SUPPLEMENTARY MATERIALS

Supplementary Methods

Study design

The workflow to identify and validate the TME risk score and TME subtypes in gastric cancer is depicted in Fig. S1. We first estimated the absolute abundance levels of the major stromal and immune cell types in the TME using bulk gene expression data, and assessed the prognostic effect of these cells in a discovery cohort. Next, we constructed a TME risk score and validated it in two independent gene expression validation cohorts and three immunohistochemistry validation cohorts. Finally, we stratified patients into four TME subtypes and examined their genomic and molecular features and relation to established molecular subtypes.

Gene expression data

To explore the prognostic landscape of the TME, we used four gene expression profile (GEP) datasets of resected gastric cancer patients with publicly available clinical information, namely, ACRG (GSE62254) (1), GSE15459 (2), GSE84437 (3), and TCGA stomach adenocarcinoma (STAD). Specifically, the raw microarray data in the ACRG cohort and GSE15459 were retrieved from the Gene Expression Omnibus (GEO), and normalized by the RMA algorithm (package affy) using custom chip definition files (Brainarray version 23 (4)) that convert Affymetrix probesets to Entrez gene IDs. For GSE84437 dataset, which was measured by the Illumina platform, we downloaded the normalized gene expression profile from the GEO (package GEOquery (5)). The Illumina probes were also mapped to Entrez genes. For multiple probes mapping to the same Entrez gene, we selected the one with the maximum mean expression level as the surrogate for the Entrez gene using the function of collapseRows (6) (package WGCNA). The normalized gene expression data in the TCGA STAD cohort was downloaded from the TCGA PanCanAtlas data portal (https://gdc.cancer.gov/about-data/publications/pancanatlas). All gene expression levels were log2 transformed and batch effects were removed using the Combat algorithm (7) (package SVA). The clinical and treatment information for ACRG cohort was retrieved from the original publication (1). The clinical outcome data for GSE15459 and GSE84437 cohorts were downloaded from the GEO. For the TCGA STAD cohort, the clinical information and genomics features were retrieved from the PanCanAtlas data portal.

Patients in the IHC cohorts

For validated purposes, we analyzed IHC data for gastric cancer patients who were treated with surgical resection at the Nanfang Hospital of Southern Medical University (SMU) and the First Affiliated Hospital of Sun Yat-Sen University (SYSU), Guangzhou, China. Written-informed consent was obtained from all of the enrolled patients. All procedures were performed in accordance with the Declaration of Helsinki and approved by the Institutional Review Boards at the two participating centers. Inclusion criteria were diagnosis of primary, biopsy-confirmed GC, surgical resection, and the availability of hematoxylin and eosin (H&E) slides with invasive tumor components, follow-up data, and clinicopathologic characteristics. No patients had received previous treatment. We included three independent cohorts of 753 patients in this study, including 247 and 234 consecutive patients treated at SMU between January 2005 and December 2007, January 2008 and December 2009, and 272 patients treated at SYSU from January 2005 to December 2007. The three cohorts were denoted as SMU1, SMU2 and SYSU. The patient characteristics are listed in Table S1.

IHC data

FFPE samples were cut into 4-μm thick sections, which were then processed for immunohistochemistry staining as previously described (8, 9). The samples were de-waxed in xylene and rehydrated in decreasing concentrations of ethanol. Prior to staining, the sections were subjected to endogenous peroxidase blocking in 1% H_2O_2 solution diluted in methanol for 10 minutes and then heated in a microwave for 30 minutes with 10 mmol/L citrate buffer (pH 6.0). Serum blocking was performed using 10% normal rabbit serum for 30 minutes. The slides were incubated overnight with an antibody against human CD57 (1:100 dilution; NeoMarker, clone NK1) for NK cells, CD34 (1:200 dilution; Abcam, ab81289) for endothelial cells, or α-SMA (smooth muscle actin; 1:100 dilution; Abcam, ab5694) for fibroblasts at 4°C, followed by incubation with an amplification system with a labeled polymer/HRP (EnVision™, DakoCytomation, Denmark) at 37°C for 30 min. The reaction was visualized using diaminobenzidine (DAB)+ chromogen, and nucleus was counterstained using hematoxylin. In all assays, we included negative control slides with the primary antibodies omitted. Every staining run contained a slide of positive control. And all slides were stained with DAB dyeing for the same time for each antibody.

IHC staining was independently assessed by two pathologists who were blinded to patient characteristics. A third pathologist was consulted when different opinions arose between the two primary pathologists and discrepancies were resolved by consensus. Under 200 magnification, photographs of 5 representative fields were captured using an inverted research microscope (model DM IRB; Leica Germany); identical settings were used for each photograph. The nucleated stained cells in each area were quantified and expressed as the number of cells per field. For the CD57 or CD34 staining in tumor tissue, the number of positive cells or microvessels was calculated in each field and expressed as the mean value of the five fields (cells per field or microvessels per field). Intensity of fibroblast cell staining (α-SMA) was graded as 0 (negative staining), 1 (weak staining), 2 (moderate staining) and 3 (strong staining); staining extent was graded as 0 (0%–4%), 1 (5%–24%), 2 (25%–49%), 3 (50%–74%) and 4 (>75%). Values of the intensity and the extent were multiplied as the level of α-SMA.

Identification of prognostic cell types in the TME

For the gene expression cohorts, the absolute abundance levels of major cell types within the TME were computed for each patient by the Microenvironment Cell Populations-counter (MCP-counter) algorithm (10). Here, the abundance levels were estimated for 8 immune and 2 stroma cell types by averaging pre-selected sets of marker genes (Table S2). The immune cells include T cells, CD8 T cells, cytotoxic lymphocytes, NK cells, B lineage, monocytic lineage, myeloid dendritic cells, and neutrophils. The stroma cells are fibroblasts and endothelial cells. To investigate the association of each TME cell type with overall survival (OS), we selected the ACRG cohort (n = 300) as the discovery cohort, because detailed clinical and treatment outcome information with sufficient follow-up is available. To assess robustness of the prognostic significance, we used two different methods, namely the Cox regression analysis (package survival) and random survival forest model (package randomForestSRC(11)). Specifically, we conducted univariate Cox regression analysis of each cell type with or without controlling for stage and treatment in the ACRG cohort, and examined the corresponding Wald test *P* values. To implement the random forest algorithm, we generated 500 stratified bootstrap resamples of approximately the same size (n = 296) of the ACRG cohort, controlling for treatment and stage. We used the 'logrank' metric as the splitting rule. The hyperparameter of the random forest model, namely mtry (number of cell types randomly sampled at each split) and nodesize (average

number of patients in a terminal node) were determined based on the out-of-bag error. The variable importance was calculated from the resulting model, which provided a ranking order of the features.

Development and validation of a TME-based risk score

Through the above analyses, NK cells, endothelial cells, and fibroblasts were identified as the most robust prognostic markers. We first explored their pairwise Pearson correlation in the ACRG cohort. Considering a high correlation and similar prognostic effects, we defined a Stroma score as the geometric mean of the estimated abundance of endothelial cells and fibroblasts to reflect the overall stromal quantity (12). To further validate the stroma score, the correlations of an EMT signature (1, 13), two fibroblast signatures (14) and the StromalScore provided by the ESTIMATE algorithm (15) were calculated regarding to proposed stroma score in each GEP cohort. The abundance of NK cells were correlated with those of T or CD8 T cells as well as a T-cell inflamed gene expression signature (16). Similarly, we assessed the correlation and prognostic independence between the NK cells and Stroma score. Given the low correlation between the NK cell and Stroma score as well as the opposite prognostic effect, we defined a TME-based risk score as the ratio of Stroma score to NK cell abundance, of which a higher value indicates an elevated risk of death.

Using the same formula, we constructed the Stroma score and the TME risk score for the independent GEP and IHC validation cohorts. Again, for the two GEP validation cohorts, the Stroma and TME risk score were derived based on the absolute abundance level of NK, endothelial cells and fibroblast, which were estimated by MCPcounter algorithm. For the IHC cohorts, the Stroma and TME risk score were constructed based on the IHC staining levels of CD57 (NK cell), CD34 (endothelial cell) and α-SMA (fibroblast). By applying a bivariate Cox regression, we confirmed the prognostic independence between the NK cells and Stroma score in the merged validation GEP (GSE84437 and GSE15459) and IHC cohorts, respectively. Then, we tested the association between the continuous TME risk score and OS in each of the independent GEP and IHC validation cohorts. TCGA STAD cohort was excluded for survival analysis because of a short follow-up time (median: <2 years vs. >5 years in other cohorts).

