha, V.K., et al. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267-273.
57. Subramanian, A., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550.
58. Gu, Z., Eils, R., and Schlesner, M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847-2849.
59. Hanzelmann, S., Castelo, R., and Guinney, J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7.
60. Boegel, S., et al. 2012. HLA typing from RNA-Seq sequence reads. *Genome Med* 4:102.
61. Hoof, I., et al. 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61:1-13.
62. Nielsen, M., and Andreatta, M. 2016. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Medicine* 8:33.
63. Taguchi, T., et al. 2002. Conventional and molecular cytogenetic characterization of a new human cell line, GIST-T1, established from gastrointestinal stromal tumor. *Lab Invest* 82:663-665.
64. Tuveson, D.A., et al. 2001. STI571 inactivation of the gastrointestinal stromal tumor c-KIT oncoprotein: biological and clinical implications. *Oncogene* 20:5054-5058.

Figure 1

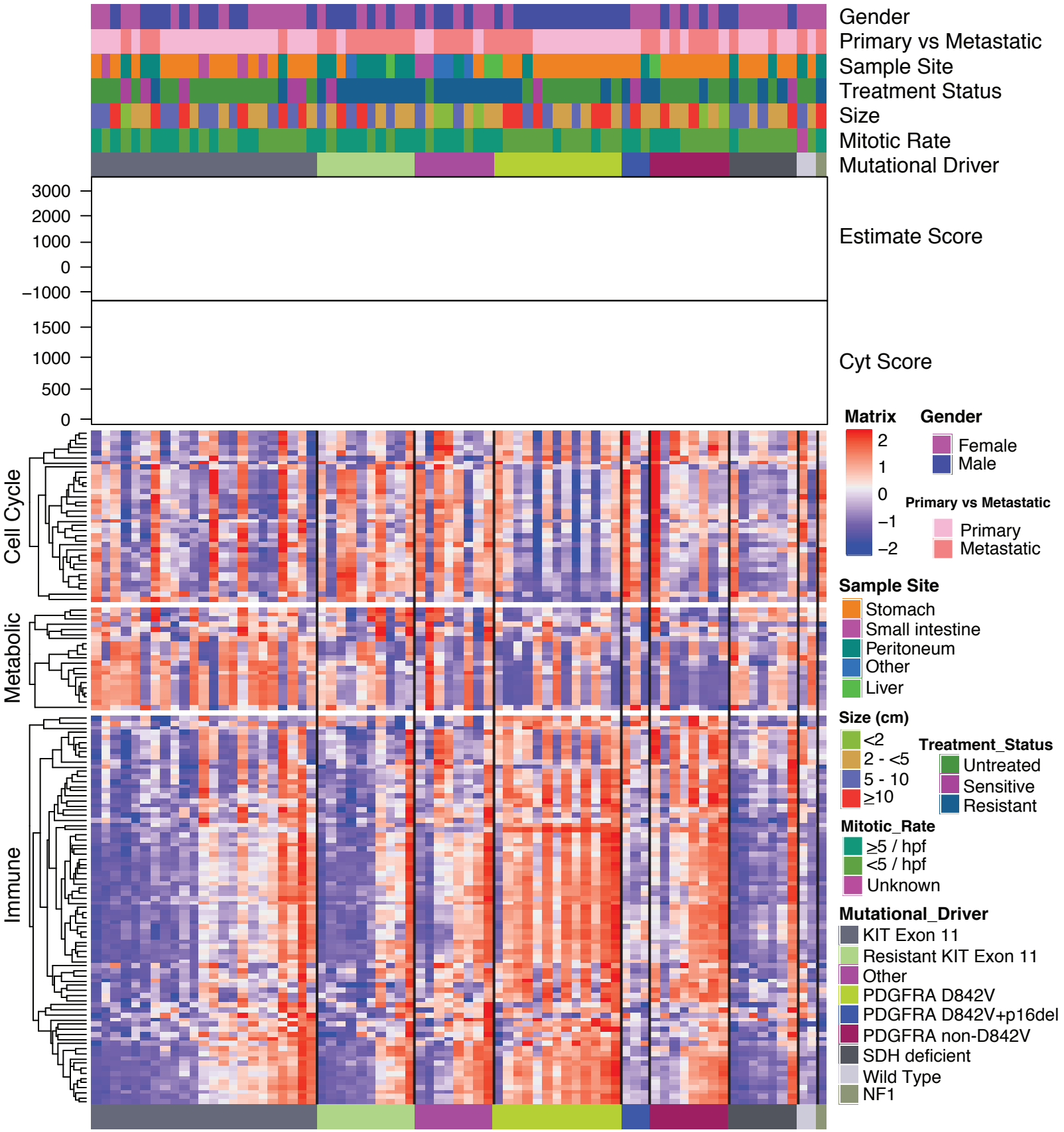


Figure 1. ssGSEA identifies immune cell pathway enrichment in *PDGFRA*-mutant GIST.
Single sample Gene Set Enrichment Analysis (ssGSEA) of 75 GIST specimens, organized by mutational driver and increasing ESTIMATE score. Unsupervised row clustering grouped gene sets into three major categories based on cell cycle pathways, metabolic pathways, and immune pathways. Clinicopathologic characteristics of the 75 GIST specimens are shown in the annotation above the heat map and in Supplemental Table 1.

