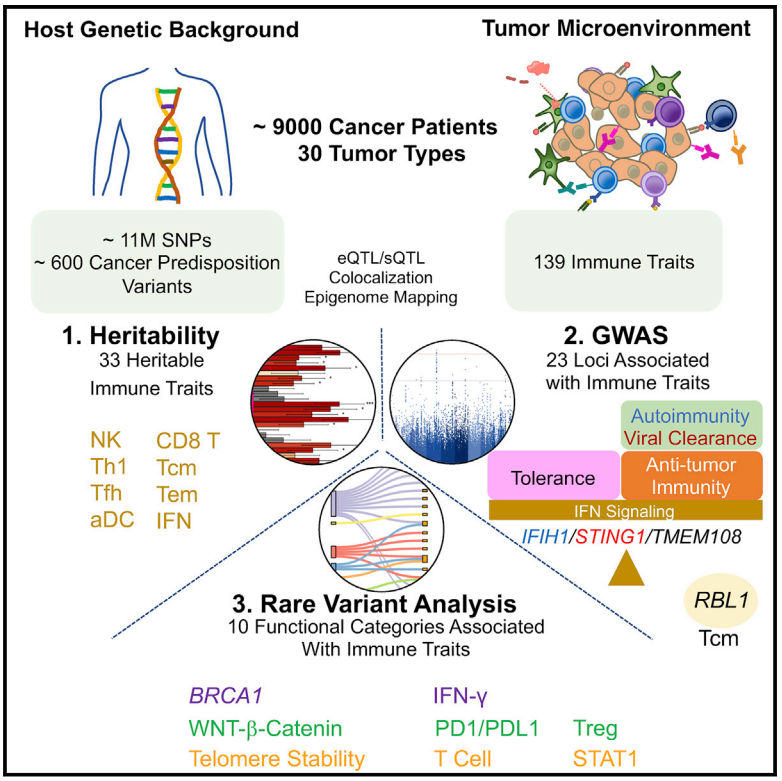
Immunity

Germline genetic contribution to the immune landscape of cancer

Graphical Abstract



Highlights

- 15–20% of intratumoral variation of interferon signaling and cytotoxic cells is heritable
- Common variants of *IFIH1*, *STING1*, and *TMEM108* affect cancer IFN signaling
- Common variants of *RBL1* are associated with differential T cell infiltration
- Rare cancer predisposition variants affect different immunomodulatory pathways

Authors

Rosalyn W. Sayaman, Mohamad Saad, Vésteinn Thorsson, ..., Jérôme Galon, Elad Ziv, Davide Bedognetti

Correspondence

rwsayaman@gmail.com (R.W.S.), elad.ziv@ucsf.edu (E.Z.), dbedognetti@sidra.org (D.B.)

In Brief

Sayaman, Saad et al. investigate the effect of common and rare germline variants on 139 well-defined immune traits in ~9000 cancer patients enrolled in TCGA, providing evidence for the impact of germline genetics on the composition and functional orientation of the tumor immune microenvironment.

Sayaman et al., 2021, Immunity *54*, 367–386
February 9, 2021
https://doi.org/10.1016/j.immuni.2021.01.011

CellPress

Immunity

CellPress

Resource

Germline genetic contribution to the immune landscape of cancer

Rosalyn W. Sayaman,[1,2,3,31,32,*] Mohamad Saad,[4,5,31] Vésteinn Thorsson,[6] Donglei Hu,[7] Wouter Hendrickx,[8,9] Jessica Roelands,[8,10] Eduard Porta-Pardo,[11,12] Younes Mokrab,[8,9,13] Farshad Farshidfar,[14,15,16,17] Tomas Kirchhoff,[18] Randy F. Sweis,[19] Oliver F. Bathe,[14,15,20] Carolina Heimann,[6] Michael J. Campbell,[21] Cynthia Stretch,[14,15] Scott Huntsman,[7] Rebecca E. Graff,[22] Najeeb Syed,[8,23] Laszlo Radvanyi,[24] Simon Shelley,[25] Denise Wolf,[2] Francesco M. Marincola,[26] Michele Ceccarelli,[27,28] Jérôme Galon,[29] Elad Ziv,[7,33,*] and Davide Bedognetti[8,9,30,33,*]

[1]Department of Population Sciences, Beckman Research Institute, City of Hope Comprehensive Cancer Center, Duarte, CA 91010, USA
[2]Department of Laboratory Medicine, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Biological Sciences and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[4]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[5]Neuroscience Research Center, Faculty of Medical Sciences, Lebanese University, Beirut, Lebanon
[6]Institute for Systems Biology, Seattle, WA 98109, USA
[7]Department of Medicine, Institute for Human Genetics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA
[8]Research Branch, Sidra Medicine, PO Box 26999 Doha, Qatar
[9]College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar
[10]Department of Surgery, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[11]Barcelona Supercomputing Center (BSC)
[12]Josep Carreras Leukaemia Research Institute (IJC), Badalona, 08034 Barcelona, Catalonia, Spain
[13]Weill Cornell Medicine, Doha, Qatar
[14]Department of Oncology, University of Calgary, Alberta AB T2N 4N1, Canada
[15]Arnie Charbonneau Cancer Institute, Calgary, Alberta AB T2N 4N1, Canada
[16]Department of Biomedical Data Science and Institute for Stem Cell Biology and Regenerative Medicine, School of Medicine, Stanford University, Stanford, CA 94305, USA
[17]Tenaya Therapeutics, South San Francisco, CA 94080, USA
[18]Perlmutter Cancer Center, New York University School of Medicine, New York University Langone Health, New York, NY 10016, USA
[19]Department of Medicine, Section of Hematology/Oncology, Committee on Clinical Pharmacology and Pharmacogenomics, Committee on Immunology, University of Chicago, Chicago, IL 60637, USA

(Affiliations continued on next page)

SUMMARY

Understanding the contribution of the host's genetic background to cancer immunity may lead to improved stratification for immunotherapy and to the identification of novel therapeutic targets. We investigated the effect of common and rare germline variants on 139 well-defined immune traits in ~9000 cancer patients enrolled in TCGA. High heritability was observed for estimates of NK cell and T cell subset infiltration and for interferon signaling. Common variants of *IFIH1, TMEM173 (STING1),* and *TMEM108* were associated with differential interferon signaling and variants mapping to *RBL1* correlated with T cell subset abundance. Pathogenic or likely pathogenic variants in *BRCA1* and in genes involved in telomere stabilization and Wnt-β-catenin also acted as immune modulators. Our findings provide evidence for the impact of germline genetics on the composition and functional orientation of the tumor immune microenvironment. The curated datasets, variants, and genes identified provide a resource toward further understanding of tumor-immune interactions.

INTRODUCTION

Immunotherapy with monoclonal antibodies that target immune inhibitory signaling (immune checkpoints) (Ishida et al., 1992; Leach et al., 1996) has emerged as the standard of care for many solid tumors, with an objective response rate up to ~40% in some cancer types (e.g., melanoma) (Chamoto et al., 2020; Sweis and Luke, 2017). However, overall, it has been estimated that fewer than ~15% of cancer patients might currently respond to such treatments (Haslam and Prasad, 2019). The density,

[20]Department of Surgery, University of Calgary, Calgary, Alberta AB T2N 4N1, Canada
[21]Department of Surgery, University of California, San Francisco, San Francisco, CA 94143, USA
[22]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94143, USA
[23]Department of Science and Technology, University of Sannio, 82100 Benevento, Italy
[24]Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada
[25]Department of Research and Development, Leukemia Therapeutics, LLC, Hull, MA 02045, USA
[26]Refuge Biotechnologies Inc., Menlo Park, CA 94025, USA
[27]Department of Electrical Engineering and Information Technology, University of Naples "Federico II," 80128 Naples, Italy
[28]Istituto di Ricerche Genetiche "G. Salvatore," Biogem s.c.ar.l., 83031 Ariano Irpino, Italy
[29]INSERM, Laboratory of Integrative Cancer Immunology, Equipe Labellisée Ligue Contre Le Cancer, Centre de Recherche de Cordeliers, Université de Paris, Sorbonne Université, Paris, France
[30]Department of Internal Medicine and Medical Specialties (Di.M.I.), University of Genoa, 16132 Genoa, Italy
[31]These authors contributed equally
[32]Lead contact
[33]Co-senior authors
*Correspondence: rwsayaman@gmail.com (R.W.S.), elad.ziv@ucsf.edu (E.Z.), dbedognetti@sidra.org (D.B.)
https://doi.org/10.1016/j.immuni.2021.01.011

location, and functional orientation of tumor infiltrating leukocytes have been associated with prognosis (Galon and Bruni, 2020), evolution of metastases in space and time (Angelova et al., 2018), and responsiveness to immunotherapy (Bruni et al., 2020), especially to checkpoint inhibition (Cristescu et al., 2018; Tumeh et al., 2014). Understanding the mechanisms underlying anti-tumor immune response will help stratify patients and may lead to the development of more effective therapies.

Cancer-cell intrinsic features such as somatic genetic alterations that activate specific oncogenic pathways (Bedognetti et al., 2016; Kalbasi and Ribas, 2020), the mutational load (Samstein et al., 2019; Snyder et al., 2014), the presence and degree of microsatellite instability (Mandal et al., 2019), and aneuploidy (Davoli et al., 2017) can differentially influence cancer immune responsiveness (Roelands et al., 2020; Rooney et al., 2015; Thorsson et al., 2018). Modifiable host factors such as the microbiome may also modulate the immune microenvironment and affect response to immunotherapy (Helmink et al., 2019; Iida et al., 2013). But host genetic factors have not been explored in depth as modulators of cancer immune responsiveness (Bedognetti et al., 2019; Havel et al., 2019).

Genome-wide association studies (GWAS) have identified germline variants that exert strong effects on circulating leukocyte counts (Keller et al., 2014) and fractions (Orrù et al., 2013), severity of immune-mediated tissue rejection in transplantation (Yang and Sarwal, 2017), and autoimmune diseases (Ye et al., 2018). Candidate gene studies in patients on immunotherapy have reported association between treatment responsiveness and immune-related genes such as *IRF5* (Uccellini et al., 2012) and *CCR5* (Bedognetti et al., 2013; Ugurel et al., 2008) in the pre-checkpoint inhibition era; and *CTLA4* (Queirolo et al., 2017), *HLA* (Chowell et al., 2018), Fc-gamma receptor (Arce Vargas et al., 2018), *IL2,* and *IL21* (Chat et al., 2019) in the context of CTLA-4 or PD-1 blockade therapy.

The Cancer Genome Atlas (TCGA) project has increased our understanding of cancer pathogenesis (Hutter and Zenklusen, 2018). A comprehensive analysis of immune signatures in TCGA identified features of the immune response that predict survival across many tumor types (Thorsson et al., 2018). Here, we used TCGA to perform a pan-cancer evaluation of the contribution of germline variation to anti-tumor immune response. First, we calculated genome-wide heritability of immune traits from common variants. Next, we performed GWAS for heritable immune traits to identify the loci with the strongest effects. Finally, we examined the contribution of rare germline variants in known cancer susceptibility genes to the cancer immune microenvironment.

Our findings provide evidence for the impact of the host's genetic background on the composition and functional orientation of the tumor immune microenvironment. The curated datasets herein generated and the list of variants and genes are intended to serve as a resource for future studies in the field of cancer germline immunogenetics and immunotherapy. Results are also presented through the CRI iAtlas portal for interactive exploration and visualization (https://www.cri-iatlas.org).

RESULTS

Overview of the discovery approach and description of the immune traits

To examine the contribution of germline genetic variation to the functional orientation of the immune microenvironment, we conducted heritability analysis, GWAS (N = 9,603), and rare variant analysis (N = 9,138) across 30 non-hematological cancer types characterized by the TCGA (Figure 1 top; Table S1**)**. Unless otherwise indicated, all analyses were adjusted for cancer type, age at diagnosis, sex, and the first seven components from principal component analysis (PCA) done on single nucleotide polymorphism (SNP) data, which largely capture genetic ancestry (see Methods). We considered 139 well-characterized immune traits estimated in the TCGA immune analysis (Thorsson et al., 2018) (Figure 1 bottom panel; Table S2). We divided the traits into six categories based on the approach used to derive them and the parameters they intend to measure: (1) leukocyte subset enrichment score (ES), which estimates leukocyte subset abundance within the tumor based on the coordinated regulation of lineage-specific genes using single sample gene set enrichment analysis (ssGSEA) (Bindea et al., 2013); (2) leukocyte subset proportion (%), which estimates the leukocyte subset proportion within infiltrating leukocytes using CIBERSORT deconvolution (Gentles et al., 2015); (3) overall proportion, which includes measures of leukocyte infiltration and stromal contents; (4) adaptive receptor, quantifying T cell receptor (TCR) and B cell receptor (BCR) diversity (TCR and BCR Shannon entropy and

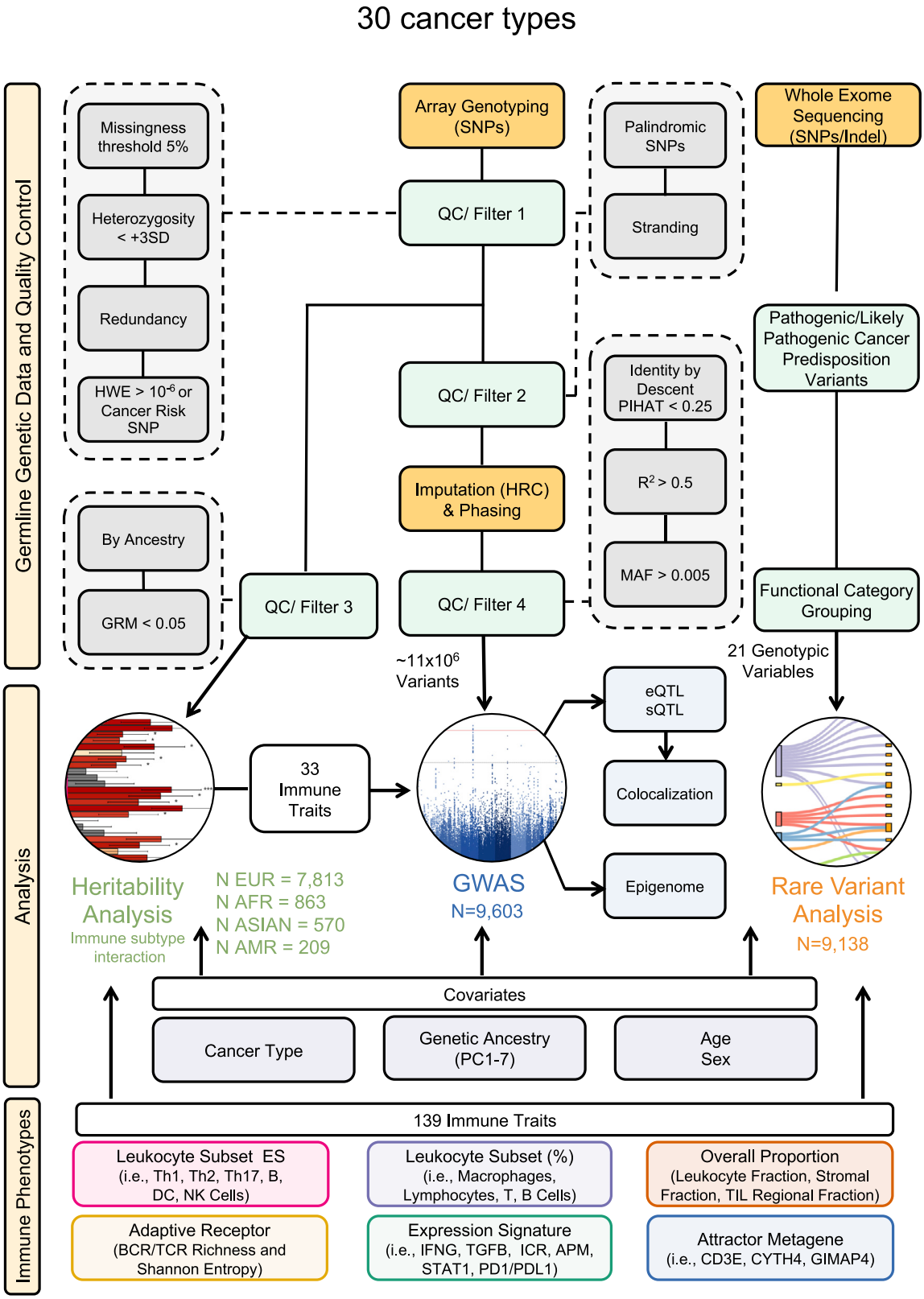


Figure 1. Overview of the discovery approach

Flowchart showing analytic workflows, source of genetic germline data, quality control filtering, and immune traits used in the analysis.
See also Figure S1, Tables S1 and S2.

