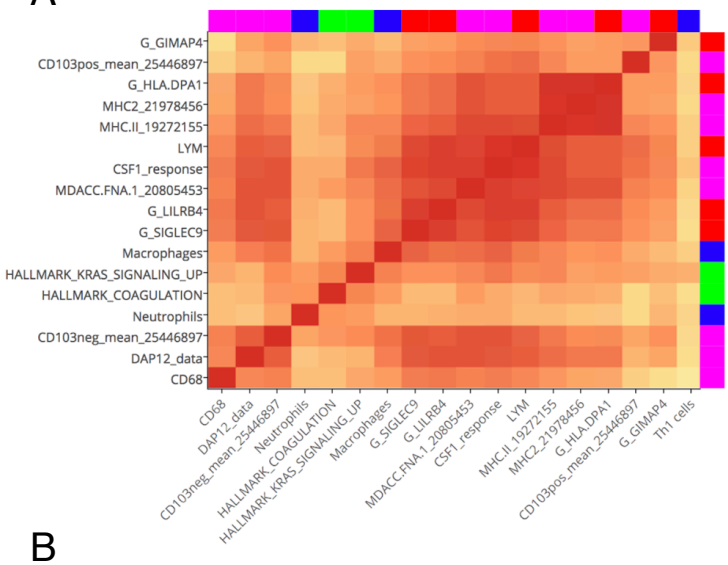
A



B

D

	Basal	Her2	LumA	LumB	Normal
C1	59	32	174	62	42
C2	104	36	116	89	46
C3	2	2	147	7	33
C4	6	2	42	30	11
C5	0	0	0	0	0
C6	1	1	29	3	6

C

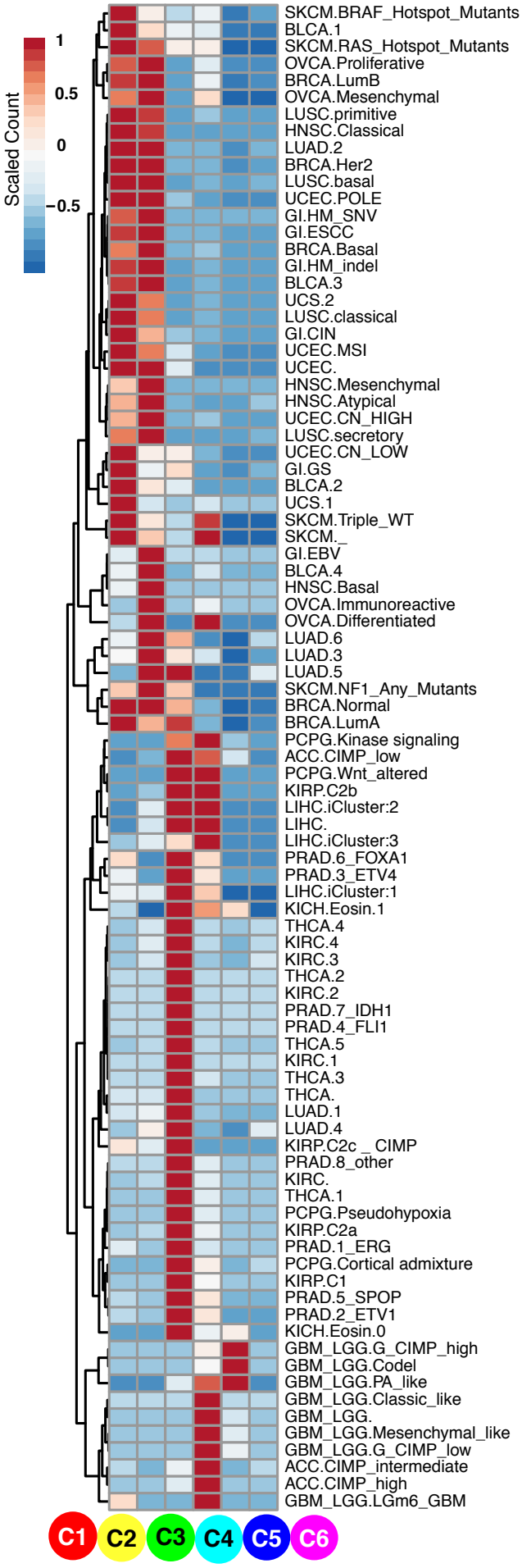
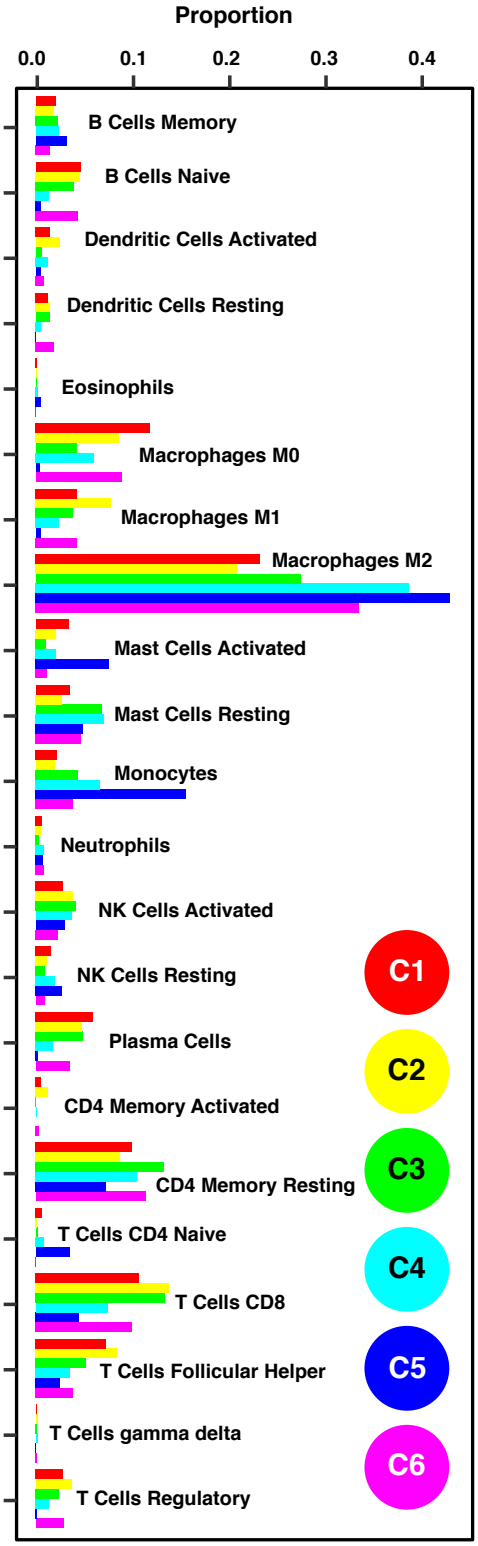
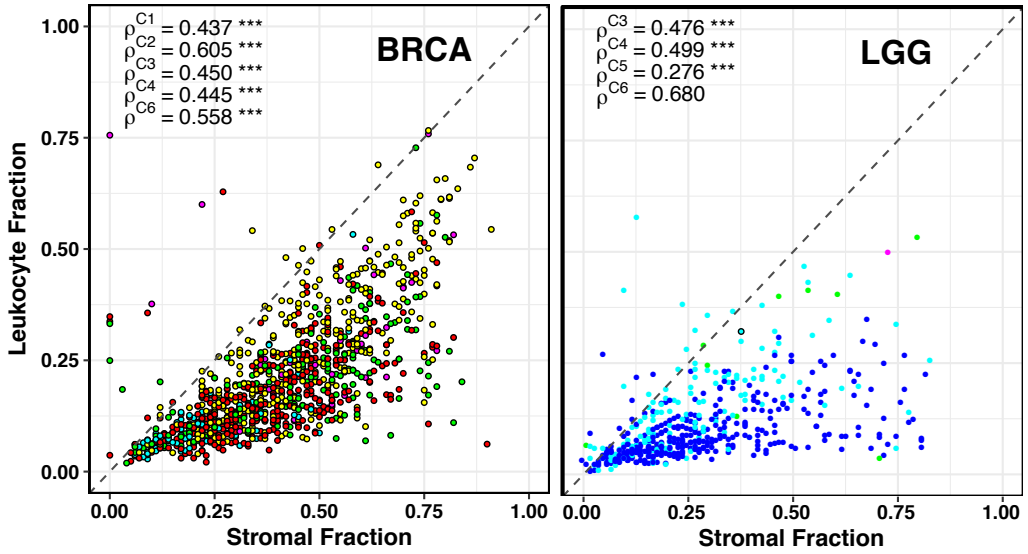


