Germline variants that influence the tumor immune microenvironment also drive response to immunotherapy

Authors: Meghana Pagadala[1], Victoria H. Wu[4], Eva Pérez-Guijarro[12], Hyo Kim[15], Andrea Castro[2], James Talwar[2], Cristian Gonzalez-Colin[3], Steven Cao[5], Benjamin J. Schmiedel[3], Rany M. Salem[5], Gerald P. Morris[6], Olivier Harismendy[7,10], Sandip Pravin Patel[8], Jill P. Mesirov[9,10], Maurizio Zanetti[10,11], Chi-Ping Day[12], Chun Chieh Fan[13], Wesley K. Thompson[14], Glenn Merlino[12], J. Silvio Gutkind[4], Pandurangan Vijayanand[3], Hannah Carter[9,10,*]

*Correspondence to: hkcarter@health.ucsd.edu

Affiliations:

[1]Biomedical Science Program, University of California San Diego, La Jolla, CA 92093, USA

[2]Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA, 92093, USA

[3]La Jolla Institute for Immunology, La Jolla, CA, 92037, USA

[4]Department of Pharmacology, UCSD Moores Cancer Center, La Jolla, CA, 92093, USA

[5]Division of Epidemiology, Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, 92093, USA

[6]Department of Pathology, University of California San Diego, La Jolla, CA 92093, USA

[7]Division of Biomedical Informatics, Department of Medicine, University of California San Diego School of Medicine, La Jolla, CA, 92093, USA

[8]Center for Personalized Cancer Therapy, Division of Hematology and Oncology, UC San Diego Moores Cancer Center, San Diego, CA, 92037, USA

[9]Division of Medical Genetics, Department of Medicine, University of California San Diego, La Jolla, CA, 92093, USA

[10]Moores Cancer Center, University of California San Diego, La Jolla, CA, 92093, USA

[11]The Laboratory of Immunology and Department of Medicine, University of California San Diego, La Jolla,

CA 92093, USA

[12]National Cancer Institute, National Institutes of Health (NIH), Bethesda, MD, 20892, USA

[13]Department of Cognitive Science, University of California San Diego, La Jolla, CA, 92093, USA

[14]Division of Biostatistics, Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, 92093, USA

[15] Undergraduate Bioengineering Program, Jacobs School of Engineering, University of California San Diego, La Jolla, CA, 92093, USA

Abstract

With the continued promise of immunotherapy as an avenue for treating cancer, understanding how host genetics contributes to the tumor immune microenvironment (TIME) is essential to tailoring cancer risk screening and treatment strategies. Using genotypes from over 8,000 European individuals in The Cancer Genome Atlas (TCGA) and 137 heritable tumor immune phenotype components (IP components), we identified and investigated 482 TIME associations and 475 unique TIME-associated variants. Many TIME-associated variants influence gene activities in specific immune cell subsets, such as macrophages and dendritic cells, and interact to promote more extreme TIME phenotypes. TIME-associated variants were predictive of immunotherapy response in human cohorts treated with immune-checkpoint blockade (ICB) in 3 cancer types, causally implicating specific immune-related genes that modulate myeloid cells of the TIME. Moreover, we validated the function of these genes in driving tumor response to ICB in preclinical studies. Through an integrative approach, we link host genetics to TIME characteristics, informing novel biomarkers for cancer risk and target identification in immunotherapy.

Keywords: germline variation; tumor immune microenvironment; GWAS; cell-type effects; TCGA; immune checkpoint blockade; immunotherapy

Significance

A systematic screen for common germline variants associated with the tumor immune microenvironment across > 8000 tumors reveals novel immune SNPs that predict disease risk, prognosis and potential to respond to immune checkpoint therapy.

Introduction

Cancer is a disease characterized by heterogeneous somatic and germline mutations that promote abnormal cellular growth, evasion from the immune system, dysregulation of cellular energetics, and inflammation[1–4]. Immunotherapy has emerged as a promising treatment; however, response rates are low and the determinants of response remain elusive[5,6]. The potential of galvanizing the immune system is still unmet due to an incomplete understanding of the complex tumor immune microenvironment. In particular, knowledge of germline factors and other intrinsic factors that interact with characteristics of tumors to render them sensitive to host-immunity or immunotherapy is lacking.

Germline variation is responsible for a considerable proportion of variation in immune traits in healthy populations[7,8]. In the context of tumors, germline variants have the potential to influence immune infiltration, antigen presentation and immunotherapy responses[9,10]. Autoimmune germline variants modify ICB response and variants underlying leukocyte genes predict tumor recurrence in breast cancer patients[11,12]. The common single nucleotide polymorphism (SNP) rs351855 in *FGFR4* was found to suppress cytotoxic CD8+ T cell infiltration and promote higher immunosuppressive regulatory T cell levels via increased *STAT3* signaling in murine models of breast and lung cancer[13]. Normal genetic variation underlying major histocompatibility complex molecules, MHC-I and MHC-II, dictate which mutations in an individual's tumor can elicit immune responses, and play a role in antigen-driven host anti-tumor immune activity that influences tumor genome evolution through immune selection[14,15]. Polymorphic variation in these regions has also been linked to treatment outcomes[16–18]. Recent literature highlights polymorphisms in other immune-related genes such as *CTLA4*[19], *IRF5*[20] and *CCR5*[21,22] that also affect treatment outcomes.

Efforts to identify germline variation influencing anti-tumor immune responses have pointed to effects on

immune infiltration levels and immune pathways, such as *TGF-β* and *IFN-γ*[9,10,23]. eGenes (genes with significant cis-eQTLs) in the TCGA are enriched for immune-related genes and associated with immune cell abundance within the tumor immune microenvironment (TIME)[24]. These studies provide evidence that variants may act through specific effects on immune cells. eQTL profiling of 15 sorted immune cell subsets from healthy individuals found that the effects of many eQTLs were specific to immune cell subsets[25]. Understanding mechanisms and cell-type effects of TIME host genetic interactions could identify needed prognostic biomarkers for ICB response, implicating targetable cell types and molecules to boost ICB response rates.

Here, we present a comprehensive pan-cancer approach to study germline determinants underlying the tumor immune microenvironment. We used genotype data from the Cancer Genome Atlas (TCGA) to identify and investigate 482 TIME associations and 475 TIME-associated variants. These variants were predictive of immunotherapy response in melanoma, non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC) human cohorts and implicated 15 ICB-response genes that specifically alter macrophage polarization. Through preclinical murine studies, we discovered genes such as *LILRB2, LILRB4, CTSS* and *LAIR1* as potential therapeutic targets to modify ICB response. The study design is summarized in **Figure 1.**

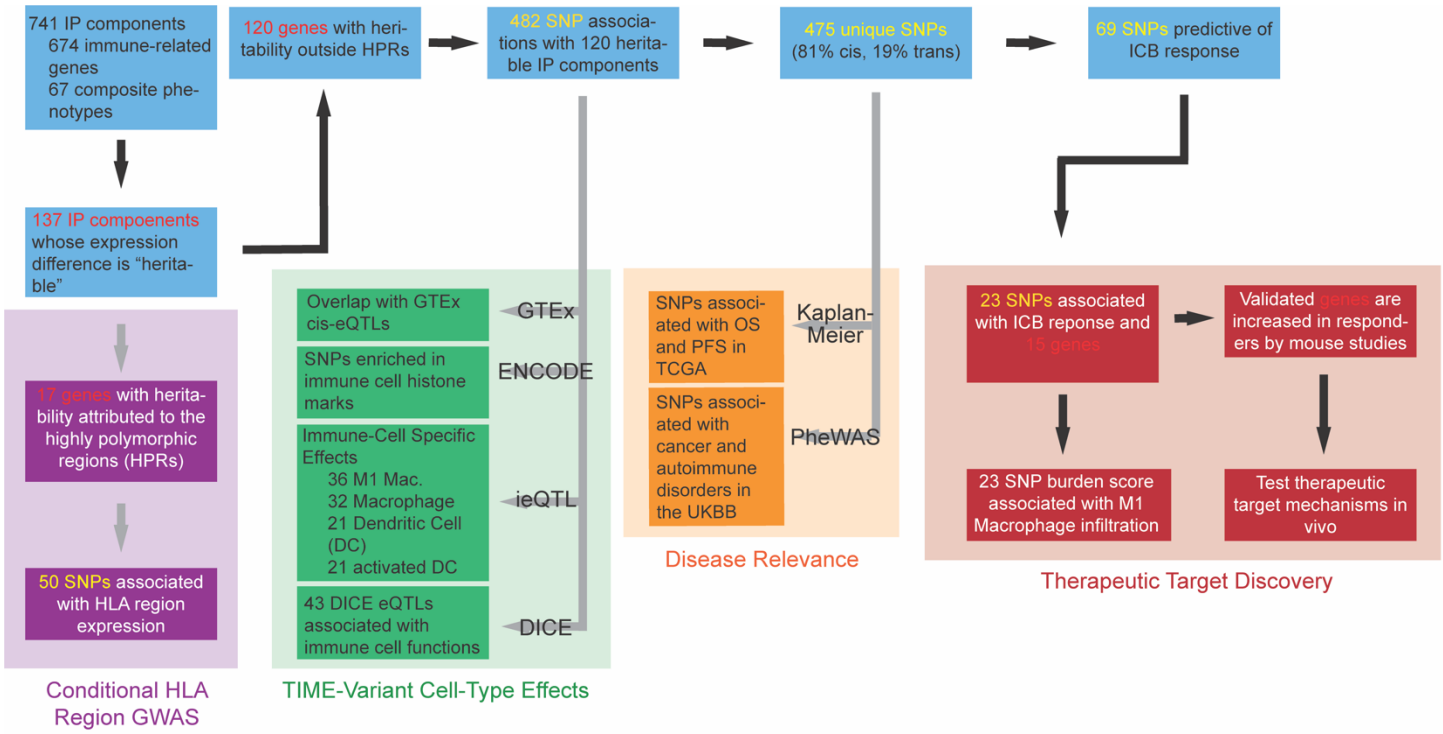

Figure 1: Overview of Analyses. Flowchart of TCGA genotype and phenotype analysis for TIME germline variant discovery and identification of ICB response genes.

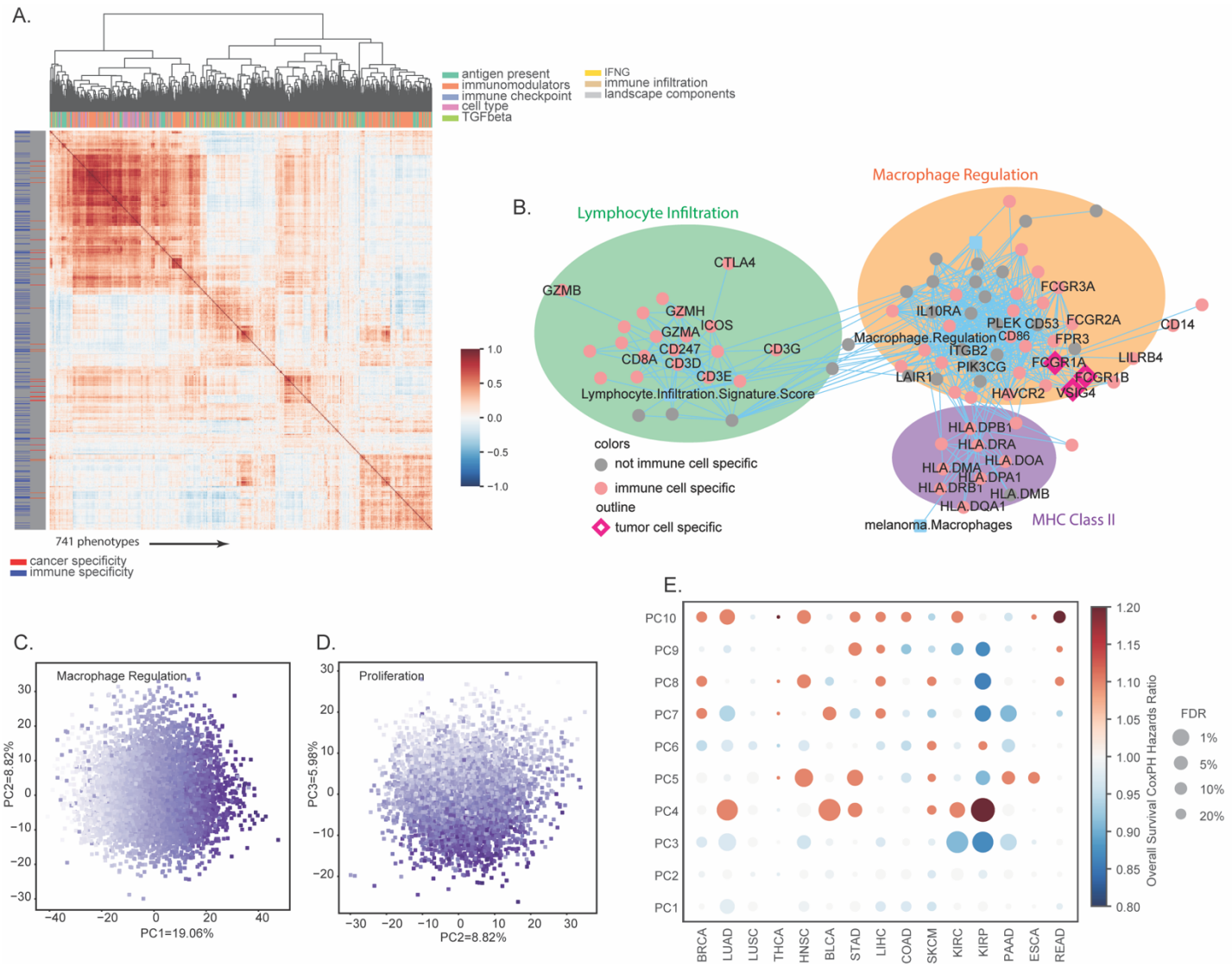

Figure 2: Characterization of Tumor Immune Microenvironment (TIME) IP components. (A) Clustermap of Pearson correlation analysis of 741 curated immune phenotypes. **(B)** Network plot of most correlated phenotypes; in coral=immune cell specific, outlined by pink diamonds=cancer specific. **(C)** Scatterplot of PC1 and PC2 for TCGA cohort colored by inverse rank normalized Thorsson et al. Macrophage Regulation phenotype score. (**D)** Scatterplot of PC2 and PC3 colored by inverse rank normalized Thorsson et al. Proliferation phenotype score. **(E)** Grid plot of hazards ratios and false discovery rate (FDR) of CoxPH analysis of 10 PCs with overall survival by TCGA cancer type.

Characterization of Immune Phenotype Components (IP components) in the Tumor Immune Microenvironment

To characterize the TIME, we conducted analysis with 724 immune-related genes and 17 composite phenotypes describing immune states and infiltration across 30 TCGA tumor types (immune phenotype "IP" components) **(Table S1)**. Immune-related genes were derived from 5 signatures involving activation of macrophages/monocytes[26], overall lymphocyte infiltration[27], *TGF-β* response[28], *IFN-γ* response[29], and wound healing[30] . To assess relatedness, we calculated pairwise Pearson correlations across the 741 inverse rank normalized IP components on TCGA dataset and performed hierarchical clustering (**Figure 2A**). Several IP components were highly correlated or anti-correlated with each other, including those curated from distinct literature or computational sources (**Figure S1B**). Among the modules of correlated IP components in tumors, the largest one was related to macrophage and lymphocyte regulation (**Figure 2B)**. Other correlated modules included *TGF-β* and *IFN-γ* pathway activities; anti-correlated modules included wound healing and *CREBBP*/*EP300* activity.