Figure 2

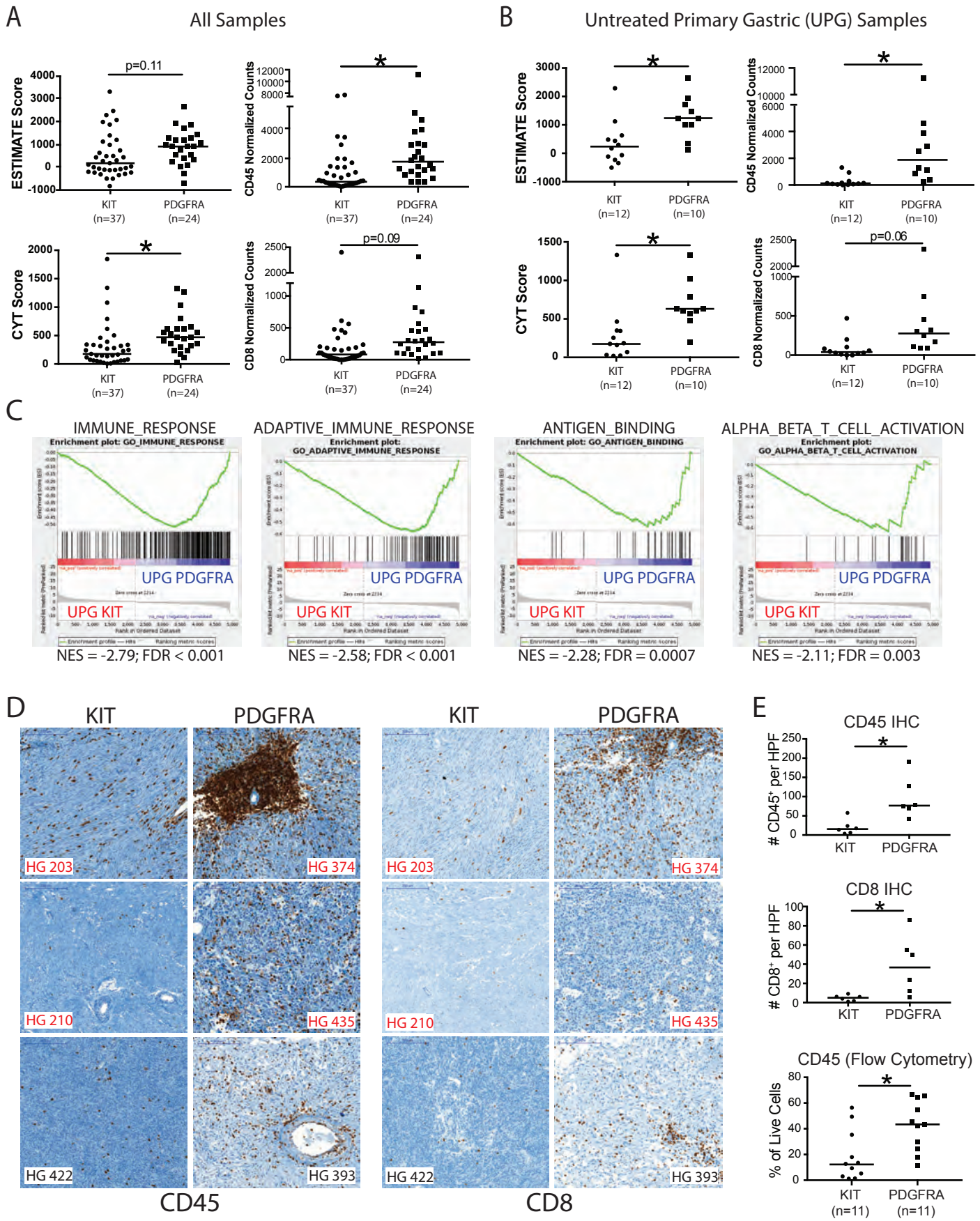


Figure 2. *PDGFRA*-mutant GIST is more immunologically active compared to *KIT*-mutant GIST.
(**A**) ESTIMATE and Cyt scores (left) and CD45 and CD8 normalized counts (right) of all *KIT* and *PDGFRA*-mutant GISTs (n=61; Supplemental Table 3). (**B**) ESTIMATE and Cyt scores (left) and CD45 and CD8 normalized counts (right) of UPG *KIT* and *PDGFRA*-mutant GISTs (n=22; Supplemental Table 4). (**C**) Gene set enrichment analysis showing multiple immune pathways enriched in UPG *PDGFRA*-mutant compared to UPG *KIT*-mutant GISTs. NES = Normalized Enrichment Score, FDR = False Discovery Rate. (**D**) 10x magnification of CD45 and CD8 immunohistochemical staining in *KIT* and *PDGFRA*-mutant GISTs. Red font color indicates specimen was included in RNA sequencing cohort, while samples with black font were not included. Representative samples shown of n=6 per group. (**E**) CD45 (top) and CD8 (middle) quantification of immunohistochemical staining. n=6 per group. The number of $CD45^+$ and $CD8^+$ cells per HPF was calculated by examining 5 HPFs per tumor and plotting the average per tumor. (bottom) CD45 expression by flow cytometry (for specimens in which flow cytometry data were available). UPG = untreated, primary, gastric *p<0.05, t-test. Bars, median.