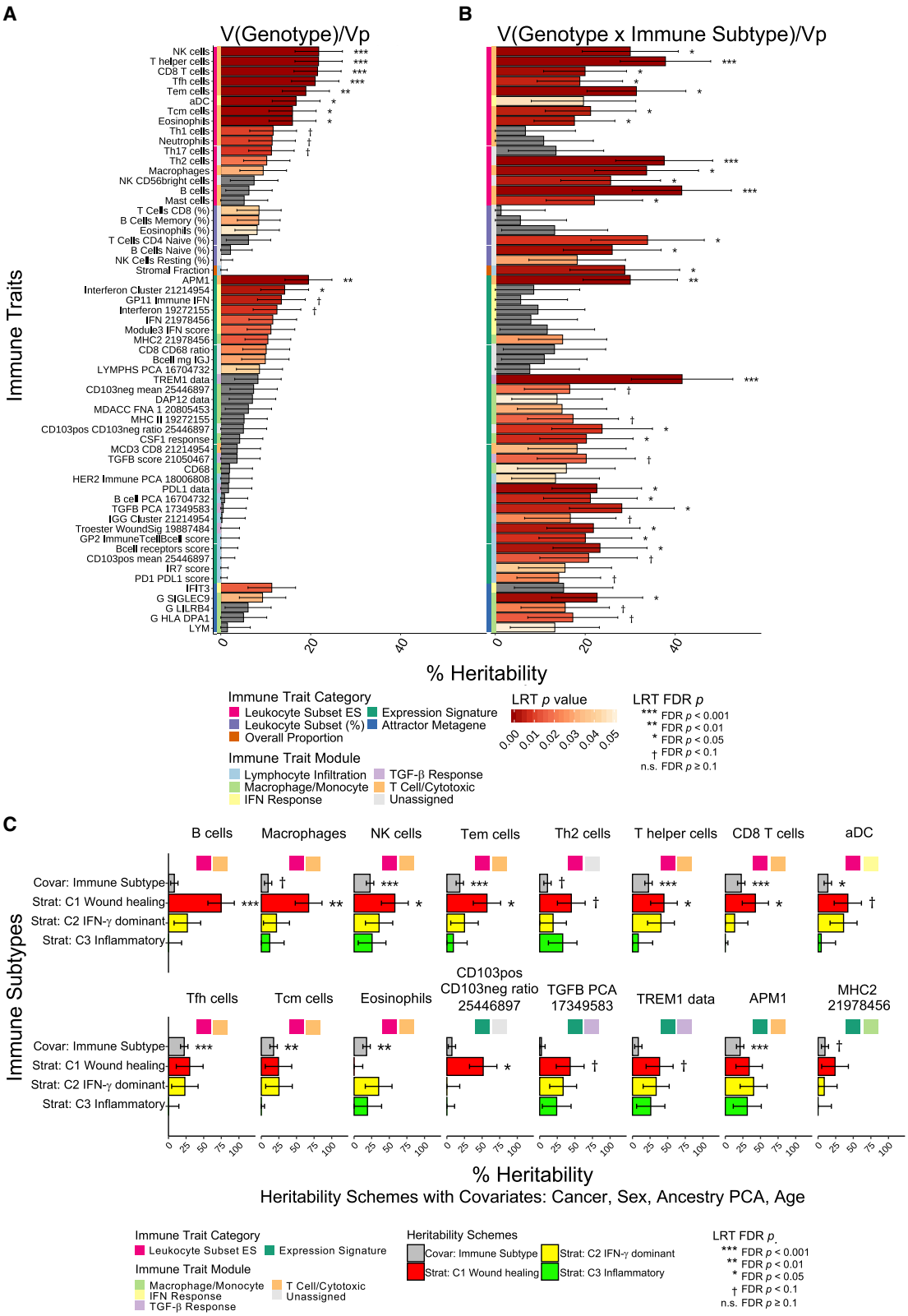


Figure 2. Genome-wide heritability of immune traits

GCTA GREML estimates of the percentage of phenotypic variance explained by all common variants. Error bars represent the standard errors and all *p* values are derived from likelihood ratio tests (LRT).

(legend continued on next page)

richness); (5) expression signature, consisting of a collection of annotated functional signatures summarizing different immune-related biological processes (e.g., wound healing, interferon (IFN) and tumor growth factor beta (TGF-β) signaling, antigen-presenting machinery etc.), and (6) attractor metagene, which includes co-expression signatures (metagene attractors) derived from the whole TCGA dataset (Cheng et al., 2013a, 2013b). All of the immune traits were derived from RNA-seq, with the exception of overall proportion estimates (which include hematoxylin and eosin [H&E] tissue imaging and DNA methylation array data).

The 139 traits were clustered based on their Pearson correlation coefficients, and six groups of correlated traits were defined and referred to here as "modules" (Figure S1 and Table S2). The first five modules recapitulate the immune modules defined in Thorsson et al. (Thorsson et al., 2018) with highly concordant module membership. The largest group included traits that were highly correlated with leukocyte fraction and lymphocyte infiltration estimates (lymphocyte infiltration module). Traits capturing monocyte function and macrophages infiltration, and MHC-related traits formed a second module (monocyte/macrophage module). Traits capturing IFN signaling were highly correlated and formed a third module (IFN response module). The next two modules included traits associated with TGF-β signaling (TGF-β response module) and wound healing (wound healing module), respectively. The last and previously undescribed module mostly included leukocyte subset ES that were excluded from the clustering analysis in Thorsson et al., such as T helper, CD8 cytotoxic, and natural killer (NK) cells (T cell/cytotoxic module) (Thorsson et al., 2018). Although most traits fit within these modules, a subset (N = 42) did not cluster within any module. In addition, there was substantial correlation between traits across modules, particularly between the lymphocyte infiltration and monocyte/macrophage modules, and between these and the IFN response modules. A subset of traits clustering within the T cell/cytotoxic module were significantly correlated with the IFN response, lymphocyte infiltration, and monocyte/macrophage modules.

GWAS were performed on traits that we found to have significant heritability since these would be most likely driven by common genetic variants. For rare variant association tests, we examined all 139 traits, and focused on well-annotated pathogenic or likely pathogenic variants within high penetrance cancer susceptibility genes (Huang et al., 2018).

Genome-wide heritability of immune traits

We performed heritability analysis on 139 traits using a mixed-model approach implemented in genome-wide complex trait analysis (GCTA) genomic-relatedness-based restricted maximum-likelihood (GREML) method (Yang et al., 2010, 2011) to calculate the proportion of immune trait variation that is attributable to common genetic variants (Zaitlen and Kraft, 2012). Heritability analyses were conducted separately within each ancestry subgroup ($N_{European}$ = 7,813, $N_{African}$ = 863, N_{Asian} = 570, and $N_{American}$ = 209 individuals), which were derived from ancestry analysis (Carrot-Zhang et al., 2020) (Figures S2A and S2B).

In the pan-cancer analysis, we found 10 immune traits with significant heritability after correction for multiple hypothesis testing (false discovery rate (FDR) $p < 0.05$) and 23 other traits with nominally significant heritability ($p < 0.05$) in at least one ancestry group (Figure 2A; Figure S2C and Table S3). Within the European ancestry group, 28 traits had at least nominally significant heritability. The most heritable traits, with ~15%–20% heritability and FDR $p < 0.05$, were members of the T cell/cytotoxic module (Figure S1A) and represent T cell subsets estimated using ES including CD8 T cells, T helper cells, T follicular helper (Tfh) cells, T effector memory (Tem) cells, T central memory (Tcm) cells, NK cells, and eosinophils (Figure 2A). An Antigen Presenting Machinery signature (APM1) clustered within the T cell/cytotoxic module (Figure S1A) and was also ~20% heritable (FDR $p < 0.05$). T helper 1 (Th1) cell ES, which was also part of this module (Figure S1), showed lower but nominally significant heritability (Figure 2A). The second most heritable traits, with ~15% heritability, included highly correlated IFN-related signatures (Interferon Cluster 21214954, GP11 Immune IFN, Interferon 19272155, IFN 21978456, Module3 IFN score, and IFIT3), and activated dendritic cells (aDC) ES, which cluster within the IFN response module (Figure S1A). Among these, Interferon Cluster 21214954 and aDC ES were significant after correction for multiple hypothesis testing (FDR $p < 0.05$). In addition, we detected nominally significant heritability for T helper 2 (Th2) and T helper 17 (Th17) cell ES, and proportions of T cells CD8+, memory B cells, and eosinophils within leukocytes (Leukocyte Subset %). The CD8+ T cell/CD68+ ratio, a B cell metagene score (B cell mg IGJ), macrophage ES, neutrophil ES, and other signatures belonging to the macrophage/monocyte module, including major histocompatibility complex class-II (MHC2 21978456) and siglec-regulation (G SIGLEC9), an attractor metagene, showed nominally significant heritability.

Despite the limited cohort size, specific immune traits were also heritable in the African and Asian ancestry groups at a nominal significance (Figure S2C). In the African ancestry group, NK CD56dim and cytotoxic cell ES, single gene immune therapy target *PDCD1* expression (PD1 data), and TGF-β immunomodulatory signaling (TGFB PCA 17349583), showed nominally significant heritability. In the Asian ancestry group, Tcm cell ES and

(A) Twenty-eight of 139 immune traits analyzed in the European ancestry group (N = 7,813) showed nominally significant level of genome-wide heritability (V(Genotype)/Vp) (LRT $p < 0.05$), 10 traits (FDR $p < 0.05$) and 15 traits (FDR $p < 0.1$) showed significant heritability after correction for multiple hypothesis testing. (B) Percentage of variance of immune traits accounted for by the interaction between germline genotypes and immune subtypes (V(GxImmune Subtype)/Vp). In the subset of individuals with immune subtype information (n = 6,586), 44 immune traits showed nominally significant heritability of interaction effects ($p < 0.05$), and 26 traits showed significant heritability of interaction effects (FDR $p < 0.05$). Heritability estimates with standard errors are shown in (A) and (B) for each of the 59 immune traits identified. Immune trait categories and corresponding immune trait modules are annotated.

(C) Immune subtype-specific heritability analysis conducted for immune traits with significant G x Immune Subtype interaction. Heritability was calculated in three of the six immune subtype groups with sufficient cohort size: C1 (N = 1,752), C2 (N = 1,813), and C3 (N = 1,737), as well as with immune subtype as an additional covariate. Stratified analysis of the 44 traits with at least nominally significant G x Immune Subtype interaction effects showed 16 traits with significant V(G)/Vp heritability in at least one of the immune subtypes or with immune subtype as a covariate (FDR $p < 0.1$). GREML analyses were performed with covariates as described in Methods.

See also Figure S2 and Table S3.

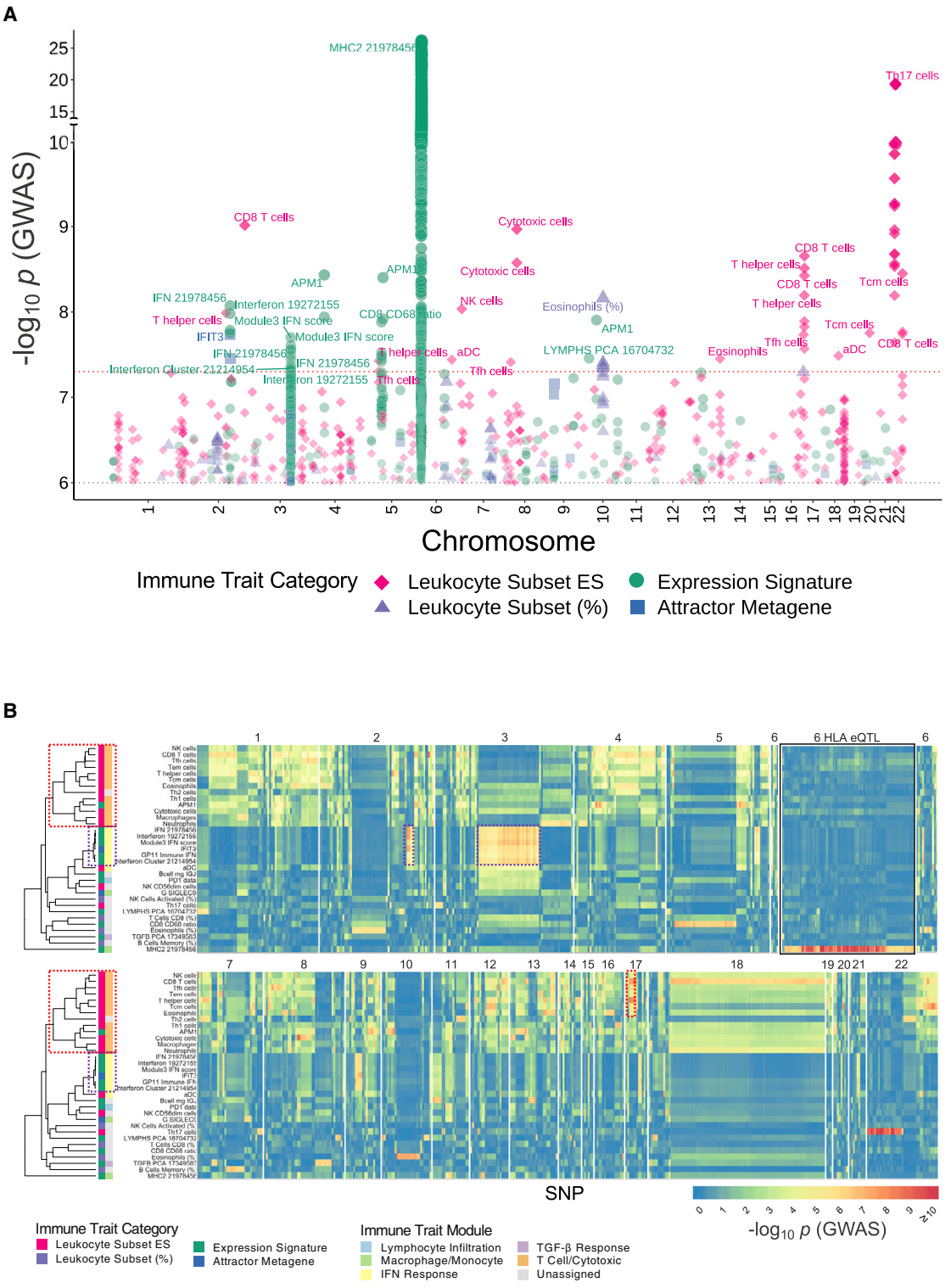


Figure 3. Genome-wide associations for variants affecting immune traits

GWAS performed on the 33 immune traits with genome-wide heritability in the ancestry clusters identify 23 loci with 598 genome-wide significant associations between single SNPs and immune disposition in 10 immune traits ($p < 5x10^{-8}$), and an additional 1,196 suggestive associations in 33 traits ($p < 1x10^{-6}$). (A) Combined Manhattan plot representing -$\log_{10}$ p of the significant and suggestive GWAS hits by chromosomal position across the 33 immune traits encompassing four phenotypic categories.

(legend continued on next page)

the proportion of NK cells activated (%) showed nominally significant heritability.

Variation of heritability of immune traits across immune subtypes

Since there was considerable heterogeneity between tumor types, we investigated whether heritability varies among previously defined tumor immune subtypes (Thorsson et al., 2018). These six distinct immune subtypes include: wound healing (C1), IFN-γ dominant (C2), inflammatory (C3), lymphocyte depleted (C4), immunologically quiet (C5), and TGF-β dominant (C6) (Figure S1B). These immune subtypes formed larger subsets than any of the individual tumor types, facilitating heritability analyses that require large numbers (Visscher et al., 2014). We performed statistical interaction analysis using GREML in the European ancestry group (Figure 2B). We found 26 immune traits with significant variance of genotype-immune subtype interaction effects at FDR $p < 0.05$ and 18 other nominally significant traits ($p < 0.05$). These interactions suggest that the contribution of genotype to immune traits differs among immune subtypes. We found substantial overlap between immune traits with nominally significant heritability and those with interaction effects. Eight of the heritable traits from the T cell/cytotoxic module had significant interactions with immune subtype including: ES of CD8 T cells, T helper cells, Tfh cells, Tem cells, Tcm cells, NK cells, eosinophils, and APM1 expression signature. In addition, ES of NK CD56 bright cells, B cells, and mast cells showed significant interaction effects of genotype-immune subtype. The majority of immune traits with significant and nominally significant genotype-immune subtype interaction belonged to the T cell/cytotoxic, macrophage/monocyte, overall lymphocyte infiltration, and TGF-β response modules. On the other hand, traits that were part of the IFN response module showed no genotype-immune subtype interaction, suggesting that the genotypic contribution to these traits was independent of immune subtypes.

To understand how immune subtype influences heritability of the 44 immune traits with at least nominally significant genotype-immune subtype interaction effects, we performed heritability analyses stratified by immune subtype. We performed these only in subtypes with sufficient sample size: C1 (N_{C1} = 1,752), C2 (N_{C2} = 1813), and C3 (N_{C3} = 1737). We found significant heritability estimates (FDR from $p < 0.1$ to $p < 0.001$ significance levels) largely within the C1 immune subtype, but not within the more immune-active subtypes, C2 and C3 (Figure 2C).

Genome-wide association for variants affecting immune traits

We selected the 33 immune traits with nominally significant heritability ($p < 0.05$) in at least one ancestry group to perform GWAS, and tested the association between each of these traits and 10,955,441 variants that passed the quality and frequency thresholds. We identified 598 genome-wide significant ($p < 5x10^{-8}$) associations at 23 loci for 10 immune traits. We also identified an additional 1,196 suggestive ($p < 1x10^{-6}$) associations for 33 traits (Figure 3A; Table S4). Summary statistics for all GWAS analyses are available on Figshare (https://doi.org/10.6084/m9.figshare.13077920).

Two of the 23 loci with the strongest associations ($p < 1x10^{-10}$ to $p < 1x10^{-25}$) included SNPs that map within ± 50 KB (or 1 MB in the case of *HLA*) of the genes that comprise the signature of the associated immune trait. These included SNPs at the *HLA* locus, which are associated with the MHC2 expression signature, and SNPs at the *IL17RA* locus, which makes up the Th17 cell ES. We concluded that these SNP associations represented simple expression quantitative trait loci (eQTLs), and we did not consider them further.

In contrast, the remaining 21 loci were not proximal to genes comprising the associated signatures; therefore, they likely represent SNPs affecting the overall immune trait. The majority of these 21 loci are associated with leukocyte subset enrichment and IFN signaling. At these 21 loci, we found 59 genome-wide significant associations with 17 traits, 45 of which were represented by unique SNPs, and 10 of which had significant association with at least two traits. Excluding the *HLA* and *IL17RA* loci, we identified 841 suggestive associations, represented by 667 unique SNPs, 70 of which had multiple suggestive hits in at least two traits, suggesting pleiotropic associations of significant SNPs (Table S4).