As IP components included expression of genes more specifically expressed in immune cells than tumors, we expected a number of genes to be differentially expressed in distinct types of immune cells. Indeed, multiple genes showed immune cell type-associated differences in expression; the largest proportions were differentially expressed in monocytes and NK cells (**Figure S1C).** This supports the potential of our IP components to identify germline modifiers that are putatively mediated by a broad array of immune cell types. As tumor-immune interactions can further be influenced by sex and age[31], we evaluated expression levels of collected immune-related genes with respect to these variables; the majority of phenotypes had higher expression values in females versus males and sex was included as a covariate for all subsequent analyses (**Figure S1D)**.

Given the high degree of correlation among phenotypes, we evaluated their potential to capture independent TIME characteristics that associate with survival outcomes in cancer. Principal Component Analysis (PCA) was conducted on the expression of IP components in TCGA tumors to obtain principal components (PCs) of orthogonal immune variance; the first 10 principal components (PCs) each contributed to at least 1% of

variance across TCGA samples (**Figure S1E)**. We observed a significant association of all 10 PCs with overall and progression free survival in different cancer types (**Figure S1F and S1G, respectively)** (Cox Proportional-Hazards analysis; Methods). The Thorsson et al Macrophage Regulation score[32] was a major contributor to the first principal component (PC1), which also explained ~19% of variance across tumors (**Figure 2C).** Thorsson et al Proliferation scores contributed greatly to PC3, which explained ~6% of variation **(Figure 2D)**. PC4 had strong effects on overall survival in renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), and urothelial bladder carcinoma (BLCA); this principal component was driven by *TGF-β* response. While all PCs were generally associated with survival outcomes across most tumors, hazards ratios for some components seemed to have large effects in a subset of tumor types, indicating that different aspects of the TIME may be particularly relevant to outcomes in those diseases (**Figure 2E**). As many IP components can interact with age and sex, we conducted Ordinary Least Squares (OLS) regression with PCs and age and sex. PC2, driven by *EP300*/proteasome expression, was significantly associated with age of diagnosis **(Table S2)**. Most principal components were not significantly associated with sex, although PC7 and PC4 had suggestive associations. Taken together, these results showed that the expression of IP components play a role in cancer prognosis and orthogonal variance of these components capture different features of the TIME.

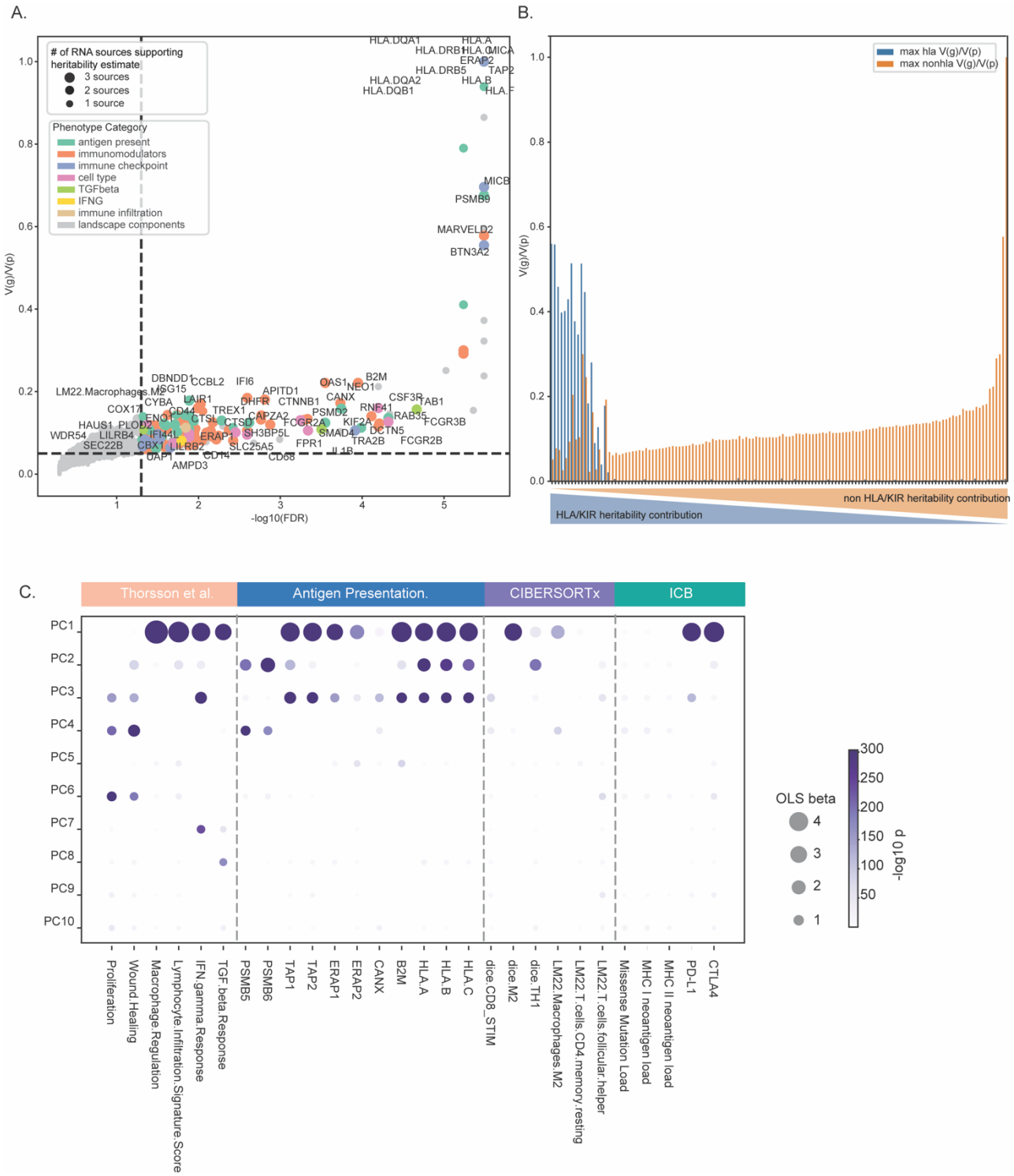

Figure 3: Heritability Analysis of IP components. (A) Scatterplot of -log10 FDR and V(g)/V(p) for 741 immune phenotypes in the TCGA cohort of European ancestry (Cutoffs: Vg/Vp > 0.05; FDR < 0.05). **(B)** Barplot of V(g)/V(p) values for 137 heritable immune phenotypes using a two-state Genome-wide Complex Trait Analysis (GCTA) model: highly polymorphic regions (HPRs) (blue), rest of the genome (orange). **(C)** Grid plot of beta values from Ordinary Least Squares (OLS) regression of principal components with TIME characteristics.

Characterization of Heritable IP Components in Cancer

To focus our analyses on IP components with strong genetic components, we performed heritability analysis **(Step 1, Figure 1)**. We identified 137 IP component genes whose expression variance is "heritable" (>5% of expression variance attributable to genetic variance; i.e. Vg/Vp; FDR < 0.05; **Figure 3A**). Of these 137 heritable IP component genes, 70 were immunomodulators, 41 were antigen presentation genes, 11 were immune checkpoint genes, 10 were immune cell-type specific genes, 3 were genes in the *TGF-β* pathway, 1 was a gene in the *IFN-γ* pathway, and 1 described immune cell infiltration.

Genes in the HLA region were highly heritable, with genetics explaining a large proportion of variance in gene expression. To assess the well-known contribution of highly polymorphic regions (HPRs), such as HLA and KIR regions, to all 137 heritable IP component genes, we compared variance explained by 1) the HPRs and 2) the rest of the genome without HPRs. Among the 137 heritable IP component genes, 17 had a larger fraction of heritability attributable to the HPRs whereas heritability of 120 of the 137 heritable IP genes originated outside of these regions (**Figure 3B**). The 17 heritable IP genes for which HPRs had higher contribution were all on chromosome 6 and within the HLA class I/II genomic regions. These genes included HLA class I and II genes in addition to genes critical for antigen presentation, such as *MICA, MICB, PSMB9, BTN3A2* and *TAP2.* We employed a separate pipeline to analyze HLA region variants as the region is highly polymorphic **(HLA Region Conditional GWAS, Figure 1)**.

We repeated clustering and PCA analysis with the 137 heritable IP component genes and observed modules of IP component genes **(Figure S2A)**; the largest clustering module is a subnetwork of the large macrophage regulation network from Figure 2B (**Figure S2B).** Comparing PCs obtained for the 741 IP components and 137 heritable IP components, the first two PCs were highly correlated, corresponding to macrophage regulation and *EP300*/proteasome gene expression[33] respectively (**Figure S2C).** Cox Proportional-Hazards analysis conducted with PCs based on 137 heritable IP component genes showed similar results (**Figure S2D,E);** stronger effects of PC1 on overall survival were observed for melanoma (SKCM) and stronger effects of PC2 on progression-free survival in urothelial bladder carcinoma (BLCA). PC3 and PC4 were highly correlated

with *IFN-γ* response and Thorsson et al Wound Healing, respectively**.** *PD-L1* expression was highly correlated with PC1 (beta=3.7, p=2.2e-308) and PC3 (beta=0.66, p=2.6e-115) (**Figure 3C).** These results confirm that the heritable IP component genes we identified are related to immune pathways, such as macrophage regulation, TGF-beta expression and IFN-gamma expression, which are known to modify the TIME, with the predominant TIME characteristic being related to macrophage regulation.

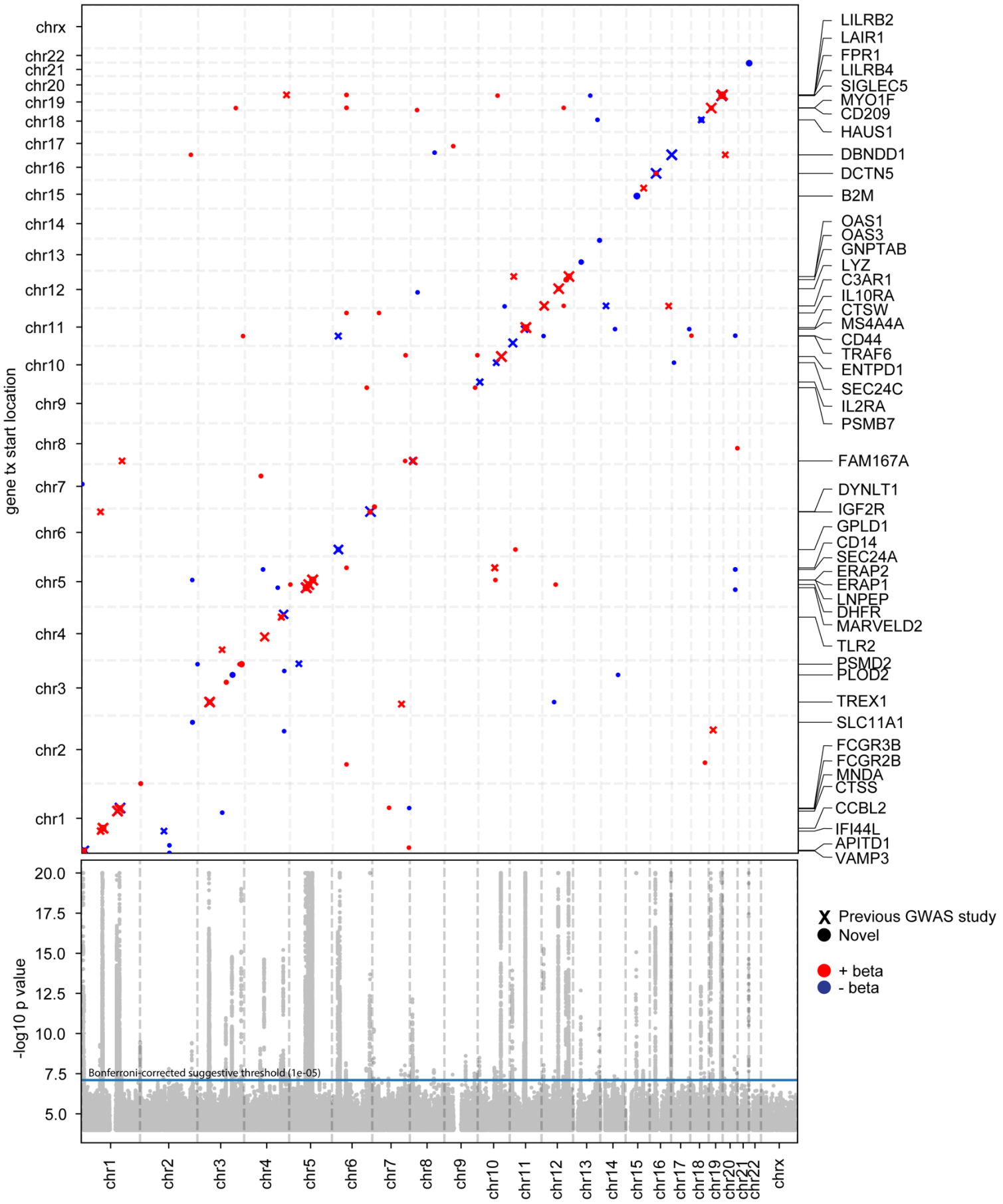
gene tx start location
chrx
chr22
chr21
chr20
chr19
chr18
chr17
chr16
chr15
chr14
chr13
chr12
chr11
chr10
chr9
chr8
chr7
chr6
chr5
chr4
chr3
chr2
chr1
LILRB2
LAIR1
FPR1
LILRB4
SIGLEC5
MYO1F
CD209
HAUS1
DBNDD1
DCTN5
B2M
OAS1
OAS3
GNPTAB
LYZ
C3AR1
IL10RA
CTSW
MS4A4A
CD44
TRAF6
ENTPD1
SEC24C
IL2RA
PSMB7
FAM167A
DYNLT1
IGF2R
GPLD1
CD14
SEC24A
ERAP2
ERAP1
LNPEP
DHFR
MARVELD2
TLR2
PSMD2
PLOD2
TREX1
SLC11A1
FCGR3B
FCGR2B
MNDA
CTSS
CCBL2
IFI44L
APITD1
VAMP3
-log10 p value
20.0
17.5
15.0
12.5
10.0
7.5
5.0
Bonferroni-corrected suggestive threshold (1e-05)
Previous GWAS study
Novel
+ beta
- beta

Figure 4: Identification of Novel TIME-associated Variants. (A). eQTL grid of SNP position and gene association position. Phenotypes without gene positions are not shown; "X" =loci identified in previous GWAS studies, size=-log10 p-value. (**B).** Manhattan plot of GWAS analysis conducted on 120 heritable IP component genes with low HPR contribution.

Identification of Novel TIME-associated Variants

As the next step (**Step 2, Figure 1),** we performed genome-wide association analyses for the 120 heritable IP component genes where HPR regions did not contribute significantly to heritability in the TCGA **(Figure S3A)**. Only common germline variants with minor allele frequency > 1% were considered and imputation quality was accounted for to ensure high accuracy (**Figure S3B)**. Using the linkage and distance-based clumping[34], we identified 482 associations across the 120 heritable IP component genes at a threshold of 8.3×10^{-7} (Bonferroni-corrected suggestive threshold of 1×10^{-5}) with 442 of the 482 (91.7%) passing a significance threshold of 5×10^{-8} and 296 of the 482 (61.4%) passing a Bonferroni-corrected genome-wide significant threshold of 4.2×10^{-10} (**Figure 4**). Of the 482 associations, 475 unique TIME-associated variants were identified. The majority (81.1%) of associations were *cis* (+/-1Mb from immune gene transcription start site[35]), while 18.9% of the associations were *trans*. *ERAP2* (69), *CCBL2* (35), *MARVELD2* (35), *DHFR* (27) *and LAIR1* (24) had the most germline associations **(Figure S3C)**. Beta values were robust to differences in bioinformatic pipelines used to process RNA sequencing data (**Figure S3D**).