Figure 3

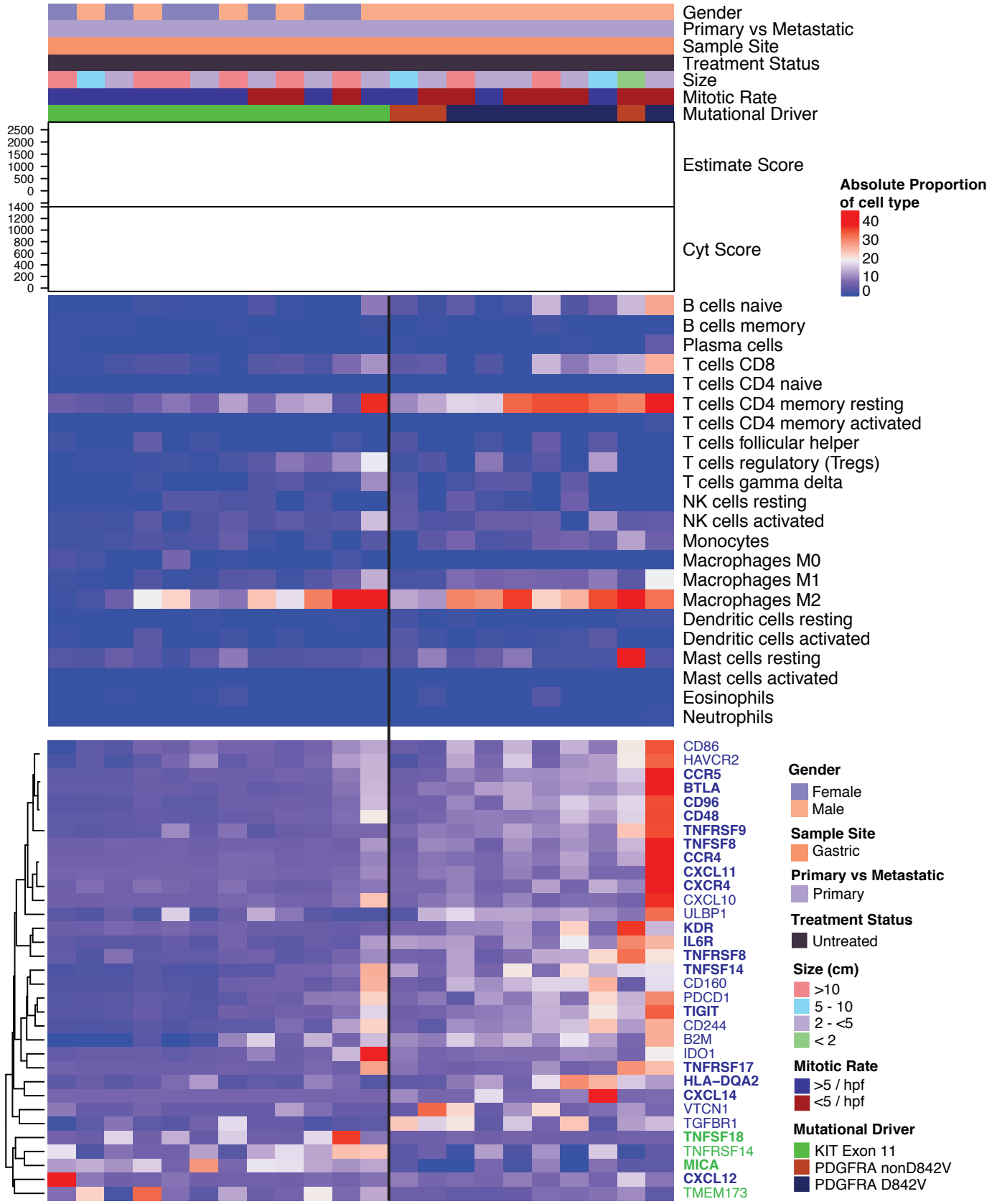