To examine pleiotropy, we clustered all 33 immune traits (Figure S3A) based on the association p values of all significant and suggestive SNPs found to be associated with at least one trait (Figure S3B). The results generally recapitulated clustering based on the correlation of their phenotypic values, with traits with common associations tending to have similar overall expression. (Figures S3A and S3B). To understand pleiotropic effects at individual loci, we visualized the strength of each association between each immune trait and each significant and suggestive SNPs across the genome (Figure 3B). SNPs that were associated with one of the T cell ES tended to be associated with multiple other T cell subsets. For example, significant and suggestive SNPs in chromosome (chr) 17 associated with CD8 T and T helper cell ES that were part of the T cell/cytotoxic-dominant cluster (Figure 3B, top cluster highlighted in red) showed nominally significant associations ($1x10^{-6} < p < 1x10^{-4}$) across a number of other traits within this cluster, and largely no associations across other traits. Similarly, SNPs associated with one of the traits from the IFN response module were usually associated with the other traits within the IFN-dominant cluster, but largely distinct from traits from other clusters. For example, SNPs at chr 2 were associated with traits in the IFN-dominant cluster, but not other traits (Figure 3B, second cluster highlighted in purple). However, a few of the top loci were associated with traits from both the IFN and T cell/cytotoxic dominant clusters. For example, significant SNPs in chr 3 associated with

(B) Heatmap demonstrating pleiotropy of the top associations across 33 immune traits. The heatmap is facetted by chr and shows the GWAS -$\log_{10}$ p of the significant and suggestive SNPs across each of the 33 immune traits. Immune traits are clustered based on the Pearson correlation of the GWAS -$\log_{10}$ p, and immune trait categories and modules are annotated. The SNPs at the *HLA* loci are shown condensed. GWAS were performed by regression analyses using covariates as described in Methods. The T cell/cytotoxic-dominant cluster and an example locus on chr 17 are indicated by dotted lines in red; the IFN-dominant cluster and example loci on chr 2 and chr 3 are indicated by dotted lines in purple.
See also Figures S3, S4, and Table S4.

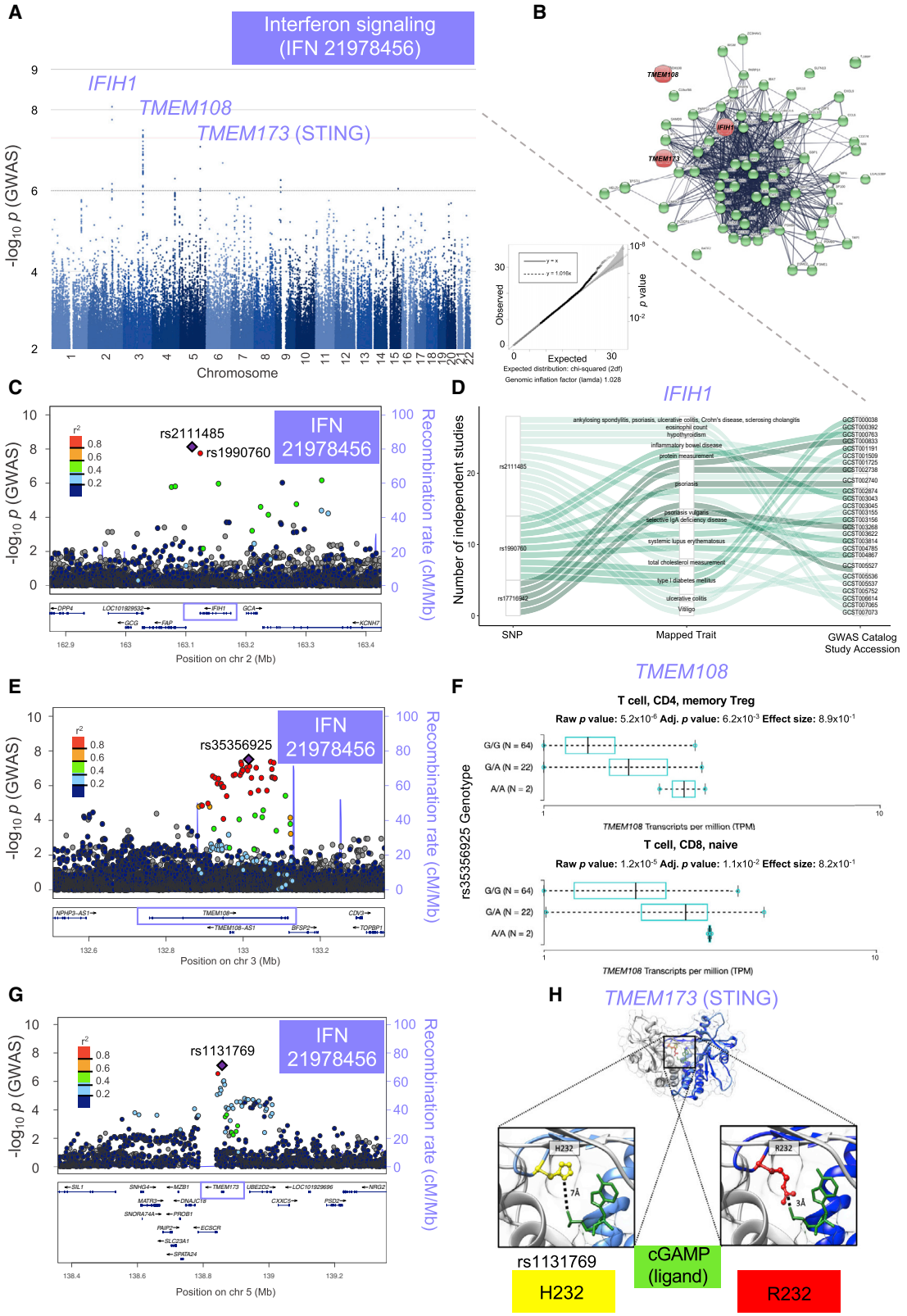


Figure 4. Genetic variants and candidate genes associated with differential IFN signaling

GWAS identified 17 associations between seven SNPs and five IFN signatures reaching genome-wide significance ($p < 5x10^{-8}$), and additional 152 suggestive associations between 29 SNPs and 6 IFN expression signatures ($p < 1x10^{-6}$).

(legend continued on next page)

IFN immune traits also showed nominally significant associations with eosinophil, Th1 cell, and cytotoxic T cell ES within the T cell/cytotoxic module. Finally, the loci for MHC2 (driven by *HLA* expression) on chr 6 and Th17 ES (driven by *IL17RA* expression) on chr 22 showed largely no association outside of their respective associated traits, consistent with being simple eQTLs.

To understand if significant and suggestive SNPs are associated with specific epigenetic regulatory regions, we mapped the SNPs to annotated chromatin states using an expanded model of 18 chromatin states (Roadmap Epigenomics Consortium et al., 2015), which takes into account six chromatin marks across 98 epigenomes, including 25 that are specifically immune related (Figure S4A and Table S4). Excluding the *HLA* and *IL17RA* loci, we found a substantial fraction of the genome-wide significant loci to have at least one SNP that maps to a potential regulatory site across multiple epigenomes including: weak transcription chromatin active states, weak enhancer active states, zinc finger (ZNF) genes/repeats active and heterochromatin inactive states, and weak repressed polycomb states (Figure S4B). These findings suggest that many of the loci we identified may act directly on gene transcription or through alteration of gene regulatory regions, including at distant enhancer regions, across multiple epigenomes including immune-related ones.

Genetic variants and candidate genes associated with IFN signaling

We found two loci associated with IFN signaling traits (IFN response module), one on chr 2 and another on chr 3, and a third locus very close to the genome-wide significant cut-off on chr 5 (Figure 4A; Table S4). These loci map to *IFIH1*, *TMEM108*, and *TMEM173*, respectively (Figures 4C, 4E, and 4G). The direction of the effect of these SNPs on the IFN traits was consistent across the majority of cancer types (Figures S5A–S5C).

Locus 2 (chr 2) is represented by SNPs rs2111485 and rs1990760, which were both significantly associated with three of the IFN traits and had suggestive associations with two other IFN traits (Table S4). These two SNPs are in high linkage disequilibrium (LD) with each other ($r^2 > 0.8$) (Figure 4C) and map to the *IFIH1* gene, which is induced by IFN and acts as an RNA-dependent ATPase (Kang et al., 2002; Yoneyama et al., 2005). Rs2111485 and rs1990760 have been reported to be associated with a number of autoimmune diseases such as psoriasis, vitiligo, systemic lupus erythematosus, ulcerative colitis, Crohn's disease, and type I diabetes mellitus (T1DM) in the GWAS catalog (Figure 4D) (Buniello et al., 2019). Rs1990760 results in an amino-acid change (A946T). The 946T allele, which was associated with higher risk of T1DM and vitiligo but decreased risk of inflammatory bowel disease and psoriasis was associated with higher IFN trait values in our analyses. This allele has been associated with higher basal and inducible production of type I IFNs in human peripheral blood mononuclear cells and enhanced antiviral response in transgenic mice (Gorman et al., 2017; Rice et al., 2014). Our analyses identified at least one additional suggestive ($p < 10^{-6}$) association at this locus characterized by rs17716942, a SNP which is not in LD with the lead SNPs ($r^2 < 0.1$) and has been associated with psoriasis (Tsoi et al., 2012).

Locus 4 (chr 3) included six genome-wide significant SNPs associated with four of the six IFN signaling traits (Figure 4A; Figure S5D and Table S4). These SNPs were nominally associated with several other cellular signatures including one for aDC and one for the fraction of NK cells (Figure 3A). These SNPs map to the *TMEM108* gene, which overlaps with *TMEM108-AS1* Antisense RNA1 gene, and are in high LD with each other ($r^2 > 0.8$) (Figure 4E). EQTL and splice quantitative trait loci (sQTL) analysis followed by colocalization with eCAVIAR demonstrate that *TMEM108, CDV3*, and *RP11-91K8.5* all plausibly colocalize in the Genotype-Tissue Expression (GTEx) data with traits of the IFN response module (Table S5). In the expanded region analyses (within ± 1MB for eQTL and ± 500 KB for the sQTL, see Methods) we did not find counter-evidence for colocalization for any of these genes (i.e., no SNPs with higher eQTL or sQTL and lower GWAS signal; expanded region plots are available at https:/figshare.com see Methods). Therefore, we considered the evidence of colocalization as strong. Of these three genes, we found that *TMEM108* (colocalization shown in Figure S5D and S5E) is also significantly associated with the top SNPs in the Dataset of Immune Cell Expression (DICE) (Schmiedel et al., 2018), particularly CD8 T cell and regulatory T cell (Treg cell) subsets (Figure 4F), suggesting it is the most likely causal gene at this locus. The alleles associated with increased expression of IFN response module traits in tumors are associated with decreased *TMEM108* expression (Table S5), implying that it may

(A) Manhattan plot (left) of GWAS -$\log_{10} p$ for a representative IFN signature, IFN 21978456, shows three main peaks on chr 2, 3, and 5. The related QQ plot (right) shows deviation of the observed p values from the expected distribution from a theoretical χ^2 distribution. Genomic inflation factor, lambda, is calculated. The red dotted line represents the threshold of genome-wide significance ($p < 5\text{x}10^{-8}$).
(B) Protein-protein interaction network (String-db minimum interaction score confidence ≥ 0.7, PPI enrichment $p < 1.0\text{x}10^{-16}$) between the 69 genes comprising the 6 IFN-related signatures, and *IFIH1*, *TMEM108 and TMEM173*. (C) LocusZoom plot of the association results maps the genomic location of the two significant and two suggestive SNPs on chr 2 to the *IFIH1-GCA-KCNH7* locus, with the two most significant SNPs (rs2111485 and rs1990760) annotated. All LocusZoom plots show linkage disequilibrium estimates (r^2 color map) and recombination rates (blue line) around the genome-wide significant loci.
(D) Three of the SNPs in the *IFIH1-GCA-KCNH7* locus were mapped to 13 traits which are predominantly autoimmune-related in 26 independent studies in the GWAS Catalog (Buniello et al., 2019).
(E) LocusZoom plot maps the genomic location of significant and suggestive SNPs on chr 3 to the *TMEM108* locus, with the most significant SNP (rs35356925) annotated.
(F) Boxplots of *TMEM108* gene expression according to rs35356925 genotype in two T cell subsets, retrieved from DICE (Schmiedel et al., 2018);p values were retrieved from DICE and were derived from regression models and permutations.
(G) LocusZoom plot maps the genomic location of the two suggestive SNPs on chr 5 to the *TMEM173* locus, with the most significant SNP (rs1131769) annotated.
(H) Overlay of the crystal structure of the R232 monomer (dark blue and red) with one of the chains of the rs1131769 H232 monomer (cyan and yellow). The R232 allele (red) interacts much more directly with the cGAMP ligand (green) than the H232 allele (yellow). In the zoom-in panel, part of the protein is removed in order to make the ligand more visible.
See also Figure S5, Tables S4, and S5.

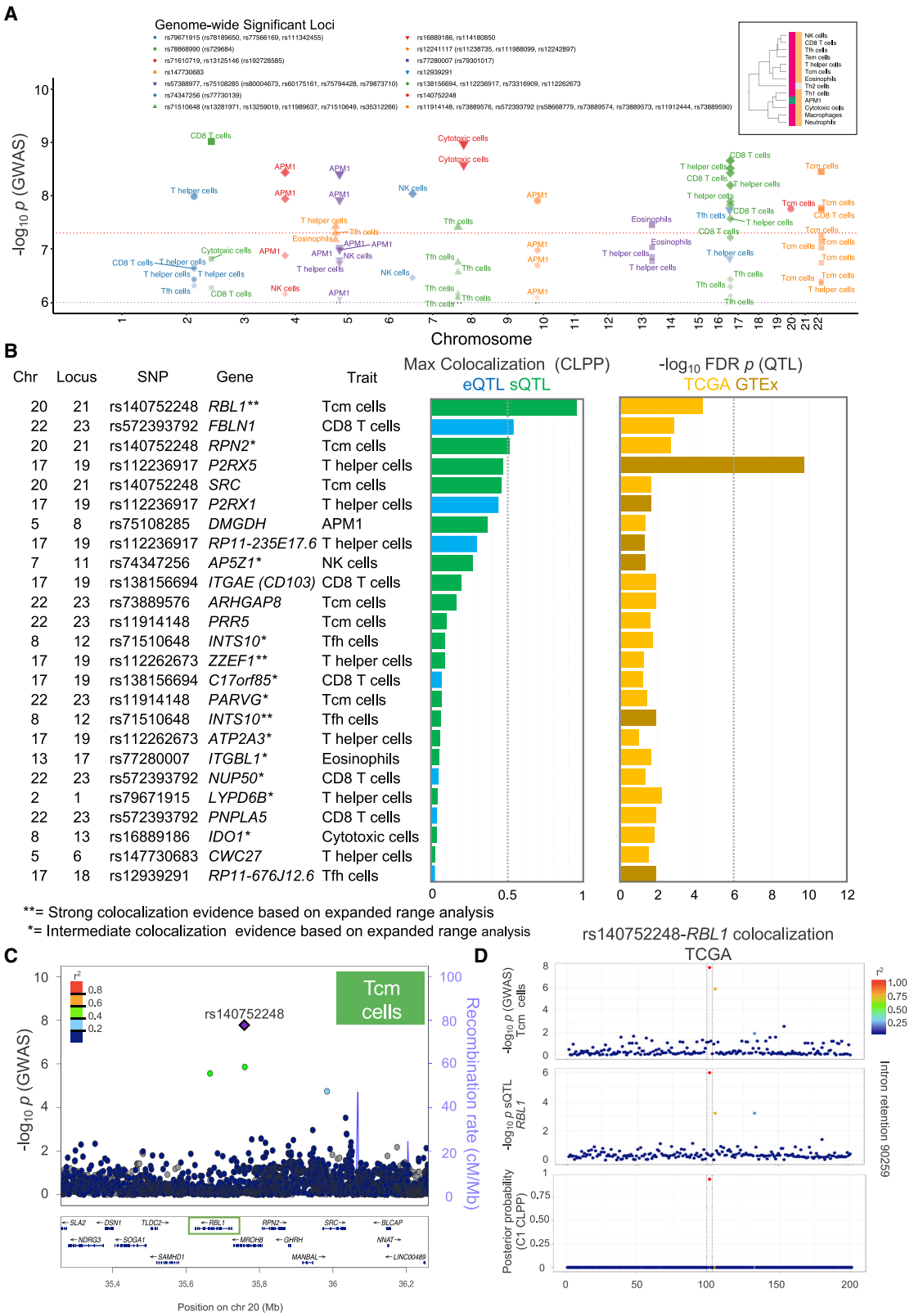


Figure 5. Genetic variants and candidate genes associated with T cell subset ES

Clustering of GWAS -log$_{10}$*p* identified highly correlated associations of SNPs across 13 immune traits predominantly associated with T cell subset ES in the T cell/cytotoxic module (inset).

(legend continued on next page)

negatively regulate IFN signaling. No other GWAS significant SNPs were associated with eQTL in DICE.

We also identified a third locus including a SNP with nearly genome-wide significance ($p = 8.27x10^{-8}$) on chr 5 (Figure 4G). This SNP (rs1131769) is a missense variant in *TMEM173*, causing an Arginine to Histidine substitution at position 232 of the STING protein, a major mediator of the innate response against virus and cancer cells (Flood et al., 2019; Patel and Jin, 2019). The minor allele T (H232) was associated with lower IFN signaling values, as compared to the major allele C (R232) (Table S4). The H232 allele results in reduced IFN-α (Kennedy et al., 2020) and IFN-β (Zhang et al., 2013) production following stimuli. The causal association between this variant and impaired IFN response has been corroborated by molecular simulations, suggesting that the ligand-binding loops were more rigid for R232 compared to H232 (Kennedy et al., 2020). We performed an *in-silico* evaluation of the effect of this amino acid substitution on the protein by superimposing H232 and R232 STING structures and found that H232 interacts less closely with its ligand cGAMP (Figure 4H), in agreement with its lower activity.

Based on known protein-protein interactions (PPI), as annotated in String database (minimum interaction score confidence ≥ 0.7), we found that *IFIH1* and *TMEM173* formed a highly connected PPI network with the majority of genes that define the traits in the IFN response module (Figure 4B). In contrast, *TMEM108* showed no direct PPI with the IFN-associated genes, suggesting an indirect immune modulatory mechanism.

Genetic variants and candidate genes associated with differential immune infiltration

Fourteen of the genome-wide significant loci were associated with eight distinct traits derived from ES that cluster within the T cell/cytotoxic module (Figure 3B**,** red dotted boxes, and Figure 5A inset**)**. We show each of the genome-wide significant and suggestive SNPs mapping to these 14 loci in a combined Manhattan plot (Figure 5A; Table S4).