We found that 144 of the 475 unique TIME-associated variants have been implicated in previous GWAS studies[36]. These 144 variants were associated with 48 of the 120 heritable IP component genes tested. Of these 144 variants, most were associated with immune traits or diseases, such as white blood cell count, neutrophil count, Crohn's disease, ulcerative colitis, and chronic inflammatory conditions (**Figure S3E).** Additionally, a number of variants were associated with risk of prostate, breast and colorectal cancer.

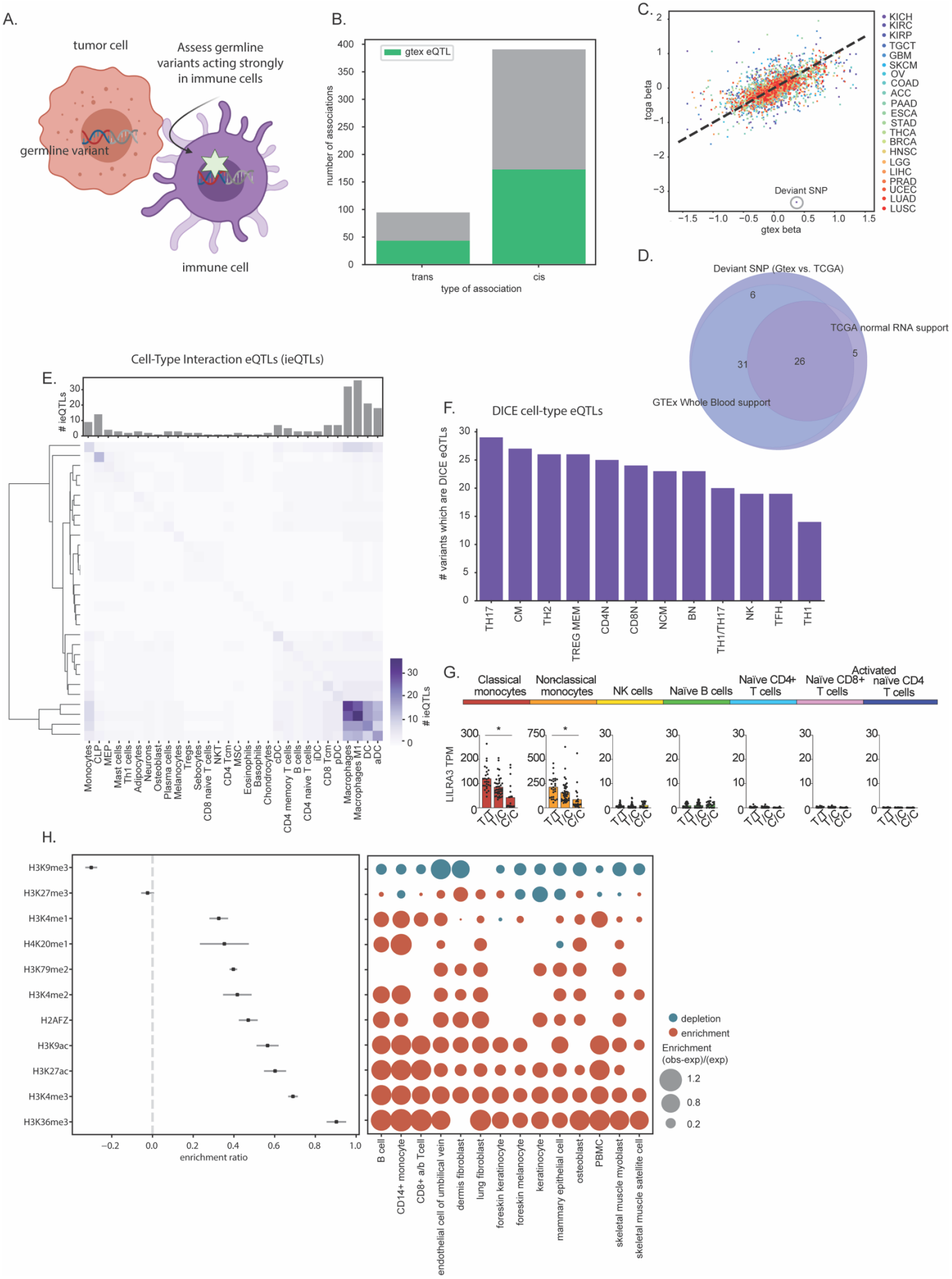
A.
tumor cell
Assess germline variants acting strongly in immune cells
germline variant
immune cell
B.
gtex eQTL
number of associations
trans
cis
type of association
C.
tcga beta
gtex beta
Deviant SNP
KICH
KIRC
KIRP
TGCT
GBM
SKCM
OV
COAD
ACC
PAAD
ESCA
STAD
THCA
BRCA
HNSC
LGG
LIHC
PRAD
UCEC
LUAD
LUSC
D.
Deviant SNP (Gtex vs. TCGA)
TCGA normal RNA support
GTEx Whole Blood support
6
31
26
5
E.
Cell-Type Interaction eQTLs (ieQTLs)
ieQTLs
Monocytes
CLP
MEP
Mast cells
Th1 cells
Adipocytes
Neurons
Osteoblast
Plasma cells
Melanocytes
Tregs
Sebocytes
CD8 naive T cells
NKT
CD4 Tcm
MSC
Eosinophils
Basophils
Chondrocytes
cDC
CD4 memory T cells
B cells
CD4 naive T cells
iDC
CD8 Tcm
pDC
Macrophages
Macrophages M1
DC
aDC
F.
DICE cell-type eQTLs
variants which are DICE eQTLs
TH17
CM
TH2
TREG MEM
CD4N
CD8N
NCM
BN
TH1/TH17
NK
TFH
TH1
G.
Classical monocytes
Nonclassical monocytes
NK cells
Naïve B cells
Naïve CD4+ T cells
Naïve CD8+ T cells
Activated naïve CD4 T cells
LILRA3 TPM
T/T
T/C
C/C
H.
H3K9me3
H3K27me3
H3K4me1
H4K20me1
H3K79me2
H3K4me2
H2AFZ
H3K9ac
H3K27ac
H3K4me3
H3K36me3
enrichment ratio
B cell
CD14+ monocyte
CD8+ a/b Tcell
endothelial cell of umbilical vein
dermis fibroblast
lung fibroblast
foreskin keratinocyte
foreskin melanocyte
keratinocyte
mammary epithelial cell
osteoblast
PBMC
skeletal muscle myoblast
skeletal muscle satellite cell
depletion
enrichment
Enrichment (obs-exp)/(exp)
1.2
0.8
0.2

Figure 5: Analysis of cell-type effects of TIME-associated variants reveals macrophage and dendritic cells as important modifiers of the TIME. (A) Conceptual model to show that variants can have cell-type effects, acting more strongly in immune or tumor cells specifically. **(B)** Barplot of number of *cis* and *trans* TIME associations which were significant GTEx eQTLs. **(C)** Scatter plot of *cis* association beta values in GTEx versus TCGA. **(D)** Venn diagram of SNPs with "deviant" behavior in GTEx versus TCGA variant analysis and whether behavior is supported by TCGA normal germline behavior or GTEx Whole Blood germline behavior. **(E)** Barplot of cell-type interaction eQTLs (ieQTLs) for each cell-type deconvoluted by xCell[37] with Clustermap of overlap between ieQTLs. **(F)** Barplot of number of TIME-associated variants (475) which are DICE immune cell type eQTLs. **(G)** rs410852 effects on *LILRA3* expression across 7 immune cell types profiled in DICE. **(H)** Mean enrichment ratio of immune microenvironment variants in histone marks with corresponding enrichment ratios in specific cell types

Analysis of Cell-Type Effects of TIME-associated Variants Reveal Macrophage and Dendritic cells as Important Modifiers of the TIME.

As variants can act more strongly in specific cells **(Figure 5A),** we sought to examine the role of TIME-associated variants with respect to tissue type using GTEx *cis*-eQTL data. Of 482 associations, 44.0% of *cis* associations and 45.1% of *trans* associations were significant GTEx eQTLs in at least 1 tissue type (**Figure 5B)**. To determine how much *trans* associations contribute to IP component expression variance, we conducted heritability analysis using only *cis* associations and then all associations; *trans* associations modestly increased the variance explained in general but appeared to be a major contributor in certain cases (**Figure S4A).** Strengths of *cis*-eQTL SNP associations varied across tissue types in GTEx data, and the strongest effects most often appeared in whole blood (**Figure S4B)**, suggesting that many of our identified TIME-associated variants could have strong effects within immune cells. To determine if *cis*-eQTLs were affecting gene expression specifically in either tumor or immune cells, we compared *cis* associations significant in Whole Blood eQTLs versus those in other tissue types (**Figure S4C).** Variants associated with *B2M*, *LILRB2*, *LILRB4, LNPEP*, and *OAS1* were

exclusively Whole Blood eQTLs, whereas variants associated with *COX17, DBNDD1*, and *GPLD1* were exclusively Breast Mammary Tissue eQTLs. This suggests that variants may act exclusively on the immune or tumor axis to modify cancer progression.

We then compared beta coefficients of TIME-associated variants between TCGA and GTEx to determine whether variants exhibit different behavior in a normal tissue context or tumor tissue context. Generally, beta coefficients from tumor and normal tissues correlated well (**Figure 5C)**; however, we noted a subset of 68 SNPs for which effects sizes were significantly different between the two datasets (Methods). To determine whether these deviations in behavior are due to differences in immune infiltration, we compared beta coefficients of TIME-associated variants to coefficients in TCGA tissue-matched normal RNA and from GTEx Whole Blood *cis*-eQTL information. 57 of the 68 deviant eQTLs had behavior similar to GTEx Whole Blood cis-eQTLs, suggesting that theses variants may be acting more specifically in immune cell types (**Figure S4D**). 26 of these 57 deviant eQTLs exhibited consistent behavior in TCGA tissue-matched normal samples and GTEx datasets, suggesting that different variant behavior in tumors may be due to biological rather than technical differences between datasets **(Figure 5D)**.

To obtain more information about variant cell-type specific effects **(TIME-Variant Cell-Type Effects, Figure 1)**, we used the GTEx consortium cell type interaction eQTL (ieQTL) discovery pipeline (Methods). Of the 482 associations, 36 were M1 macrophage ieQTLs, 32 were macrophage ieQTLs, 21 were dendritic cells ieQTLs, and 18 were activated dendritic cell ieQTLs **(Figure 5E)**. There were also ieQTLs identified for non-immune cell types, such as adipocytes and melanocytes. To further assess cell-specificity of our immune microenvironment associations, we analyzed our variants against the DICE database of immune cell-type specific eQTLs[38]. 43 of 475 unique SNPs were eQTLs in 1 or more of the 12 immune cell types profiled **(Figure 5F)**, and 25.6% (n=11) were specific to a single cell type **(Figure S4E).** rs410852 is an eQTL for *LILRA3* in classical and non-classical monocytes, but not other immune cell types **(Figure 5G)**. TIME phenotypes that were eQTLs in the DICE database included *TREX1, ERAP1, CTSS,* and genes tightly correlated to Macrophage Regulation such as *LILRB4, PLEK, LILRB2* (**Figure S4F)**. These findings reveal a subset of variants that are specifically modifying

immune rather than tumor behavior.

Furthermore, several TIME-associated variants were within non-coding genomic regions and could be located in regulatory elements. We conducted GREGOR[41] enrichment analysis with publicly available ENCODE ChIP-seq experiments. Variants associated with TIME IP components were generally enriched in regions with histone marks H3K36me3, H3K4me3, H3K27ac, H3K9ac, H2AFZ and depleted in H3K9me3 histone marks **(Figure 5H).** Such enrichment patterns were preferentially associated with specific immune cell types than others. The strongest enrichment in *cis*-regulatory elements was seen in PBMCs, CD14+ monocytes, B cells and CD8+ T cells. TIME-associated variants were also enriched at binding sites for important transcription factors (**Figure S5A).** These results further support cell-type specific behavior of our identified TIME-associated variants.

Our study identified a number of *trans* associations with IP components that could also potentially act in a cell-type specific manner. In GTEx, we found that *trans* associations are often *cis*-eQTLs for nearby genes, suggesting *trans* associations may act through mediator genes that influence downstream IP components. Most mediator genes are protein-coding with a sizable fraction related to regulatory RNAs, such as lincRNAs, and pseudogenes. *LINC00680,* which has previously been implicated in poor prognosis in glioblastoma[39,40] and non-small cell lung cancer[39], was identified as a mediator gene for 5 IP components. rs2693076 was significantly associated with increased *LINC00680* expression and downregulation of several immune genes related to macrophage regulation, such as *LAIR1, MYO1F,* and *PLEK* **(Figure S4G)** and served as an eQTL for several immune cell types, such as TH1 cells, naive CD8+ T cells, regulatory T cells and naive B cells (**Figure S4H)**.

Thus far, our analyses have focused on individual TIME-associated variants, but variants can interact to more strongly influence disease phenotypes. Our PCA analysis of 137 heritable IP component genes suggest that there exist multiple independent immune variance components across TCGA tumors. We investigated the relationship of the 475 TIME-associated variants by running associations with the first 10 principal components. PC1 correlated strongly with macrophage regulation **(Figure 2C)** and was associated with the largest number of variants (n=28). Two of our 28 PC1-correlated SNPs were DICE eQTLs for *LINC00680* and *IL10RA* in classical monocytes and non-classical monocytes, respectively. We quantified the similarity of the effects of the 28 PC1-

correlated SNPs across all 120 heritable IP component genes with explained heritability outside the HPRs. Clustering SNPs according to their effects suggested that a subgroup of these SNPs act very similarly **(Figure S5B)**. We constructed a network from the SNPs and their associated IP components, removing variants with beta values less than 0.25. This revealed a group of SNPs with similar behavior but broad effects on IP components. We also noted SNPs that have stronger effects on expression of only a few genes, such as *ERAP2, SIGLEC5,* and *LAIR1.* (**Figure S5C).** To evaluate synergistic effects, we calculated burden scores for TCGA individuals based on PC1-associated SNPs; individuals with a higher burden score had higher Thorsson et al. macrophage regulation scores (**Figure S5D).**

PC2 was driven by the *EP300*/proteasome axis. PC2-associated SNPs were mutually exclusive with the 28 PC1-associated SNPs. One of the 13 PC2-associated SNPs was a DICE immune cell eQTL for *ERAP1* in TH17 cells. Clustering of the Pearson correlation values showed that several of these SNPs are related to each other, but not in LD (R2>0.5) (**Figure S5E).** Network analysis of SNP-IP component associations filtered for beta values of 0.25 again showed SNPs that have strong effects on IP components and SNPs that have strong effect on a specific gene, such as *LILRB4*, *GLPD1* and *ERAP1* (**Figure S5F)**. Higher PC2-associated SNP burden scores were associated with more extreme values of *EP300* (**Figure S5G**). Genes strongly affected by variants seemed to be distinct between PC1 and PC2, indicating that underlying pathways and biological processes may differ. Taken together, these results demonstrate the complex epistatic interactions of TIME-associated variants and underlying pathways driving TIME components.