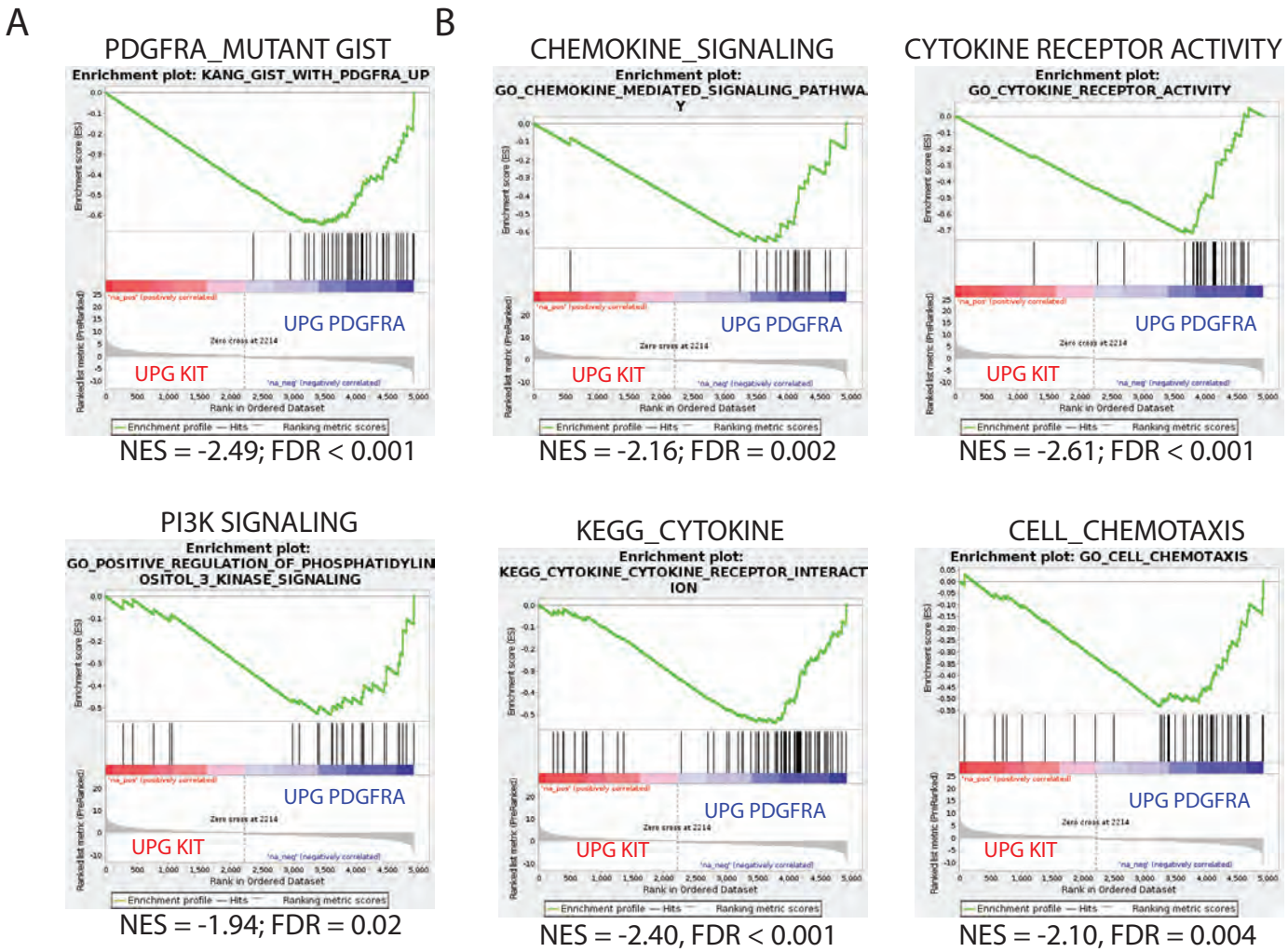
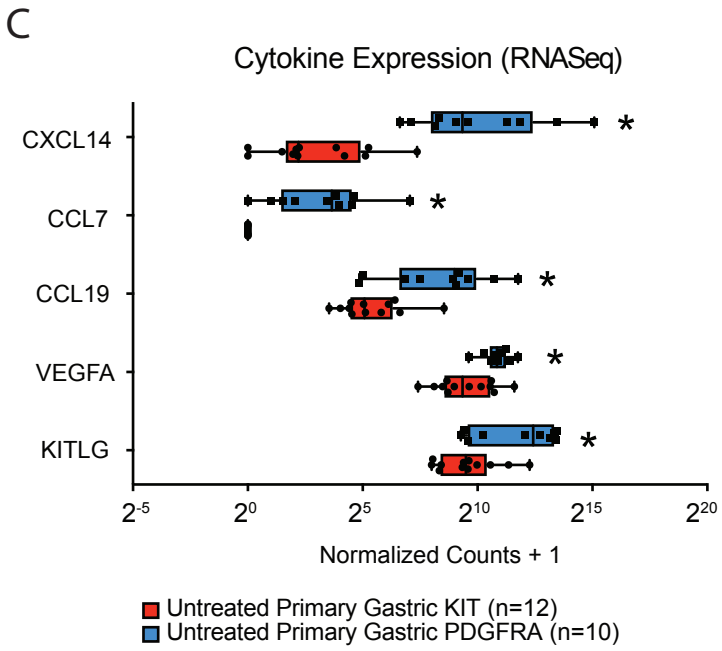