Figure S1. Identification and Characterization of Immune Subtypes derived from Model Based Clustering. Related to Figure 1. A Macrophage co-expression cluster. A detail of the lower left of Figure 1A, illustrating macrophage/monocyte/MHC cluster module. Signatures are labeled as rows and columns, with red denoting high intercorrelation among signature scores over all TCGA tumors. Colored side bars reflect different published sources of immune signatures. **B.** Relative contribution of tumor types to immune subtype composition. Bars indicate relative proportions of each tumor type (rows) in immune subtypes (columns). Column width indicates relative numbers of tumors comprising of each immune subtype. **C.** TCGA subtypes and immune subtypes**.** Each row corresponds to a TCGA subtype, and the distribution of immune subtypes within that subtype, standardized to mean 0 and unit variance is shown. Red indicates greater proportion of that immune subtype is present in a given TCGA subtype. **D.** TCGA breast cancer subtypes (columns) and corresponding immune subtypes (rows).

A



B



C

Study	ρ	p-value
BLCA	0.45	5.2×10^{-16}
BRCA	0.41	3.5×10^{-38}
CESC	0.37	7.2×10^{-9}
COAD	0.25	2.4×10^{-7}
LUAD	0.40	5.5×10^{-17}
LUSC	0.42	9.3×10^{-18}
PAAD	0.20	0.019
PRAD	0.21	0.00041
READ	0.35	2.2×10^{-5}
SKCM	0.45	1.0×10^{-20}
STAD	0.26	7.0×10^{-6}
UCEC	0.33	4.2×10^{-13}
UVM	0.10	0.42

D

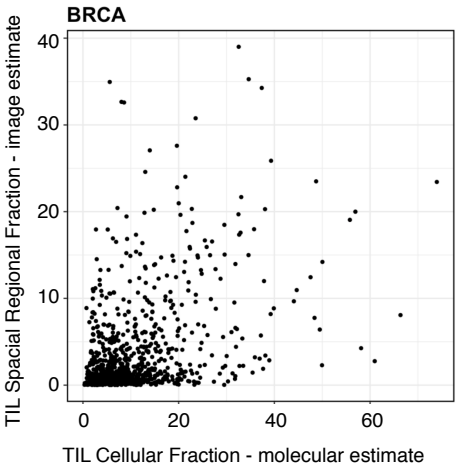


Figure S2. Composition of the Immune Infiltrate. Related to Figure 2. A. Relative proportions of all leukocyte classes within each immune subtype based on CIBERSORT analysis, displayed as mean cell type propensity. **B.** LF for BRCA and LGG (y-axes) is compared with the non-tumor cellular stromal fraction (SF) in the tumor microenvironment (x-axes). Immune subtypes within BRCA and LGG differ in leukocyte proportions (ρ=LF/SF) as evidenced in statistical differences in the estimated slopes as designated in the insets. The immune subtype of tumor sample points is indicated with the color scheme seen in A, with *** denoting linear regression significance at $p<0.001$. **C & D.** The fraction of tumor infiltrating lymphocytes estimated from digital analysis of H&E slide images is compared with the estimated lymphocytic content (calculated from CIBERSORT and methylation data). The table (S2C) summarizes the correspondence (Spearman rho) for the 13 TCGA tumor types analyzed, and the image (S2D) illustrates the comparison for BRCA tumor samples, with TIL spatial fraction from slides on the y-axis, and TIL fraction molecular estimates on the x-axis.