Comprehensive eQTL and sQTL analyses and colocalization revealed several candidate causal genes at these loci (Figure 5B). The strongest evidence for colocalization was with locus 21 (chr 20), which was associated with Tcm cell ES and splicing of *RBL1* (eCaviar colocalization posterior probability (CLPP) = 0.95) (Figure 5B; Figures S6A, S6B, and Table S5). The top SNP at this locus (rs140752248) mapped to the region 5′ of *RBL1* and two additional SNPs in moderate LD with the top SNP mapped to introns of *RBL1* (Figure 5C). This sQTL also displayed consistent evidence for expanded region colocalization as there was no evidence for a secondary signal for splicing separate from the signal from the association with Tcm cell ES in the expanded range analyses (Figure 5D; Table S5). The association of this locus with Tcm cell ES was fairly consistent across cancer subsets (Figure S6B). *RBL1* has homology with the tumor suppressor *RB* (Ng et al., 2020), which is also involved in immune regulatory functions (Garfin et al., 2013; Jerby-Arnon et al., 2018). We also found weaker evidence for colocalization at this locus with sQTLs for *RPN2* and *SRC*, but both of these results also had stronger secondary associations with the splicing event, with *RPN2* having intermediate evidence and SRC having negative evidence in the expanded region colocalization. We found strong evidence for colocalization of a SNP on chr 8 (locus 12, rs71510648) associated with Tfh cell ES and with an sQTL of *INTS10* in GTEx and in TCGA (Figure 5B; Figures S6C–S6F and Table S5). Examining the expanded region evidence for colocalization, we found strong evidence in GTEx and intermediate evidence in TCGA (Figure 5B). *INTS10* has been identified in a GWAS of persistent hepatitis B infection and was found to mediate its effect via *IRF3*, an IFN regulatory transcription factor (Li et al., 2016).

At locus 19 on chr 17, we identified multiple genome-wide significant and suggestive associations with three immune traits associated with T cell subsets: rs112236917, rs112262673, rs73316909, and rs138156694 were significantly associated with T helper cell ES; rs112236917, rs73316909, and rs138156694 had significant associations with CD8 T cell ES; and rs112236917, rs73316909, and rs138156694 had suggestive associations with Tfh cell ES (Figure 5A;Table S4). Colocalization of eQTL and sQTL data in GTEx identified an eQTL of *P2RX1* and sQTLs of *P2RX5* and *RP11-235E17.6* as potential candidates (Figure 5B; Table S5). Of these, *P2RX1* and *P2RX5,* purigenic receptors which function as ATP-gated ion channels, are known to have an effect on multiple cells in the immune system, including lymphocytes (Junger, 2011). Colocalization in TCGA data identified *ATP2A3, C17orf85, ITGAE,* and *ZZEF1* as additional candidate genes (Figure 5B; Table S5). Of these, *ITGAE* (also known as CD103) is known to have a function in T cell homing to epithelial tissue and tumors (Kim et al., 2019). However, only rs112262673-*ZZEF1* splicing met our criteria for evidence of colocalization in the expanded range analyses (see Methods; Table S5). *ZZEF1* (ZNF ZZ-Type And EF-Hand Domain

(A) Combined Manhattan plot showing 14 distinct loci and 26 genome-wide significant associations ($p < 5x10^{-8}$) between 22 SNPs and eight immune traits within the T cell subset-dominant cluster: T helper, CD8 T, Tfh, Tcm, cytotoxic, NK cells, and eosinophil ES and APM1 expression signature. Of these, 13 loci have multiple significant or suggestive ($p < 1x10^{-6}$) hits, and 10 have hits in the same region in more than one immune trait.

(B) Colocalization analysis in GTEx and TCGA for the SNPs associated with T cell subset-dominant cluster (max CLPP > 0.01). Plot of the maximum CLPP (left) either assuming one causal SNP, two causal SNPs, or the regional CLPP in GTEx and TCGA. For each eQTL, the gene-SNP pair with the highest CLPP is represented. For each sQTL, the SNP-splicing event pair with the highest CLPP is represented. For GTEx, results from the tissue with the highest max CLPP is represented. Plot of the eQTL and sQTL $-\log_{10} p$ (right). For each SNP-gene expression or SNP-gene splicing event pair, 200 SNPs (± 100 SNPs) around the index SNP were included in the analysis. Evidence for the expanded range analysis (± 1MB for eQTL and ± 500KB for sQTL) is labeled as intermediate (*) or strong (**). Pseudogenes are not shown.

(C) LocusZoom plot of the association results for Tcm cell ES maps the genomic location of a significant SNP (rs140752248) on chr 20 to the *MROH8-RBL1-RPN2* locus.

(D) 3-level plot displaying colocalization of rs140752248 with *RBL1* splicing in TCGA. The plot shows the GWAS $-\log_{10} p$ for Tcm cell ES in TCGA, the sQTL $-\log_{10} p$ for *RBL1* splicing in TCGA, and the CLPP assuming one causal SNP; the index SNP rs140752248 and the nearest 100 upstream and 100 downstream SNPs are represented.

See also Figure S6, Tables S4, and S5.

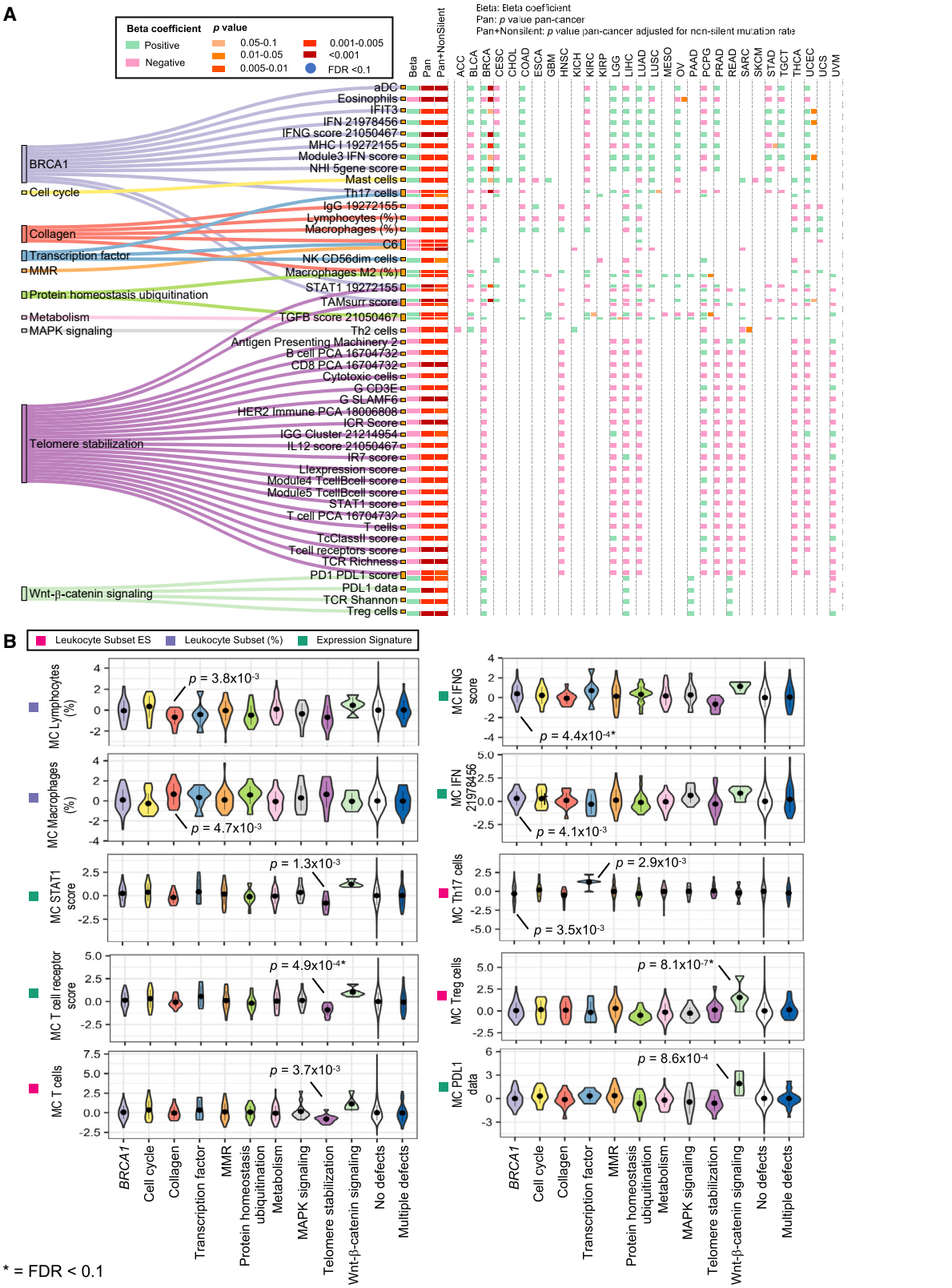


Figure 6. Cancer predisposition variants modulating immune traits

(A) Suggestive associations ($p \leq 0.005$ and FDR $p < 0.25$) between germline pathogenic or likely pathogenic cancer predisposition variants (rare variants) extracted from whole-exome data (Huang et al., 2018), grouped by curated mutually exclusive functional categories (*left nodes*), and immune traits (*right nodes*), as

(legend continued on next page)

Containing 1) is involved in calcium ion binding and SNPs at the *ZZEF1* locus have been associated with adiposity and type 2 diabetes mellitus (Mahajan et al., 2018). However, these SNPs, most of which are in LD with each other, are not in LD with rs112262673 ($r^2 < 0.05$). At locus 23 on chr 22, we identified three SNPs with genome-wide significant associations including two SNPs, rs73889576 and rs11914148, associated with Tcm cell ES (Figure 5A; Table S4). A third SNP, rs572393792, with a suggestive association with Tcm cell ES, had a genome-wide significant association with CD8 T cell ES and suggestive association with T helper cell ES. EQTL and sQTL analyses coupled with colocalization in TCGA identified *ARHGAP8*, *FBLN1, NUP50, PARVG,* and *PRR5* as possible candidates (Figure 5B; Table S5). Among them, only *NUP50* (associated with CD8 T cell ES), a component of the nuclear pore complex interacting with cell cycle regulatory proteins (Ogawa et al., 2010), was supported by intermediate evidence in the expanded range colocalization analyses.

In addition, we found potential evidence for colocalization with several other genes at CLPP > 0.01 (Figure 5B; Table S5). These included some genes with known functional effect on the immune system and/or tumor microenvironment such as *ID01* (Munn and Mellor, 2016) on chr 8 (rs16889186, CLPP = 0.036), associated with cytotoxic cell ES. However, these results should be interpreted with caution, since the effect on gene expression was also modest (Figure 5B). The full list of eQTL and sQTL with FDR $p < 0.1$ for all suggestive and significant SNPs either in TCGA or GTEx, the results of the eCaviar colocalization analysis for all of them, and the results of the expanded range analyses for all the significant SNPs with CLPP > 0.01 are reported in Table S5.

Associations with cancer predisposition variants

We performed association analyses between germline pathogenic and likely pathogenic cancer predisposition variants (referred to here as rare variants) (Huang et al., 2018) in high penetrance susceptibility genes, and immune traits and immune subtypes (Figure 6). Since mutations in most of the genes were rare, when possible, we collapsed genes into categories summarizing different biological processes or functions (Figure S7A). However, as mutations within the homologous recombination repair (HR) genes *BRCA1* and *BRCA2* were more common (82 and 79 events, respectively), we analyzed these genes separately. Overall, 21 genotypic variables were used (Figures S7A and S7B).

In an analysis across all cancer types, we found significant associations (FDR $p < 0.1$) between at least one immune trait and germline mutations in *BRCA1* and genes in Wnt-β-catenin and telomere-stabilization pathways. We found suggestive associations ($p < 0.005$, FDR p from 0.1 to 0.25) between at least one immune trait and seven other categories: cell cycle, collagen, transcription factor, mismatch repair (MMR), protein homeostasis ubiquitination, metabolism, and MAPK signaling (Figures 6A and 6B). These associations were only minimally altered by correction for the somatic mutational load (Figure 6A). Mutations in genes involved in telomere stabilization were correlated with features associated with T cell exclusion, including low abundance of T cell and cytotoxic cells and diminished IFN-related signatures (including STAT1), and TCR diversity (TCR richness) (Figure 6B). Germline mutations of a collagen-related gene (COL7A1) had nominally significant associations with increased macrophage infiltration and decreased lymphocyte infiltration (Figure 6B). Lastly, we found the Wnt-β-catenin germline mutations to be consistent with the presence of an inflammatory phenotype accompanied by counter-regulatory mechanisms such as the activation of the PD-L1 signaling and the recruitment of Treg cells (Figure 6B).

We also tested the relation between rare variants and somatic DNA alterations (Figures S7C, S7D and Table S6). However, since germline mutations in MMR genes are known to affect the somatic mutation rate, which predicts immunotherapy responsiveness, we performed more detailed analyses of these mutations. As expected, MMR germline variants were associated with a higher mutational/neoantigen load and a higher microsatellite instability score (MANTIS) (Figure S7C) (Middha et al., 2017). These associations were significant in colon adenocarcinoma (COAD) and uterine corpus endometrial carcinoma (UCEC). MMR germline mutations were associated with higher leukocyte infiltration only in colon cancer (Figure 7A; Table S6). Overall, a higher leukocyte fraction and non-silent mutation rate in MMR germline mutated samples were confined to tumors with microsatellite instability (MSI-H) (Figure 7B). Interestingly, among MSI-H tumors, the ones driven by MMR germline mutations tended to have a higher leukocyte infiltration, although this comparison was not significant ($p > 0.05$) (Figure 7B). A similar trend was observed for other immune signatures (data not shown). Germline mutations in Fanconi Anemia (FA) and in *BRCA1/2* genes were associated with a higher HR defect score (Figure S7C). However, only mutations in *BRCA1* were associated with favorable immunologic parameters such as higher values of MHC and IFN response module traits including IFN-related signatures and aDC ES. These associations were driven by breast invasive carcinoma (BRCA) samples (Figures 6A and 7A). When the analysis in BRCA samples was adjusted or stratified for intrinsic molecular subtypes (basal-like versus non-basal-like tumors), these associations were no longer significant (Table S6).

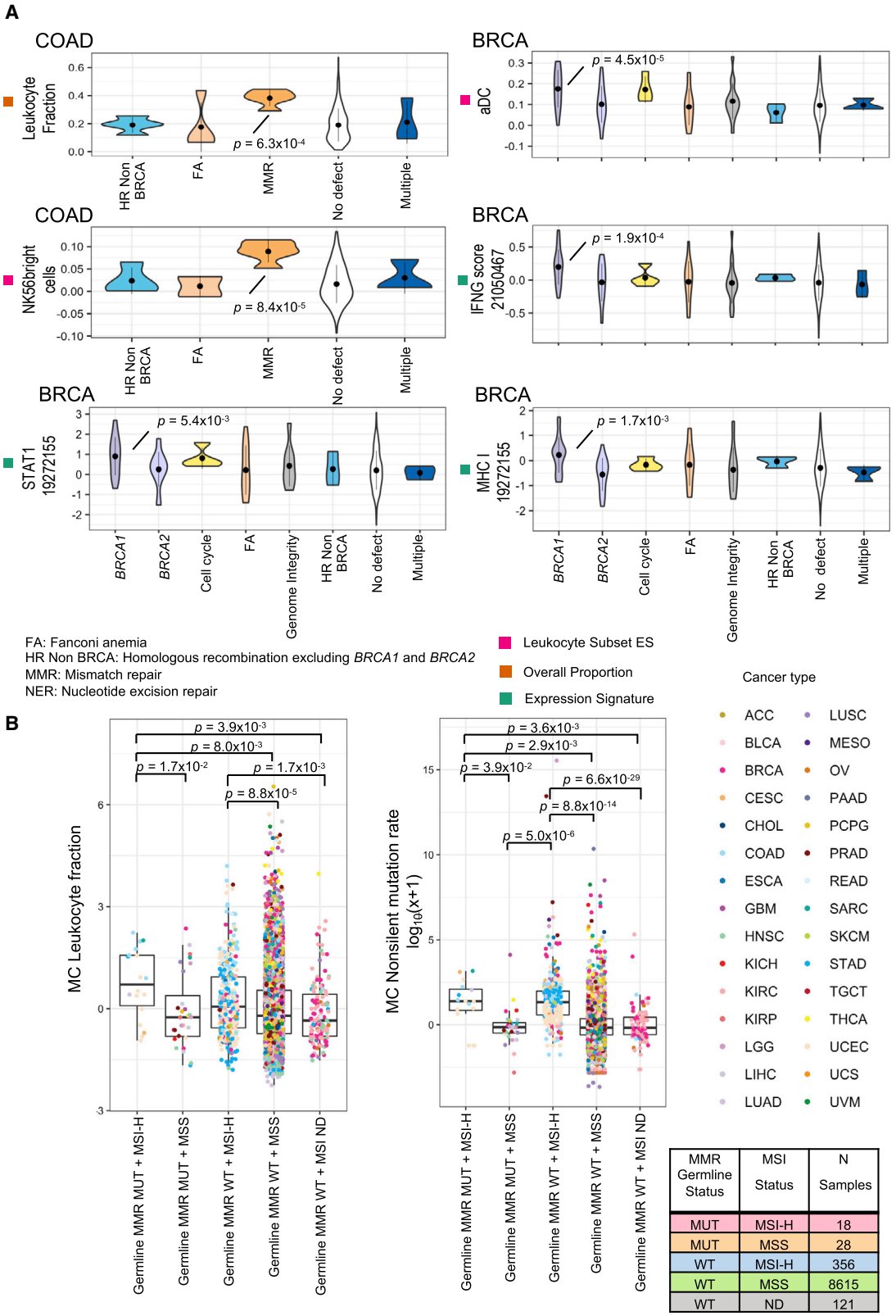
Discussion

We conducted a comprehensive pan-cancer analysis of the germline genetic contribution to the tumor immune microenvironment by evaluating common variant heritability, performing

identified in pan-cancer regression models adjusted for standard covariates. Significant associations ($p < 0.005$ and FDR $p < 0.1$) are highlighted with blue dots. The adjusted p for the non-silent mutation rate is also shown. Beta coefficients and significance level are visualized pan-cancer and per cancer (*right side*). The Beta coefficient is shown irrespectively of the significance and number of events.