Figure 3. CIBERSORT and DGE analysis identify unique immune signatures in GIST.
CIBERSORT (middle) and immune gene expression (bottom) of UPG *KIT* and UPG *PDGFRA*-mutant GISTs, organized by mutational driver and increasing ESTIMATE score (n=22). Unsupervised row normalized clustering of genes shows grouping of genes into distinct groups, suggestive of oncogene driven immune profiles. Blue gene lettering indicates significantly enriched in UPG *PDGFRA* compared to UPG *KIT*-mutant, while green gene lettering indicates significant enrichment in UPG *KIT* compared to UPG *PDGFRA*-mutant. Enrichment was considered as an adjusted p-value <0.1 as calculated by DESeq2 for R, while bold gene lettering indicates a significant difference with an adjusted p-value of <0.05. UPG = untreated, primary, gastric. Clinicopathologic characteristics of the UPG GIST specimens are shown in the annotation above the heat map and in Supplemental Table 4.

Figure 4





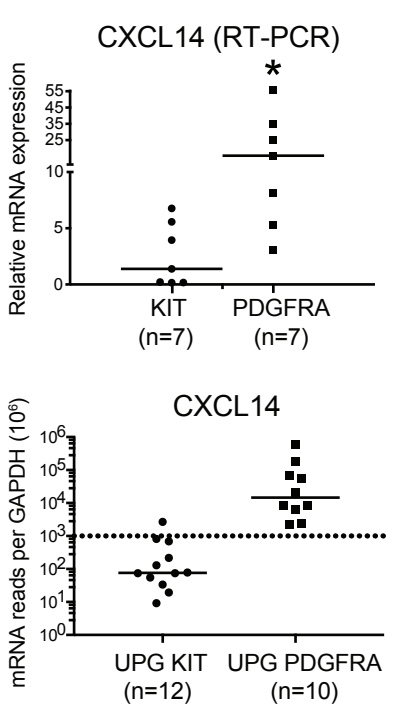


Figure 4. *PDGFRA* and *KIT*-mutant GISTs have distinct signaling and cytokine signatures.
Gene set enrichment analyses showing enrichment of (**A**) *PDGFRA* signaling, *PI3K* signaling, and (**B**) cytokine signaling pathways in UPG *PDGFRA* compared to UPG *KIT*-mutant GISTs (n=22). NES = Normalized Enrichment Score, FDR = False Discovery Rate. (**C**) Distribution of cytokines between UPG *KIT* and UPG *PDGFRA*-mutant GISTs, by RNAseq. Adjusted p-values <0.05 as calculated by DESeq2. (**D**) (top) Relative *CXCL14* mRNA expression by RT-PCR in *KIT* (n=7) and *PDGFRA*-mutant (n=7) GISTs from the RNAseq cohort, compared to GIST-T1 cell line (expression set at 1, not shown). (bottom) *CXCL14* mRNA expression relative to *GAPDH* x 10^6 in UPG *KIT* and *PDGFRA*-mutant GISTs. Horizontal dotted line represents *CXCL14* mRNA expression needed to induce tumor regression from Cicchini et. al. 2016. UPG = untreated, primary, gastric. *p<0.05, t-test. Bars, mean + SEM.