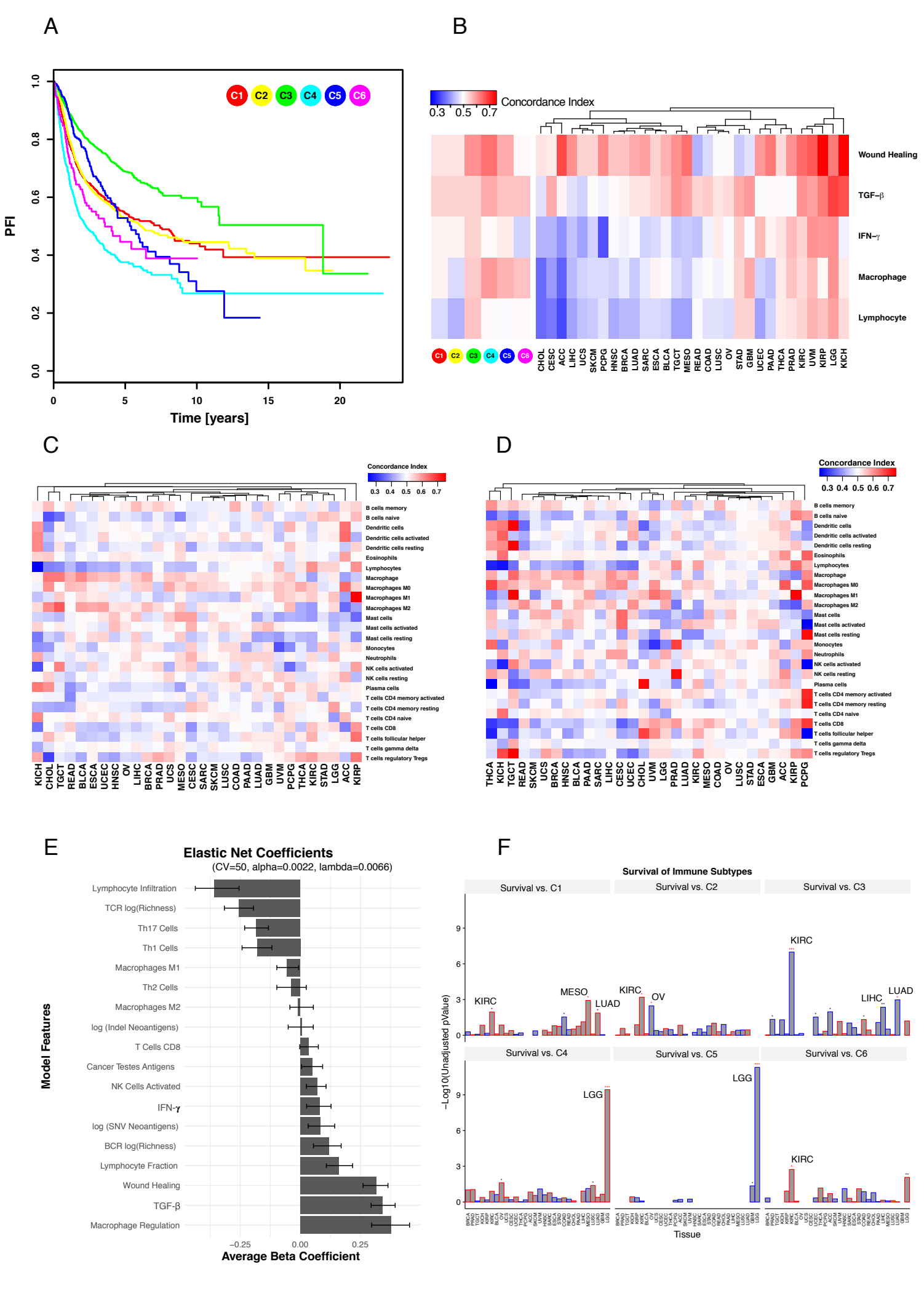

A
PFI
Time [years]
C1
C2
C3
C4
C5
C6
B
Concordance Index
Wound Healing
TGF−β
IFN−γ
Macrophage
Lymphocyte
C
Concordance Index
D
Concordance Index
E
Elastic Net Coefficients
(CV=50, alpha=0.0022, lambda=0.0066)
Model Features
Lymphocyte Infiltration
TCR log(Richness)
Th17 Cells
Th1 Cells
Macrophages M1
Th2 Cells
Macrophages M2
log (Indel Neoantigens)
T Cells CD8
Cancer Testes Antigens
NK Cells Activated
IFN-γ
log (SNV Neoantigens)
BCR log(Richness)
Lymphocyte Fraction
Wound Healing
TGF-β
Macrophage Regulation
Average Beta Coefficient
F
Survival of Immune Subtypes
Survival vs. C1
Survival vs. C2
Survival vs. C3
Survival vs. C4
Survival vs. C5
Survival vs. C6
−Log10(Unadjusted pValue)
Tissue

Figure S3. Immune Response and Prognostics. Related to Figure 3. A. Progression free interval (PFI) among immune subtypes. **B.** Concordance Index (CI) for five characteristic immune expression signature scores (rows) in relation to PFI, for immune subtypes and TCGA tumor types (columns). Red indicates an association of immune signature with higher risk (shorter PFI) and blue lower risk (longer PFI) within the specific immune subtype or tumor type. **C.** CI for cell type fractions (rows) estimated using CIBERSORT in relation to PFI, for TCGA tumor types (columns). Red indicates an association of immune cell type with higher risk (shorter PFI) and blue lower risk (longer PFI) within the specified tumor type. **D.** As for C, with OS. **E.** Coefficients of elastic net model. The box and whisker plot indicates the average coefficient value of the elastic net model features at the optimal alpha and lambda parameters. Positive coefficients indicate higher hazard ratio (i.e., decreased survival) and negative coefficients indicate lower hazard ratio (increased survival). Error bars indicate the 95^{th} confidence interval across cross validation iterations. **F.** Prognostic significance of immune subtypes for OS within each tumor type was evaluated using CoxPH models. Bar heights indicate the negative log10 transformed p values of the immune subtype in predicting OS in each tumor type. Color outlines indicate the direction of the coefficient (blue: negative coefficient, increased survival; red: positive coefficient, decreased survival). Stars indicate significant p values. Black stars indicate significance with uncorrected p value. Red stars indicate significance with p values BH-corrected for FDR, with 1, 2 and 3 stars indicating FDR corrected values below 0.05, 0.01 and 0.001, respectively.