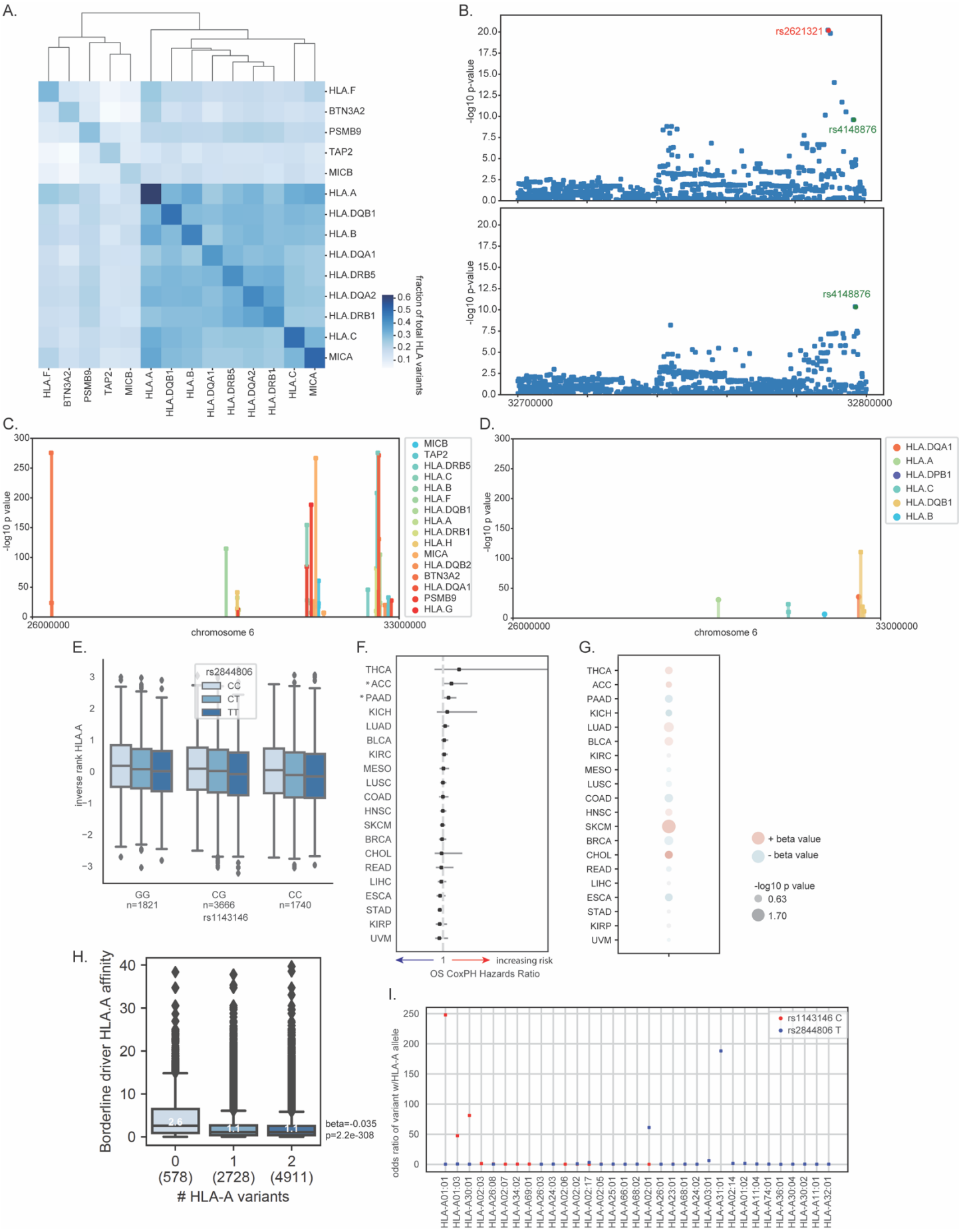
A.
HLA.F
BTN3A2
PSMB9
TAP2
MICB
HLA.A
HLA.DQB1
HLA.B
HLA.DQA1
HLA.DRB5
HLA.DQA2
HLA.DRB1
HLA.C
MICA
fraction of total HLA variants
B.
rs2621321
rs4148876
-log10 p-value
32700000
32800000
C.
-log10 p value
chromosome 6
26000000
33000000
MICB
TAP2
HLA.DRB5
HLA.C
HLA.B
HLA.F
HLA.DQB1
HLA.A
HLA.DRB1
HLA.H
MICA
HLA.DQB2
BTN3A2
HLA.DQA1
PSMB9
HLA.G
D.
HLA.DQA1
HLA.A
HLA.DPB1
HLA.C
HLA.DQB1
HLA.B
E.
rs2844806
CC
CT
TT
inverse rank HLA.A
GG
n=1821
CG
n=3666
CC
n=1740
rs1143146
F.
THCA
*ACC
*PAAD
KICH
LUAD
BLCA
KIRC
MESO
LUSC
COAD
HNSC
SKCM
BRCA
CHOL
READ
LIHC
ESCA
STAD
KIRP
UVM
increasing risk
OS CoxPH Hazards Ratio
G.
+ beta value
- beta value
-log10 p value
0.63
1.70
H.
Borderline driver HLA.A affinity
beta=-0.035
p=2.2e-308
0
(578)
1
(2728)
2
(4911)
HLA-A variants
I.
rs1143146 C
rs2844806 T
odds ratio of variant w/HLA-A allele
HLA-A01:01
HLA-A01:03
HLA-A30:01
HLA-A02:03
HLA-A26:08
HLA-A02:07
HLA-A34:02
HLA-A69:01
HLA-A26:03
HLA-A24:03
HLA-A02:06
HLA-A02:02
HLA-A02:17
HLA-A02:05
HLA-A25:01
HLA-A66:01
HLA-A68:02
HLA-A02:01
HLA-A26:01
HLA-A23:01
HLA-A68:01
HLA-A24:02
HLA-A03:01
HLA-A31:01
HLA-A02:14
HLA-A01:02
HLA-A11:04
HLA-A74:01
HLA-A36:01
HLA-A30:04
HLA-A30:02
HLA-A11:01
HLA-A32:01

Figure 6: Variant Discovery in the HLA region. (A) Confusion matrix of HLA region associations **(B)** Example conditional GWAS for *TAP2* expression (top) Iteration 1 identifies rs2621321 as significant variant (bottom) Iteration 2 identified rs4148876 as significant variant after conditioning on rs2621321. **(C)** Significant variants identified through conditional analysis of HLA expression quantified through traditional alignment. **(D)** Significant variants identified through conditional analysis of HLA expression quantified through allele-specific alignment. **(E)** *HLA-A* expression (Firebrowse) by rs1143146 and rs2844806 genotype. **(F)** Overall survival CoxPH Hazards Ratio by TCGA cancer type for patients with 0, 1, 2 of *HLA-A* variants. **(G)** OLS regression with age of diagnosis by TCGA cancer type for patients with 0, 1, 2 of *HLA-A* variants. **(H)** Predicted peptide binding affinity for borderline presentable cancer-driver mutations for patients with 0, 1, 2 of *HLA-A* variants. **(I)** Odds Ratio plot of *HLA-A* alleles significantly associated with rs1143146 and rs2844806 variants.

Variant Discovery in the HLA Region

Given the polymorphic nature of the HLA region, we employed a separate analysis pipeline for the 17 heritable IP component genes for which HPRs contributed most to phenotypic variance. Heritability estimates of HLA region gene expression were >50%, yet high linkage disequilibrium within this region makes analysis difficult. Indeed, clustering suggested many genes shared significant variants; MHC class II genes showed the highest overlaps **(Figure 6A)**. To filter out HLA region associations solely attributable to LD structure, conditional GWAS analysis was conducted until no SNPs with significance threshold < $1e^{-06}$ remained. Conditional analysis with all 3 RNA processing pipelines retained two SNPs associated with *TAP2* expression, rs4148876 and rs2621321 (**Figure 6B)**. Both have previously been implicated in GWAS studies; rs4148876 has been associated with type 2 diabetes risk, and rs2621321 with strep throat, psoriatic arthritis, platelet distribution width and thyroid peroxidase antibody positivity **(Figure S6A)**. In total, 40 independent SNPs were associated with HLA region expression **(Figure 6C)**; *HLA-DQB1* and *MICA* each had 4 significant LD-independent SNPs. Generally, LD-independent SNPs clustered by genomic regions with *HLA-A, HLA-B, HLA-C* variants in MHC Class I genomic region and *HLA-DQB1, HLA-DQA1, HLA-DPB1, HLA-DRB5* variants in MHC Class II genomic

region. Alignment to a general HLA gene reference may introduce error into expression level estimates due to the highly polymorphic nature of these genes. We therefore revisited SNP associations with gene expression estimates derived from allele-specific RNA alignments **(Methods)**. GWAS analysis was performed on combined expression of called TCGA patient HLA alleles. This approach identified 10 variants associated with HLA expression, of which 4 were associated with *HLA-DQB1* (**Figure 6D).**

As antigen presentation via the major histocompatibility complex plays a central role in host-tumor immune responses, we sought to investigate whether SNPs modulating *HLA-A* gene expression affected other TIME characteristics. The minor allele of rs1143146 was associated with decreased *HLA-A* allele expression and was also associated with decreased infiltration of CD8 and CD4 T cells **(Figure S6B).** The major allele of rs2844806 was associated with decreased *HLA-A* expression; patients with both *HLA-A* variants tended to have lower *HLA-A* expression (**Figure 6E)**. To evaluate whether patients with both *HLA-A* variants had differential survival outcomes, we conducted Cox Proportional Hazards analysis with overall survival; pancreatic adenocarcinoma (PAAD) and adrenocortical carcinoma (ACC) patients with both expression-reducing variants had worse survival outcome (**Figure 6F)**, in line with previous literature that increased expression of HLA corresponds with better survival and prognosis[42]. Ordinary Least Squares (OLS) regression with age of diagnosis showed the melanoma (SKCM) patients with both expression-reducing variants had later ages of diagnoses (**Figure 6G)**.

Our previous work suggests that effective immune surveillance is also dependent on binding affinities of somatically altered peptides for the MHC[14]. Patients with both *HLA-A* variants had higher affinities (scores closer to 0) for peptides overlapping driver mutations (Methods), suggesting that the HLA alleles associated with these SNPs also differ (**Figure 6H)**. Further analysis identified specific *HLA-A* alleles that were more commonly observed in association with the minor allele (**Figure 6I)**. Thus, rs1143146 and rs2844806 were associated not only with differences in *HLA-A* expression, but also in alleles and peptide binding specificities, potentially explaining variable association with clinical characteristics across tumor types.

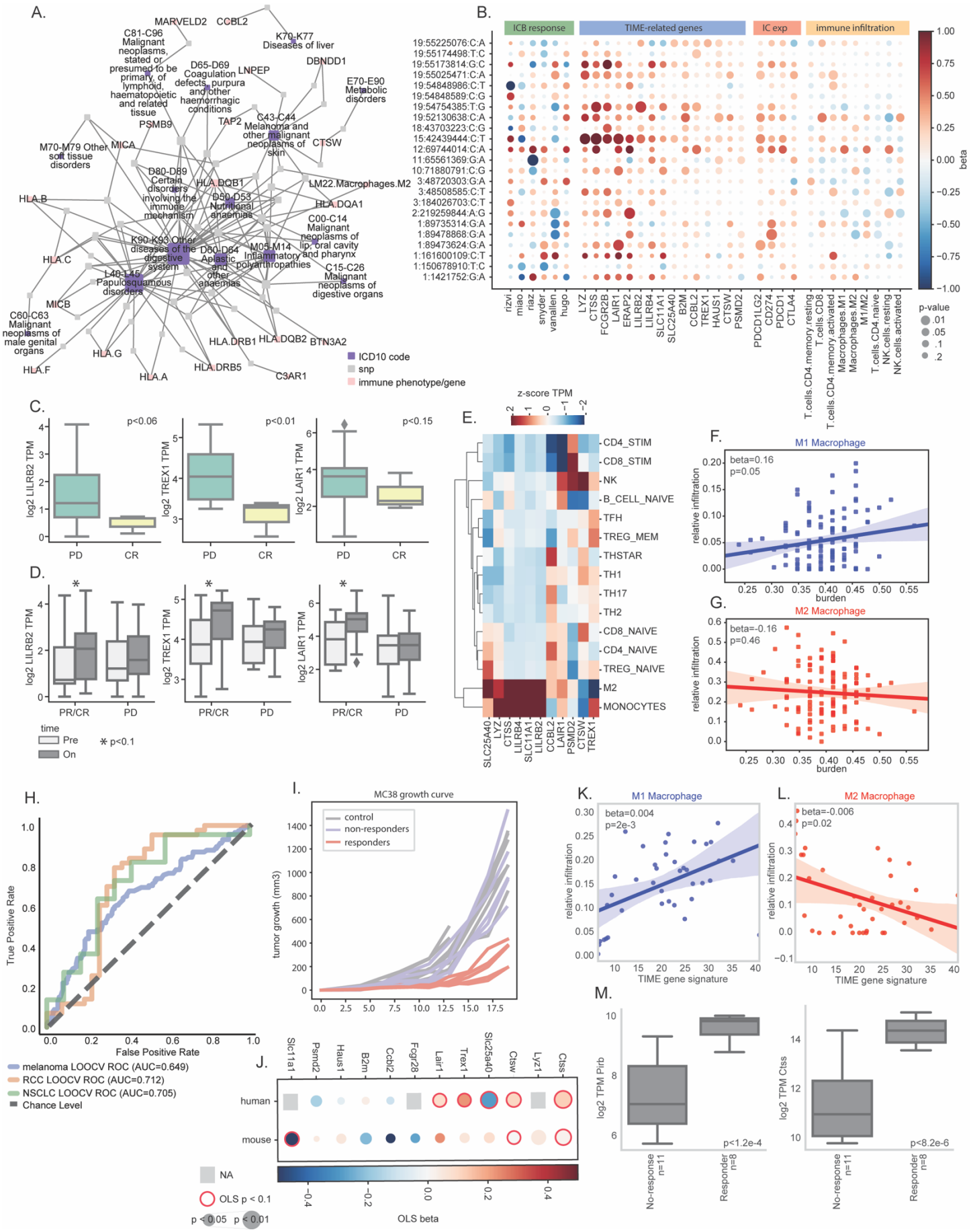
A.
MARVELD2
CCBL2
C81-C96 Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue
K70-K77 Diseases of liver
DBNDD1
D65-D69 Coagulation defects, purpura and other haemorrhagic conditions
LNPEP
E70-E90 Metabolic disorders
PSMB9
TAP2
C43-C44 Melanoma and other malignant neoplasms of skin
CTSW
M70-M79 Other soft tissue disorders
MICA
D80-D89 Certain disorders involving the immune mechanism
HLA.DQB1
LM22.Macrophages.M2
HLA.B
D50-D53 Nutritional anaemias
HLA.DQA1
C00-C14 Malignant neoplasms of lip, oral cavity and pharynx
K90-K93 Other diseases of the digestive system
D60-D64 Aplastic and other anaemias
M05-M14 Inflammatory polyarthropathies
HLA.C
L40-L45 Papulosquamous disorders
C15-C26 Malignant neoplasms of digestive organs
MICB
C60-C63 Malignant neoplasms of male genital organs
HLA.DRB1
HLA.DQB2
BTN3A2
HLA.G
HLA.F
HLA.A
HLA.DRB5
C3AR1
ICD10 code
snp
immune phenotype/gene
B.
ICB response
TIME-related genes
IC exp
immune infiltration
19:55225076:C:A
19:55174498:T:C
19:55173814:G:C
19:55025471:C:A
19:54848986:C:T
19:54848589:C:G
19:54754385:T:G
19:52130638:C:A
18:43703223:C:G
15:42439444:C:T
12:69744014:C:A
11:65561369:G:A
10:71880791:C:G
3:48720303:G:A
3:48508585:C:T
3:184026703:C:T
2:219259844:A:G
1:89735314:G:A
1:89478868:G:A
1:89473624:G:A
1:161600109:C:T
1:150678910:T:C
1:1421752:G:A
rizvi
miao
riaz
snyder
vanallen
hugo
LYZ
CTSS
FCGR2B
LAIR1
ERAP2
LILRB2
LILRB4
SLC11A1
SLC25A40
B2M
CCBL2
TREX1
HAUS1
CTSW
PSMD2
PDCD1LG2
CD274
PDCD1
CTLA4
T.cells.CD4.memory.resting
T.cells.CD8
T.cells.CD4.memory.activated
Macrophages.M1
Macrophages.M2
M1/M2
T.cells.CD4.naive
NK.cells.resting
NK.cells.activated
beta
p-value
C.
p<0.06
p<0.01
p<0.15
log2 LILRB2 TPM
log2 TREX1 TPM
log2 LAIR1 TPM
PD
CR
D.
PR/CR
PD
time
Pre
On
* p<0.1
E.
z-score TPM
CD4_STIM
CD8_STIM
NK
B_CELL_NAIVE
TFH
TREG_MEM
THSTAR
TH1
TH17
TH2
CD8_NAIVE
CD4_NAIVE
TREG_NAIVE
M2
MONOCYTES
F.
M1 Macrophage
beta=0.16
p=0.05
relative infiltration
burden
G.
M2 Macrophage
beta=-0.16
p=0.46
H.
True Positive Rate
False Positive Rate
melanoma LOOCV ROC (AUC=0.649)
RCC LOOCV ROC (AUC=0.712)
NSCLC LOOCV ROC (AUC=0.705)
Chance Level
I.
MC38 growth curve
control
non-responders
responders
tumor growth (mm3)
J.
Slc11a1
Psmd2
Haus1
B2m
Ccbl2
Fcgr28
Lair1
Trex1
Slc25a40
Ctsw
Lyz1
Ctss
human
mouse
NA
OLS p < 0.1
p < 0.05
p < 0.01
OLS beta
K.
M1 Macrophage
beta=0.004
p=2e-3
TIME gene signature
L.
M2 Macrophage
beta=-0.006
p=0.02
M.
log2 TPM Pirb
p<1.2e-4
No-response n=11
Responder n=8
log2 TPM Ctss
p<8.2e-6