Additionally, cutoff values for TME risk scores were derived separately for the gene expression and IHC cohorts, given the different measurement platforms. For gene expression cohorts, 100 quantile cut points were generated covering the 10th percentile to 90th percentile of the TME risk score in the ACRG cohort. We dichotomized patients in the ACRG cohort based on each cut point and recorded the corresponding Wald test *P* value of Cox regression with stage and treatment controlled as strata. We chose the cutoff value with the minimum *P* value as the optimal prognostic threshold to differentiate patients into high or low TME risk groups. We used the same criteria to select the cutoff value for the IHC cohort based on the SMU1 cohort. Then, we examined the OS differences between different TME risk groups in each of the GEP and IHC cohorts.

Independent prognostic effect of TME risk score to clinicopathologic factors

We used multivariable Cox regression to assess the independent prognostic value of the continuous TME risk score by adjusting for clinicopathologic factors including age at diagnosis, gender, pathological stage, Lauren histology, and treatment by chemotherapy in the ACRG and merged IHC cohort. In the GEP cohorts, in addition to clinicopathologic factors, the inferred MSI status (1, 17) was also included in the multivariate Cox regression. To assess the statistical significance of additive prognostic effect of TME risk score to stage, which was the strongest

prognostic clinical factor, we compared Cox regression models that included stage with or without the continuous TME risk score based on the likelihood-ratio test for nested models. The strength of the additive effect of the continuous TME risk score to pathological stage was assessed by continuous net reclassification index analyses (package survIDINRI) (18) at 5 years.

We also examined the prognostic effect of the TME risk groups in relatively more homogenous sub-populations. Since the ACRG, GSE15459 and all IHC cohorts provide the patient-level stage information, we assessed the OS difference between different TME risk groups in the merged GEP (ACRG and GSE15459) and merged IHC cohorts. Similarly, since the ACRG and three IHC cohorts provide patient-level treatment information, we also examined the prognostic effect of TME risk groups in the patients who only received surgery.

Identification of TME subtypes and genomic correlates

To elucidate the molecular underpinnings among tumors with different TME characteristics, we stratified patients based on the median levels of NK cells and Stroma score of the merged GEP cohort. This led to four TME subtypes, namely NK low & Stroma low, NK high & Stroma low, NK low & Stroma high, and NK high & Stroma high. Similarly, TME subtypes in the IHC cohorts were defined based on the median levels of the NK cells and Stroma score of the merged IHC cohort. We assessed the survival differences among the TME subtypes in the merged GEP (except TCGA STAD) and merged IHC cohorts.

We compared our TME subtypes with the molecular subtypes proposed by the ACRG (ACRG subtype) and the intrinsic subtypes for gastric cancer (19) in the GEP cohorts. To evaluate the complementary prognostic effect of our TME subtypes to the ACRG subtypes, we examined the survival differences among TME subtypes within each individual ACRG subtype in the merged GEP cohort (except TCGA STAD). Since the ACRG subtyping was only available for patients in the ACRG cohort, we generated the ACRG subtype labels for the remaining 3 cohorts (1). Briefly, we generated the signature scores for microsatellite instability (MSI), epithelial-to-mesenchymal transition (EMT) and TP53 activity by averaging the expression levels of genes in the corresponding signature gene list. The cutoff values for MSI, EMT and TP53 activity were computed by maximizing the Youden index in the ACRG cohort (package OptimalCutpoints (20)). Then, the patients without the ACRG subtype annotation were assigned to appropriate ACRG subtypes based on these derived cutoff values. We also constructed the intrinsic subtypes for gastric cancer based on the expression level of 171 genes (19) using the nearest template prediction algorithm (21).

In the TCGA STAD cohort, we assessed the genomic characteristics of patients in different TME subtypes. These genomic features can be grouped into 4 major categories: 1) the genomic subtypes designed for gastrointestinal tract adenocarcinomas (GIACs); 2) status for key driver genes (the mutation status of *TP53* and *PIK3CA*, *HER2* amplification status, and the epigenetic silencing status of *CDKN2A*, *MLH1*, *BRCA1* and *RAD51C*); 3) genome instability measurement (aneuploidy score, ploidy level, whole genome doubling status [WGD], clonal deletion score [CDS], the number of homozygous deletions, the number of arm level copy number events, the number of focal amplification copy number events, the number of focal deletion copy number events, the number of overall focal copy number events and the chromosomal instability [CIN] Focal vs Broad classification); and 4) mutational burden (SNV density, indel density, overall mutation density, and hypermutation status). Chi-squared and Mann-Whitney tests were used to assess the difference between TME subgroups regarding categorical and continuous features, respectively. The

Benjamini-Hochberg method was used to compute the false discovery rate (FDR) to adjust for multiple testing.

The following is a brief description of the molecular features used in this study. The molecular subtype for GIACs (GIAC subtype) stratify patients of the TCGA STAD cohort into five groups, namely Epstein-Barr virus-positive (EBV), hypermutated-SNV (HM-SNV), MSI, CIN and genome stable (GS) subtypes (22). Mutation status for *TP53* and *PIK3CA* were derived from the MC3 mutation annotation file. *HER2* amplification status was derived from the gene-level copy number data. The epigenetic silencing status of *CDKN2A*, *MLH1*, *BRCA1*, and *RAD51C*, which indicates a hyper-methylated promoter region and a reduced gene expression level, was inferred from the matched DNA methylation and gene expression profiles. The aneuploidy score, ploidy level, WGD, CDS, and the number of homozygous deletions were estimated using the ABSOLUTE algorithm. The aneuploidy scores were calculated as the sum total of amplified or deleted arms (23). The ploidy level reflects the distribution of total copy number. The presence of WGD was determined based on whether the fraction of genome with duplicated alleles was higher than 0.5. The CDS quantifies the number of clonally deleted genomic regions in each tumor's genome. Other copy number alterations were identified from segmented data using GISTIC 2.0 algorithm. The CIN Focal classification represents a higher quantity and intensity of high-amplitude focal DNA amplifications relative to the CIN Broad classification. The mutation density was defined as the number of corresponding mutations per megabase. The hypermutation was defined as mutation density larger than 10. More details about these molecular features can be found in (22). Molecular features with FDR < 0.01 were reported. To excluding confounding effects of the MSI or EBV status, similar analyses were conducted in the subgroup of patients with CIN in the TCGA cohort. Similar analyses were conducted on the subgroups of patients with the same MSI (1, 17) or T-cell inflamed signatures (16).

Identification of gene expression, molecular pathways, and cytokines correlated with NK cell abundance and stroma score

We assessed the Pearson correlation between the expression levels of each individual gene with the NK cell abundance in a meta-analysis of the four GEP cohorts. Specifically, we included genes measured by at least two cohorts and removed the marker genes of NK cells in the MCPcounter algorithm, which led to 21,616 unique genes. To minimize false positive findings, we summarized the overall correlation strength over multiple cohorts for a certain gene using a fixed-effect model based on Fisher's z transformation of correlation (Package meta (24)).

Next, we computed the pathway activity score of the 50 hallmark gene sets (MsigDB) for each patient in the GEP cohorts, using the single sample Gene Set Enrichment Analysis (ssGSEA) algorithm (package GSVA) (25), The correlation between pathway activity and NK cell abundance as well as Stroma score were assessed with a similar meta-analysis framework described above.

Soluble factors such as cytokines are key modulating factors of TME. Thus, we analyzed the differentially expressed (DE) cytokines according to the status of NK cells and Stroma score. We focused on 171 cytokine genes in the human cytokine-cytokine receptor interaction pathway (has:04060) of the KEGG database, out of which 110 cytokine genes were measured by all the GEP platforms. We used the limma (26) package to calculate the significant DE cytokine genes in the merged GEP cohort based on the NK cell and Stroma status. Genes with FDR < 0.01 and fold change > 1.5 were defined as the DE cytokines.