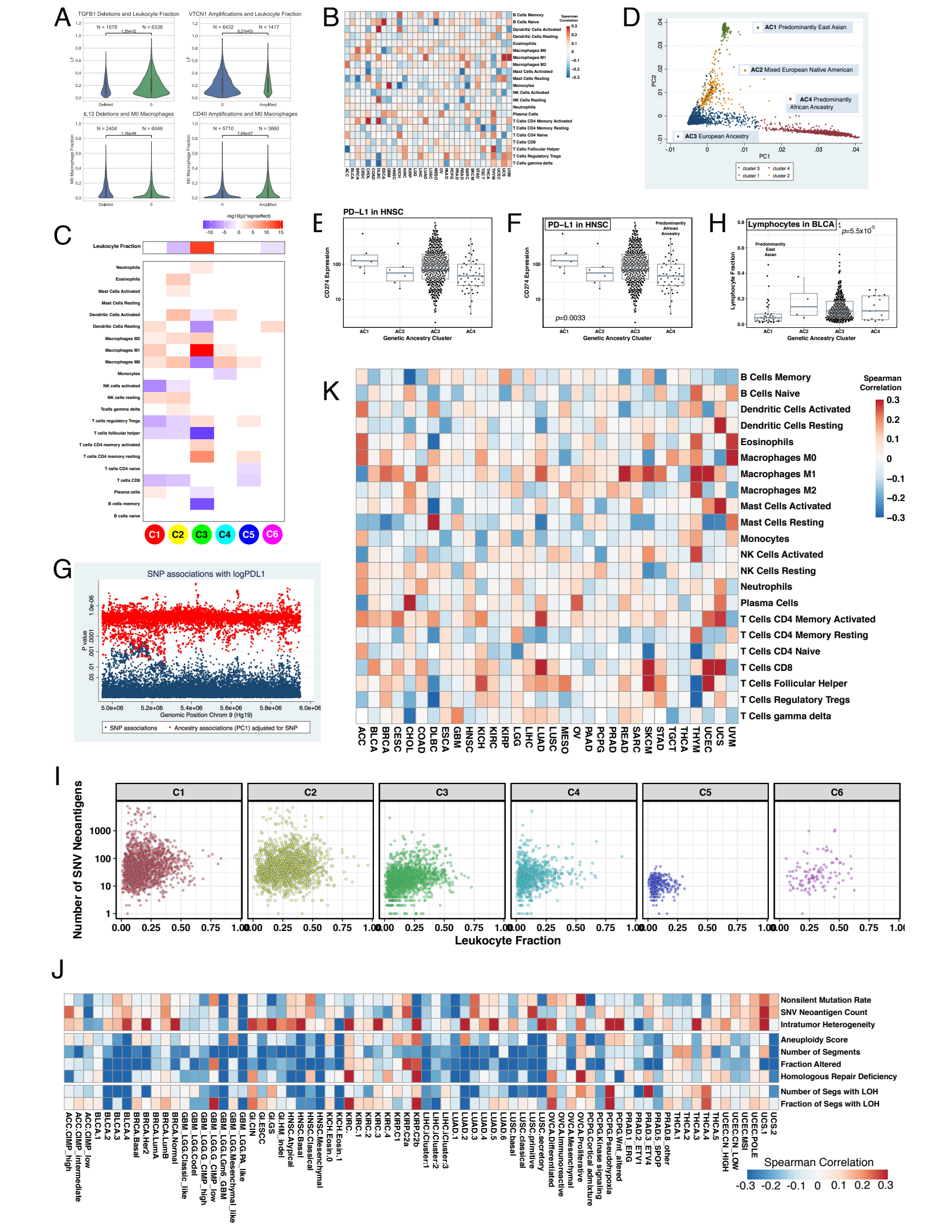

A
TGFB1 Deletions and Leukocyte Fraction
VTCN1 Amplifications and Leukocyte Fraction
IL13 Deletions and M0 Macrophages
CD40 Amplifications and M0 Macrophages
Deleted
Amplified
B
Spearman Correlation
B Cells Memory
B Cells Naive
Dendritic Cells Activated
Dendritic Cells Resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast Cells Activated
Mast Cells Resting
Monocytes
NK Cells Activated
NK Cells Resting
Neutrophils
Plasma Cells
T Cells CD4 Memory Activated
T Cells CD4 Memory Resting
T Cells CD4 Naive
T Cells CD8
T Cells Follicular Helper
T Cells Regulatory Tregs
T Cells gamma delta
D
AC1 Predominantly East Asian
AC2 Mixed European Native American
AC4 Predominantly African Ancestry
AC3 European Ancestry
PC1
PC2
cluster 1
cluster 2
cluster 3
cluster 4
C
-log10(p)*sign(effect)
Leukocyte Fraction
Neutrophils
Eosinophils
Mast Cells Activated
Mast Cells Resting
Dendritic Cells Activated
Dendritic Cells Resting
Macrophages M2
Macrophages M1
Macrophages M0
Monocytes
NK cells activated
NK cells resting
Tcells gamma delta
T cells regulatory Tregs
T cells follicular helper
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
Plasma cells
B cells memory
B cells naive
C1
C2
C3
C4
C5
C6
E
PD–L1 in HNSC
CD274 Expression
Genetic Ancestry Cluster
F
PD–L1 in HNSC
Predominantly African Ancestry
p=0.0033
H
Lymphocytes in BLCA
p=5.5x10-5
Predominantly East Asian
Lymphocyte Fraction
G
SNP associations with logPDL1
P value
Genomic Position Chrom 9 (Hg19)
SNP associations
Ancestry associations (PC1) adjusted for SNP
K
Spearman Correlation
B Cells Memory
B Cells Naive
Dendritic Cells Activated
Dendritic Cells Resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast Cells Activated
Mast Cells Resting
Monocytes
NK Cells Activated
NK Cells Resting
Neutrophils
Plasma Cells
T Cells CD4 Memory Activated
T Cells CD4 Memory Resting
T Cells CD4 Naive
T Cells CD8
T Cells Follicular Helper
T Cells Regulatory Tregs
T Cells gamma delta
I
C1
C2
C3
C4
C5
C6
Number of SNV Neoantigens
Leukocyte Fraction
J
Nonsilent Mutation Rate
SNV Neoantigen Count
Intratumor Heterogeneity
Aneuploidy Score
Number of Segments
Fraction Altered
Homologous Repair Deficiency
Number of Segs with LOH
Fraction of Segs with LOH
Spearman Correlation