Figure 5

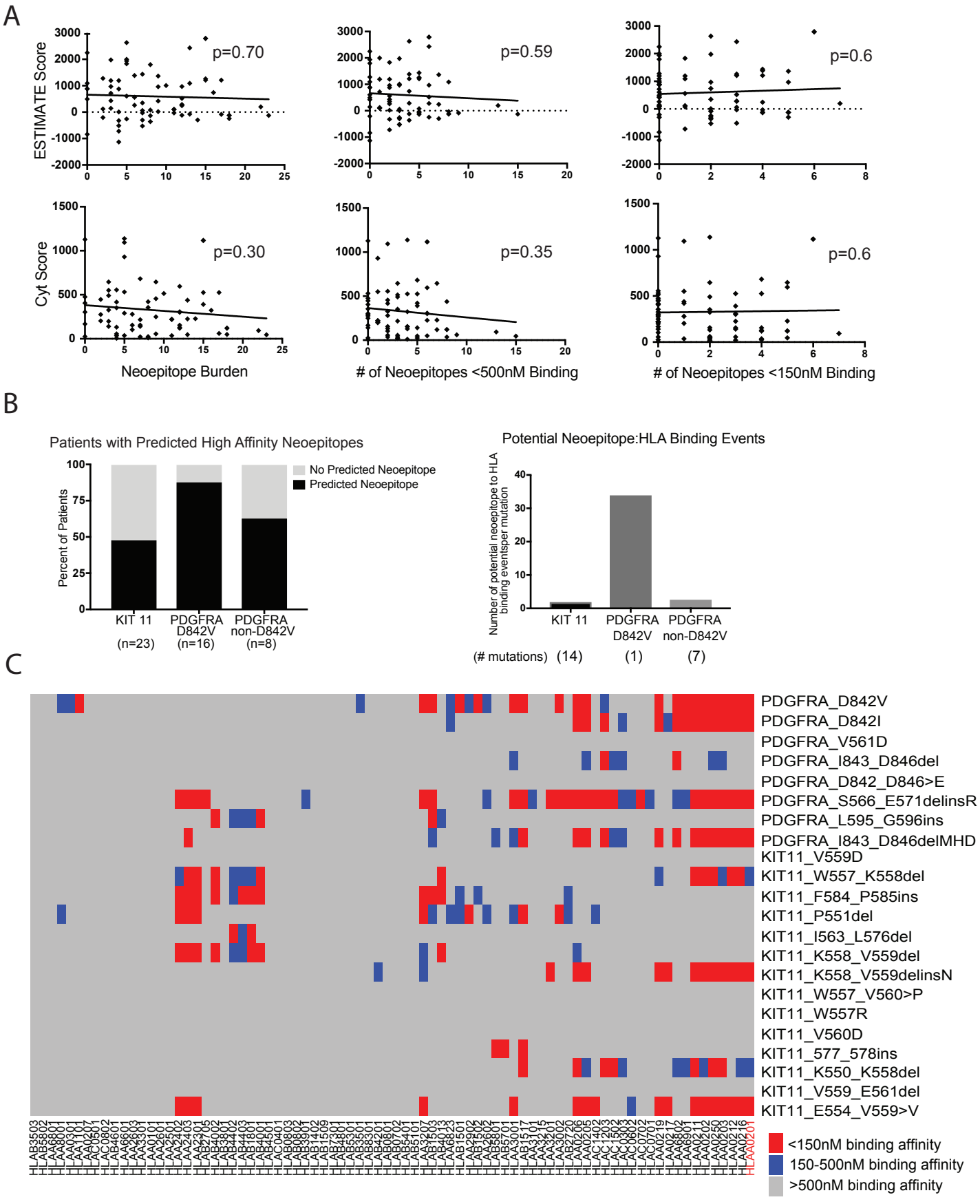


Figure 5. *PDGFRA* mutation produces multiple, HLA-diverse, strong binding neoepitopes.
(**A**) Pearson correlation of ESTIMATE (top) and Cyt (bottom) scores with total neoepitope burden (left), number of high affinity neoepitopes (middle), and number of very high affinity neoepitopes (right) among all GIST samples (n=75). (**B**) (left) Percent of patients with indicated mutation whose mutation produced a predicted high affinity neoepitope. (right) Number of potential neoepitope:HLA binding events produced by mutation type, averaged over the number of mutations per group. (**C**) Heat map of the binding affinities of *KIT* and *PDGFRA* mutation specific neoepitopes to all validated NetMHC3.0 HLA-types. Clinicopathologic characteristics of the GIST specimens are shown in Supplemental Table 3. Additional details regarding mutations, neoepitopes, and HLA types used to create this heat map are shown in Supplemental Tables 6 and 7.