(B) Values of representative immune traits (mean centered by cancer type, (MC), to visually reflect the cancer-type covariate used in the model) are displayed across samples with mutations in genes related to the defined functional categories. Regression analyses were performed with covariates as described in Methods. See Methods for cancer type abbreviations.

See also Figure S7 and Table S6.



MMR Germline Status	MSI Status	N Samples
MUT	MSI-H	18
MUT	MSS	28
WT	MSI-H	356
WT	MSS	8615
WT	ND	121

Figure 7. Representative immune traits modulated by cancer predisposition variants in specific cancers

(A) Representative associations between pathogenic or likely pathogenic variants, grouped by functional categories, and representative immune traits within COAD and BRCA.

(legend continued on next page)

GWAS paired with colocalization analyses, and assessing the effect of rare variants in cancer predisposition genes.

Heritability analyses revealed that common genetic variants explain up to one fifth of the variance of some immune traits. About 25% of the traits (33 of 139) were heritable. The traits most strongly influenced by germline genetics include estimates of the abundance of cytotoxic T, NK, Tfh cells (heritability ~20%), and IFN signaling (heritability ~15%), which have been associated with favorable prognosis and/or responsiveness to immunotherapy (Galon and Bruni, 2020). The magnitude of the heritabilities that we identified were similar to the ones observed for complex traits in humans, such as body mass index (Shi et al., 2016). The leukocyte subset ES were highly correlated (all part of the T cell/cytotoxic module) and had largely overlapping GWAS loci. However, ES for each of these cell types were calculated based on distinct gene sets suggesting that the genetic correlations between these traits were unlikely due to traditional eQTLs but rather represented similar genetic mechanisms affecting the cellular infiltrates and/or the activation states of these cells that likely reflect an overall coordination of the immune response. The heritabilities were partially dependent on immune subtypes that characterize the overall patterns of immune response seen in cancer (Thorsson et al., 2018). For instance, in the immune traits that showed interactions with immune subtypes, the heritabilities were higher in the wound healing subtype, suggesting that the effect of common genetic variants might be most pronounced in highly proliferative and poorly immune infiltrated tumors.

In GWAS analyses, we found two significant loci for IFN-gamma signatures. One locus at chr 2 included SNPs within the *IFIH1* gene, previously associated with multiple autoimmune disorders (Buniello et al., 2019; Rice et al., 2014), demonstrating a link between autoimmunity and the immune response to cancer. Since IFN signatures have been associated with responsiveness to immunotherapy (Ayers et al., 2017; Cristescu et al., 2018), these SNPs may also be associated with efficacy among patients receiving immunotherapy, which is consistent with preliminary data in melanoma patients receiving checkpoint inhibitors (Chat et al., 2019). Our analyses also identified an IFN-associated cluster of SNPs on chr 3 near *TMEM108*. *TMEM108* is not known to have an effect on IFN signaling but may signal through the Wnt-β-catenin pathway (Yu et al., 2019). Our eQTL and sQTL analyses found colocalization with *TMEM108* in GTEx, and DICE, but not in TCGA, suggesting that the mechanism of action is via expression on either the immune infiltrating cells and/or normal tissue surrounding the tumor. In addition, we observed a near genome-wide significant association between IFN signaling and a well-characterized, functional missense variant of *TMEM173* (rs1131769), which encodes for the STING protein. Since STING is a key modulator of IFN-mediated response against viruses and cancer cells, the association with IFN signaling traits is consistent with expectation. STING pathway agonists are in early clinical trials and show promising results (Flood et al., 2019). Our results suggest that patients' responsiveness to these agents might differ according to the *TMEM173* genotype. STING genotypes may also affect immune checkpoint blockade (Wang et al., 2017) and radiotherapy-induced tumor immunogenicity (Vanpouille-Box et al., 2017), as mice models suggest that STING activation is necessary for their efficacy.

We also found 14 genomic loci significantly associated with traits that are part of the T cell/cytotoxic module. Of these, we found the strongest evidence for colocalization between splicing with *RBL1* and a locus associated with Tcm cell ES. *RBL1* has homology with the well-known tumor suppressor *RB* and may also act as a tumor suppressor (Ng et al., 2020). *RB* and *RBL1* affect DNA repair through non-homologous end-joining (Cook et al., 2015), and thus the effect we detected may be an indirect effect of this splice variant on DNA repair in tumors. However, the RB family of genes may also have direct effects on the immune system (Garfin et al., 2013) by controlling proliferation of T lymphocytes (Mulligan et al., 1998), and expression of Toll-like receptors (Taura et al., 2012). Moreover, *RB1* has been implicated in T cell exclusion and immune checkpoint resistance (Jerby-Arnon et al., 2018).

Another study also performed GWAS of immune traits in TCGA (Shahamatdar et al., 2020), but only analyzed a limited number of traits (N = 17) and focused on European ancestry subjects (N = 5,788). In addition to the association between *IL17RA* locus and Th17 cell ES which we determined is likely an eQTL, the study also identified another locus, with lead SNP rs3366, associated with Tfh cell proportion, inferred by CIBERSORT. This trait was not included in our GWAS because it was not significantly heritable. Shahamatdar et al. also found a relationship between polygenic risk for autoimmune disorders and immune infiltration into tumors (Shahamatdar et al., 2020). Our findings also support the overlap between autoimmunity and cancer immune response, specifically, with the identification of the *IFIH1* locus.

Rare variant analyses demonstrated intriguing associations between genetic variants related to cancer development and intratumoral immune response. Among HR gene categories, only *BRCA1* mutations were associated with higher levels of favorable immune parameters. These effects were restricted to BRCA samples and driven by the higher rate of basal-like phenotype among *BRCA1* mutation carriers. Triple negative/basal-like BRCA samples have previously been shown to be more likely to have robust immune infiltration compared with other breast cancer subtypes (Hendrickx et al., 2017; Jézéquel et al., 2015; Miller et al., 2016). Our analyses suggest that *BRCA1* may mediate its tissue-restricted effect on immune response by uniquely modulating oncogenic pathways captured by the intrinsic molecular subtype classification. We also found that mutations in genes in the MMR pathway were associated with a more robust immune response. MMR deficiency is a strong predictor of response to checkpoint inhibitors (Le et al., 2017) and is one of the FDA indications for treatment with these agents regardless of tissue of origin (Marcus et al., 2019). However, the observation

(B) Leukocyte fraction and non-silent mutation rate (mean centered by cancer type, (MC)) by combined germline mutation status across MMR genes and somatic MSI status: MSI High (MSI-H) versus microsatellite stable (MSS), as identified by MANTIS score (threshold = 0.4, (Middha et al., 2017). MSI ND: MSI status not determined. Samples are colored by cancer type. Significance $p < 0.05$ are annotated from regression analyses using covariates as described in the Methods. See Methods for cancer type abbreviations.
See also Figure S7 and Table S6.

that tumors from patients with germline MMR mutations have, overall, a strong immune infiltration only when they display a MSI-H phenotype suggests that MMR germline mutations, alone, might not be sufficient to accurately predict response to immunotherapy.

Patients with germline mutations in telomere-stabilization genes (*DKC1* and *POT1*) had lower lymphocytic infiltration. This could be the effect of a reduced lymphocyte proliferative capacity following antigen recognition due to a low telomere length (Rosenberg et al., 2011). However, this interpretation should be taken with caution since lower telomere length has only been demonstrated for mutations in *DKC1* (Aubert and Lansdorp, 2008; Calado and Young, 2009) and not in *POT1* (Rice et al., 2017).

The associations between germline mutations of genes in the Wnt-β-catenin pathways and increased levels of T cell and counter-regulatory mechanisms (PD-L1 and Treg cells), support the critical role of this pathway in modulating anti-tumor immunity (Luke et al., 2019; Spranger et al., 2015). While deleterious germline mutations of Wnt-β-catenin negative regulators *APC* and *PTCH1* might predict pathway activation, it is possible that such alterations induce the pathway's downregulation at the somatic level by triggering compensatory mechanisms.

In summary, our analysis demonstrated that both common and rare germline genetic variants can shape the functional orientation of the tumor microenvironment and identified potential modulatory genes and mechanisms involved in this process. The extended and curated list of variants, candidate genes, and pipelines provided here as a resource for the scientific community, might be exploited in the context of immunotherapy, spurring studies that could lead to the development of personalized therapeutic strategies.

LIMITATIONS OF THE STUDY

While our cohort was large, it is composed of 30 tumor types, which may lead to a loss of true signals due to heterogeneity. Larger studies focusing on specific cancers could expand and refine our observations. Our heritability analyses only used common variants which likely underestimates heritability (Shi et al., 2016), and heritability using whole-genome sequencing will be likely higher. Many of the traits we analyzed were highly correlated, thus the number of effectively independent immune features is fewer than the sum of the individual traits. However, this correlation reflects a natural feature of the immune system (Orrù et al., 2013). Our fine mapping efforts focused on gene expression, which only explains ~15% of heritability (Yao et al., 2020). The remainder of non-coding variants associated with complex traits may mediate their effects via conditional effects on gene expression, or via effects on a specific cell lineage that would not be detectable in bulk tissue analyses. Future studies of these loci in larger datasets of homogeneous cell types and in studies using single cell sequencing may identify more candidate genes, and mechanistic experiments might further elucidate their function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.immuni.2021.01.011.

ACKNOWLEDGMENTS

We are grateful to the Society for Immunotherapy of Cancer (SITC) for the logistical support of the investigator meeting within the SITC Cancer Immune Responsiveness Workshop (San Francisco CA, US, April 2018; Houston, TX, US, May 2019). We are also grateful to Noah Zaitlen and Andy Dahl for useful discussions on heritability interaction analyses. This work was funded in part by the National Institutes of Health R01CA227466 and K24CA169004 to E.Z. and T32CA221709 Cancer Metabolism Training Program Postdoctoral Fellowship to R.W.S., Qatar National Research Fund (QNRF) NPRP11S-0121-180351 grant and the Sidra Precision Medicine Program internal grant (SDR100035 and SDR400023) to D.B., Associazione Italiana per la Ricerca sul Cancro (AIRC) grant IG2018-21846 to M.C., the Cancer Research Institute funding to V.T., and the Fundació Bancaria La Caixa, a La Caixa Junior Leader Fellowship (LCF/BQ/PI18/11630003) to E.P-P.

AUTHOR CONTRIBUTIONS

Conceptualization: V.T., D.W., F.M.M., E.Z., D.B. Design: R.W.S., M.S., V.T., Y.M., T.K., R.F.S., O.F.B., E.P.P., S.S., D.W., E.Z., D.B. Analysis, Computation, and Software: R.W.S., M.S., V.T., W.H., J.R., Y.M., F.F., R.F.S., O.F.B., C.H., D.H., S.H., N.S., E.Z., D.B. Interpretation: R.W.S., M.S., V.T., W.H., J.R., Y.M., F.F., T.K., R.F.S., O.F.B., C.H., E.P.P., M.Ca., C.S., D.H., S.H., R.E.G., N.S., L.R., M.L.D, S.S., D.W., F.M.M., J.G., M.Ce, E.Z., D.B. Writing- Review & Editing: R.W.S., M.S., V.T., W.H., J.R., Y.M., F.F., T.K., R.F.S., O.F.B.,

CellPress

C.H., E.P.P., M.Ca., C.S., D.H., S.H., R.E.G., N.S., L.R., M.L.D, S.S., D.W., F.M.M., J.G., M.Ce, E.Z., D.B. Supervision & Funding Acquisition: E.Z., D.B.

DECLARATIONS OF INTERESTS

R.F.S. has received consulting honoraria from Aduro, Astellas, AstraZeneca, BMS, EMD Serono, Exelixis, Eisai, Janssen, Mirati, Pfizer, Puma, and Seattle Genetics. R.F.S. has received research support (to institution) from Abbvie, Aduro, Bayer, BMS, CytomX, Eisai, Eli Lilly, Genentech/Roche, Immunocore, Novartis, Merck, Mirati, Moderna, and QED therapeutics. F.M.M. is an employee of Refuge Biotechnologies. J.G. has patents associated with ''*in vitro* method for the prognosis of progression of a cancer'' (PCT/IB2006/003168, PCT/EP2013/062405)'' and received royalties from INSERM. J.G. is Co-founder and chairman of the scientific advisory board of HalioDx. J.G. has Collaborative Research Agreement grants with Akoya, IObiotech, MedImmune, AstraZeneca, Janssen, and Imcheck Therapeutics. J.G. participated to Scientific Advisory Boards and is consultant for BMS, Novartis, Merck Serono, IObiotech, Nanostring, Illumina, Northwest Biotherapeutics, Actelion, Amgen, Merck MSD, Lunaphore, Catalym, and Sanofi.

Received: January 20, 2020
Revised: October 14, 2020
Accepted: January 13, 2021
Published: February 9, 2021

REFERENCES

Amara, D., Wolf, D.M., van 't Veer, L., Esserman, L., Campbell, M., and Yau, C. (2016). Co-expression modules identified from published immune signatures reveal five distinct immune subtypes in breast cancer. Breast Cancer Res. Treat. *161*, 41–50.

Angelova, M., Mlecnik, B., Vasaturo, A., Bindea, G., Fredriksen, T., Lafontaine, L., Buttard, B., Morgand, E., Bruni, D., Jouret-Mourin, A., et al. (2018). Evolution of Metastases in Space and Time under Immune Selection. Cell *175*, 751–765.e16.

Arce Vargas, F., Furness, A.J.S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., Solomon, I., Lesko, M.H., Ruef, N., Roddie, C., et al.; TRACERx Melanoma; TRACERx Renal; TRACERx Lung consortia (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. Cancer Cell *33*, 649–663.e4.

Aubert, G., and Lansdorp, P.M. (2008). Telomeres and aging. Physiol. Rev. *88*, 557–579.

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. J. Clin. Invest. *127*, 2930–2940.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.; MC3 Working Group; Cancer Genome Atlas Research Network (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18.

Bedognetti, D., Hendrickx, W., Ceccarelli, M., Miller, L.D., and Seliger, B. (2016). Disentangling the relationship between tumor genetic programs and immune responsiveness. Curr. Opin. Immunol. *39*, 150–158.

Bedognetti, D., Ceccarelli, M., Galluzzi, L., Lu, R., Palucka, K., Samayoa, J., Spranger, S., Warren, S., Wong, K.-K., Ziv, E., et al.; Society for Immunotherapy of Cancer (SITC) Cancer Immune Responsiveness Task Force and Working Groups (2019). Toward a comprehensive view of cancer immune responsiveness: a synopsis from the SITC workshop. J. Immunother. Cancer *7*, 131.

Bedognetti, D., Spivey, T.L., Zhao, Y., Uccellini, L., Tomei, S., Dudley, M.E., Ascierto, M.L., De Giorgi, V., Liu, Q., Delogu, L.G., et al. (2013). CXCR3/CCR5 pathways in metastatic melanoma patients treated with adoptive therapy and interleukin-2. Br. J. Cancer *109*, 2412–2423.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782–795.

Bruni, D., Angell, H.K., and Galon, J. (2020). The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. Nat. Rev. Cancer *20*, 662–680.

Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47* (D1), D1005–D1012.

Calado, R.T., and Young, N.S. (2009). Telomere diseases. N. Engl. J. Med. *361*, 2353–2365.

Carrot-Zhang, J., Chambwe, N., Damrauer, J.S., Knijnenburg, T.A., Robertson, A.G., Yau, C., Zhou, W., Berger, A.C., Huang, K., Newberg, J.Y., et al. (2020). Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. Cancer Cell *37*, 639–654.e6.

Chamoto, K., Hatae, R., and Honjo, T. (2020). Current issues and perspectives in PD-1 blockade cancer immunotherapy. Int. J. Clin. Oncol. *25*, 790–800.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

Chat, V., Ferguson, R., Simpson, D., Kazlow, E., Lax, R., Moran, U., Pavlick, A., Frederick, D., Boland, G., Sullivan, R., et al. (2019). Autoimmune genetic risk variants as germline biomarkers of response to melanoma immune-checkpoint inhibition. Cancer Immunol. Immunother. *68*, 897–905.

Cheng, W.-Y., Ou Yang, T.H., and Anastassiou, D. (2013a). Biomolecular events in cancer revealed by attractor metagenes. PLoS Comput. Biol. *9*, e1002920.

Cheng, W.-Y., Yang, T.-H.O., and Anastassiou, D. (2013b). Development of a Prognostic Model for Breast Cancer Survival in an Open Challenge Environment. Sci. Transl. Med. *5*, https://doi.org/10.1126/scitranslmed.3005974.

Chowell, D., Morris, L.G.T., Grigg, C.M., Weber, J.K., Samstein, R.M., Makarov, V., Kuo, F., Kendall, S.M., Requena, D., Riaz, N., et al. (2018). Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. Science *359*, 582–587.

Cook, R., Zoumpoulidou, G., Luczynski, M.T., Rieger, S., Moquet, J., Spanswick, V.J., Hartley, J.A., Rothkamm, K., Huang, P.H., and Mittnacht, S. (2015). Direct involvement of retinoblastoma family proteins in DNA repair by non-homologous end-joining. Cell Rep. *10*, 2006–2018.

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., Sher, X., Liu, X.Q., Lu, H., Nebozhyn, M., et al. (2018). Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science *362*, https://doi.org/10.1126/science.aar3593.

Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*, https://doi.org/10.1126/science.aaf8399.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics, 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 1184–1191.

Ergun, S.L., Fernandez, D., Weiss, T.M., and Li, L. (2019). STING Polymer Structure Reveals Mechanisms for Activation, Hyperactivation, and Inhibition. Cell *178*, 290–301.e10.

Flood, B.A., Higgs, E.F., Li, S., Luke, J.J., and Gajewski, T.F. (2019). STING pathway agonism as a cancer therapeutic. Immunol. Rev. *290*, 24–38.

Galon, J., and Bruni, D. (2020). Tumor Immunology and Tumor Evolution: Intertwined Histories. Immunity *52*, 55–81.