Figure S4 - Immune Response and Genomic State. Related to Figure 4.
A. Correlation of representative CN alterations with immune infiltrate. Four representative IM genes are shown, demonstrating associations between copy number deletion (TGFB1 and IL13) or amplification (VTCN1 and CD40), x-axes, and LF (TGFB1 and VTCN1) or fractional level of M0 macrophages (IL13 and CD40), y-axes. Welch t-test p values are indicated above each comparison. **B.** Intratumoral heterogeneity correlation with individual immune cell types in TCGA tissues Values displayed are the Spearman correlation of immune cell fractions (rows) with intratumoral heterogeneity (subclonal genome fraction) within each tumor type (columns). Red indicates positive correlation (increasing proportion of indicated cell type with increasing ITH), and blue negative correlation. **C.** Association of LF and individual immune cell fractions with somatic alterations (mutation and copy number) in the TGF-β pathway. The association score is based on the BH-adjusted Student *t*-test.Red indicates increasing proportion of immune cell type with pathway disruption, and blue decreased proportion with pathway disruption. **D.** Genetic ancestry principal components and ancestry clusters We used the germline genetic data (TCGA SNPs) to make genotype calls on which we performed principal components analysis (PCA). We inferred how the principal components (PCs) related to the continental ancestry by comparing them with patient self- reported of race/ethnicity. We then clustered genetic ancestry into 4 clusters (derived ancestry clusters (AC1-AC4) by performing K means clustering on genotype PCs PC1, PC2 and PC3. The first two PCs are shown. The cluster with highest PC1 values correlates with African ancestry and the cluster with highest PC2 values correlates with Asian ancestry. AC1-AC4, correspond to predominately East Asian, European/Native American, European, and predominately African ancestry, respectively, as indicated by colored dots representing individual patients. **E.** PD-L1 expression differences between females and males in KIRP. Box plots show that in KIRP females have higher median CD274/PD-L1 expression than males. **F.** PD-L1 in HNSC. In head and neck squamous cell carcinoma, CD274/PD-L1 gene expression is relatively low in individuals of African ancestry, who are predominant in PC ancestry cluster 4 (AC4). **G.** SNP Associations with PD-L1 expression. The graph represents SNP associations with log2(PDL1) expression levels, for SNPs at each position in a 1 megabase region around the PDL1 transcriptional start site. The X-axis represents genomic position and Y-axis represents the negative log p value of associations. Below the X-axis is a map of genes in that region including PDL1 (CD274). The dashed line represents the threshold for significance for SNPs, adjusting for the effective number of independent SNPs at the locus. The blue dots represent the association statistics between SNPs and PDL1 expression. The red dots represent the association statistic for PC1 (African ancestry) for each of the SNPs. Where the red dots become less significant, the corresponding SNP is leading to some attenuation between ancestry and PDL1 levels, suggesting that SNP may explain some of the association. **H.** Lymphocytes in BLCA. In bladder cancer, lymphocytic infiltrate is relatively low in individuals of Asian ancestry, the predominant group in ancestry cluster 1 (AC1). **I.** Leukocytic relation with neoantigen burden in immune subtypes. Scatter plots for each immune subtype, with Spearman correlation coefficient and associated p-value on right. Dots represent individual tumors within each immune subtype. **J.** Spearman correlation (red: positive;blue:negative), between different measures of DNA damage (rows) and LF for TCGA tumor subtypes (columns) with more than 10 tumors in each. **K.** Neoantigen burden correlates with individual immuncell types in TCGA tissues. Values displayed are the Spearman correlation of immune cell fractions (rows) with neoantigen burden within each tumor type (columns).Red indicates positive correlation (increasing proportion of indicated cell type with increasing neoantigen burden), and blue negative correlation.