References:

1. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med.* 2015;21(5):449-56.
2. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *Plos Genet.* 2009;5(10):e1000676.
3. Cheong J-H, Yang H-K, Kim H, Kim WH, Kim Y-W, Kook M-C, et al. Predictive test for chemotherapy response in resectable gastric cancer: a multi-cohort, retrospective analysis. *The Lancet Oncology.* 2018.
4. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 2005;33(20):e175.
5. Davis S, and Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23(14):1846-7.
6. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, et al. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics.* 2011;12:322.
7. Johnson WE, Li C, and Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8(1):118-27.
8. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg.* 2018;267(3):504-13.
9. Jiang Y, Xie J, Han Z, Liu W, Xi S, Huang L, et al. Immunomarker Support Vector Machine Classifier for Prediction of Gastric Cancer Survival and Adjuvant Chemotherapeutic Benefit. *Clin Cancer Res.* 2018;24(22):5574-84.
10. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17(1):218.
11. Ishwaran H, Kogalur UB, Blackstone EH, and Lauer MS. Random survival forests. *The annals of applied statistics.* 2008;2(3):841-60.
12. Rooney MS, Shukla SA, Wu CJ, Getz G, and Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160(1):48-61.
13. Loboda A, Nebozhyn MV, Watters JW, Buser CA, Shaw PM, Huang PS, et al. EMT is the dominant program in human colon cancer. *BMC medical genomics.* 2011;4(1):9.
14. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov.* 2019;9(8):1102-23.
15. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications.* 2013;4:2612.
16. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science.* 2018;362(6411).
17. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012;487(7407):330-7.
18. Pencina MJ, D'Agostino Sr RB, and Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in medicine.* 2011;30(1):11-21.
19. Tan IB, Ivanova T, Lim KH, Ong CW, Deng N, Lee J, et al. Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology.* 2011;141(2):476-85, 85 e1-11.

20. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, and Gude-Sampedro F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw.* 2014;61(8):1-36.
21. Hoshida Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PloS one.* 2010;5(11).
22. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer cell.* 2018;33(4):721-35 e8.
23. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity.* 2018;48(4):812-30 e14.
24. Schwarzer G. meta: An R package for meta-analysis. *R news.* 2007;7(3):40-5.
25. Hänzelmann S, Castelo R, and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics.* 2013;14(1):7.
26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research.* 2015;43(7):e47-e.

Supplementary Tables

Table S1. Marker genes to calculate absolute abundance levels of TME cell types

Gene symbols	Cell population	ENTREZID	Gene symbols	Cell population	ENTREZID
CD28	T cells	940	*BANK1*	B lineage	55024
CD3D	T cells	915	*CD19*	B lineage	930
CD3G	T cells	917	*CD22*	B lineage	933
CD5	T cells	921	*CD79A*	B lineage	973
CD6	T cells	923	*CR2*	B lineage	1380
CHRM3-AS2	T cells	100506915	*FCRL2*	B lineage	79368
CTLA4	T cells	1493	*IGKC*	B lineage	3514
FLT3LG	T cells	2323	*MS4A1*	B lineage	931
ICOS	T cells	29851	*PAX5*	B lineage	5079
MAL	T cells	4118	***CD160***	**NK cells**	**11126**
MGC40069	T cells	348035	***KIR2DL1***	**NK cells**	**3802**
PBX4	T cells	80714	***KIR2DL3***	**NK cells**	**3804**
SIRPG	T cells	55423	***KIR2DL4***	**NK cells**	**3805**
THEMIS	T cells	387357	***KIR3DL1***	**NK cells**	**3811**
TNFRSF25	T cells	8718	***KIR3DS1***	**NK cells**	**3813**
TRAT1	T cells	50852	***NCR1***	**NK cells**	**9437**
CD8B	CD8 T cells	926	***PTGDR***	**NK cells**	**5729**
CD8A	Cytotoxic lymphocytes	925	***SH2D1B***	**NK cells**	**117157**
EOMES	Cytotoxic lymphocytes	8320	*ADAP2*	Monocytic lineage	55803
FGFBP2	Cytotoxic lymphocytes	83888	*CSF1R*	Monocytic lineage	1436
GNLY	Cytotoxic lymphocytes	10578	*FPR3*	Monocytic lineage	2359
KLRC3	Cytotoxic lymphocytes	3823	*KYNU*	Monocytic lineage	8942
KLRC4	Cytotoxic lymphocytes	8302	*PLA2G7*	Monocytic lineage	7941
KLRD1	Cytotoxic lymphocytes	3824	*RASSF4*	Monocytic lineage	83937
CD1A	Myeloid dendritic cells	909	*TFEC*	Monocytic lineage	22797
CD1B	Myeloid dendritic cells	910	***ACVRL1***	**Endothelial cells**	**94**
CD1E	Myeloid dendritic cells	913	***APLN***	**Endothelial cells**	**8862**
CLEC10A	Myeloid dendritic cells	10462	***BCL6B***	**Endothelial cells**	**255877**
CLIC2	Myeloid dendritic cells	1193	***BMP6***	**Endothelial cells**	**654**
WFDC21P	Myeloid dendritic cells	645638	***BMX***	**Endothelial cells**	**660**
CA4	Neutrophils	762	***CDH5***	**Endothelial cells**	**1003**
CEACAM3	Neutrophils	1084	***CLEC14A***	**Endothelial cells**	**161198**
CXCR1	Neutrophils	3577	***CXorf36***	**Endothelial cells**	**79742**
CXCR2	Neutrophils	3579	***EDN1***	**Endothelial cells**	**1906**
CYP4F3	Neutrophils	4051	***ELTD1***	**Endothelial cells**	**64123**
FCGR3B	Neutrophils	2215	***EMCN***	**Endothelial cells**	**51705**
HAL	Neutrophils	3034	***ESAM***	**Endothelial cells**	**90952**
KCNJ15	Neutrophils	3772	***ESM1***	**Endothelial cells**	**11082**
MEGF9	Neutrophils	1955	***FAM124B***	**Endothelial cells**	**79843**
SLC25A37	Neutrophils	51312	***HECW2***	**Endothelial cells**	**57520**
STEAP4	Neutrophils	79689	***HHIP***	**Endothelial cells**	**64399**
TECPR2	Neutrophils	9895	***KDR***	**Endothelial cells**	**3791**
TLE3	Neutrophils	7090	***MMRN1***	**Endothelial cells**	**22915**
TNFRSF10C	Neutrophils	8794	***MMRN2***	**Endothelial cells**	**79812**
VNN3	Neutrophils	55350	***MYCT1***	**Endothelial cells**	**80177**
COL1A1	**Fibroblasts**	**1277**	***PALMD***	**Endothelial cells**	**54873**
COL3A1	**Fibroblasts**	**1281**	***PEAR1***	**Endothelial cells**	**375033**
COL6A1	**Fibroblasts**	**1291**	***PGF***	**Endothelial cells**	**5228**
COL6A2	**Fibroblasts**	**1292**	***PLXNA2***	**Endothelial cells**	**5362**
DCN	**Fibroblasts**	**1634**	***PTPRB***	**Endothelial cells**	**5787**
GREM1	**Fibroblasts**	**26585**	***ROBO4***	**Endothelial cells**	**54538**

PAMR1	**Fibroblasts**	**25891**	***SDPR***	**Endothelial cells**	**8436**
TAGLN	**Fibroblasts**	**6876**	***SHANK3***	**Endothelial cells**	**85358**
			SHE	**Endothelial cells**	**126669**
			TEK	**Endothelial cells**	**7010**
			TIE1	**Endothelial cells**	**7075**
			VEPH1	**Endothelial cells**	**79674**
			VWF	**Endothelial cells**	**7450**

288

Table S2. Multivariable Cox regression analysis of overall survival using the TME risk score, clinicopathologic factors and MSI status

Variables	ACRG cohort		GSE15459		GSE84437	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value	HR (95% CI)	*P* value
TME risk score	1.39 (1.19 - 1.63)	4.70×10^{-5} ***	1.44 (1.14 - 1.83)	0.0026 **	1.41 (1.21 - 1.64)	1.20×10^{-5} ***
Age	1.03 (1.01 - 1.04)	0.0017 **	1.01 (0.99 - 1.03)	0.23	1.02 (1.01 - 1.03)	0.0039 **
Gender						
Female	1.00	-	1.00	-	1.00	-
Male	1.23 (0.87 - 1.76)	0.24	0.84 (0.52 - 1.36)	0.47	1.29 (0.95 - 1.75)	0.098
Stage						
I	1.00	-	1.00	-	-	-
II	1.49 (0.57 - 3.89)	0.42	2.47 (0.76 - 8.05)	0.13	-	-
III	3.02 (1.18 - 7.79)	0.022 *	9.10 (3.17 - 26.09)	4.00×10^{-5} ***	-	-
IV	6.96 (2.73 - 17.74)	4.90×10^{-5} ***	23.64 (8.05 - 69.38)	8.50×10^{-9} ***	-	-
Lauren classification						
Diffuse/Mixed	1.00	-	1.00	-	-	-
Intestinal	0.70 (0.49 - 1.00)	0.049 *	1.22 (0.77 - 1.92)	0.41	-	-
MS status						
Stable	1.00	-	1.00	-	1.00	-
Instable	0.75 (0.47 - 1.21)	0.23	1.51 (0.86 - 2.66)	0.15	1.07 (0.74 - 1.56)	0.72
Chemotherapy						
No	1.00	-	-	-	-	-
Yes	0.48 (0.34 - 0.70)	8.60×10^{-5} ***	-	-	-	-