Figure 7: TIME-associated Variants are Implicated in Disease Risk and Response to Immune Checkpoint Blockade (ICB). (A) Disease-Phenotype network of UK Biobank PheWAS results; purple=ICD10 diagnosis code, gray=variant, pink= IP component. **(B)** General Linear Model (GLM) regression beta coefficients/log(odds ratio) and -log10 p-value for 23 ICB-associated variants with ICB response, immune gene expression and macrophage infiltration; color=coefficient value, size=-log10 p-value. **(C)** Boxplot of pre-treatment *LILRB2, TREX1* and *LAIR1* expression in Riaz et al. ICB melanoma cohort. **(D)** Boxplot of pre-treatment and post-treatment *LILRB2, TREX1* and *LAIR2* expression in Riaz et al. ICB melanoma cohort responders vs nonresponders. **(E)** Clustermap of z-scored expression of ICB-response genes across 15 profiled immune cells in DICE database **(F)** Scatterplot of 23 ICB-associated variant burden score and M1 Macrophage infiltration. **(G)** Scatterplot of 23-ICB associated variant burden score and M2 Macrophage infiltration. **(H)** AUC curves for optimized polygenic risk models (Optuna framework, Leave One Out cross-validation) constructed from 69 variants for melanoma, renal cell carcinoma and non-small cell lung cancer cohorts. **(I)** Tumor growth curves of responders or non-responders to anti-PD-1 in MC38 tumor-bearing mice. **(J)** Grid plot of OLS results (beta, p-values) for on-treatment RNA for ICB responders and non-responders in humans (Riaz et al) and mice (MC38); color=OLS beta value, size=-log10 association p-value, red outline=significant p-values (< 0.1), square=expression levels not available. **(K)** Boxplot of post-treatment *Pirb* (human *LILRB* ortholog) and *Ctss* in M4 mice treated with anti-CTLA4. **(L)** Correlation of 12-gene mouse TIME signature with M1 macrophage infiltration. **(M)** Correlation of 12-gene mouse TIME signature with M2 macrophage infiltration. Two-sided Mann-Whitney *U*-tests were used for Panel C; paired T-tests were used for Panel D.

TIME-associated variants are implicated in disease risk and response to Immune Checkpoint Blockade (ICB)

To assess disease relevance of TIME-associated variants, we evaluated their effects on disease progression, risk and response to therapy. First, we ran Kaplan-Meier analysis on TCGA overall and progression-free survival for all 475 TIME-associated variants comparing minor allele and major allele homozygotes (**Disease**

Relevance, Figure 1). Immune microenvironment variants were associated with overall and progression-free survival in multiple tumor types **(Figure S7A)**. Eight variants were associated with overall survival in testicular germ cell tumors (TGCT); 5 were associated with progression-free survival in head and neck squamous cell carcinoma (HNSC). The rs2098952 minor allele was associated with *LILRB4* expression and poor prognosis in rectum adenocarcinoma (READ) **(Figure S7B).** *LILRB4* is an immunoinhibitory receptor, specifically expressed on monocytes, macrophages, and dendritic cells[43] and is being investigated as a potential therapeutic target for monocytic acute myeloid leukemia[44]. *CCBL2*-modifying variant rs1932757 was associated with progression-free survival in stomach adenocarcinoma (STAD). *CCBL2* is part of the kynurenine pathway, which some studies suggest can play a role in immune evasion **(Figure S7C)**[45].

To assess the role of TIME-associated variants in disease risk, we performed PheWAS with ICD10 codes in the UKBB for our 475 unique TIME-associated variants and 50 HLA region-associated variants. Eighty-four TIME-associated variants were significantly associated with cancer-related or immune-related phenotypes in the UKBB (**Figure S7D)**. rs7751376 linked to infiltration of tumors by M2 Macrophages was associated with malignant neoplasms of digestive organs, malignant neoplasms of skin, and aplastic and nutritional anemias, suggesting a macrophage-mediated etiology to not only cancer risk but other immune-related disorders (**Figure 7A)**. We also identified 2 novel SNPs modulating *MARVELD2* expression, that are associated with cancer risk in the UKBB cohort. *MARVELD2* is involved in regulating epithelial tight junctions and has been implicated in epithelial-mesenchymal transition (EMT), which greatly changes the tumor immune microenvironment[46]. These two analyses demonstrate the effects of TIME-associated variants on the occurrence and progression of cancer and immune-related diseases.

To investigate the implication of TIME variants for ICB response (**Therapeutic Target Discovery, Figure 1)**, we collected sequencing and phenotype information for 383 patients with melanoma, renal cell carcinoma, and non-small cell lung cancer treated with ICB-blockade from 6 studies[47–52], and imputed SNPs from exome sequencing data. Accuracy of imputation was assessed based on TCGA which provides both exome sequencing and SNP data; most variants were accurately imputed aside from variants on chromosome 6 within

HLA gene regions (**Figure S7E).** SNP associations with response were evaluated for each ICB study and combined for meta-analysis[53]. 69 of the 475 TIME-associated variants with minor allele frequency > 0.01 in all ICB cohorts were considered. Of these variants, 23 were significantly associated with ICB response, implicating genes such as *CCBL2, LAIR1, LYZ, SLC11A1, TREX1, PSMD2, CTSW, LILRB2, SLC25A40, LILRB4*, *CTSS, B2M,* and *SIGLEC5* **(Figure 7B)**. Indeed, several of these genes behaved differently based on ICB response status. Pre-treatment, *LILRB2, TREX1* and *LAIR1* were expressed at lower levels in patients with iRECIST complete response compared to progressive disease **(Figure 7C).** Comparing pre- and on-treatment measurements, we noted a significant increase in *LAIR1, TREX1,* and *LILRB2* expression among tumors with partial and complete response, but no change among patients with progressive disease (**Figure 7D).** These increases are consistent with remodeling of the composition of the TIME in responders, as was recently reported in another immunotherapy study with on-treatment profiling[54]. Moreover, they imply these genes may drive responses to ICB in cancer patients.

As several ICB response genes implicated by TIME-associated variants are upregulated in M2 macrophages and monocytes **(Figure 7E)**, we hypothesized that TIME-associated variants influence ICB response through modulation of properties of the myeloid landscape. For example, variant 19:54754385:T:G was associated with decreased *LILRB2* expression, better ICB response, and increased M1 macrophage infiltration. Indeed, other ICB-associated variants, such as 12:69744014:C:A, 1:89735314:G:A, and 19:52130638:C:A were also associated with increased M1 infiltration **(Figure 7B)**. We revisited analysis to determine whether these variants interact to promote more extreme TIME characteristics, such as observed with Macrophage Regulation score strength and *EP300* expression. We constructed a burden score with the 23 ICB-associated variants by summing the alleles that were associated with better response in melanoma and ran OLS regression with M1 macrophage infiltration. Indeed, ICB patients with a higher burden score had increased M1 infiltration **(Figure 7F)**, but not necessarily increased M2 infiltration (**Figure 7G)**. These findings suggest that TIME-associated variant effects on ICB response may be mediated through effects on macrophage behaviors.

To determine the value of TIME-associated variants as prognostic biomarkers for ICB, we evaluated

TIME-associated variants using an optimized classification framework. With germline variants alone, we achieved ROC AUCs of 0.65, 0.71 and 0.71 for melanoma, non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC), respectively **(Figure 7H)**. Variants with strongest odds ratios differ between cancer types; in renal cell carcinoma, B2M variant 15:42439444:C:T had stronger effects compared to melanoma and non-small cell lung cancer. In contrast to melanoma and NSCLC, RCC TMB is typically low and its link to ICB response is less clear[55]. Underlying TIME-associated variants may shed light on mechanisms mediating ICB response in RCC.

Through a genome-wide approach, we have thus identified germline variants predictive of response to ICB, as well as genes implicated by them. We further validated the functional relevance of our identified ICB-response genes in a preclinical study. C57Bl/6 mice were implanted with syngeneic murine MC38 cells, followed by treatment with anti-PD-1. Interestingly, the tumors showed diverse responses, a clinical situation frequently observed in cancer patient cohorts receiving ICB treatment. We classified responders as those with tumor volumes <500 mm^3 after mice receiving the last dose of anti-PD-1, and those exhibiting growth kinetics similar to controls as non-responders (**Figure 7I)**. At post-inoculation day 20, the tumors were harvested for RNA extraction. We quantified expression through qPCR of 12 of the 15 human genes for which we identified mouse orthologs. The results were compared with human on-treatment RNA-seq data from Riaz *et al.* Similar to human responders, mouse ICB responders had significantly higher levels of *Ctsw* and *Ctss* **(Figure 7J)** and higher levels of *Lair1* and *Trex1* post treatment. Increased *CTSS* expression in dendritic cells can alter antigen presentation and result in a more diverse T_{FH} repertoire[56]. *CTSW* plays an important role in cytotoxic immune response[57] and has been associated with immune infiltration in breast cancer[58].

To confirm the value of the identified genes as biomarkers for different ICB therapies, we analyzed the RNA sequencing data generated from a mouse "M4" melanoma model receiving anti-CTLA-4 treatment[59]. Gene expression levels (TPM) were analyzed by deconvolution with immune gene signatures to obtain immune cell infiltration levels. As expected, across the post-treatment tumors, the expression of the 12-gene mouse TIME signature correlated positively with inflammatory M1 macrophage levels **(Figure 7K)** and negatively with

immunosuppressive M2 macrophage levels (**Figure 7L).** Since *LILRB* and *CTSS* are involved in phenotype switching of macrophages, we examined their expression in the post-treatment regressing (responder), stable and growing (non-responder) tumors. The responder and stable tumors expressed significantly higher levels of *Pirb* (human LILRB ortholog) and *Ctss* than non-responder tumors **(Figure 7M)**, suggesting these may serve as therapeutic targets for enhancing the ICB efficacy in cancer treatment.

Discussion

Understanding the host genetics that influence the characteristics of the TIME may suggest strategies for better risk stratification and response prediction. In this study, we present a comprehensive analysis of multiple genomic and clinical databases to understand germline variants underlying characteristics of the tumor immune microenvironment, disease progression and response to therapy. We identified 482 TIME associations and 475 TIME-associated variants in the TCGA from analysis of 137 heritable IP components. IP components mainly characterized macrophage behavior, *EP300*/proteasome expression, *IFN-γ*. Integration of GTEx, DICE, and ENCODE gave us insight into roles of TIME-associated variants in normal tissue setting, cell-type manner, and epigenetic landscape. Finally, we evaluated the role of TIME-associated variants in disease risk and ICB response, achieving predictive value for response in ICB cohorts of 3 different cancer types. Ultimately, we believe this study will pave the path for better understanding how individual heterogeneity contributes to the TIME and aid precision medicine initiatives.

The characteristics of the TIME that varied most across tumors were macrophage regulation and activity of *EP300* and the proteasome. We note that macrophages are typically the most abundant cell type in immune cell infiltrates, thus our study design may have been best powered to detect SNPs associated with macrophage characteristics. We identified variants modifying macrophage-related genes, such as *CD53, FCGR2A, FCGR3A, VSIG4, LILRB4, LILRB2*, in a cell-type specific manner. Of these, we found that variants associated with *LILRB2* and *LILRB4* expression were associated with response to immune checkpoint blockade therapy. Not only are these genes involved in macrophage infiltration, but also macrophage polarization. *LILRB2* mediates polarization

of macrophages to inflammatory phenotypes and may be a myeloid immune checkpoint[60]. *In vivo* studies in syngeneic mice suggested that targeting *LILRB2* could reprogram macrophages and improve antitumor immune responses; additionally phase 2 clinical trials are in development to target *LILRB2* and boost anti-PD1 response[60]. *LILRB2* may not only serve as a promising therapeutic target, but also a useful biomarker for ICB responders. *LILRB4* is upregulated on acute myeloid leukemia (AML) cells[44]. Our study did not include hematologic malignancies but suggests that *LILRB4* may be relevant to determining TIME characteristics in solid tumors as well. *LILRB3* has also been implicated as a potent myeloid checkpoint[61]; however, we did not find evidence that genetic variation affecting this gene influenced the TIME in the current cohort. We also note that the Macrophage Regulation score is tightly correlated with the *TGF-β* score described in Thorsson et al; *TGF-β* plays a role in M2 Macrophage polarization[62]. Indeed, we found also that these variants have cell-type specific behavior through integration of DICE eQTL dataset and interaction term linear regression models; variants associated with *LILRB2, LILRB4, and LAIR1* were specific to monocytes or macrophages.

The *EP300/*proteasome axis was the second most variable characteristic of the TIME and associated with infiltration of tumors by TH1 cells. *EP300* and the proteasome work together to regulate stress responses through modulation of Heat Shock Factor 1 (HSF1) activity[33]. A shift in Th1/Th2 cytokines underlies stress responses in mouse colon cancer models[63]. We detected 2 variants which were significantly associated with expression of *MARVELD2 and LILRB2* specifically in Th1 cells; however, a better understanding of cell-types is needed. Thus, *EP300* activity may broadly contribute to an immunosuppressive tumor microenvironment via effects on several cell types[64]. Importantly, *EP300* has also been implicated as an oncogene in tumor cells[65–68], thus it will be important to understand the specific roles of *EP300* in each cell type to effectively target its activity therapeutically.

We also investigated SNP associations in the HLA region, using two strategies to deal with the high levels of polymorphism in this region. Using conditional analysis and allele-specific expression greatly reduced the number of significant associations. These two approaches returned slightly different results; however, it is difficult to determine whether this was due to the loss of sample size when acquiring allele-specific expression,

or a higher error rate when aligning to a general reference. We identified several variants not only influencing MHC class I and II genes, but other genes implicated in the antigen presentation pathway, such as *TAP2*, *TAP1*, and *PSMB9*. SNP rs1143146 influenced expression levels of *HLA-A*, with no effect on *HLA-B* or *HLA-C* expression and interacted with rs2844806. Patients with both variants had poorer outcomes in pancreatic and rectal adenocarcinoma; further tumor samples may be needed to implicate variants in other cancer types. Patients with both variants also had differences in potential to present borderline driver mutations and biases for certain *HLA-A* alleles, suggesting variants may affect multiple aspects of MHC function and more investigation must be done to determine which are important for effective immune responses. Several variants identified through HLA analysis were also associated with disease risk in UK Biobank; this suggests that immune-related dysregulation could not only have a role in cancer, but other diseases.