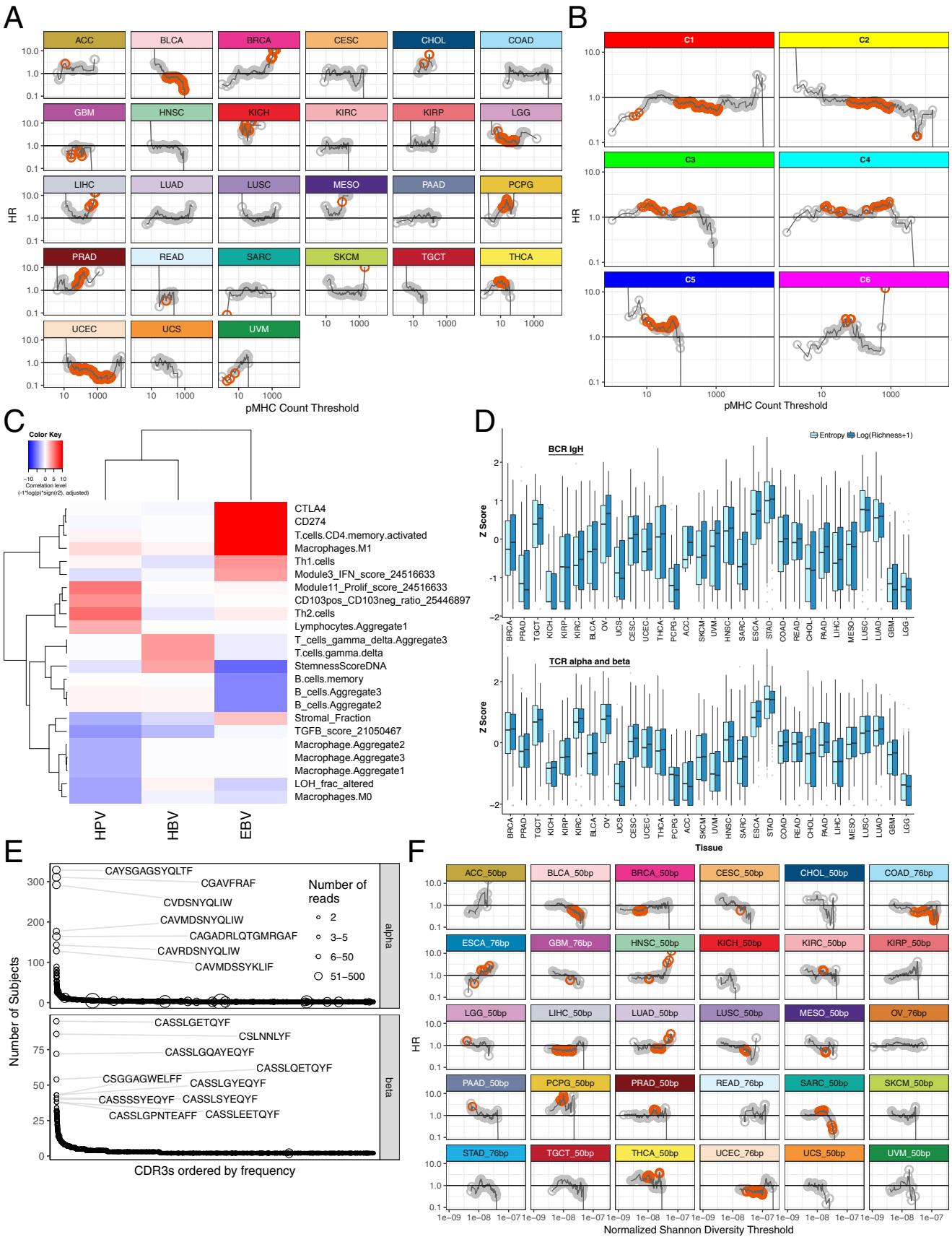

A
ACC
BLCA
BRCA
CESC
CHOL
COAD
GBM
HNSC
KICH
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
MESO
PAAD
PCPG
PRAD
READ
SARC
SKCM
TGCT
THCA
UCEC
UCS
UVM
HR
pMHC Count Threshold
B
C1
C2
C3
C4
C5
C6
HR
pMHC Count Threshold
C
Color Key
Correlation level
(-1*log(p)*sign(r2), adjusted)
CTLA4
CD274
T.cells.CD4.memory.activated
Macrophages.M1
Th1.cells
Module3_IFN_score_24516633
Module11_Prolif_score_24516633
CD103pos_CD103neg_ratio_25446897
Th2.cells
Lymphocytes.Aggregate1
T_cells_gamma_delta.Aggregate3
T.cells.gamma.delta
StemnessScoreDNA
B.cells.memory
B_cells.Aggregate3
B_cells.Aggregate2
Stromal_Fraction
TGFB_score_21050467
Macrophage.Aggregate2
Macrophage.Aggregate3
Macrophage.Aggregate1
LOH_frac_altered
Macrophages.M0
HPV
HBV
EBV
D
Entropy
Log(Richness+1)
BCR IgH
TCR alpha and beta
Z Score
Tissue
E
CAYSGAGSYQLTF
CGAVFRAF
CVDSNYQLIW
CAVMDSNYQLIW
CAGADRLQTGMRGAF
CAVRDSNYQLIW
CAVMDSSYKLIF
Number of reads
2
3–5
6–50
51–500
alpha
beta
Number of Subjects
CASSLGETQYF
CSLNNLYF
CASSLGQAYEQYF
CASSLQETQYF
CSGGAGWELFF
CASSLGYEQYF
CASSSYEQYF
CASSLSYEQYF
CASSLGPNTEAFF
CASSLEETQYF
CDR3s ordered by frequency
F
ACC_50bp
BLCA_50bp
BRCA_50bp
CESC_50bp
CHOL_50bp
COAD_76bp
ESCA_76bp
GBM_76bp
HNSC_50bp
KICH_50bp
KIRC_50bp
KIRP_50bp
LGG_50bp
LIHC_50bp
LUAD_50bp
LUSC_50bp
MESO_50bp
OV_76bp
PAAD_50bp
PCPG_50bp
PRAD_50bp
READ_76bp
SARC_50bp
SKCM_50bp
STAD_76bp
TGCT_50bp
THCA_50bp
UCEC_76bp
UCS_50bp
UVM_50bp
HR
Normalized Shannon Diversity Threshold

Figure S5 - The Tumor Immune Interface. Related to Figure 5. A & B. Impact of Immunogenicity Features on Prognostic Associations Across TCGA Tumor Types (A) and Immune Subtypes (B). CoxPH models were built with age, gender, LF, and tumor type (if applicable) as covariates. Every possible cutoff for number of predicted neoantigens (pMHC) is tested, and Hazard Ratios (HRs) for PFI are recorded for "greater than cutoff". Significant HRs are indicated in orange. Cutoff number for pMHC increases from left to right; tumors where HR hovers around 1 (e.g. HNSC) show no correlation between pMHC counts and PFI, while BLCA shows HR consistently < 1 across a range of thresholds. **C.** Correlations with Viral Load. Plot shows correlation between increasing viral read counts for specified viruses (rows) and immune features (columns) in a regression model adjusted for cancer type. Color key: red=positively correlated, blue=negatively correlated, and intensity= -1*log(p-val), with a ceiling at p-value < 10^{-10}. **D.** TCR and BCR heavy chain diversity measures across tumor types**.** Diversity indices Shannon entropy and species richness (i.e. number of unique clonotypes) were calculated by MiTCR. Richness was log-transformed and both entropy and richness were standardized with Z-scores calculated to place values on the same scale prior to plotting. **E.** Recurrence of TCR CDR3 sequences in RNA-seq datasets. Of 108,185 distinct alpha and 181,787 distinct beta sequences, 16,971 alpha and 12,079 beta are found in at least two patients. The plot indicates that very few recurrent sequences are shared by >20 patients. Sequences with higher recurrence are labelled. Number of reads is tallied across all samples.Larger circles (number of reads) indicate highly abundant TCRs in one or more samples. **F.** Impact of TCR diversity on prognostic associations. CoxPH models built with age, gender, and leukocyte fraction as covariates. Every possible (normalized Shannon diversity) cutoff is tested to evaluate the effect of increasing TCR diversity on PFI, and Hazard Ratios for PFI is recorded (for "greater than cutoff"). Significant HRs are indicated in orange. For each category, 76 bp datasets used unless unavailable, in which case 50 bp was used.