Table S3. Top 20 genes positively correlated with NK cell abundance in a meta-analysis of the combined GEP cohorts

Gene symbol	Pearson's correlation	Gene name	Alias
KLRC1	0.73	killer cell lectin like receptor C1	*CD159A, NKG2, NKG2A, KLRC1*
KIR3DL2	0.69	killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2	*3DL2, CD158K, KIR-3DL2, NKAT-4, NKAT4, NKAT4B, p140, KIR3DL2*
FASLG	0.68	Fas ligand	*ALPS1B, APT1LG1, APTL, CD178, CD95-L, CD95L, FASL, TNFSF6, TNLG1A, FASLG*
SLA2	0.67	Src like adaptor 2	*C20orf156, MARS, SLAP-2, SLAP2, SLA2*
KLRC3	0.66	killer cell lectin like receptor C3	*NKG2-E, NKG2E, KLRC3*
PRF1	0.66	perforin 1	*HPLH2, P1, PFP, PRF1*
KLRC2	0.64	killer cell lectin like receptor C2	*CD159c, NKG2-C, NKG2C, KLRC2*
CD244	0.63	CD244 molecule	*2B4, NAIL, NKR2B4, Nmrk, SLAMF4, CD244*
GZMA	0.63	granzyme A	*CTLA3, HFSP, GZMA*
TARP	0.62	TCR gamma alternate reading frame protein	*CD3G, TCRG, TCRGC1, TCRGC2, TCRGV, TARP*
IFNG	0.62	interferon gamma	*IFG, IFI, IFNG*
APOBEC3H	0.62	apolipoprotein B mRNA editing enzyme catalytic subunit 3H	*A3H, ARP-10, ARP10, APOBEC3H*
KLRD1	0.61	killer cell lectin like receptor D1	*CD94, KLRD1*
GBP5	0.61	guanylate binding protein 5	*GBP-5, GBP5*
IL12RB1	0.60	interleukin 12 receptor subunit beta 1	*CD212, IL-12R-BETA1, IL12RB, IMD30, IL12RB1*
ZNF683	0.60	zinc finger protein 683	*Hobit, ZNF683*
GZMB	0.60	granzyme B	*C11, CCPI, CGL-1, CGL1, CSP-B, CSPB, CTLA1, CTSGL1, HLP, SECT, GZMB*
GNLY	0.59	granulysin	*D2S69E, LAG-2, LAG2, NKG5, TLA519, GNLY*
LAG3	0.59	lymphocyte activating 3	*CD223, LAG3*

Table S4. Top pathways correlated with NK cell abundance or Stroma score in a meta-analysis of the combined GEP cohorts

Pathways positively correlated with NK cell abundance			Pathways positively correlated with stroma score		
Rank	**r**	**Pathways**	**Rank**	**r**	**Pathways**
1	0.52	INTERFERON_GAMMA_RESPONSE	1	0.91	EPITHELIAL_MESENCHYMAL_TRANSITION
2	0.51	ALLOGRAFT_REJECTION	2	0.85	UV_RESPONSE_DN
3	0.48	INTERFERON_ALPHA_RESPONSE	3	0.84	APICAL_JUNCTION
4	0.44	IL6_JAK_STAT3_SIGNALING	4	0.83	MYOGENESIS
5	0.40	COMPLEMENT	5	0.76	ANGIOGENESIS
6	0.38	INFLAMMATORY_RESPONSE	6	0.73	KRAS_SIGNALING_UP
7	0.33	PI3K_AKT_MTOR_SIGNALING	7	0.68	COAGULATION
8	0.29	APOPTOSIS	8	0.65	HYPOXIA
9	0.29	IL2_STAT5_SIGNALING	9	0.62	IL2_STAT5_SIGNALING
10	0.29	TNFA_SIGNALING_VIA_NFKB	10	0.62	TGF_BETA_SIGNALING
Pathways negatively correlated with NK cell abundance			**Pathways positively correlated with stroma score**		
Rank	**r**	**Pathways**	**Rank**	**r**	**Pathways**
1	-0.28	HEDGEHOG_SIGNALING	1	-0.61	E2F_TARGETS
2	-0.27	MYOGENESIS	2	-0.61	MYC_TARGETS_V2
3	-0.24	WNT_BETA_CATENIN_SIGNALING	3	-0.59	G2M_CHECKPOINT
4	-0.20	UV_RESPONSE_DN	4	-0.56	DNA_REPAIR
5	-0.20	ADIPOGENESIS	5	-0.55	MYC_TARGETS_V1
6	-0.20	BILE_ACID_METABOLISM	6	-0.51	SPERMATOGENESIS
7	-0.19	NOTCH_SIGNALING	7	-0.40	UNFOLDED_PROTEIN_RESPONSE
8	-0.18	PANCREAS_BETA_CELLS	8	-0.38	MTORC1_SIGNALING
9	-0.14	TGF_BETA_SIGNALING	9	-0.37	OXIDATIVE_PHOSPHORYLATION
10	-0.13	EPITHELIAL_MESENCHYMAL_TRANSITION	10	-0.33	PEROXISOME

Supplementary Figures

Figure S1. The workflow of this study.

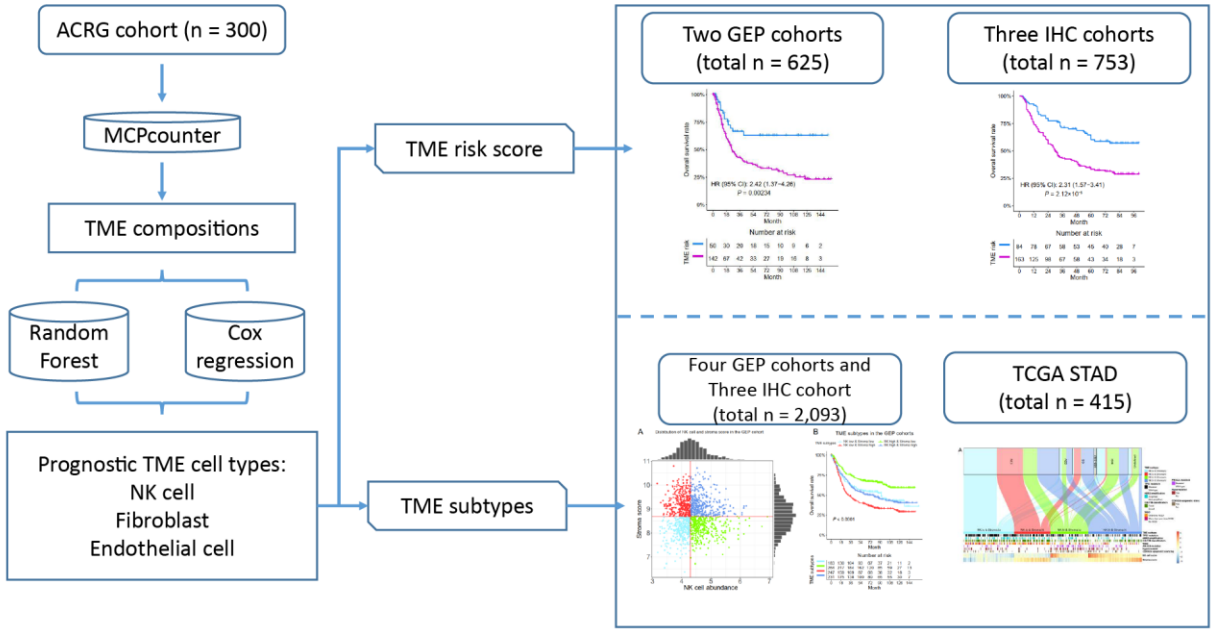


Figure S2. Prognostic effects of the major cellular components of the TME and their pairwise correlation in the ACRG discovery cohort (A) Univariate Cox regression analysis revealed that the absolute abundance levels of NK cells, fibroblasts, and endothelial cells were prognostic of overall survival with a pre-specified statistical significance $P < 0.05$. (B) Univariate Cox regression analysis with stage and chemotherapy adjusted revealed that the absolute abundance levels of NK cells, endothelial cells, and fibroblasts were prognostic of overall survival with a pre-specified statistical significance $P < 0.05$. (C) The abundance levels of NK cells, fibroblasts and endothelial cells were the most important features in the random survival forest model for predicting overall survival. Turquoise color corresponds to positive feature importance and coral color corresponds to negative features importance. (D) Endothelial cells, fibroblasts, and the stroma score were all highly correlated with each other. By contrast, NK cells were uncorrelated to endothelial cells, fibroblasts, and the stroma score. Hazard ratios (HR), confidence intervals (CI) and *P* values in A and B were estimated by Cox regression.