The tumor immune microenvironment is a determinant of response to immunotherapy, with both the number and quality of T cells implicated as being determinants of immune response[69]. Other studies have implicated *IFN-γ* response, cytotoxicity, T cell activation and T cell exhaustion signatures[51,70–73]. Non-T cell factors, such as *PD-L1* expression, neoantigen load, antigen presentation pathways, mesenchymal transition, and wound healing have been implicated in ICB response as well[50,52,74–77]. We therefore evaluated the predictive potential of our TIME-associated variants for ICB response. Of the 23 variants implicated in ICB response, many modulate genes involved in myeloid activation, such as *LILRB2, LAIR1, LYZ, PSMD2, SLC11A1,* and *CTSS*. Genes, such as *LILRB4* and *LILRB2,* are overexpressed in macrophages and typically have an immunoinhibitory role. Our studies suggest that genetic variation affecting these immunoinhibitory molecules may aid in risk stratification for cancer survival and ICB response. Specifically, we were able to achieve an AUC of 71% in renal cell carcinoma; the link between tumor mutation burden (TMB) and ICB response in RCC is not clear, in contrast to high TMB diseases like Melanoma and NSCLC where higher TMB is associated with better responses[52,78–80]. Possibly, in a setting with low TMB such as in RCC, host genetics have more value as prognostic biomarkers. In this study, we only assessed predictive value of germline variation to ICB response, but these predictive models could potentially be augmented with somatic features, such as TMB, PD-L1 positivity, and neoantigen load.

Additionally, most of the SNPs tested for ICB response had minor allele frequency > 10% in TCGA; through the incorporation of rare variants, this AUC could be boosted[81,82].

Discovery of TIME influential SNPs is dependent on the availability of paired genomic and transcriptomic data from tumors, which is currently available only for a few cohorts. Effect sizes associating genetic variants with cellular phenotypes are likely to be larger than those linking genetic variants to diseases[83–85], however the number of associations detected may still be limited by available sample sizes and the limited population diversity thereof. We were able to impute a subset of our SNPs into existing ICB study cohorts that had only exome sequencing, but others falling outside of exonic regions could not be analyzed in this context. Studies focused on tumor exomes and transcriptomes could include SNP chips to allow more effective integration into future studies of germline genomic variation.

Conclusion

Interindividual differences encoded by germline variation contribute to considerable population variation in heritable traits, including in the immune system. Our analysis highlights the utility of studying this variation in the context of the tumor immune microenvironment. Identification of SNPs associated with major axes of immune variation provide clues as to the major factors influencing anti-tumor immune responses. TIME-associated variants associated with cancer risk or response to immunotherapy merit further investigation as possible entry points for future efforts to engineer the tumor immune microenvironment.

Acknowledgements

This work was supported by Emerging Leader Award from The Mark Foundation for Cancer Research, grant #18-022-ELA to HC, NIH grant R01CA220009 to HC and MZ, U01CA196406 to OH, 1F30CA247168-01 and T32CA067754 to MP, and NIH Intramural Research Program to CPD, GM, EPG. UKBB data was retrieved under project ID 37671. The results shown here are in large part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga and Genotype-Tissue Expression (GTEx) Project:

https://gtexportal.org/home/. GTEx was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 10/10/20.

Author Contributions

M.P. and H.C. conceived the work and designed and analyzed the experiments; W.T. and C.F. assisted in statistical analyses; M.P. and H.C. wrote the paper with assistance from P.V., W.T., M.Z., J.M., S.P., O.H., C.F., G.P.M. and A.C.; A.C. and G.P.M. advised on HLA region analysis; J.T. performed optimization of risk score analysis; B.J.S. and C.G.C. performed DICE annotation under supervision of P.V.; H.K. generated network plots in epistatic analysis and conducted UKBB PheWAS analysis; S.C. and R.M.S. performed UKBB preprocessing used in PheWAS analysis; V.W., S.G., C.P.D, E.P.G, G.M. performed mouse experimental validation and analyzed results.

Declaration of Interests

Dr. Patel receives scientific advisory income from: Amgen, AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Genentech, Illumina, Merck, Rakuten, Paradigm, Tempus. Dr. Patel's university receives research funding from: Bristol-Myers Squibb, Eli Lilly, Incyte, AstraZeneca/MedImmune, Merck, Pfizer, Roche/Genentech, Xcovery, Fate Therapeutics, Genocea, Iovance.

Resource Availability

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Hannah Carter (hkcarter@health.ucsd.edu).

Data Availability

The data that support the findings of this study are available on request from the corresponding authors H.C. and

M.P. The data are not publicly available as they were accessed through dbGaP applications and are controlled data that could lead to reidentification of individuals. For Data Access to processed genotyping, transcriptomic and mutation data, contact corresponding authors with proof of access to dbGaP studies and requests will be completes

Code Availability

All code used for analysis and figure generation are available at https://github.com/meghanasp21/TIMEgermline.

Supplementary Tables

Supplementary Table 1: Genome-wide Complex Trait Analysis (GCTA) results
Supplementary Table 2: Significant Associations from GWAS Analysis of 120 Immune Phenotypes
Supplementary Table 3: Significant Associations from HLA variant analysis
Supplementary Table 4: METAL Immune-Checkpoint Blockade (ICB) association analysis results

Method Details

Key Resources Table

Public Data Sets			
Raw and processed clinical, array, and sequence data	NCI Genomic Data Commons	https://portal.gdc.cancer.gov/	SCR_014514
DICE	Schmiedel et al. 2018	https://dice-database.org/	SCR_018259

ChIP-seq	Sloan et al. 2016	https://www.encodeproject.org/	SCR_015482
RCC ICB trial data	Miao et al 2018		
Melanoma ICB trial data	Hugo et al. 2016, Van Allen et al. 2015, Riaz et al, Snyder et al		
NSCLC ICB trial data	Rizvi et al.		
Software and Algorithms			
bam2fq	NA	http://www.htslib.org/	SCR_002105
Bedtools	Quinlan et al. 2010	https://github.com/arq5x/bedtools2	SCR_006646
Birdsuite	Korn et al. 2008	https://gaow.github.io/genetic-analysis-software/b/birdsuite/	SCR_001794

BWA (v0.7.17-r1188)	Li et al. 2010	http://bio-bwa.sourceforge.net/	SCR_010910
CaviarBF	Chen et al. 2015	https://bitbucket.org/Wenan/caviarbf	SCR_000502
CIBERSORTx	Newman et al. 2019	https://cibersortx.stanford.edu/	SCR_016955
Cytoscape	Shannon et al. 2003	https://cytoscape.org/	SCR_003032
DeepVariant (v0.10.0-gpu)	Poplin et al. 2018; Yun et al. 2020	https://github.com/google/deepvariant	
fastq-pair	Edwards et al. 2019	https://github.com/linsalrob/fastq-pair	
GATK (v3.8-1-0)	McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013	https://gatk.broadinstitute.org/hc/en-us	SCR_001876

GCTA (1.93.0beta)	Yang et al. 2011; Yang et al. 2010	https://cnsgenomics.com/software/gcta/#Overview	
GREGOR (v1.4.0)	Schmidt et al. 2015	https://genome.sph.umich.edu/wiki/GREGOR	SCR_009165
HLApers	Aguiar et al. 2019	https://github.com/genevol-usp/HLApers	
Michigan Imputation Server	Das et al. 2016	https://imputationserver.sph.umich.edu/index.html#!	SCR_017579
NetMHCpan4.0	Jurtz et al. 2017	http://www.cbs.dtu.dk/services/NetMHCpan-4.0/	
Networkx	Aric et al. 2018	https://networkx.org/documentation/stable/index.html	SCR_016864
Picard	NA	https://broadinstitute.github.io/picard/	SCR_006525
Plato (v2.0.0)		https://ritchielab.org/plato	
PLINK	Purcell et al. 2007	http://zzz.bwh.harvard.edu/plink/	SCR_001757
RSEM (v1.2.21)	Li et al. 2011	http://deweylab.biostat.wisc.edu/rsem/	SCR_013027

Samtools			SCR_002105
STAR (v2.4.1d)	Dobin et al. 2013	https://github.com/alexdobin/STAR	SCR_015899
VEP (Ensembl Variant Effect Predictor) v87		http://www.ensembl.org/info/docs/tools/vep/index.html	SCR_007931

TCGA Subject Details

The Cancer Genome Atlas (TCGA) consists of tumor and matched normal samples for over 11,000 patients. The Genomic Data Commons (GDC) legacy archive contains germline data for 11,542 samples from 10,875 unique individuals. Samples with TCGA project IDs: DLBC, LAML, THYM were excluded as they represent cancers derived from immune cells. Pairs of individuals with estimated KING kinship coefficient > 0.177, which represents first-degree relatedness were excluded. TCGA individuals were consented for general research use and no attempts were made to reidentify or contact subjects. Both females and males were included, using sex along with individual age as a covariate. Experimenters were not blinded and randomization of subjects was not relevant to the study. We did not check for sample sizes using a power analysis because our sample sizes were over 10,000 individuals.

TCGA Genotypes

Normal (non-tumor) level 2 genotype calls generated from Affymetrix SNP6.0 array intensities using BIRDSUITE (RRID: SCR_001794) software[86] were retrieved from TCGA GDC Legacy Portal. In these files, each of 906600 SNPs was annotated with an allele count (0=AA, 1=AB, 2=BB, -1=missing) and confidence score between 0 and 1. Genotypes with a score larger than 0.1 (error rate > 10%) were set to missing and data were reformatted for PLINK (RRID:SCR_001757)[34]. We discarded 322 SNPs with probe names that did not match the

hg19 UCSC Genome Browser (RRID:SCR_005780) Affymetrix track (track: SNP/CNV Arrays, table:snpArrayAffy6). Allele counts were converted to alleles using the definitions in metadata distributed with BIRDSUITE (RRID: SCR_001794) and negative strand genotypes were flipped to the positive strand using PLINK (PLINK, RRID:SCR_001757).

Pre-imputation processing of autosomal and X chromosome genotypes consisted of the following steps:

1. SNPs with call rate <90% were removed
2. SNPs with minor allele frequency (MAF) <1% were removed
3. Individuals with genotype coverage <90% were removed
4. Individuals with conflicting gender assignments were flagged
5. Heterozygous haploid SNPs were set to missing.

After applying these filters, the remaining 800644 autosomal and 32809 X chromosome SNPs were input to the secure Michigan Imputation Server[87]. Additional SNPs were imputed with Minimac3/Minimac4 and ancestry-matched reference panel -- European (HRC Version r1.1 2016 reference, Eaglev2.3), African (CAAPA, Eaglev2.3), and Asian (Genome Asia Pilot).

Post-imputation processing of genotypes included:

1. SNPs with MAF <1% were removed
2. Autosomal SNPs with Hardy-Weinberg Equilibrium <1e-9 were removed
3. Individuals with high heterozygosity rates (>3 SDs of mean) were removed
4. Pairs of individuals with

Rsq values from INFO files were extracted to assess genotyping quality. The final genotyping data included 8217 individuals and 7,884,718 variants.

TCGA Population Stratification

Ancestry filtering was applied using two techniques: (1) k-means clustering and (2) outlier identification. For k-means clustering, TCGA and HapMap Phase III populations were combined. HapMap Phase III genotypes were

obtained from the NCBI HapMap ftp site and lifted to hg19 using the liftOver utility[88]. Genotypes were merged and reduced to a set of 33,675 independent SNPs determined previously[85,88] through linkage-based filtering using PLINK (RRID:SCR_001757). Pairwise identity-by-state (IBS) between all individuals was calculated and the resulting IBS matrix was used for PCA analysis. K-means clustering trained on HAPMAP Phase III separated individuals into the following groups: (1) TSI, CEU, (2) JPT, CHD, CHB, (3) MEX, (4) GIH, (5) MKK, (6) YRI, ASW, LWK. This trained model was used to predict groups in TCGA. Cluster (1) were identified as European individuals.

We ran aberrant R package v1.0 with lambda 20 for outlier identification[89]. Intersection of k-means clustered individuals and non-outlier individuals from outlier identification analysis was used to include TCGA individuals in the European ancestry discovery cohort.

TCGA RNA-seq processing

Three TCGA RNA sources were considered for TCGA phenotype characterization. One was processed RNAseq data from a previous analysis[90], where RNA-seq sequencing reads obtained from TCGA were realigned and quantified using Sailfish v.0.7.660[91] with default parameters. The others included Level_3 RSEM_genes_normalized firebrowse data (version 2016_01-28), from the firebrowse portal (http://firebrowse.org/, date: 06/12/20) and PanCanAtlas RNA data from GDC PanCanAtlas Publications Supplemental Data (https://gdc.cancer.gov/about-data/publications/pancanatlas). Only primary tumors (barcode: 01A/01B/01C) were considered in our analysis. Corresponding clinical metadata were obtained from the GDC Portal (https://tcga-data.nci.nih.gov/docs/publications/tcga/).

PCA analysis

Inverse-ranked phenotypes from 3 RNA-processing pipelines -- Firebrowse, Sailfish, and Pancan Atlas -- were combined by taking the median of values. Individuals with > 10% phenotypes missing were removed from analysis. Sklearn was used to scale values (Standard Scaler) and conduct principal component analysis. Ordinary

least squares (OLS) regression was conducted with age, sex and tissue type as covariates. Variance explained by each component were computed and first 10 principal components were considered for analysis.

TCGA Phenotype Data

The following phenotypes were extracted or generated using all 3 TCGA RNA sources.

1. **Immunomodulators:** 436 genes used to define immune states from Thorsson et al.
2. **Immune checkpoint molecules:** 78 immune checkpoint stimulatory and inhibitory molecules from Thorsson et al.
3. **Antigen Presentation:** 231 antigen presentation genes from Gene Ontology [GO_REF:0000022]
4. **Immune cell markers:** 60 immune cell type markers from Danaher et al.
5. ***IFN-γ*:** *IFN-γ* genes retrieved from Biocarta [Systematic Name: M18933]
6. ***TGF-β*:** *TGF-β* genes retrieved from Biocarta [Systematic Name: M22085)
7. **Immune states:** Individual level scores for 6 immune states [wound healing, *IFN-γ* dominant, inflammatory, lymphocyte depleted, immunologically quiet, and *TGF-β* dominant] from Thorsson et al.
8. **Immune infiltration levels:** 61 relative immune infiltration estimates from CIBERSORTx[92] using 5 signature matrices: LM22, DICE, melanoma, Non-Small cell Lung Cancer (NSCLC), Head and Neck Squamous Cell Carcinoma (HNSCC).

Phenotypes with greater than 10% zero values were excluded and rank-based inverse normal transformation (**Figure S1)** was applied to each tissue type using Eqn 1[93].

qnorm((rank(x,na.last="keep")-0.5)/sum(!is.na(x))) Eqn 1

A total of 741 phenotypes remained for preliminary analyses.

Antigen Affinity scores for all TCGA individuals were calculated for missense and in-frame indel somatic mutations using NetMHCpan4.0[94] and NetMHCIIpan3.2[95] to predict peptide-MHC binding affinities for all peptides spanning the novel mutated protein sequence. For a more in-depth description see Marty et al. 2017[14], Castro et al. 2019[96].

Heritability Estimates

Heritability estimates were calculated with the genomic-relatedness-based restricted maximum-likelihood (GREML) approach implemented in GCTA (Genome-wide Complex Trait Analysis)[97,98]. Genetic relationship matrices (GRMs) which measure genetic similarity of unrelated individuals (GRM <0.05) were constructed for the autosomal and X chromosomes for the European cohort. Benjamini-Hochberg false discovery rates (FDR) were calculated using statsmodels[99]. Immune traits were considered sufficiently heritable if the V(g)/V(p) value was >0.05 based on at least 2 of 3 RNA sources and FDR was <0.05 for at least 1 of 3 RNA sources.