Fuchsberger, C., Abecasis, G.R., and Hinds, D.A. (2015). minimac2: faster genotype imputation. Bioinformatics *31*, 782–784.

Galon, J., Angell, H.K., Bedognetti, D., and Marincola, F.M. (2013). The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. Immunity *39*, 11–26.

Gao, P., Ascano, M., Zillinger, T., Wang, W., Dai, P., Serganov, A.A., Gaffney, B.L., Shuman, S., Jones, R.A., Deng, L., et al. (2013). Structure-Function Analysis of STING Activation by c[G(2′,5′)pA(3′,5′)p] and Targeting by Antiviral DMXAA. Cell *154*, 748–762.

Garfin, P.M., Min, D., Bryson, J.L., Serwold, T., Edris, B., Blackburn, C.C., Richie, E.R., Weinberg, K.I., Manley, N.R., Sage, J., and Viatour, P. (2013). Inactivation of the RB family prevents thymus involution and promotes thymic function by direct control of Foxn1 expression. J. Exp. Med. *210*, 1087–1097.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. *21*, 938–945, advance online publication.

Gorman, J.A., Hundhausen, C., Errett, J.S., Stone, A.E., Allenspach, E.J., Ge, Y., Arkatkar, T., Clough, C., Dai, X., Khim, S., et al. (2017). The A946T variant of the RNA sensor IFIH1 mediates an interferon program that limits viral infection but increases the risk for autoimmunity. Nat. Immunol. *18*, 744–752.

Haslam, A., and Prasad, V. (2019). Estimation of the Percentage of US Patients With Cancer Who Are Eligible for and Respond to Checkpoint Inhibitor Immunotherapy Drugs. JAMA Netw. Open *2*, e192535.

Havel, J.J., Chowell, D., and Chan, T.A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. Nat. Rev. Cancer *19*, 133–150.

Helmink, B.A., Khan, M.A.W., Hermann, A., Gopalakrishnan, V., and Wargo, J.A. (2019). The microbiome, cancer, and cancer therapy. Nat. Med. *25*, 377–388.

Hendrickx, W., Simeone, I., Anjum, S., Mokrab, Y., Bertucci, F., Finetti, P., Curigliano, G., Seliger, B., Cerulo, L., Tomei, S., et al. (2017). Identification of genetic determinants of breast cancer immune phenotypes by integrative genome-scale analysis. OncoImmunology *6*, https://doi.org/10.1080/2162402X.2016.1253654.

Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet. *99*, 1245–1260.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.

Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al.; Cancer Genome Atlas Research Network (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. Cell *173*, 355–370.e14.

Hutter, C., and Zenklusen, J.C. (2018). The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. Cell *173*, 283–285.

Iida, N., Dzutsev, A., Stewart, C.A., Smith, L., Bouladoux, N., Weingarten, R.A., Molina, D.A., Salcedo, R., Back, T., Cramer, S., et al. (2013). Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science *342*, 967–970.

Ishida, Y., Agata, Y., Shibahara, K., and Honjo, T. (1992). Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. EMBO J. *11*, 3887–3895.

Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.-J., Melms, J.C., Leeson, R., Kanodia, A., Mei, S., Lin, J.-R., et al. (2018). A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell *175*, 984–997.e24.

Jézéquel, P., Loussouarn, D., Guérin-Charbonnel, C., Campion, L., Vanier, A., Gouraud, W., Lasla, H., Guette, C., Valo, I., Verrièle, V., and Campone, M. (2015). Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. Breast Cancer Res. *17*, 43.

Junger, W.G. (2011). Immune cell regulation by autocrine purinergic signalling. Nat. Rev. Immunol. *11*, 201–212.

Kahles, A., Lehmann, K.-V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., and Rätsch, G.; Cancer Genome Atlas Research Network (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell *34*, 211–224.e6.

Kalbasi, A., and Ribas, A. (2020). Tumour-intrinsic resistance to immune checkpoint blockade. Nat. Rev. Immunol. *20*, 25–39.

Kang, D.C., Gopalkrishnan, R.V., Wu, Q., Jankowsky, E., Pyle, A.M., and Fisher, P.B. (2002). mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. Proc. Natl. Acad. Sci. USA *99*, 637–642.

Keller, M.F., Reiner, A.P., Okada, Y., van Rooij, F.J.A., Johnson, A.D., Chen, M.-H., Smith, A.V., Morris, A.P., Tanaka, T., Ferrucci, L., et al.; CHARGE Hematology; COGENT; BioBank Japan Project (RIKEN) Working Groups (2014). Trans-ethnic meta-analysis of white blood cell phenotypes. Hum. Mol. Genet. *23*, 6944–6960.

Kennedy, R.B., Haralambieva, I.H., Ovsyannikova, I.G., Voigt, E.A., Larrabee, B.R., Schaid, D.J., Zimmermann, M.T., Oberg, A.L., and Poland, G.A. (2020). Polymorphisms in STING affect human innate immune responses to poxviruses. Front Immunol. *11*, https://doi.org/10.3389/fimmu.2020.567348.

Kim, Y., Shin, Y., and Kang, G.H. (2019). Prognostic significance of CD103+ immune cells in solid tumor: a systemic review and meta-analysis. Sci. Rep. *9*, 3808.

Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., Cherniack, A.D., Fan, H., Shen, H., Way, G.P., Greene, C.S., et al.; Cancer Genome Atlas Research Network (2018). Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. *23*, 239–254.e6.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Le, D.T., Durham, J.N., Smith, K.N., Wang, H., Bartlett, B.R., Aulakh, L.K., Lu, S., Kemberling, H., Wilt, C., Luber, B.S., et al. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science *357*, 409–413.

Leach, D.R., Krummel, M.F., and Allison, J.P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. Science *271*, 1734–1736.

Li, Y., Si, L., Zhai, Y., Hu, Y., Hu, Z., Bei, J.-X., Xie, B., Ren, Q., Cao, P., Yang, F., et al. (2016). Genome-wide association study identifies 8p21.3 associated with persistent hepatitis B virus infection among Chinese. Nat. Commun. *7*, 11664.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al.; Cancer Genome Atlas Research Network (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*, 400–416.e11.

Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016a). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.

Loh, P.-R., Palamara, P.F., and Price, A.L. (2016b). Fast and accurate long-range phasing in a UK Biobank cohort. Nat. Genet. *48*, 811–816.

Luke, J.J., Bao, R., Sweis, R.F., Spranger, S., and Gajewski, T.F. (2019). WNT/b-catenin pathway activation correlates with immune exclusion across human

cancers. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. https://doi.org/10.1158/1078-0432.CCR-18-1942.

Mahajan, A., Wessel, J., Willems, S.M., Zhao, W., Robertson, N.R., Chu, A.Y., Gan, W., Kitajima, H., Taliun, D., Rayner, N.W., et al.; ExomeBP Consortium; MAGIC Consortium; GIANT Consortium (2018). Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat. Genet. *50*, 559–571.

Mandal, R., Samstein, R.M., Lee, K.-W., Havel, J.J., Wang, H., Krishna, C., Sabio, E.Y., Makarov, V., Kuo, F., Blecua, P., et al. (2019). Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. Science *364*, 485–491.

Marcus, L., Lemery, S.J., Keegan, P., and Pazdur, R. (2019). FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clin. Cancer Res. *25*, 3753–3758.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

Middha, S., Zhang, L., Nafa, K., Jayakumaran, G., Wong, D., Kim, H.R., Sadowska, J., Berger, M.F., Delair, D.F., Shia, J., et al. (2017). Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. JCO Precis. Oncol. https://doi.org/10.1200/PO.17.00084.

Miller, L.D., Chou, J.A., Black, M.A., Print, C., Chifman, J., Alistar, A., Putti, T., Zhou, X., Bedognetti, D., Hendrickx, W., et al. (2016). Immunogenic Subtypes of Breast Cancer Delineated by Gene Classifiers of Immune Responsiveness. Cancer Immunol. Res. *4*, 600–610.

Mulligan, G.J., Wong, J., and Jacks, T. (1998). p130 is dispensable in peripheral T lymphocytes: evidence for functional compensation by p107 and pRB. Mol. Cell. Biol. *18*, 206–220.

Munn, D.H., and Mellor, A.L. (2016). IDO in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance. Trends Immunol. *37*, 193–207.

Ng, S.R., Rideout, W.M., 3rd, Akama-Garren, E.H., Bhutkar, A., Mercer, K.L., Schenkel, J.M., Bronson, R.T., and Jacks, T. (2020). CRISPR-mediated modeling and functional validation of candidate tumor suppressor genes in small cell lung cancer. Proc. Natl. Acad. Sci. USA *117*, 513–521.

Ogawa, Y., Miyamoto, Y., Asally, M., Oka, M., Yasuda, Y., and Yoneda, Y. (2010). Two isoforms of Npap60 (Nup50) differentially regulate nuclear protein import. Mol. Biol. Cell *21*, 630–638.

Orrù, V., Steri, M., Sole, G., Sidore, C., Virdis, F., Dei, M., Lai, S., Zoledziewska, M., Busonero, F., Mulas, A., et al. (2013). Genetic variants regulating immune cell levels in health and disease. Cell *155*, 242–256.

Patel, S., and Jin, L. (2019). TMEM173 variants and potential importance to human biology and disease. Genes Immun. *20*, 82–89.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

Queirolo, P., Dozin, B., Morabito, A., Banelli, B., Piccioli, P., Fava, C., Leo, C., Carosio, R., Laurent, S., Fontana, V., et al. (2017). 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*, https://doi.org/10.3389/fimmu.2017.00386.

Rashkin, S.R., Graff, R.E., Kachuri, L., Thai, K.K., Alexeeff, S.E., Blatchins, M.A., Cavazos, T.B., Corley, D.A., Emami, N.C., Hoffman, J.D., et al. (2019). Pan-Cancer Study Detects Novel Genetic Risk Variants and Shared Genetic Basis in Two Large Cohorts. BioRxiv. https://doi.org/10.1101/635367.

Rice, G.I., Del Toro Duany, Y., Jenkinson, E.M., Forte, G.M.A., Anderson, B.H., Ariaudo, G., Bader-Meunier, B., Baildam, E.M., Battini, R., Beresford, M.W., et al. (2014). Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. Nat. Genet. *46*, 503–509.

Rice, C., Shastrula, P.K., Kossenkov, A.V., Hills, R., Baird, D.M., Showe, L.C., Doukov, T., Janicki, S., and Skordalakes, E. (2017). Structural and functional analysis of the human POT1-TPP1 telomeric complex. Nat. Commun. *8*, 14928.

Roelands, J., Hendrickx, W., Zoppoli, G., Mall, R., Saad, M., Halliwill, K., Curigliano, G., Rinchai, D., Decock, J., Delogu, L.G., et al. (2020). Oncogenic states dictate the prognostic and predictive connotations of intratumoral immune response. J. Immunother. Cancer *8*, https://doi.org/10.1136/jitc-2020-000617.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

Rosenberg, S.A., Yang, J.C., Sherry, R.M., Kammula, U.S., Hughes, M.S., Phan, G.Q., Citrin, D.E., Restifo, N.P., Robbins, P.F., Wunderlich, J.R., et al. (2011). Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clin. Cancer Res. *17*, 4550–4557.

Rozenblit, M., Hendrickx, W., Heguy, A., Chiriboga, L., Loomis, C., Ray, K., Darvishian, F., Egeblad, M., Demaria, S., Marincola, F.M., et al. (2019). Transcriptomic profiles conducive to immune-mediated tumor rejection in human breast cancer skin metastases treated with Imiquimod. Sci. Rep. *9*, 8572.

Samstein, R.M., Lee, C.-H., Shoushtari, A.N., Hellmann, M.D., Shen, R., Janjigian, Y.Y., Barron, D.A., Zehir, A., Jordan, E.J., Omuro, A., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat. Genet. *51*, 202–206.

Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell *175*, 1701–1715.e16.

Şenbabaoğlu, Y., Sümer, S.O., Sánchez-Vega, F., Bemis, D., Ciriello, G., Schultz, N., and Sander, C. (2016). A Multi-Method Approach for Proteomic Network Inference in 11 Human Cancers. PLoS Comput. Biol. *12*, e1004765.

Shahamatdar, S., He, M.X., Reyna, M.A., Gusev, A., AlDubayan, S.H., Van Allen, E.M., and Ramachandran, S. (2020). Germline Features Associated with Immune Infiltration in Solid Tumors. Cell Rep. *30*, 2900–2908.e4.

Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. *99*, 139–153.

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J.M., Desrichard, A., Walsh, L.A., Postow, M.A., Wong, P., Ho, T.S., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. *371*, 2189–2199.

Spranger, S., Bao, R., and Gajewski, T.F. (2015). Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. Nature *523*, 231–235.

Sweis, R.F., and Luke, J.J. (2017). Mechanistic and pharmacologic insights on immune checkpoint inhibitors. Pharmacol. Res. *120*, 1–9.

Taura, M., Suico, M.A., Koyama, K., Komatsu, K., Miyakita, R., Matsumoto, C., Kudo, E., Kariya, R., Goto, H., Kitajima, S., et al. (2012). Rb/E2F1 regulates the innate immune receptor Toll-like receptor 3 in epithelial cells. Mol. Cell. Biol. *32*, 1581–1590.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.-H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al.; Cancer Genome Atlas Research Network (2018). The Immune Landscape of Cancer. Immunity *48*, 812–830.e14.

Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J.E., et al.; Collaborative

Association Study of Psoriasis (CASP); Genetic Analysis of Psoriasis Consortium; Psoriasis Association Genetics Extension; Wellcome Trust Case Control Consortium 2 (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat. Genet. *44*, 1341–1348.

Tumeh, P.C., Harview, C.L., Yearley, J.H., Shintaku, I.P., Taylor, E.J., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature *515*, 568–571.

Uccellini, L., De Giorgi, V., Zhao, Y., Tumaini, B., Erdenebileg, N., Dudley, M.E., Tomei, S., Bedognetti, D., Ascierto, M.L., Liu, Q., et al. (2012). IRF5 gene polymorphisms in melanoma. J. Transl. Med. *10*, 170.

Ugurel, S., Schrama, D., Keller, G., Schadendorf, D., Bröcker, E.-B., Houben, R., Zapatka, M., Fink, W., Kaufman, H.L., and Becker, J.C. (2008). Impact of the CCR5 gene polymorphism on the survival of metastatic melanoma patients receiving immunotherapy. Cancer Immunol. Immunother. *57*, 685–691.

Vanpouille-Box, C., Alard, A., Aryankalayil, M.J., Sarfraz, Y., Diamond, J.M., Schneider, R.J., Inghirami, G., Coleman, C.N., Formenti, S.C., and Demaria, S. (2017). DNA exonuclease Trex1 regulates radiotherapy-induced tumour immunogenicity. Nat. Commun. *8*, 15618.

Visscher, P.M., Hemani, G., Vinkhuyzen, A.A.E., Chen, G.-B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet. *10*, e1004269.

Wang, H., Hu, S., Chen, X., Shi, H., Chen, C., Sun, L., and Chen, Z.J. (2017). cGAS is essential for the antitumor effect of immune checkpoint blockade. Proc. Natl. Acad. Sci. USA *114*, 1637–1642.

Wolf, D.M., Lenburg, M.E., Yau, C., Boudreau, A., and van 't Veer, L.J. (2014). Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. PLoS ONE *9*, e88309.

Yang, J.Y.C., and Sarwal, M.M. (2017). Transplant genetics and genomics. Nat. Rev. Genet. *18*, 309–326.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

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

Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. *52*, 626–633.

Ye, J., Gillespie, K.M., and Rodriguez, S. (2018). Unravelling the Roles of Susceptibility Loci for Autoimmune Diseases in the Post-GWAS Era. Genes (Basel) *9*, https://doi.org/10.3390/genes9080377.

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y.-M., Gale, M., Jr., Akira, S., et al. (2005). Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. J. Immunol. *175*, 2851–2858.

Yu, Z., Lin, D., Zhong, Y., Luo, B., Liu, S., Fei, E., Lai, X., Zou, S., and Wang, S. (2019). Transmembrane protein 108 involves in adult neurogenesis in the hippocampal dentate gyrus. Cell Biosci. *9*, 9.

Zaitlen, N., and Kraft, P. (2012). Heritability in the genome-wide association era. Hum. Genet. *131*, 1655–1664.