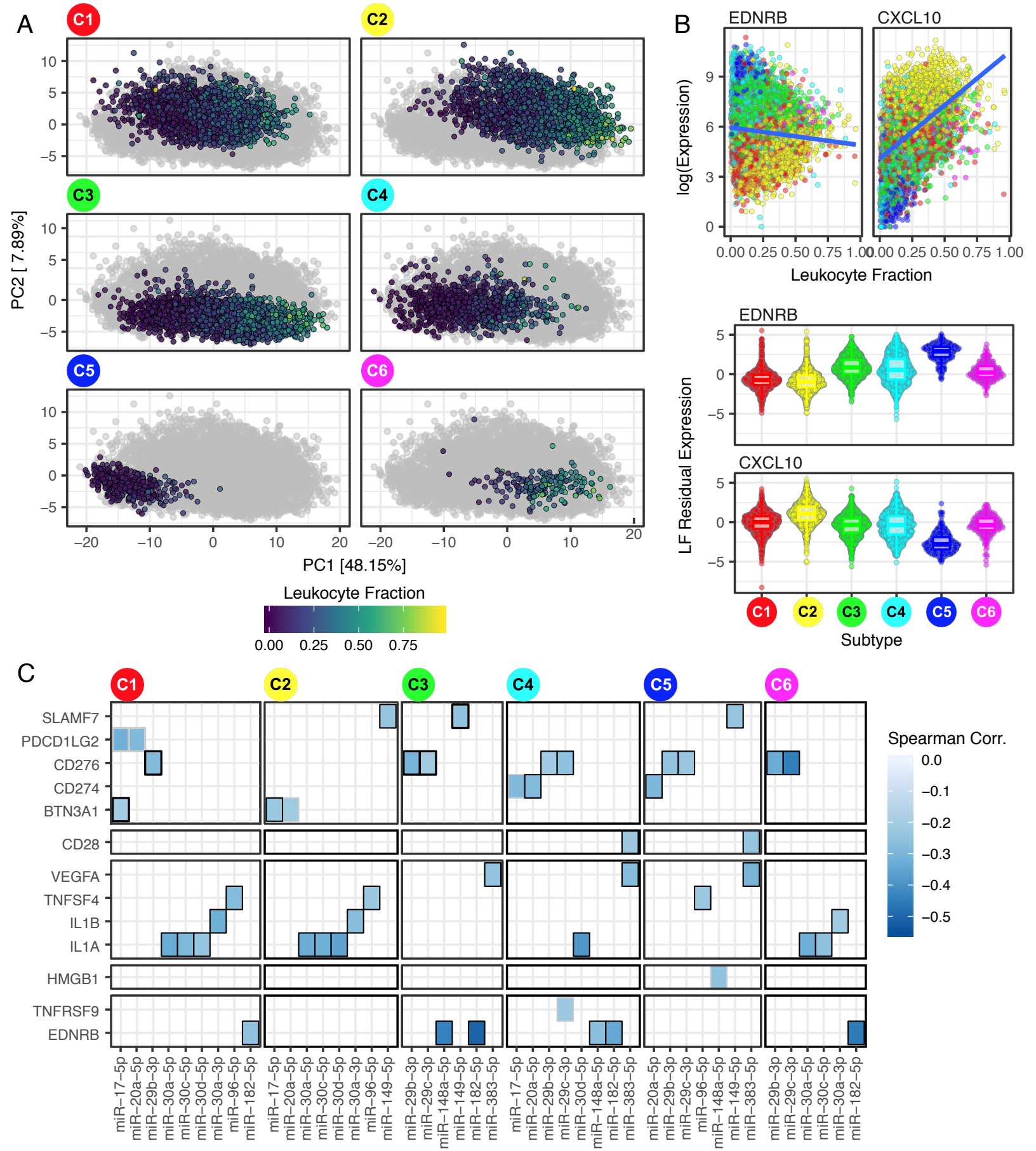

A
C1
C2
C3
C4
C5
C6
PC2 [7.89%]
PC1 [48.15%]
Leukocyte Fraction
0.00 0.25 0.50 0.75
B
EDNRB
CXCL10
log(Expression)
Leukocyte Fraction
LF Residual Expression
Subtype
C
SLAMF7
PDCD1LG2
CD276
CD274
BTN3A1
CD28
VEGFA
TNFSF4
IL1B
IL1A
HMGB1
TNFRSF9
EDNRB
Spearman Corr.
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
miR−17−5p
miR−20a−5p
miR−29b−3p
miR−30a−5p
miR−30c−5p
miR−30d−5p
miR−30a−3p
miR−96−5p
miR−182−5p
miR−149−5p
miR−29c−3p
miR−148a−5p
miR−383−5p

Figure S6. Regulation of Immunomodulators in Cancer. Related to Figure 6. A. IM expression segregates tumors by immune subtype. PCA of 9,058 TCGA samples, based on expression of 67 IM genes. Samples within a subtype are highlighted (non-gray) in each sub-panel, with points colored by leukocyte fraction (LF). Plots show that PCA using IM gene expression recapitulates immune subtypes, with each immune subtype showing relatively distinct separation along PC1 and PC2; a gradient of LF values is correlated with PC1 within each subtype. **B.** Correlation between expression and LF for IM genes with largest differences across subtypes (by Kruskal-Wallis test) (top); distribution of LF-adjusted expression levels in each subtype (bottom). **C.** Regulation of immunomodulators by miRNA. Associations are shown between commonly implicated miRNAs and immunomodulators for each immune subtype. All associations shown represent BH-adjusted p-value<0.05 and Spearman correlation ≤ -0.2; each miRNA included is negatively correlated with a gene for which it is predicted to bind in miRDB.