Figure 6

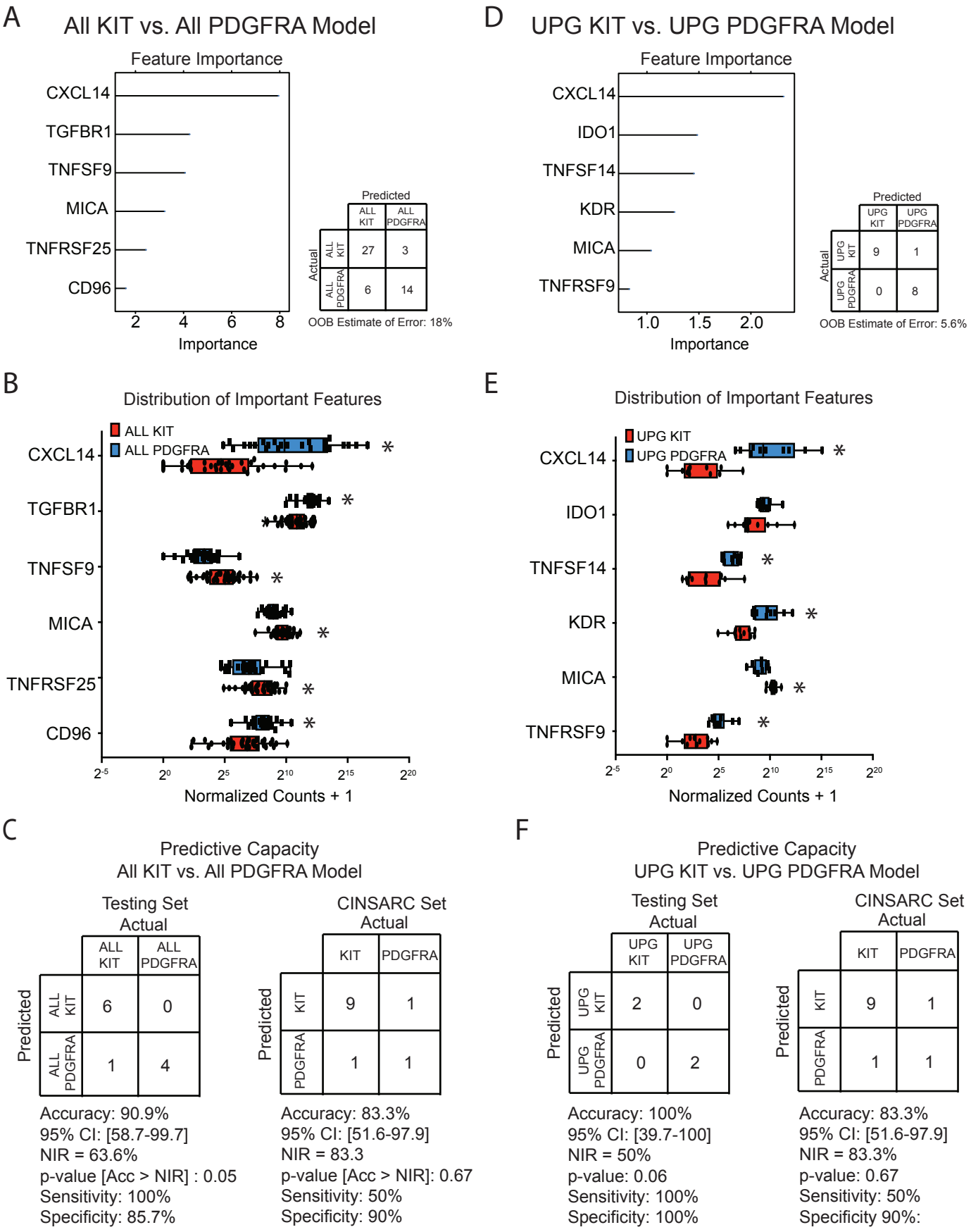


Figure 6. Machine learning identifies an immune signature predictive of *KIT* and *PDGFRA*-mutant GIST.
(**A**) Random forest modeling with 5-fold cross-validation of *KIT* and *PDGFRA*-mutant GIST specimens (training set created by partitioning 80% of *KIT* and *PDGFRA* samples from Supplemental Table 3, n=50). Confusion matrix (bottom right) indicates assessment of model fit to training set. (**B**) Distribution of top 6 features identified by random forest modeling. *adjusted p-value <0.05 from DSeq2. (**C**) Predictive capacity of model on remaining *KIT* and *PDGFRA*-mutant GIST testing set (n=11) and the CINSARC cohort (n=12). Accuracy (Acc), sensitivity, specificity, and p-value [Acc >No Information Rate (NIR)] of model are shown, calculated by caret package for R. Bars, mean + SEM. (**D**) Random forest modeling with 5-fold cross-validation of UPG *KIT* and UPG *PDGFRA*-mutant GIST specimens (training set created by partitioning 80% of UPG *KIT* and UPG *PDGFRA* samples from Supplemental Table 4, n=18). Confusion matrix (bottom right) indicates assessment of model fit to training set. (**E**) Distribution of top 6 features identified by random forest modeling. *adjusted p-value <0.05 from DSeq2. (**F**) Predictive capacity of model on remaining UPG *KIT* and UPG *PDGFRA*-mutant GIST testing set (n=4) and the CINSARC cohort (n=12) Accuracy (Acc), sensitivity, specificity, and p-value [Acc >No Information Rate (NIR)] of model are shown, calculated by caret package for R. Bars, mean + SEM.