To assess heritability contribution from each chromosome, GREML was applied to each separated chromosome GRM. To assess heritability contribution from highly polymorphic immune loci as compared to the rest of the genome, GRMs were generated separately for the HLA/KIR regions and the whole genome excluding HLA/KIR regions. Immune phenotypes were considered as having majority heritability contribution from HLA/KIR polymorphic regions if at least 2 RNA sources had great HLA/KIR V(g)/V(p) estimates than non-HLA/KIR V(g)/V(p) estimates. All analyses were adjusted for covariates age and sex.

GWAS Analysis

PLINK (RRID:SCR_001757) GLM method was used to conduct association analyses with immune phenotypes. All associations were adjusted for covariates age, sex and first ten principal components. Gene expression values, CIBERSORTx relative infiltration estimates, and immune state scores were inverse-rank normalized by tissue type to control for tissue-type expression effects. Significant associations were identified with the PLINK (RRID:SCR_001757) clumping method using primary significance threshold (1e-5)/120[100], kb threshold 1000, and R^2 threshold of 0.1. Associations with minimum p-value passing suggestive threshold ($1e10^{-5}$) Bonferroni-corrected for number of phenotypes (120) were used as the final set of significant associations.

Different thresholds were tested to determine ability to recover bonafide GWAS SNPs and functionally enriched SNPs (**Table S3).** Suggestive threshold $1e^{-06}$ recovered most functional SNPs. To reduce false positive rate, only

SNPs that were determined to be functional by another data source were considered for further evaluation.

Conditional HLA analysis

PLINK (RRID:SCR_001757) GLM method was used to run stepwise conditional analysis for identification of independent HLA associations[101]. First, associations with HLA region phenotypes were conducted identical to methods for non-HLA phenotypes. After identification of the most significant SNP with the lowest p-value, we re-ran the analysis conditioning on the most significant SNP as a covariate. Analysis was conducted until no SNPs with p-value $< 1x10^{-6}$ remained.

Conditional analysis was conducted with all 3 RNA processing pipelines -- Sailfish, Firebrowse, Pancan Atlas-- with the 17 phenotypes identified through GCTA analysis to be driven by variation in the HLA regions: HLA-A, BTN3A2, HLA-DQB2, HLA-C, TAP2, HLA-DQB1, HLA-B, HLA-DRB5, MICA, HLA-H, HLA-DQA1, HLA-G, MICB, HLA-DQA2, HLA-F, PSMB9, HLA-DRB1. Only variants on chromosome 6 and identified by at least 2 RNA processing pipelines were further investigated.

HLA Allele Specific Expression

TCGA tumor specific RNA BAMs were downloaded from the GDC using a manifest from 07/16/2019. The HLApers[102] kallisto-based pipeline was used with gencode.v30 annotations[103]. Default parameters were used and the two alleles with the highest calculated expression were retained for each HLA gene if there were more than 2 alleles reported. The top 2 highest expressed HLA alleles were averaged for input into SNP analyses. If expression for at least two alleles was not calculated, expression was set as missing for the sample.

Only primary samples (01A/01B/01B) were considered for analysis. Summed HLA allele specific expression was inverse-rank normalized by cancer type and used for downstream analyses.

Affinity analysis

A set of 723 driver mutations with potential to bind MHC-I were identified by constraining the median percentile

rank binding affinity <5 across 2915 MHC-I alleles for 1018 driver mutations previously described in Marty *et al*[14]. Binding affinity was calculated using NetMHCpan4.0[94]. *HLA-A*, *HLA-B*, and *HLA-C* gene-level driver affinity scores were obtained by taking the minimum (best) of the two allele-specific patient affinity scores using HLA alleles previously typed in Castro et al. 2019[96]. Ordinary least squares (OLS) from the python statsmodels (v0.10.1) package was used to fit patient *HLA-A*, *HLA-B*, and *HLA-C* driver affinity scores for patient genotypes for 109 SNPs identified in HLA-region conditional analyses.

DICE

DICE eqtls were obtained at https://dice-database.org/. Methods associated with DICE eQTL discovery are published in Schmiedel et al[38].

GREGOR (RRID: SCR_009165)

GREGOR (RRID: SCR_009165) was used to analyze SNP enrichment at epigenetic features. We obtained 479 bed files for 11 histone experiments and 52 cell types from ENCODE (RRID:SCR_015482) (downloaded on May 3, 2020**).** Only "stable peaks" and "replicated peaks" files were kept for analysis. If more than 1 bed file for a cell type and histone mark were available, the files were combined.

In addition, 323 bed files for 12 transcription factor binding experiments and 12 cell types were downloaded from ENCODE (RRID:SCR_015482**)** on August 4, 2020. Only "optimal IDR thresholded peaks" and "conservative IDR thresholded peaks" files were kept. If more than 1 bed file for a cell type and transcription factor were available, the files were combined.

GREGOR (RRID: SCR_009165) was run with EUR Reference files made from 1000G data with LD window size 1MB and LD r2 > 0.7. Enrichment ratios were calculated by taking the difference between observed and expected number of SNPs and dividing by the expected number of SNPs. Any files with Audit errors were excluded.

GTEx

GTEx V8 full summary statistics from cis-eqtl analysis were downloaded from Google Bucket. SNPs were extracted from compiled parquet files for each cell type. If minor allele and major allele were reversed in GTEx, beta value signs were reversed.

Significant GTEx eQTLs were retrieved from the GTEx download page. *Cis* associations were associations from which immune phenotype was the gene for which GTEx eQTL was associated with. *Trans* associations were associations for which a different gene from the immune phenotypes was the gene for which GTEx eQTL was associated with.

Z-scores (beta/SE) were calculated for all SNPs. 1000 random SNP-gene combinations were selected to conduct association analyses. Z-scores for random SNP-gene pairs were calculated. The random distribution was sampled 1000 times and the percentage of samples with greater than observed score was calculated and averaged. If the average percentage of random samples was less than 1% for a SNP by at least 2 RNA processing pipelines, variants were identified as having deviating behavior in tumor states. The same approach was repeated with Whole Blood as the matching GTEx tissue. Variants were identified as having similar behavior to Whole Blood if the average percentage of random samples was not less than 1% for a SNP. A deviating SNP was identified to have normal tissue support if the z-score (tcga beta - gtex beta)/(gtex se) was lower using normal tissue association data than tumor tissue association data.

Infiltration Interaction analysis

We followed the GTEx approach for cell type interaction eQTL discovery[35]. We ran linear regression model with an interaction term accounting for interactions between genotype and cell type enrichment:

$$p \sim g + i + g \circ i + C \qquad \text{Eqn 2}$$

where p is the immune phenotype vector, g is the genotype vector, i is the inverse normal transformed by tissue type xCell enrichment score[37], and the interaction term $g \circ i$ corresponds to pointwise multiplication of genotypes and cell type enrichment scores. The same covariates, denoted by C, were used as in regular immune

microenvironment GWAS analysis. Benjamini-Hochberg FDR was calculated for beta coefficient of interaction term and only variants with FDR < 0.1 were evaluated.

UKBB

UKBB subjects were subsetted into separate ethnic-racial groups largely following continental ancestry prior to analysis. The sub-setting was performed to generate homogenous groups and reduce potential admixture bias in the genetic analyses. To identify the European-ancestry samples, we started with directly called genotype data and identified a set of overlapping SNPs with 1000 Genomes Project and AWS (RRID:SCR_008801) (1KG) population and then merged them together. Next, we pruned the SNP set so remaining SNPs were in linkage equilibrium using PLINK (PLINK, RRID: SCR_001757)[34]. flashpca was used to calculate principal components for 1KG SNPs[104]. The UKBB samples were projected onto 1KG space using flashpca. To identify subjects of European ancestry, we utilized Aberrant to generate clusters with a broad set of lambda values (clustering thresholds) and checked that the cluster included all 1KG subjects of European ancestry and maximized the total number of UKBB subjects (lambda=8.2)[89]. Finally, we compared the self-reported race/ethnicity of subjects within this cluster and removed samples that were discordant. We identified 454,487 subjects of European ancestry. To identify the unrelated samples from the finalized European list, we used the relatedness file provided by UKBB and a custom script was used to select unrelated samples while maximizing sample counts. The final European unrelated set included 382,841 subjects.

Variant dosages extracted from imputed UKBB BGEN files were used for phewas analysis with PLATO v2.0.0[105]. ICD10 diagnosis codes associated with neoplasms and immune disorders were collapsed according to level-1 groupings used by UKBB resulting in a total of 24 groups. For example, C00-C14 is one of the groups containing ICD10 codes associated with malignant neoplasm of lip,oral cavity, and pharynx. Individuals with diagnosis code in group were coded as 1, with the remaining individuals coded as 0. Logistic regression was conducted with UKBB binary files containing HLA-immune variants, logistic phenotype file, and age, sex, and principal components 1-10 as covariates. P-values were Benjamini-Hochberg FDR adjusted.

Immunotherapy Response Validation

Raw fastq files were obtained for the following immune checkpoint trials: Hugo et al. 2016 (SRA accession: SRP090294, SRP067938; Cancer: melanoma)[47], Van Allen et al. 2015 (SRA accession: SRP011540, Cancer: melanoma)[48], Miao et al. 2018 (SRA accession: SRP128156, Cancer: clear cell renal carcinoma)[49], Riaz et al. 2017 (SRA accession: SRP095809, SRP094781; Cancer: melanoma)[51], Rizvi et al. 2015 (SRA accession: SRP064805, Cancer: non-small cell lung cancer)[50], Snyder et al. 2014 (SRA accession: SRP072934, Cancer: melanoma)[106]. Reads were aligned to UCSC hg19 coordinates using BWA (RRID:SCR_010910) v0.7.17-r1188[107]. Reads were sorted by SAMTOOLS (RRID:SCR_002105) v0.1.19[108,109], marked for duplicates with Picard (RRID:SCR_006525) Tools v2.12.3 and recalibrated with GATK (RRID:SCR_001876) v3.8-1-0[110–112]. Germline variants were called from sorted BAM files using DeepVariant v0.10.0-gpu[113,114]. The final immunotherapy cohort consisted of 70 clear cell renal carcinoma, 278 melanoma and 33 non-small cell lung cancer patients.

Overlapping DeepVariant germline variants and imputed TCGA variants were used to impute genotypes for both immunotherapy trials and TCGA European discovery cohorts. Imputation performance was confirmed by comparing TCGA imputed genotypes using full Affymetrix 6.0 variants vs. variants overlapping WXS calls. Variants with >5% mismatch in genotype calls and minor allele frequency < 1% were excluded. Only variants with at least 1% frequency in all 5 ICB cohorts were considered for analysis, leaving 69 SNPs.

Subject phenotypes were downloaded from supplementary information of ICB trial publications. PLINK (PLINK, RRID:SCR_001757) logistic regression was conducted with age and sex as covariates. Response phenotypes were determined from iRECIST criteria[115]. Patients were categorized as responders if they have iRECIST criteria: CR (complete response), PR (partial response) and SD (stable disease) with >1 year overall survival. Genome-wide association studies (GWASs) were conducted for ICB responders within each ICB-cohort using PLINK (PLINK, RRID:SCR_001757). Age and sex were included in the logistic analysis as covariates. We then used METAL[53] with a sample size weighting scheme to perform a pan-study meta-analysis for ICB response. Only SNPs with

MAF > 0.01 in all 6 studies were considered resulting in 69 SNPs for analysis. SNPs with Benjamini-Hochberg FDR less than 0.1 were reported as significant SNPs (14 SNPs).

Polygenic risk scores for melanoma, RCC and NSCLC were based on an individual's odds ratio weighted genotype. Given the limited number of samples in these ICB trials across cancer types we employed leave-one-out cross validation to assess the overall predictiveness (AUC) of our identified germline variants. Specifically, variant odds ratios were generated for n-1 samples per cancer type. These odds ratios were then utilized to obtain a weighted genotype for all n samples. Hyperoptimized models were then trained on the discovery n-1 subset before holdout evaluation. The reported final AUCs are the result of n composite runs of this cross-validation, one for each unique n-1 fold in the data set.

For each cancer type we employed Bayesian optimization for model selection and hyperparameter optimization using the Optuna framework[116]. Model configurations that were evaluated for each cohort were as follows: random forest classifiers, support vector machine (SVM), and logistic regression. Hyperparameters evaluated focused on intrinsic model properties (e.g., SVM kernel) and regularization constraints and methodologies. Logistic regression was the identified optimal framework for both melanoma and NSCLC, however for RCC SVMs yielded the best performance. This approach was repeated for unweighted and complete response evaluations as well.

ICB trial RNA

FASTQ/BAM files were downloaded for 33 RCC and 120 melanoma patients. If BAM files were downloaded, they were converted to FASTQ using bam2fq[109]. Unpaired reads were removed using fastq pair[117]. Paired reads were aligned with STAR (RRID:SCR_004463) v2.4.1d[118] to GRCh37 reference alignment. RSEM[119] was used for transcript quantification. TPM values were log2 transformed for analyses.

Mouse experiments

Wild-type C57BL/6 (RRID:IMSR_JAX:000664) were purchased from The Jackson Laboratory. All the animal

studies were approved by the Institutional Animal Care and Use Committee (IACUC) of university of California, San Diego, with protocol ASP #S15195. Mice at Moores Cancer Center, UCSD are housed in micro-isolator and individually ventilated cages supplied with acidified water and fed 5053 Irradiated Picolab Rodent Diet 20 lab diet. Temperature for laboratory mice in our facility is mandated to be between 65–75 ° F (~18–23 °C) with 40–60% humidity. All animal manipulation activities are conducted in laminar flow hoods. All personnel are required to wear scrubs and/or lab coat, mask, hair net, dedicated shoes, and disposable gloves upon entering the animal rooms. $2x10^5$ MC38 (RRID:CVCL_B288) cells were transplanted into two flanks per C57Bl/6 (RRID:IMSR_JAX:000664) mouse, aged 7-8 weeks. When tumors reached 100 mm^3, mice were randomized and treated with anti-PD-1 (Bio X Cell Cat# BE0146, RRID:AB_10949053, clone RMP1-14) or isotype control antibody through intraperitoneal injection. Mice were given 300ug of anti-PD-1 (Bio X Cell Cat# BE0146, RRID:AB_10949053, clone RMP1-14) or isotype control every 3 days for a total of 6 doses. The mice were then euthanized after the completion of the treatment (or when control-treated mice succumbed to tumor burdens, as determined by the ASP guidelines). MC38 cells were not screened using STR profiled on site.

RNA-seq and CIBERSORTx infiltration estimates for four additional murine melanoma models were obtained from GEO accession (GSE144946). Responders were mice whose size at harvest was smaller than the last dose of anti-CTLA4. RNA-seq counts were converted to TPM and log2 normalized. Out of 4 models, only M4 showed response to anti-CTLA4, thus analysis of *PirB* and *Lair1* expression relative to response was confined to M4 mice. Significance was determined by T-test. A gene signature comprising 12 genes was constructed by taking average of expression of mouse orthologs of human genes implicated by TIME-associated variants: *HAUS1, B2M, SLC25A40, TREX1, LYZ, LILRB2, LAIR1, CTSW, PSMD2, CTSS, FCGR2B and SLC11A1.* Average expression was then normalized by number of genes. Correlation was assessed using Ordinary Least Squares regression.

Reagents

PD-1 antibody (clone RMP1-14, Bio X Cell Cat# BE0146, RRID:AB_10949053) and isotype antibody (catalog #BE0091) were purchased from Bio X Cell.

RT-PCR

RNA from MC38 (RRID:CVCL_B288) tumors was extracted using the RNeasy Mini Kit (Qiagen catalog #74104). 500ng of RNA per reaction was used to prepare cDNA with the SuperScript™ VILO™ cDNA Synthesis Kit (ThermoFisher Scientific) following manufacturer's instructions. The cDNA was used to set up the RT-PCR reaction with 4 technical replicates per tumor with the Fast SYBR™ Green Master Mix (ThermoFisher Scientific) according to manufacturer's instructions. PCR quantification was conducted using the $2^{-\Delta\Delta CT}$ method and normalized to the housekeeping gene β-actin. All the primers used in this study were ordered from Sigma-Aldrich and the sequences are shown in Supplemental Table 5.