Zhang, X., Shi, H., Wu, J., Zhang, X., Sun, L., Chen, C., and Chen, Z.J. (2013). Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. Mol. Cell *51*, 226–235.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Tumor Samples	Primary tumor samples	See experimental methods for additional details.
Normal Samples	Whole blood or surrounding normal tissue	See experimental methods for additional details.
Processed Data		
Germline Genotype Data	Affymetrix 6.0 array genotype type processed via Birdseed	https://portal.gdc.cancer.gov/ RRID:SCR_014514
Germline Genotype Data	Whole exome sequencing data	https://portal.gdc.cancer.gov/ RRID:SCR_014514
Gene expression data	RNA-seq data	https://portal.gdc.cancer.gov/ RRID:SCR_014514
MANTIS score	Middha et al., 2017	https://github.com/OSU-SRLab/MANTIS
Immune traits	(Thorsson et al., 2018)	https://doi.org/10.1016/j.immuni.2018.03.023
Haplotype reference consortium	(McCarthy et al., 2016)	http://www.haplotype-reference-consortium.org/
SNP annotations	Ensembl Variant Effect Predictor	https://grch37.ensembl.org/info/docs/tools/vep/index.html
GTEx Version 8 summary statistics	GTEx website	http://www.gtexportal.org/home/index.html RRID:SCR_0f13042
GTEx Version 8 genotypes	dbGAP	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v8.p2
TCGA Splicing data	Percent Spliced In (PSI)	https://portal.gdc.cancer.gov/ RRID:SCR_014514
DICE	Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics	https://dice-database.org RRID:SCR_018259
Crystal structure of hSTING (H232) in complex with ligand	RCSB PDB (Gao et al., 2013)	PDB: 4LOH
Crystal structure of hSTING (R232) in complex with ligand	RCSB PDB (Ergun et al., 2019)	PDB: 6DNK
Software		
PLINK 1.9	Chang et al., 2015	http://zzz.bwh.harvard.edu/plink/ https://www.cog-genomics.org/plink/
McCarthy Group tools (HRC-1000G-check-bim-v4.29) – Stranding (V1.1 HRC.r1-1.GRCh37.wgs.mac5.sites.tab)		https://wwwlwell.ox.ac.uk/~wrayner/tools/
Eagle v2.3	(Loh et al., 2016b)	https://alkesgroup.broadinstitute.org/Eagle/downloads/
Minimac3 (HRC r1.1.2016 reference panel)	(Fuchsberger et al., 2015; Howie et al., 2012)	https://genome.sph.umich.edu/wiki/Minimac3
GCTA GREML 1.91.2beta	Yang et al., 2011	https://cnsgenomics.com/software/gcta/#Download
R 3.5.0		https://www.r-project.org/
R package: snplist_0.18.1		https://cran.r-project.org/web/packages/snplist/index.html
Bioconductor package: SNPlocs.Hsapiens.dbSNP144.GRCh37_0.99.20		https://bioconductor.org/packages/release/data/annotation/html/SNPlocs.Hsapiens.dbSNP144.GRCh37.html
Bioconductor package: biomaRt_2.36.1 (Host: http://grch37.ensembl.org/index.html)	(Durinck et al., 2005, 2009)	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
Bioconductor package: GenomicRanges_1.32.6	(Lawrence et al., 2013)	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LocusZoom (Genome Build:hg19; LD Population: 1000 Genomes Project, EUR, Nov 2014)	Pruim et al., 2010	http://locuszoom.org/
eCAVIAR	(Hormozdiari et al., 2016)	http://zarlab.cs.ucla.edu/tag/ecaviar/
Chimera 1.14	(Pettersen et al., 2004)	https://www.cgl.ucsf.edu/chimera/
Deposited Code and Data		
Sayaman, et al, TCGA QC HRC Imputed Genotype Data	Sayaman et al., Immunity 2021	https://gdc.cancer.gov/about-data/publications/CCG-AIM-2020
Sayaman et al., TCGA Germline-Immune GWAS Summary Statistics	Sayaman et al., Immunity 2021	https://doi.org/10.6084/m9.figshare.13077920
Sayaman et al., Code Repository	Sayaman et al., Immunity 2021	https://github.com/rwsayaman/TCGA_PanCancer_Immune_Genetics
Sayaman et al., CRI iAtlas Interactive Visualization of Results	Sayaman et al., Immunity 2021	https://www.cri-iatlas.org/

RESOURCE AVAILABILITY

Lead contact and materials availability

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Rosalyn Sayaman (rwsayaman@gmail.com). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

A total of 11,521 genotype files from participants across 33 different cancer types included in TCGA were downloaded (Downloaded May 30, 2018 from https://portal.gdc.cancer.gov/legacy-archive). The TCGA dataset has previously been described (Liu et al., 2018; Thorsson et al., 2018). We excluded all participants with hematological malignancies (diffuse large B cell lymphoma, and acute myeloid leukemia) and thymoma since these could not be characterized for immune cell infiltration based on gene expression analyses. We included all participants who had genotype data from Affymetrix array on at least one normal sample (peripheral blood or matched normal tissue). After data cleaning (see below), the dataset for imputation included 10,128 individuals (Table S1). Further removal of samples based on genetic relatedness, availability of immune traits and covariate data resulted in a dataset of 9,603 individuals. Of these individuals, there were a total of 4,585 men and 5,018 women based on self-reported sex. The participants' age ranged from 11yr to 90yr with a median age of 61yr. Self-reported race and ethnicity were available for 8,510 samples, of which 7,073, 825, 583, 20, and 9 samples were reported to be from White/Caucasian, Black/African American, Asian, American Indian or Alaska Native, and Native Hawaiian/Pacific Islanders, respectively. Self-reported ethnicity was reported on 7,351 samples, of which 295 and 7056 samples were reported to be from Hispanic/Latino, or Not Hispanic/Latino, respectively. Of these samples, 8,204 were typed on blood-derived normal, 1,397 on solid tissue normal and 2 on buccal cell normal tissue. Institutional review boards at each of the sites that provided samples and data reviewed the consent forms and approved the use of samples.

Germline genotype data

Germline genotype data for common variants used in heritability analysis and GWAS were obtained from Affymetrix Genome Wide SNP 6.0 arrays (TCGA legacy archive https://portal.gdc.cancer.gov/legacy-archive). Birdseed genotyping files representing 905,600 variants for 11,521 samples were downloaded.

Whole exome sequencing data

Germline genotype data for rare variants were based on whole exome sequencing data (TCGA archive https://portal.gdc.cancer.gov/). Processed whole exome sequencing data for pathogenic or likely pathogenic variants from (Huang et al., 2018) were considered for rare variant analysis representing 10,389 individuals of which 9,138 had all of the phenotype and covariate information for analysis (Table S1).

Immune traits

Immune traits considered for analysis were merged from two sources from (Thorsson et al., 2018): the Feature Matrix (56 immune related features selected, Table S1) and the scores for 160 genes signatures in tumor samples (160 features, Scores_160_Signatures.tsv) across 9,769 individuals (GDC manuscript publication page https://gdc.cancer.gov/about-data/publications/panimmune).

METHOD DETAILS

Affymetrix genome-wide SNP 6.0 quality control

Birdseed files were read in R v3.5.0 using the Affymetrix SNP Array 6.0 (release 35) annotation file, and 905,422 variants were successfully loaded and analyzed in PLINK version 1.9. Samples were cross-referenced against previously included genotyping samples (Thorsson et al., 2018). Based on established TCGA barcode identifiers, samples annotated with Analyte code "G" (Whole Genome Amplification) were further excluded. A final set of 10,946 included samples with Analyte code "D" (DNA) were retained for quality assessment.

Stringent quality control measures were applied to the SNP genotyping data (Figure 1, top QC panel). SNPs and individuals with greater than 5% missingness were excluded; leaving a total of 861,351 variants and 10,917 samples for subsequent analysis.

Initial PCA ancestry analysis was performed to facilitate heterozygosity calculations. PCA without LD pruning was performed in PLINK 1.9 (Chang et al., 2015), and visual examination of the concordance of the principal component plots with the self-reported race and ethnicity annotations revealed that the first 3-4 PCs captured population structure information, while PCs 5-6 captured outliers. PCA initial ancestry clusters were determined by performing both k-means and partition around medoids (PAM) clustering on either the first three or first four PCs as previously described (Carrot-Zhang et al., 2020) We computed gap statistics and average silhouette widths iteratively for the number of clusters, k = 1 to 10 for k-means and PAM methods respectively to find the optimal number of clusters for each method. We found PAM using the three PCs yielding 4 optimal clusters to show high concordance with self-reported race/ethnicity (ancestry cluster 1 = European, cluster 2 = Asian, cluster 3 = African, cluster 4 = American). Based on the initial ancestry cluster assignments, heterozygosity was calculated in PLINK 1.9 within each initial PCA-based ancestry cluster and a total of 250 samples with heterozygosity > 3*SD above the ancestry mean were removed.

Selection of a representative sample for each individual was then conducted. For individuals represented by more than one sample, blood-derived normal samples were preferentially selected; for those with more than one blood-derived samples, samples with higher call rates were retained. After these steps, a total of 10,128 unique individuals remained for subsequent analysis.

Final filtering steps for SNPs were conducted across the 10,128 unique individuals and restricted to autosomal chrs. Hardy-Weinberg Equilibrium (HWE) was calculated in PLINK 1.9 across individuals within the largest ancestry cluster (European ancestry cluster 1). SNPs that deviated from the expectation under HWE ($p < 1x10^{-6}$) within the European ancestry cluster were excluded with the exception of SNPs previously associated with any cancer as reported in the GWAS catalog ($p < 5x10^{-8}$) (Rashkin et al., 2019) since they may deviate from HWE in cancer patients. Minor allele frequency (MAF) was calculated and variants with MAF < 0.005 were excluded. Finally, duplicate SNPs with identical genomic first position were removed. A total of 838,948 autosomal chr variants for 10,128 unique individuals passed the aforementioned QC steps.

Stranding and reference panel imputation

The quality-controlled genotyping file was stranded and imputed against the Haplotype Reference Consortium (HRC) (Loh et al., 2016a) (McCarthy et al., 2016). Prior to HRC stranding, all palindromic SNPs (A/T or G/C) were removed. Stranding was then performed using the McCarthy Group tools (HRC-1000G-check-bim-v4.29), which compares our data genotyping alleles to the corresponding SNP alleles from HRC (v1.1 HRC.r1-1.GRCh37.wgs.mac5.sites.tab), leaving 680,389 correctly matched variants for imputation.

Phasing and imputation were performed using a standard pipeline on the Michigan Imputation Server (MIS). Phasing was performed using Eagle version v2.3 (Loh et al., 2016b) on the variant call file (VCF). To reduce the run time, the VCF file was divided into 22 files corresponding to individual autosomal chrs. By default, Eagle restricts analysis to bi-allelic variants that exist in both the target and reference data. Minimac3 was used to run the imputation. For each of the 22 VCF files, the MIS breaks the dataset into non-overlapping chunks prior to imputation. For HRC imputation, the HRC r1.1.2016 reference panel was selected using mixed population for QC, with a total of 39,127,678 SNPs returned after imputation.

Final ancestry calls

PCA was performed on the final quality-controlled genotyping file and final PAM-based ancestry clusters were computed as previously described (Carrot-Zhang et al., 2020) for the 10,128 individuals for optimal k = 4. We found very high concordance of initial and final ancestry assignments (99.98% matching, the 2 samples varying between initial and final ancestry cluster computation assigned to NA).

The four ancestry cluster are as follows: (1) PAM ancestry cluster 1 is concordant with European ancestry, capturing 97.27% of individuals self-reporting as White, as well as 82.16% of individuals with self-reported non-Hispanic/non-Latino ancestry and 45.96% with self-reported Hispanic/Latino ancestry; (2) ancestry cluster 2 with African ancestry, capturing of 97.53% of individuals self-reporting as Black/African-American race; (3) ancestry cluster 3 with Asian ancestry, capturing 90.88% of individuals self-reporting as Asian and 88.89% self-reporting as Native Hawaiian/Pacific Islander; and (4) ancestry cluster 4 with a subgroup of individuals with American ancestry capturing 60% of individuals self-reporting as American Indian /Alaska Native and 47.2% with self-reported Hispanic/Latino ethnicity (GDC Publication Page Figure S1-B, https://gdc.cancer.gov/about-data/publications/CCG-AIM-2020).

PC1-7 showed further population sub-structure in the Asian and European ancestry clusters (GDC Publication Page Figure S2, https://gdc.cancer.gov/about-data/publications/CCG-AIM-2020). PAM ancestry sub-clusters were computed using PC1-7 for individuals within the Asian ancestry cluster which yielded two optimal sub-clusters (GDC Publication Page Figure S2-A), and within the

European ancestry cluster which yielded three optimal sub-clusters (GDC Publication Page Figure S2-B). Of note, 72.46% of European sub-cluster 3 self-reports as Asian (15.94% have no race reported). Ancestry clusters, sub-clusters, self-reported race and ethnicity and PC1-7 are provided for each individual (Table S1).

Feature selection for analysis

Immune traits considered for analysis were merged from two sources in Thorsson et al. (Thorsson et al., 2018): the Feature Matrix (Table S1 of Thorsson et al., 56 immune related features selected) and the scores for 160 genes signatures in tumor samples (160 features, Scores_160_Signatures.tsv on GDC manuscript publication page https://gdc.cancer.gov/about-data/publications/panimmune).

The 216 features were then filtered at three levels: (i) Level 1 filtering removed redundant features based on overlap of the feature matrix and the 160 signature feature set; (ii) Level 2 filtering removed features with limited interpretability; (iii) Level 3 filtering removed features with highly skewed distributions, which would not be amenable to subsequent analyses. A final set of 139 features was used in subsequent germline analysis. Table S2 lists all 216 previously well-characterized immune traits, and the final 139 immune traits selected (Figure 1, bottom Immune Traits panel). These selected immune phenotypes encompass six broadly defined immune trait phenotype categories: (1) Leukocyte subset ES, which include 24 immune cell-specific activation scores as captured by single-sample ssGSEA (Bindea et al., 2013); (2) Leukocyte Subset Percentages (%), which include 26 immune cell relative proportion measures (22 individual cells and 4 aggregates, as estimated by CIBERSORT) (Gentles et al., 2015; Thorsson et al., 2018); (3) Overall Proportion, which includes three measures, namely leukocyte fraction, stromal fraction, and tumor infiltrating leukocyte (TIL) regional fraction; (4) Adaptive Receptor, which includes four scores related to TCR and BCR Shannon diversity and richness; (5) Expression Signature, which includes four ssGSEA scores specific to lymphatic vessels (Bindea et al., 2013), antigen-presenting machinery (APM1 and APM2) (Şenbabaoğlu et al., 2016), and angiogenesis (Şenbabaoğlu et al., 2016), a collection of 68 gene signatures related to immunomodulatory signaling including IFN signaling, TGF-β, wound healing (core serum response) and T/B cell response cataloged from earlier studies (Amara et al., 2016) (Wolf et al., 2014), and an Immunologic Constant of Rejection (ICR) signature summarizing a Th1/cytotoxic polarization of the tumor microenvironment associated with favorable prognosis and responsiveness to immunotherapy (Galon et al., 2013; Hendrickx et al., 2017; Roelands et al., 2020; Rozenblit et al., 2019); and (6) Attractor Metagene, which includes nine TCGA-based co-expression signatures (metagene attractors) (Cheng et al., 2013a, 2013b) (Table S2). Eight immune traits are represented by single genes: Treg cells (*FOXP3*), CD68, CD8A, PD1 data (*PDCD1*), PDL1 data (*CD274*), CTLA4 data (*CTLA4*), TREM1 data (*TREM1*), and DAP12 data (*TYROBP*).

We used the term immune trait ''categories'' to describe the methodological origin of the immune trait measures including the six categories above, e.g., 1. Leukocyte subset ES (ssGSEA), 2. Leukocyte Subset Percentages (%), etc. We used the term ''module'' to describe the grouping of immune traits that were generated based on clustering (Figure S1). We used the term ''immune subtype'' to describe the immune grouping of samples previously characterized by Thorsson et al., (Thorsson et al., 2018) and used the same sample assignments used in that report.

Covariate selection

For each analysis, we included age, sex, cancer type, and genetic ancestry from the PCA (PC1-7) as covariates, unless otherwise indicated. Self-reported age (age at diagnosis in years) from clinical data and PC1-7 were used as continuous covariates. Cancer type based on TCGA study assignment and curated genotype-imputed sex assignments were used as categorical covariates. Due to missing self-reported sex, and discrepancy in self-reported and genotype-type based sex in the data, we carefully curated sex assignments. We recovered sex information by using X chr homozygosity estimate (XHE) after removal of the pseudo-autosomal region of the X chr. For those with missing sex information, individuals with XHE < 0.2 were assigned as females and individuals with XHE > 0.8 were assigned as males (N = 21 males, N = 28 females). Self-reported sex assignments were curated with individuals self-reporting as males with XHE < 0.2 reassigned as female (N = 20 reassigned as females), and individuals self-reporting as females with XHE > 0.8 reassigned as males (N = 6 reassigned as males). For brevity in the remainder of the manuscript, we refer to curated genotype-imputed sex assignments simply as sex. For a subset of analyses, we also included immune subtype as a covariate, as indicated in the text.

QUANTIFICATION AND STATISTICAL ANALYSIS

Immune trait correlations and clustering

We calculated Pearson's correlation coefficients across all pairs of the 139 immune traits from 9,769 individuals. We then used hierarchical clustering, using 1-correlation as distance metric and complete agglomerative clustering method (R heatmap.2 function in gplots package), to identify modules of the immune traits.

Immune trait normalization for heritability, GWAS, and rare variant analysis

For immune phenotypes that appeared approximately normally distributed or normally distributed after correction for immune subtype, we calculated heritability using the immune phenotype without any transformation/normalization. For immune phenotypes that were highly skewed we applied a $\log_{10}$ transformation, and those that were approximately normally distributed after transformation were analyzed as log-transformed values. For immune phenotypes that could not be normalized with $\log_{10}$ transformation (usually

due to a large number of 0 values), we dichotomized them at the median (i.e., higher and lower than median). These phenotypes were treated as binary variables in subsequent analysis. Traits that underwent transformation and dichotomization are indicated in Table S2.

Germline analysis

To examine the contribution of germline genetic variation to the functional orientation of the immune microenvironment, we conducted three types of analyses: (1) parallel heritability analysis (N_{EUR} = 7,813, N_{AFR} = 863, N_{ASIAN} = 570, N_{AMR} = 209 individuals), (2) GWAS (N = 9,603), and (3) rare variant analysis (N = 9,138) across 30 different non-hematological cancer types in TCGA (Figure 1, middle Analysis panel). Number of individuals, N, for each analysis type represented the maximum number of samples, however analyses per immune trait proceeded using fewer samples if NA values for specific individuals were present in a given immune trait.

Heritability analysis

Estimates of genome-wide heritability of the 139 described immune traits were calculated using GCTA GREML approach implemented in GCTA 1.91.2beta, which simultaneously models the effect of all genetic variants (MAF > 0.01) (Yang et al., 2010, 2011). GREML calculates a genetic relatedness matrix (GRM) as a measure of the genetic similarity of unrelated individuals (GRM < 0.05) and compares it to the similarity of the measured immunological traits to calculate the total narrow-sense contribution of genotypic variance to overall phenotypic variance, V(Genotype)/V(Phenotype) (Yang et al., 2010, 2011). All GREML analyses used the default average information (AI) algorithm to run REML iterations.

Since calculation of the GRM results in biased relatedness estimates for pairs of individuals who have different ancestry, we restricted heritability analysis primarily within the European ancestry group (N_{EUR} = 7,813, M = 701,189 SNPs, K = 32,292,666 GRM elements), which constitutes the largest ancestry population; secondary analysis is performed in the smaller African (N_{AFR} = 863, M = 791,989, K = 409,060), Asian (N_{ASIAN} = 570, M = 649,768, K = 183,315), and American (N_{AMR} = 209, M = 751,507, K = 24,753) ancestry groups. The four genetic ancestry groups were derived from optimal partition around medoids (PAM) clustering (Figure S2A) of individuals in principal components 1-3 based on their genotyping data (Figure S2B).