Figure 7

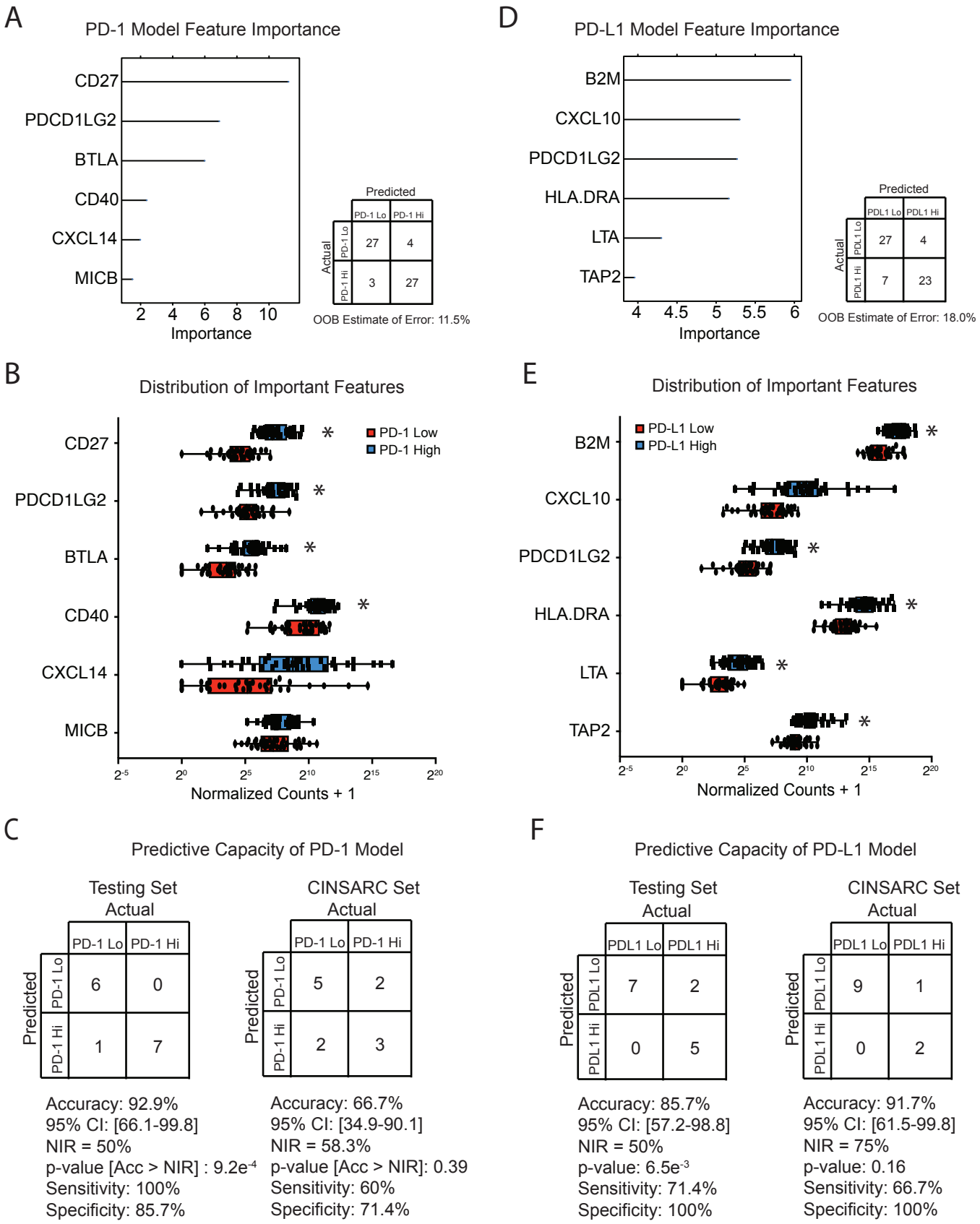


Figure 7. Machine learning identifies an immune signature predictive of PD-1 and PD-L1 expression in GIST.
(**A**) Random forest modeling of PD-1 with 5-fold cross-validation of GIST specimens (training set created by partitioning 80% of all GIST samples from Supplemental Table 1, n=61). Confusion matrix (bottom right) indicates assessment of model fit to training set. (**B**) Distribution of top 6 features identified by random forest modeling. *adjusted q-value<0.1. (**C**) Predictive capacity of model on remaining 14 GISTs (testing set) and external CINSARC GIST cohort (n=12). Accuracy (Acc), sensitivity, specificity, and p-value [Acc >No Information Rate (NIR)] of model are shown, calculated by caret package for R. (**D**) Random forest modeling of PD-L1 with 5 k-folds cross-validation of GIST specimens (training set created by partitioning 80% of all GIST samples from Supplemental Table 1, n=61). Confusion matrix (bottom right) indicates assessment of modeling fit to training set. (**E**) Distribution of top 6 features identified by random forest modeling. *adjusted q-value<0.1. (**F**) Predictive capacity of model on remaining 14 GISTs (testing set) and external CINSARC GIST cohort (n=12). Accuracy, sensitivity, specificity, and p-value [Acc >No Information Rate] of model are shown, as calculated by caret package for R. Bars, mean + SEM.