Survival Analysis

CoxPH analysis was implemented with the R Survival package in R statistical environment (version 3.6.1) using overall survival information retrieved from Liu et al[120]. The analysis was restricted to tumors in the TCGA European discovery cohort with available AJCC stage categorization, overall survival time less than 5 years and not identified as MSI-high. CoxPH with inverse-ranked phenotype values was conducted including age at diagnosis, sex and stage as covariates. Stage was collapsed into four groups: Stage I (Stage I, Stage IA, Stage IB), Stage II (Stage II, Stage IIA, Stage IIB), Stage III (Stage III, Stage IIIA, Stage IIIB, Stage IIIC), Stage IV (Stage IV, Stage IVB, Stage IVC). Categorical covariates were dummy coded (0/1) prior to model inclusion. Wald test p-values and hazard ratios for immune phenotypes were used for further analysis. Benjamini-Hochberg FDR values were calculated using the statsmodel package.

Kaplan-Meier analysis of immune microenvironment associations were conducted with overall and progression-free survival by cancer type using the lifelines package. TCGA individuals were divided into 3 groups based on genotype calls: minor allele homozygotes, heterozygotes and major allele homozygotes. Significance was conducted using the logrank test between minor allele and major allele homozygotes. SNPs with at least 1% frequency in cancers and more than 1 minor allele homozygous individual were considered for analysis. Logrank

p-values were corrected using the Benjamini-Hochberg method and only variants with FDR < 0.1 were considered.

References

1. Stratton, M. R., Campbell, P. J. & Andrew Futreal, P. The cancer genome. *Nature* vol. 458 719–724 (2009).
2. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* vol. 339 1546–1558 (2013).
3. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* vol. 100 57–70 (2000).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* vol. 144 646–674 (2011).
5. Borcoman, E., Nandikolla, A., Long, G., Goel, S. & Le Tourneau, C. Patterns of Response and Progression

to Immunotherapy. *American Society of Clinical Oncology Educational Book* 169–178 (2018) doi:10.1200/edbk_200643.

6. Miao, D. & Van Allen, E. M. Genomic determinants of cancer immunotherapy. *Current Opinion in Immunology* vol. 41 32–38 (2016).
7. Mangino, M., Roederer, M., Beddall, M. H., Nestle, F. O. & Spector, T. D. Innate and adaptive immune traits are differentially affected by genetic and environmental factors. *Nat. Commun.* **8**, 13850 (2017).
8. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013).
9. Shahamatdar, S. *et al.* Germline Features Associated with Immune Infiltration in Solid Tumors. *Cell Rep.* **30**, 2900–2908.e4 (2020).
10. Sayaman, R. W. *et al.* Germline genetic contribution to the immune landscape of cancer. doi:10.1101/2020.01.30.926527.
11. Khan, Z. *et al.* Polygenic risk for skin autoimmunity impacts immune checkpoint blockade in bladder cancer. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12288–12294 (2020).
12. Milanese, J.-S. *et al.* Germline variants associated with leukocyte genes predict tumor recurrence in breast cancer patients. *NPJ Precis Oncol* **3**, 28 (2019).
13. Kogan, D., Grabner, A., Yanucil, C., Faul, C. & Ulaganathan, V. K. STAT3-enhancing germline mutations contribute to tumor-extrinsic immune evasion. *J. Clin. Invest.* **128**, 1867–1872 (2018).
14. Marty, R. *et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* **171**, 1272–1283.e15 (2017).
15. Marty Pyke, R. *et al.* Evolutionary Pressure against MHC Class II Binding Cancer Mutations. *Cell* **175**, 1991 (2018).
16. Chowell, D. *et al.* Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nature Medicine* vol. 25 1715–1720 (2019).
17. Arce Vargas, F. *et al.* Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4

Antibodies. *Cancer Cell* **33**, 649–663.e4 (2018).

18. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* vol. 359 582–587 (2018).
19. Queirolo, P. *et al.* Association of CTLA-4 Gene Variants with Response to Therapy and Long-term Survival in Metastatic Melanoma Patients Treated with Ipilimumab: An Italian Melanoma Intergroup Study. *Front. Immunol.* **8**, 386 (2017).
20. Uccellini, L. *et al.* IRF5 gene polymorphisms in melanoma. *J. Transl. Med.* **10**, 170 (2012).
21. Bedognetti, D. *et al.* CXCR3/CCR5 pathways in metastatic melanoma patients treated with adoptive therapy and interleukin-2. *Br. J. Cancer* **109**, 2412–2423 (2013).
22. Ugurel, S. *et al.* Impact of the CCR5 gene polymorphism on the survival of metastatic melanoma patients receiving immunotherapy. *Cancer Immunol. Immunother.* **57**, 685–691 (2008).
23. Geeleher, P. *et al.* Cancer expression quantitative trait loci (eQTLs) can be determined from heterogeneous tumor gene expression data by modeling variation in tumor purity. *Genome Biol.* **19**, 130 (2018).
24. Lim, Y. W. *et al.* Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11701–E11710 (2018).
25. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* vol. 175 1701–1715.e16 (2018).
26. Beck, A. H. *et al.* The macrophage colony-stimulating factor 1 response signature in breast carcinoma. *Clin. Cancer Res.* **15**, 778–787 (2009).
27. Calabrò, A. *et al.* Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer. *Breast Cancer Res. Treat.* **116**, 69–77 (2009).
28. Teschendorff, A. E. *et al.* Improved prognostic classification of breast cancer defined by antagonistic activation patterns of immune response pathway modules. *BMC Cancer* **10**, 604 (2010).
29. Wolf, D. M., Lenburg, M. E., Yau, C., Boudreau, A. & van ’t Veer, L. J. Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. *PLoS One* **9**, e88309 (2014).

30. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, E7 (2004).

31. Castro, A. *et al.* Strength of immune selection in tumors varies with sex and age. *Nat. Commun.* **11**, 4128 (2020).

32. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **51**, 411–412 (2019).

33. Raychaudhuri, S. *et al.* Interplay of acetyltransferase EP300 and the proteasome system in regulating heat shock transcription factor 1. *Cell* **156**, 975–985 (2014).

34. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* vol. 81 559–575 (2007).

35. Consortium, T. G. & The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* vol. 369 1318–1330 (2020).

36. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

37. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).

38. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701–1715.e16 (2018).

39. Wang, H. *et al.* LINC00680 Promotes the Progression of Non-Small Cell Lung Cancer and Functions as a Sponge of miR-410-3p to Enhance HMGB1 Expression. *Onco. Targets. Ther.* **13**, 8183–8196 (2020).

40. Tang, W. *et al.* LINC00680 and TTN-AS1 Stabilized by EIF4A3 Promoted Malignant Biological Behaviors of Glioblastoma Cells. *Mol. Ther. Nucleic Acids* **19**, 905–921 (2020).

41. Schmidt, E. M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).

42. Shehata, M. *et al.* Human leukocyte antigen class I expression is an independent prognostic factor in advanced ovarian cancer resistant to first-line platinum chemotherapy. *Br. J. Cancer* **101**, 1321–1328

(2009).

43. Liu, J. *et al.* LILRB4, from the immune system to the disease target. *Am. J. Transl. Res.* **12**, 3149–3166 (2020).

44. Anami, Y. *et al.* LILRB4-targeting Antibody-Drug Conjugates for the Treatment of Acute Myeloid Leukemia. *Mol. Cancer Ther.* (2020) doi:10.1158/1535-7163.MCT-20-0407.

45. Hornigold, N. *et al.* Dysregulation at multiple points of the kynurenine pathway is a ubiquitous feature of renal cancer: implications for tumour immune evasion. *Br. J. Cancer* **123**, 137–147 (2020).

46. Jung, A. R., Jung, C.-H., Noh, J. K., Lee, Y. C. & Eun, Y.-G. Epithelial-mesenchymal transition gene signature is associated with prognosis and tumor microenvironment in head and neck squamous cell carcinoma. *Sci. Rep.* **10**, 3652 (2020).

47. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **168**, 542 (2017).

48. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).

49. Miao, D. *et al.* Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science* **359**, 801–806 (2018).

50. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).

51. Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934–949.e16 (2017).

52. Neoadjuvant PD-1 Blockade in Resectable Lung Cancer; Nivolumab and Ipilimumab in Advanced Melanoma; Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma; Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy; Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma; Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma; Nivolumab and Ipilimumab versus Ipilimumab in Untreated

Melanoma; Rapid Eradication of a Bulky Melanoma Mass with One Dose of Immunotherapy; Genetic Basis for Clinical Response to CTLA-4 Blockade; Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma; Nivolumab plus Ipilimumab in Advanced Melanoma; Safety and Tumor Responses with Lambrolizumab (Anti-PD-1) in Melanoma; Hepatotoxicity with Combination of Vemurafenib and Ipilimumab. *N. Engl. J. Med.* **379**, 2185 (2018).

53. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
54. Anagnostou, V. *et al.* Integrative Tumor and Immune Cell Multi-omic Analyses Predict Response to Immune Checkpoint Blockade in Melanoma. *Cell Rep Med* **1**, 100139 (2020).
55. Maia, M. C., Almeida, L., Bergerot, P. G., Dizman, N. & Pal, S. K. Relationship of tumor mutational burden (TMB) to immunotherapy response in metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology* vol. 36 662–662 (2018).
56. Kim, S. J. *et al.* Increased cathepsin S in Prdm1 dendritic cells alters the T cell repertoire and contributes to lupus. *Nat. Immunol.* **18**, 1016–1024 (2017).
57. Ondr, J. K. & Pham, C. T. N. Characterization of murine cathepsin W and its role in cell-mediated cytotoxicity. *J. Biol. Chem.* **279**, 27525–27533 (2004).
58. Zhang, Y. *et al.* The Cancer-Associated Genetic Variant Rs3903072 Modulates Immune Cells in the Tumor Microenvironment. *Front. Genet.* **10**, 754 (2019).
59. Pérez-Guijarro, E. *et al.* Multimodel preclinical platform predicts clinical response of melanoma to immunotherapy. *Nat. Med.* **26**, 781–791 (2020).
60. Chen, H.-M. *et al.* Blocking immunoinhibitory receptor LILRB2 reprograms tumor-associated myeloid cells and promotes antitumor immunity. *J. Clin. Invest.* **128**, 5647–5662 (2018).
61. Yeboah, M. *et al.* LILRB3 (ILT5) is a myeloid cell checkpoint that elicits profound immunomodulation. *JCI Insight* **5**, (2020).
62. Zhang, F. *et al.* TGF-β induces M2-like macrophage polarization via SNAIL-mediated suppression of a

pro-inflammatory phenotype. *Oncotarget* **7**, 52294–52306 (2016).

63. Hou, N. *et al.* A novel chronic stress-induced shift in the Th1 to Th2 response promotes colon cancer growth. *Biochem. Biophys. Res. Commun.* **439**, 471–476 (2013).
64. Zhou, Y. *et al.* Activation of NF-κB and p300/CBP potentiates cancer chemoimmunotherapy through induction of MHC-I antigen presentation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
65. Mahmud, Z. *et al.* Oncogenic EP300 can be targeted with inhibitors of aldo-keto reductases. *Biochem. Pharmacol.* **163**, 391–403 (2019).
66. Ring, A., Kaur, P. & Lang, J. E. EP300 knockdown reduces cancer stem cell phenotype, tumor growth and metastasis in triple negative breast cancer. *BMC Cancer* **20**, 1076 (2020).
67. Garcia-Carpizo, V. *et al.* CREBBP/EP300 bromodomains are critical to sustain the GATA1/MYC regulatory axis in proliferation. *Epigenetics Chromatin* **11**, 30 (2018).
68. Bi, Y. *et al.* EP300 as an oncogene correlates with poor prognosis in esophageal squamous carcinoma. *J. Cancer* **10**, 5413–5426 (2019).
69. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
70. Ayers, M. *et al.* IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
71. Prat, A. *et al.* Immune-Related Gene Expression Profiling After PD-1 Blockade in Non-Small Cell Lung Carcinoma, Head and Neck Squamous Cell Carcinoma, and Melanoma. *Cancer Res.* **77**, 3540–3550 (2017).
72. Daud, A. I. *et al.* Tumor immune profiling predicts response to anti-PD-1 therapy in human melanoma. *J. Clin. Invest.* **126**, 3447–3452 (2016).
73. Huang, A. C. *et al.* T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* **545**, 60–65 (2017).
74. Gao, J. *et al.* Loss of IFN-γ Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4

Therapy. *Cell* **167**, 397–404.e9 (2016).

75. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171**, 1259–1271.e11 (2017).
76. Sade-Feldman, M. *et al.* Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat. Commun.* **8**, 1136 (2017).
77. Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).
78. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
79. de Velasco, G. *et al.* Tumor Mutational Load and Immune Parameters across Metastatic Renal Cell Carcinoma Risk Groups. *Cancer Immunol Res* **4**, 820–822 (2016).
80. Hellmann, M. D. *et al.* Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *N. Engl. J. Med.* **378**, 2093–2104 (2018).
81. Miao, L. *et al.* PD-L1 and c-MET expression and survival in patients with small cell lung cancer. *Oncotarget* vol. 8 53978–53988 (2017).
82. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
83. Maranville, J. C. & Cox, N. J. Pharmacogenomic variants have larger effect sizes than genetic variants associated with other dichotomous complex traits. *Pharmacogenomics J.* **16**, 388–392 (2016).
84. Tin, A. *et al.* Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.* **51**, 1459–1474 (2019).
85. Carter, H. *et al.* Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. *Cancer Discov.* **7**, 410–423 (2017).
86. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).

87. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

88. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).

89. Bellenguez, C. *et al.* A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134–135 (2012).

90. Velez, D. O. *et al.* 3D collagen architecture induces a conserved migratory and transcriptional response linked to vasculogenic mimicry. *Nat. Commun.* **8**, 1651 (2017).

91. Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* **32**, 462–464 (2014).

92. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

93. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–272 (2012).

94. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology* vol. 199 3360–3368 (2017).

95. Jensen, K. K. *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**, 394–406 (2018).

96. Castro, A. *et al.* Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med. Genomics* **12**, 107 (2019).

97. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

98. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).

99. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of*

the 9th Python in Science Conference (2010) doi:10.25080/majora-92bf1922-011.

100. Zhang, Y.-B. *et al.* Genome-wide association study identifies multiple susceptibility loci for craniofacial microsomia. *Nat. Commun.* **7**, 10605 (2016).

101. Knight, J. *et al.* Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. *Hum. Mol. Genet.* **21**, 5185–5192 (2012).

102. Aguiar, V. R. C., César, J., Delaneau, O., Dermitzakis, E. T. & Meyer, D. Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. *PLoS Genet.* **15**, e1008091 (2019).

103. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

104. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).

105. Hall, M. A. *et al.* PLATO software provides analytic framework for investigating complexity beyond genome-wide association studies. *Nat. Commun.* **8**, 1167 (2017).

106. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).

107. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

108. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

109. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

110. Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* vol. 43 (2013).

111. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

112. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation

DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

113. Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).

114. Yun, T. *et al.* Accurate, scalable cohort variant calls using DeepVariant and GLnexus. doi:10.1101/2020.02.10.942086.

115. Seymour, L. *et al.* iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol.* **18**, e143–e152 (2017).

116. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019) doi:10.1145/3292500.3330701.

117. Edwards, J. A. & Edwards, R. A. Fastq-pair: efficient synchronization of paired-end fastq files. doi:10.1101/552885.

118. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

119. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* vol. 12 (2011).

120. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400–416.e11 (2018).