Heritability analyses are generally only well-powered for sample sizes of > 1000; therefore, only the European ancestry subgroup was adequately powered and were presented in the main results. However, since we have previously seen associations between several of these immune parameters with various ancestral populations (Thorsson et al., 2018), we performed the heritability analyses separately in African, Asian, and American ancestral populations and present these in the supplementary results. To reduce bias in the heritability estimates, we removed one of each pair of related individuals with A_{jk} > 0.05 (calculated from SNPs with MAF > 0.01) prior to running GREML. We calculated heritability using an unconstrained approach (allowing heritability to be < 0). Constraining the heritability to a range of 0-1 may lead to an upward bias of the low heritability values, which is likely to be worse in smaller datasets. We used the likelihood ratio test (LRT) implemented in GREML to test if heritability is different than zero for each of the immune traits analyzed and used Benjamini-Hochberg adjustment (Benjamini and Hochberg, 1995) to calculate the FDR. We present both FDR adjusted *p* values and unadjusted *p* values in the manuscript. All heritability analyses were run with age, cancer type, sex and PC 1-7 as covariates unless otherwise indicated.

We also used GREML to determine whether there are any contextual factors that interact with genome-wide common variant effects, including the major immune subtypes as determined by Thorsson et al. and somatic mutations (divided into tertiles and dichotomized at 10 MB). We implemented the gene x environment (GxE) feature calculation in the European ancestry cluster in GREML. For those immune traits for which we found nominally significant ($p < 0.05$) interactions, we calculated heritability in each stratified subset, as well as with immune subtype as an additional covariate. For GxE calculations, the LRT tests the significance of the variance of GxE interaction effects.

Genome-wide association studies (GWAS)

We selected each of the immune phenotypes that demonstrated nominally significant genome-wide heritability (N = 33) for GWAS. GWAS was conducted on all of the genotyped SNPs that passed QC and all of the imputed SNPs that had imputation R^2 > 0.5 and minor allele frequency > 0.005 in the 9,603 unrelated individuals (PLINK 1.9 identity by descent, IBD, pihat < 0.25). Minor allele frequencies were recalculated post-imputation for only the subset of 9,603 individuals (PLINK 1.9). Of the 39,127,678 SNPs available after imputation, 10,955,441 passed both imputation quality and frequency thresholds and thus were included in the association analysis.

GWAS was performed using PLINK 1.9. Immune phenotypes that were approximately normally distributed or normally distributed after stratification by covariates were tested for association with SNPs using linear regression with age, sex, cancer type, and PC1-7 as covariates. Immune traits that were dichotomized for heritability analyses were analyzed using logistic regression models, with the same covariates. For each GWAS we also calculated the genome-wide inflation coefficient (lambda). We used the traditional cutoff of $p < 5x10^{-8}$ as a cutoff for genome-wide significance and $p < 1x10^{-6}$ to denote suggestive loci. Since we only selected the subset of phenotypes that were heritable and since many of the phenotypes were highly correlated, we did not correct the GWAS for the number of phenotypes analyzed. SNPs were annotated based on spanned genomic ranges (R v3.5.0, Bioconductor package GenomicRanges_1.32.6) with rsIDs (R v3.5.0, R package snplist_0.18.1, Bioconductor package SNPlocs.Hsapiens.dbSNP144.GRCh37_0.99.20) and with genes within ± 50KB, ± 500KB, and ± 1MB of the SNP (R v3.5.0, Bioconductor package biomaRt_2.36.1 using http://grch37.ensembl.org/index.html as host). Variant annotations for all genome-wide and suggestive SNPs were determined

using the web interface of the Ensembl Variant Effect Predictor (VEP, https://grch37.ensembl.org/info/docs/tools/vep/index.html). All annotations were based on *Homo sapiens* (human) genome assembly GRCh37 (hg19) from Genome Reference Consortium. All association statistics for the GWAS are available at Figshare https://doi.org/10.6084/m9.figshare.13077920.

To plot the individual locus results we used LocusZoom (Pruim et al., 2010), which shows LD estimates (r^2 color map) and recombination rates (blue line) around the genome-wide significant loci. The online version of LocusZoom did not plot the LD information for *TMEM173*; therefore, we downloaded a standalone version and ran it locally using LD information computed in TCGA data.

Within cancer association tests for forest plots were run with age, sex and PC1-7 as covariates, except in CESC, OV, PRAD, TGCT, UCEC and UCS where only age and PC1-7 were used. The 30 cancer types included in our analyses are abbreviated as follows: ACC: Adrenocortical Carcinoma; BLCA: Bladder Urothelial Carcinoma; BRCA: Breast Invasive Carcinoma; CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma; CHOL: Cholangiocarcinoma; COAD: Colon Adenocarcinoma; ESCA: Esophageal Carcinoma; GBM: Glioblastoma; HNSC: Head and Neck Squamous Cell Carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney Renal Clear Cell Carcinoma; KIRP: Kidney Renal Papillary Cell Carcinoma; LGG: Low Grade Glioma; LIHC: Liver Hepatocellular Carcinoma; LUAD: Lung Adenocarcinoma; LUSC: Lung Squamous Cell Carcinoma; MESO: Mesothelioma; OV: Ovarian Serous Cystadenocarcinoma; PAAD: Pancreatic Adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate Adenocarcinoma; READ: Rectum Adenocarcinoma; SARC: Sarcoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach Adenocarcinoma; TGCT: Testicular Germ Cell Tumors; THCA: Thyroid Carcinoma; UCEC: Uterine Corpus Endometrial Carcinoma; UCS: Uterine carcinosarcoma; UVM Uveal Melanoma.

Rare variant analyses

For rare variant analysis, we focused on well-annotated, germline pathogenic or likely pathogenic cancer predisposition variants as previously defined (allele frequency in 1000 Genomes and ExAC (release r0.3.1) < 0.05%) (Huang et al., 2018).

Exome files related to samples for which all the covariates (age, sex, cancer type, and PC1-7) and at least one immune trait was available were retained (N = 9,138). There were 832 pathogenic/likely pathogenic SNPs/Indels events with at least one copy of rare allele in the whole exome sequencing data, corresponding to 586 distinct pathogenic SNPs/Indels mapping to 99 genes.

We performed a pathway burden analysis using selected pre-defined biological pathways such as DNA damage repair and oncogenic processes, pan-cancer and per cancer (Bailey et al., 2018; Huang et al., 2018; Knijnenburg et al., 2018). These pathways were manually curated to generate a list of mutually exclusive categories (genotypic variables). The only genes that were not collapsed into pathways were *BRCA1* and *BRCA2* for which a sufficient number of events across cancers exist. Overall, 21 genotypic variables were used for analyses (Figure S7A). In the pan-cancer analysis, we only included genotypic variables with number of events (mutations) greater than 4 across cancers, including a total of 90 genes. In the per-cancer analysis, we only included genes or pathways with at least 3 events in the analyses. For each pathway, variants that fall within its selected set of genes were collapsed based on the presence or absence of any rare variant (i.e., 0 if no rare variant was present and 1 if there is at least one variant). We conducted regression analyses (linear or logistic, as done for GWAS) to assess the association between the pathways' burden of rare variants and immune traits. Traits assessed in these analyses were the same as the ones used for heritability analyses, with the addition of the immune subtypes (C1, C2, etc.), DNA-alteration related metrics such as the mutational load, the neoantigen load, the degree of copy number alterations (Thorsson et al., 2018) and the MANTIS score (threshold = 0.4, Middha et al., 2017) (full list in Table S2). All pan-cancer regression models included the following covariates: age, sex, cancer type, and PC1-7. The same covariates were used for per-cancer regression analyses, except for CESC, OV, PRAD, TGCT, UCEC and UCS where only age and PC1-7 were used.

To check whether the results were driven by the mutational load, we ran regression models that include this variable as a covariate in the regression model. Additionally, in the BRCA cohort, additional analyses were performed stratifying for basal-like and non-basal-like subtypes and adding the molecular subtype (basal-like and non-basal like) as a covariate (see Table S6).

In the pan-cancer analysis, we used a Benjamini-Hochberg FDR (Benjamini and Hochberg, 1995) to correct for multiple hypothesis testing, accounting for all 21 genes and pathways tested and 154 phenotypes (139 immune traits, 9 DNA related metrics, and 6 immune subtypes). We used a cutoff of FDR $p < 0.1$ to identify significant gene/pathway-immune trait associations and a threshold of nominal $p < 0.005$ (FDR $p < 0.25$) to identify suggestive associations. We used a more permissive cut-off in these analyses than the ones used in the heritability and GWAS to reduce type II error due to the low number of events (germline mutations). In addition, leukocyte fraction and non-silent mutation rate were compared by categories defined by combining germline mutation status across MMR genes (MMR mutated versus MMR wild-type) and somatic MSI status (MSI-H versus MSS, as identified by MANTIS score, threshold = 0.4 (Middha et al., 2017)). For such comparisons, regression analyses were adjusted for sex, age, and PC1-7. All rare variant analyses were performed using R (http://www.R-project.org/).

Epigenome chromatin states

Mnemonic bed files for the Expanded-18-state model, which takes into account six chromatin marks from ChIP-seq datasets (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3, H3K27ac) (Roadmap Epigenomics Consortium et al., 2015), were downloaded for 98 annotated epigenomes (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#exp_18state). We identified 25 of these epigenomes to be specifically immune-related, including those associated with primary immune cells, as well as bone marrow-derived mesenchymal cells, leukemia-associated cells, thymus, and spleen (Table S4). The first 12 states were defined to be active states associated with expressed genes, while the last six states were defined to be inactive or repressed states. Epigenome IDs were mapped to corresponding Standardized Epigenome Names (https://docs.google.com/spreadsheets/d/

1yikGx4MsO9Ei36b64yOy9Vb6oPC5IBGlFbYEt-N6gOM/edit#gid=15). The chromatin state at each significant and suggestive SNP was extracted based on positional overlaps with each epigenome.

In silico Analysis of Non-synonymous amino acid substitutions

We used Chimera 1.14 to display the structure of the STING protein with the H232 and R232 alleles and to align these to the ligand (Pettersen et al., 2004). We aligned the A chains of the structures from the PDB files 4LOH (containing the H232 allele) (Gao et al., 2013) and 6DNK (containing the R232 allele) (Ergun et al., 2019) using the Matchmaker function from Chimera.

Gene expression and splice quantitative trait locus analysis, and Colocalization

We performed eQTL and sQTL analyses in TCGA and used summary statistics from the GTEx datasets to search for potential candidate genes. We excluded the *HLA* and *IL17RA* loci since SNPs at these loci are known eQTLs for genes that are part of the immune trait. For the significant and suggestive SNPs, we tested all genes within ± 1MB for eQTL and all transcipts within ± 500KB for sQTL. We used a shorter range for sQTLs with the assumption that SNPs affecting splicing are likely to act at a shorter distance.

TCGA dataset

RNA-seq gene expression and splicing data were downloaded from the NIH Genomics Data Commons (https://gdc.cancer.gov/about-data/publications/pancanatlas and https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Splicing-2018 (Kahles et al., 2018). For the sQTL analysis, we considered the following splicing categories: 3′, 5′, exome skipping, intron retention, and mutually exclusive exon events quantified by the Percent Spliced In (PSI) (Kahles et al., 2018). Only splicing events with more than 800 non-missing observations (~10% of the total data) were considered. Association analyses between either gene expression or PSI and the imputed SNPs were performed using linear regression using age, sex, cancer type, and PC1-7 as covariates. We calculated FDR for each SNP separately, under the assumption that the SNP was already either significant or suggestive, and thus we had to correct for each of the genes at the locus but not all of the other SNPs (Table S5). We then selected the SNP-gene expression (eQTL) or SNP-gene splicing (sQTL) pairs with FDR $p < 0.1$ for further colocalization analysis.

GTEx dataset

We downloaded all summary statistics for expression quantitative loci (eQTL - GTEx_Analysis_v8_eQTL_all_associations), and splicing quantitative loci (sQTL - GTEx_Analysis_v8_sQTL_all_associations) from GTEx project (https://console.cloud.google.com/storage/browser/gtex-resources) using the results from the latest version of the GTEx database (Version 8). For each SNP that had a genome-wide significant or suggestive association with one of the 33 immune traits by GWAS, we extracted all of the association statistics from the summary statistics for eQTLs within ± 1MB and for sQTLs within ± 500 KB from all tissues in the GTEx summary statistics dataset. We then calculated FDR for each SNP, correcting for all of the genes at the locus across all tissues as we did for TCGA. For eQTLs and sQTLs that had FDR $p < 0.1$, we pursued colocalization as below. TCGA GWAS summary statistics are annotated in Build 37, GTEx QTL summary stats are annotated in Build 38, when appropriate, liftover from Build 38 to 37 are provided using R/Bionconductor packages AnnotationHub (v2.12.1) (AH14150 chain file) and rtracklayer (v1.40.6). In the GTEx summary file (Table S5) we annotated both Build 37 and Build 38 positions.

Colocalization analysis

We performed colocalization analysis using eCAVIAR (Hormozdiari et al., 2016) on both TCGA and GTEx results. eCAVIAR computes a posterior probability of causality (CLPP) based on association data and LD structure for the eQTL or sQTL and the trait GWAS and then calculates the joint probability of both of these being causal. It requires both summary statistics from GWAS and from the eQTL or sQTL analysis and the LD matrix of SNPs used in both analyses. For TCGA, we began with all SNPs that had FDR $p < 0.1$ with at least one gene and/or transcript and computed the eQTL and sQTL associations for the surrounding SNPs from the index SNP for that same gene/transcript using the same approach as outlined above. For GTEx, we began with SNP-gene expression or SNP-gene splicing pairs that met our FDR $p < 0.1$ criteria and extracted the eQTL and sQTL results for the surrounding SNPs from the summary results. For the GWAS and TCGA analyses, we calculated the genotype correlation (r) at each locus from the genotype data. For the GTEx analysis, we downloaded the individual genotype data from dbGAP for GTEx participants (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v8.p2) and calculated genotypic correlation between SNPs in R. We then ran eCAVIAR separately for each FDR $p < 0.1$ eQTL and sQTL association from TCGA and GTEx, considering models at each locus that assume one or two causal variants. For each SNP-gene expression or SNP-gene splicing category pair, 200 SNPs (± 100 SNPs) around the index SNP were included in colocalization analysis. The CLPP of each SNP being causal was calculated, and also a regional CLPP by summing all 201 SNP CLPPs. We used a posterior probability of > 0.01 to consider plausible colocalization, including both the 1 and 2 locus model and considering the sum of the posterior probability SNPs in the colocalization results.

Expanded region criteria for colocalization

Since eCAVIAR identified multiple genes at the same locus for many loci that have plausible colocalization within a ± 100 SNP boundary, we sought stronger evidence for colocalization at the loci where eCAVIAR found colocalization by examining an expanded region. We reasoned that a gene or transcript that is causal for the immune trait should not be more strongly associated with another SNP in the region that has little or no evidence of association with the immune trait. Therefore, for each gene or splice variant that had

plausible colocalization by eCAVIAR, we performed an expanded region search (± 1MB for eQTL and ± 500KB for sQTL) to see if we can identify one or more SNPs that had a stronger effect in the eQTL or sQTL analysis in the same tissue/dataset, which we called ''counter-evidence'' SNPs. If eCAVIAR produced plausible evidence of colocalization (posterior probability > 0.01) and we could find no SNPs that met our counter-evidence criteria in the expanded region, we considered the expanded region evidence for colocalization as strong. If we did find SNPs that met our counter-evidence criteria in the expanded region, then we compared the significance level for the eQTL or sQTL association of the counter-evidence SNP versus the eQTL or sQTL association with the index SNP (associated with the immune trait). If the counter-evidence SNP association with eQTL or sQTL had a $-\log_{10} p$ value that was $\leq$1.5 higher than the index SNP (GWAS significant SNP for the immune trait), then we considered the expanded region evidence as intermediate. If the difference in $-\log_{10} p$ values was > 1.5, we considered the expanded region analysis to be negative.

To visualize the colocalization in the expanded region, we generated plots that show the $-\log_{10} p$ QTL versus $-\log_{10} p$ GWAS for all of the GWAS significant SNPs with CLPP > 0.01. The plots included the association p values for all of the SNPs at ± 1MB for eQTL and at ± 500KB for sQTL from the gene which had a CLPP > 0.01. These plots are available at Figshare (GTEX expanded region analysis plots: https://doi.org/10.6084/m9.figshare.13089341; TCGA expanded region analysis plots: https://doi.org/10.6084/m9.figshare.13090031. We color-coded these plots with the LD, based on the LD matrix from the TCGA. Counter-SNPs are found in the top left corner of these plots (i.e., strong association with the eQTL or sQTL but no association with the immune trait). Conversely if there were no counter-SNPs, then the strongest SNPs for association with the immune trait were also the strongest SNPs for the association with the eQTL or sQTL.

DATA AND CODE AVAILABILITY

The TCGA Genome Wide SNP 6.0 birdseed genotyping data and clinical data can be found at the legacy archive of the GDC (https://portal.gdc.cancer.gov/legacy-archive). Access to the TCGA original birdseed and pre-processed quality controlled genotyping data imputed to HRC generated in the current manuscript ("Sayaman et al TCGA QC HRC Imputed Genotyping Data") requires GDC controlled access permission approval. The quality controlled and HRC imputed genotyping data were contributed towards ancestry analyses (Carrot-Zhang et al., 2020) and are accessible at the GDC publication page (https://gdc.cancer.gov/about-data/publications/CCG-AIM-2020). Please cite the current manuscript (Sayaman, Saad et al., Immunity 2021) when using the quality controlled and HRC imputed genotyping data. The summary statistics from the GWAS have been deposited to FigShare (https://doi.org/10.6084/m9.figshare.13077920). Details for software availability are in the Key Resources Table. The code generated during this study has been deposited to github (https://github.com/rwsayaman/TCGA_PanCancer_Immune_Genetics). Interactive visualization is available at CRI iAtlas (https://www.cri-iatlas.org/).