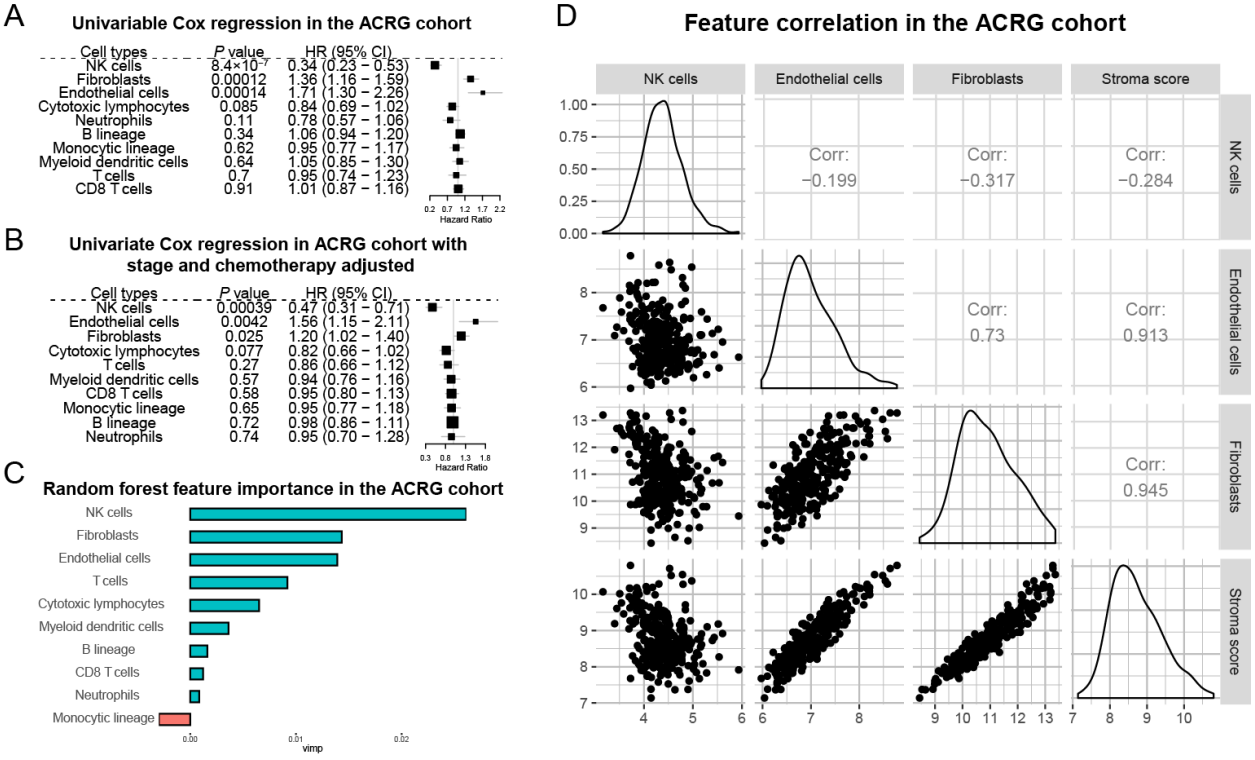
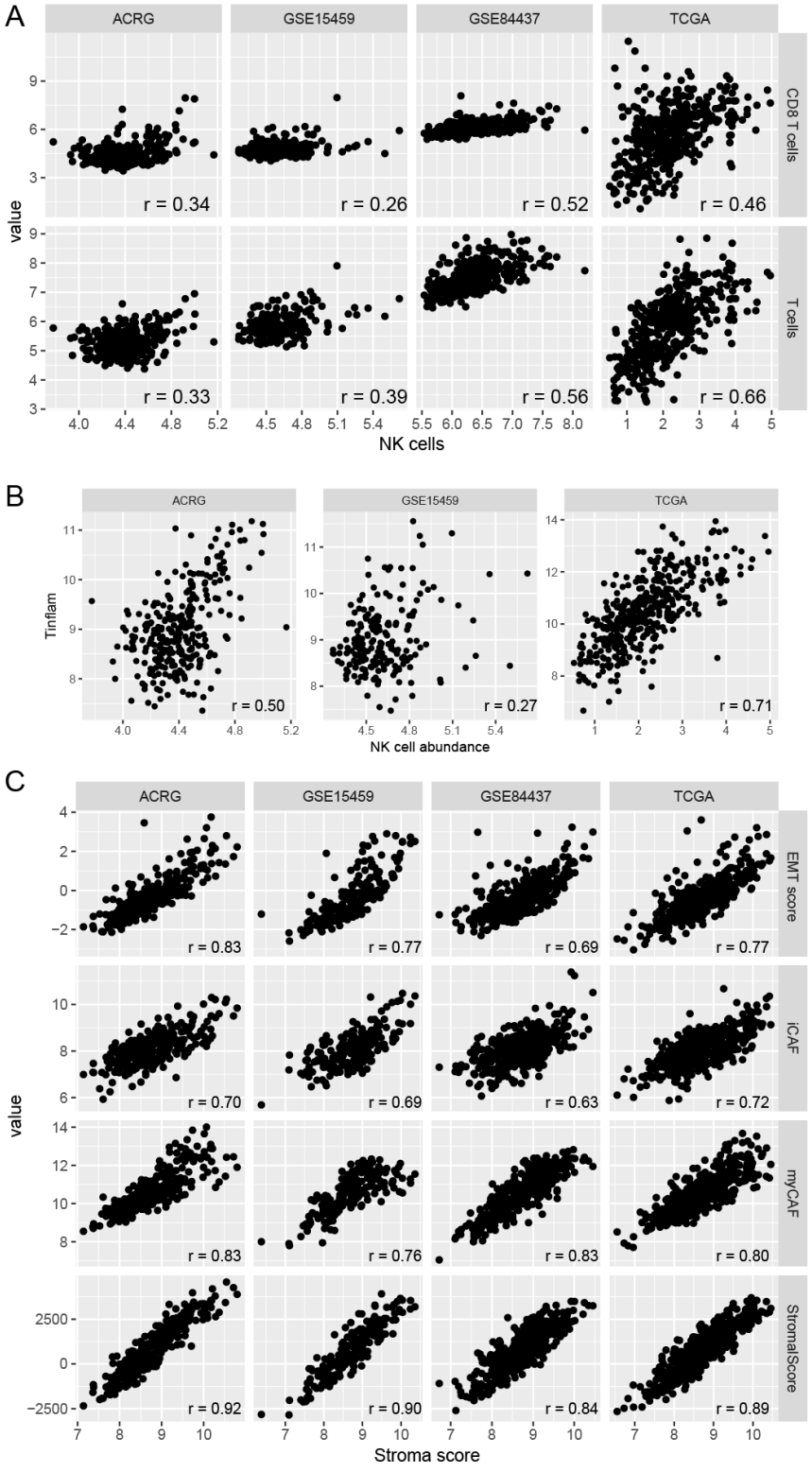


Figure S3. Correlation of NK cells or stroma scores with T cells and other gene signatures



Figures S4. Bivariate Cox regression on the merged GEP and IHC cohorts. Bivariate Cox regression indicated that NK cell abundance and stroma score were independent prognostic factors in the merged GEP validation cohorts (GSE15459 and GSE84437) (A) and merged IHC validation cohorts (B). Hazard ratios (HR), confidence interval (CI) and *P* values were estimated by Cox regression.