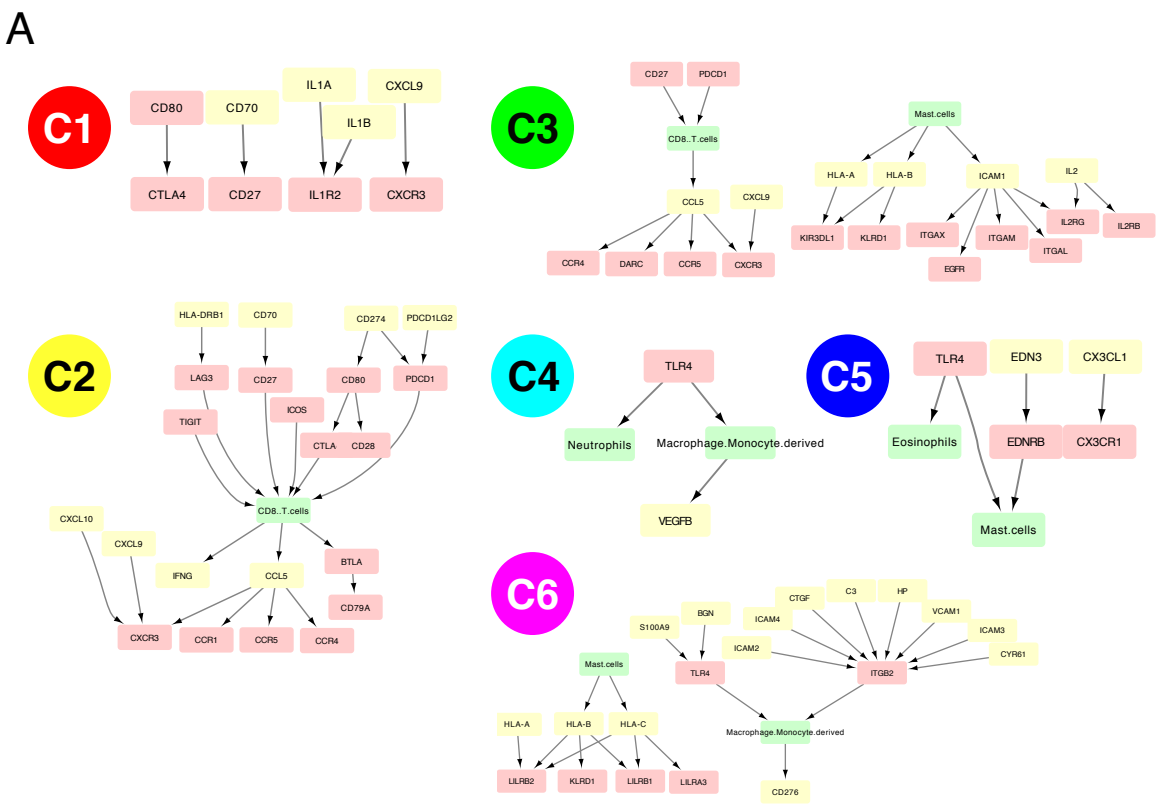

A
C1
CD80
CD70
IL1A
CXCL9
IL1B
CTLA4
CD27
IL1R2
CXCR3
C3
CD27
PDCD1
CD8..T.cells
CCL5
CXCL9
CCR4
DARC
CCR5
CXCR3
Mast.cells
HLA-A
HLA-B
ICAM1
IL2
KIR3DL1
KLRD1
ITGAX
ITGAM
ITGAL
IL2RG
IL2RB
EGFR
C2
HLA-DRB1
CD70
CD274
PDCD1LG2
LAG3
CD27
CD80
PDCD1
TIGIT
ICOS
CTLA
CD28
CD8..T.cells
CXCL10
CXCL9
IFNG
CCL5
BTLA
CD79A
CXCR3
CCR1
CCR5
CCR4
C4
TLR4
Neutrophils
Macrophage.Monocyte.derived
VEGFB
C5
TLR4
EDN3
CX3CL1
Eosinophils
EDNRB
CX3CR1
Mast.cells
C6
CTGF
C3
HP
S100A9
BGN
ICAM4
VCAM1
ICAM3
ICAM2
CYR61
TLR4
ITGB2
Mast.cells
HLA-A
HLA-B
HLA-C
Macrophage.Monocyte.derived
LILRB2
KLRD1
LILRB1
LILRA3
CD276
B
PanCanSYGNAL TRN
43,000 significantly
co-expressed biclusters
Immunomod.
enriched
biclusters
Leukocyte
associated
biclusters
18
171
6,496
Leukocyte associated
& immunomod. enriched
biclusters
PanImmuneSYGNAL TRN
Tumor types
33
Tumor type
Mutations
56
Somatic mutation
TFs
19
TF
Immuno.
27
Immunomodulators
C
NFE2L2
PREX2
BAP1
NOTCH1
FGFR3
DNAH17
ARID1B
CIC
MKI67
RB1
FOXA1
AKAP9
TP53
DNER
ABI3BP
CUX1
KEAP1
RHOB
PTPRC
EIF1AX
SPOP
KRAS
CCND1
CENPF
PBRM1
SMARCA4
CTCF
GRM7
IVL
HRAS
MARK3
NEB
SACS
ATM
ZBTB20
KDM6A
PTEN
COL3A1
FUBP1
CR1
AXIN1
ZC3H13
BCOR
CREBBP
ARID1A
CDKN2A
ELF3
NF1
NRAS
EIF5B
ARID5B
FBXW7
CTNNB1
MEN1
FMN2
ARHGAP35
IRF3
NFKB2
ETS1
IRF7
IRF9
SPI1
IRF4
NFKB1
IRF2
PRDM1
IRF8
STAT5A
ZNF683
FLI1
IRF5
ETV7
KLF2
STAT4
ELF4
HLA-A
VSIR
HLA-DQA1
HAVCR2
HLA-DRB1
CX3CL1
TNFRSF14
HLA-B
BTN3A1
CXCL9
ITGB2
GZMA
HLA-DQA2
HLA-DRA
PRF1
HLA-DRB5
CXCL10
HLA-DPA1
BTN3A2
SLAMF7
HLA-DPB1
ICAM1
HLA-DQB1
CD40
IDO1
CCL5
HLA-C

Figure S7. Intra and Extracellular Networks Modulating Immune Response to Tumors. Related to Figure 7. A. Cytokine immune subnetworks. Cell types are shown in green, ligands in yellow, and receptors in red. Edges connect nodes if each is found to be abundant within an immune subtype. Small example sub-networks from each immune subtype are shown where edges are filtered to be in the top 25%, by concordance value. Subtype C4 did not have any high scoring edges Diversity of cytokine signalling networks between immune subtypes emphasizes differential regulation of immune function within the TME of each subtype. **B.** Pan-Immune SYGNAL Network inference. PanCanSYGNAL network was filtered to 43,000 significantly co-expressed biclusters. We then identified biclusters significantly associated with leukocyte infiltration (leukocyte associated), biclusters enriched with IM genes, and biclusters with both associations. PanImmuneSYGNAL transcriptional regulatory network (TRN) was discovered as the co-regulated biclusters from the 171 leukocyte associated and immunomodulatory enriched biclusters. The resulting TRN is a predictive map with interactions from all 33 tumor types that implicate specific somatic mutations in modulating the expression of a TF(s) that in turn regulate IM gene(s) (within a bicluster). A summary of the counts for each feature in the PanImmuneSYGNAL TRN are shown on the left. **C.** Regulation of IMs according to the Pan-Immune SYGNAL network. Legend for node types is shown above in the rightmost section of panel B. The network contains mechanistic regulatory interactions linking 56 somatic mutations through 19 transcription factors (ELF4, ETS1, ETV7, FLI1, IRF2/3/4/5/7/8/9, KLF2, NFKB1/2, PRDM1, SPI1, STAT4/5A, and ZNF683) to drive the expression of 43 of the 171 IM enriched and leukocyte infiltration associated biclusters, which includes 27 IMs.