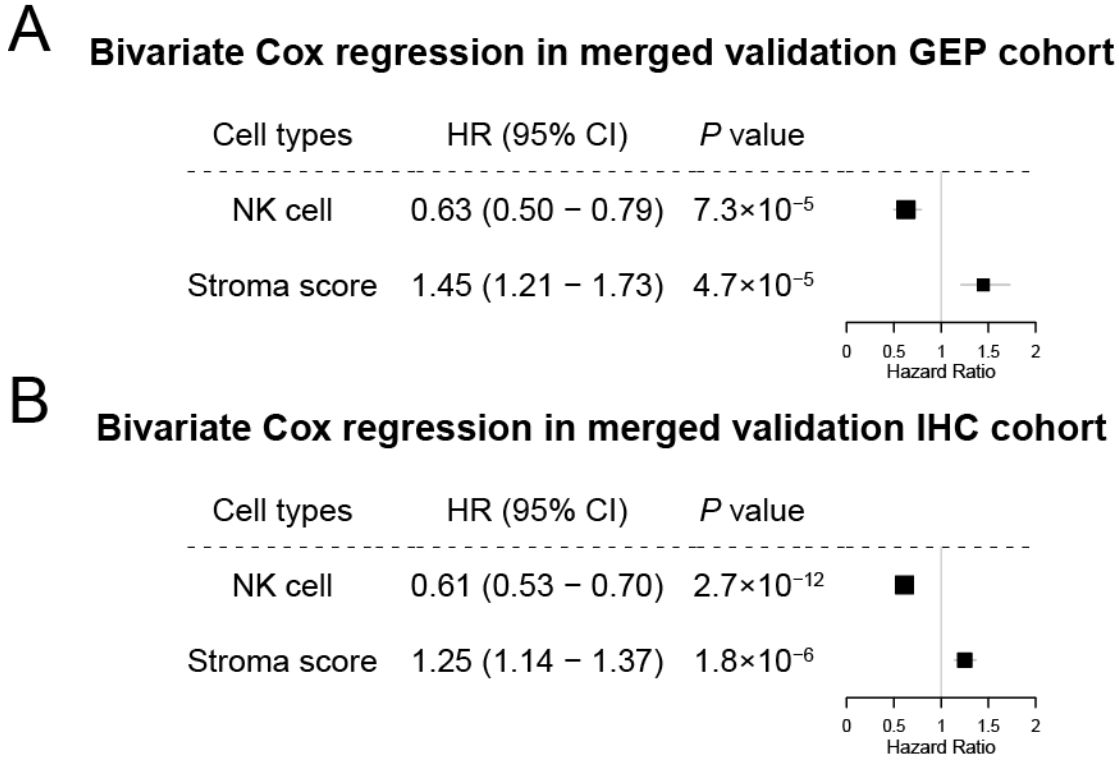
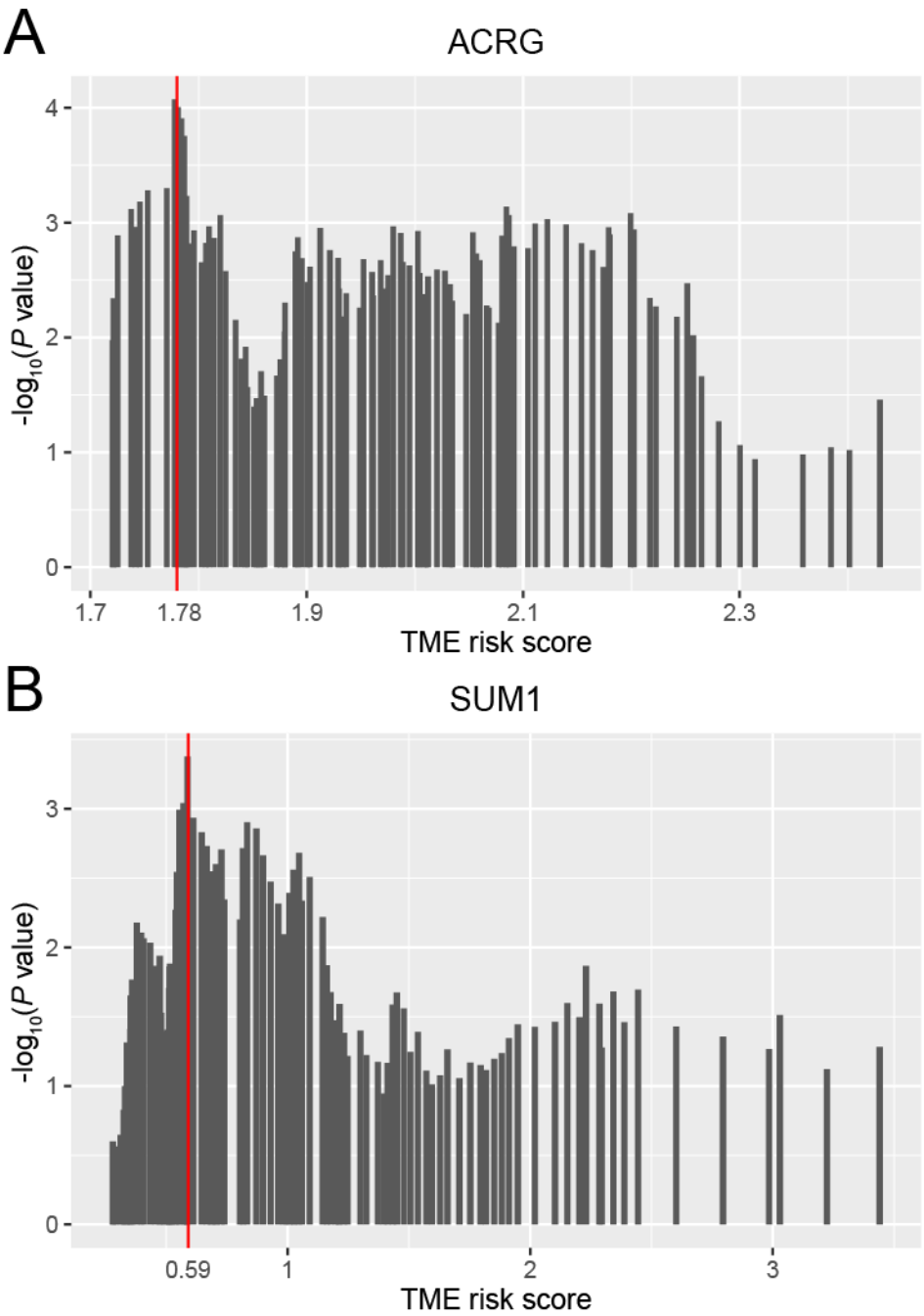


Figure S5. Cox regression *P* values at different cutoffs of the TME risk score. (A) The TME risk score of 1.78 (vertical red line) was chosen as the cutoff for GEP cohorts, based on the minimal Cox regression *P* value in the ACRG cohort. (B) The TME risk score of 0.59 (vertical red line) was chosen as the cutoff for IHC cohorts, based on the minimal Cox regression *P* value in the SMU1 cohort.



Figures S6. The prognostic effects of the TME risk score in patients within each pathological stage in the combined GEP cohorts. A high TME risk score was consistently associated with worse overall survival in patients with stage I (A), stage II (B), stage III (C), and stage IV (D) GC in the merged GEP cohorts (ACRG and GSE15459). Hazard ratios (HR) and confidence intervals (CI) were estimated by Cox regression. *P* values were generated by log-rank test.

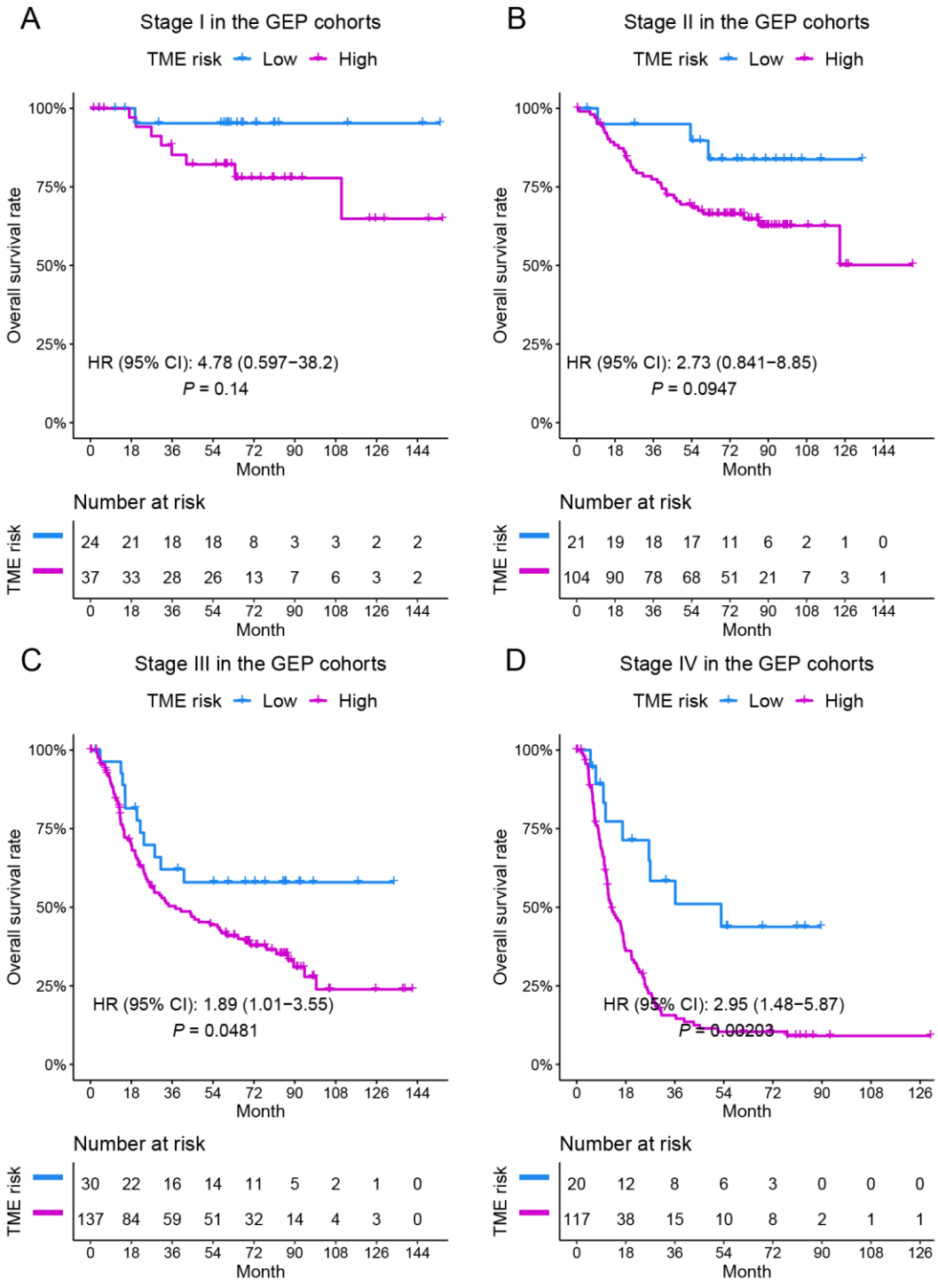


Figure S7. The prognostic effect of the TME risk score in patients treated with surgery only and without chemotherapy. A high TME risk score was associated with worse overall survival in patients who received surgery alone in the ACRG (A), SMU1 (B), SMU2 (C), and SYSU (D) cohorts. HR and confidence intervals were estimated by Cox regression. *P* values were generated by log-rank test.

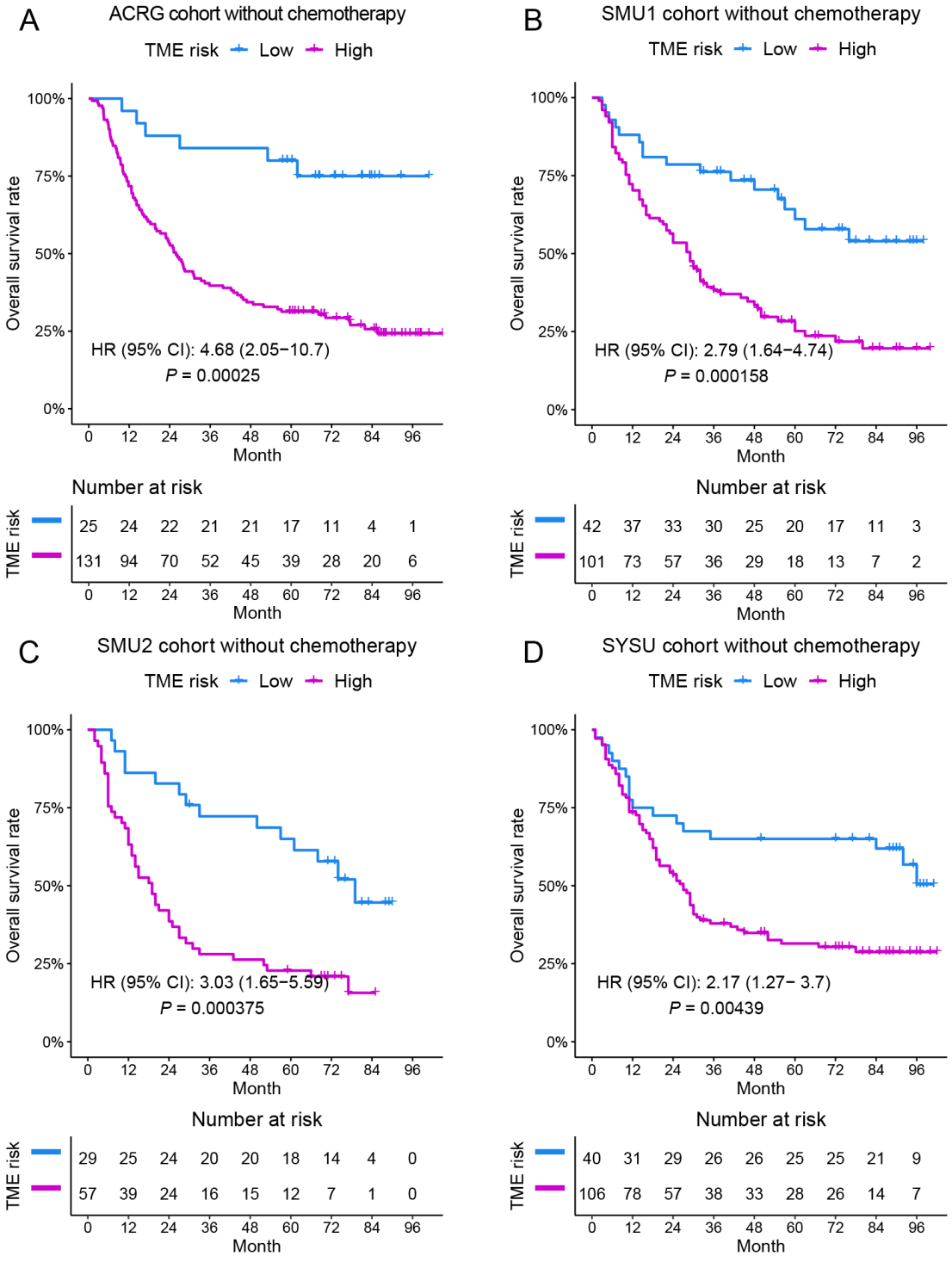


Figure S8. Predictive relevance of the TME risk score for the benefit of chemotherapy in unmatched stage I-III gastric cancer. (A) Patients with a high TME risk score derived a significant survival benefit from adjuvant chemotherapy at 5 years. However, patients with a low TME risk score did not benefit from adjuvant chemotherapy (B). Hazard ratios (HR) and confidence intervals (CI) were estimated by Cox regression. *P* values were generated by log-rank test. The *P* value for the interaction between the TME risk group and adjuvant chemotherapy was 0.0965.

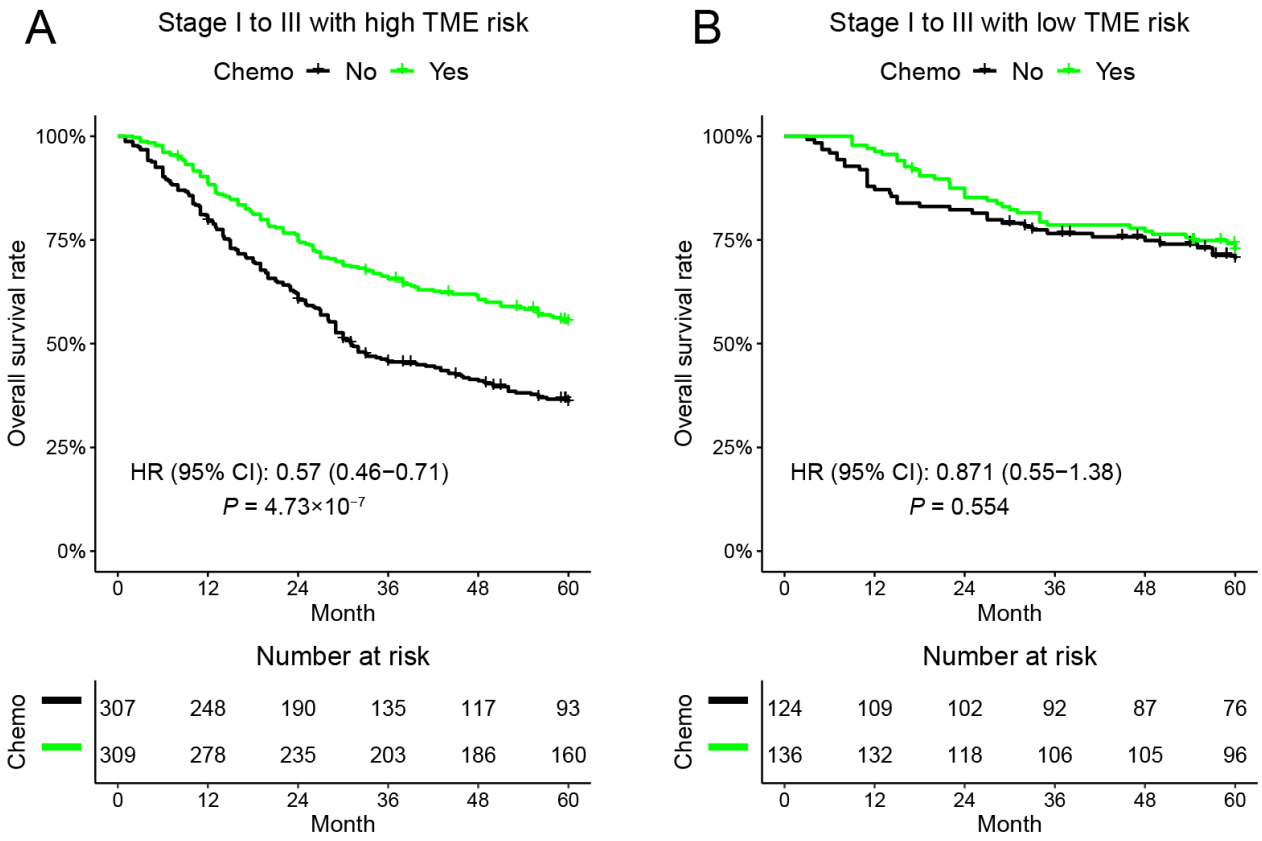


Figure S9. Complementary prognostic value of the TME subtypes to the intrinsic subtypes for gastric cancer. *P* values were generated by log-rank test.

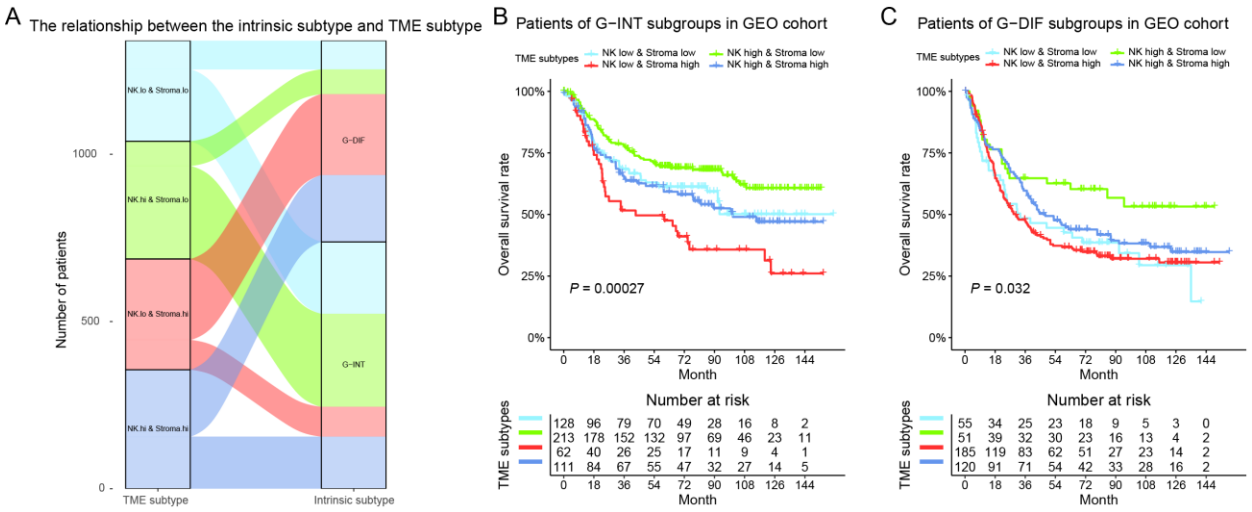


Figure S10. Genomic features significantly associated with NK cell infiltration status.

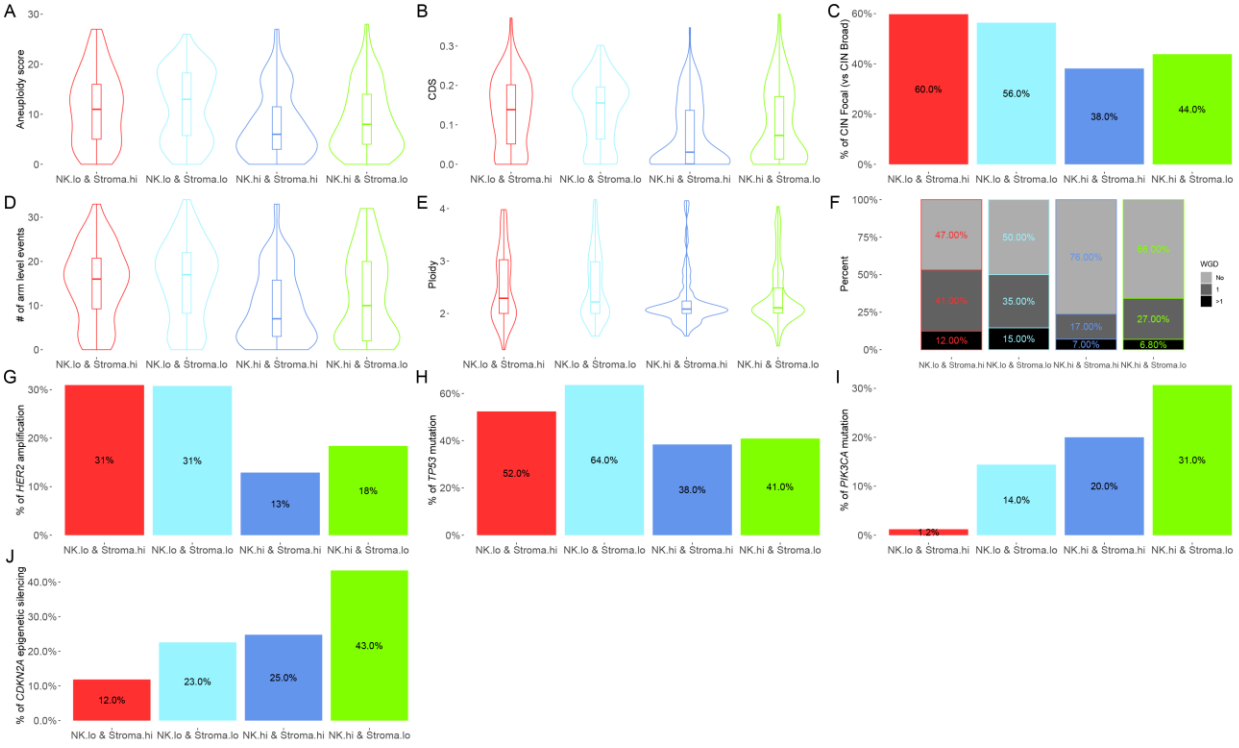


Figure S11. Differential mutation status of TP53 and PIK3CA in patients with different MSI or T-cell inflamed levels. *P* values were generated by Chi-squared tests.

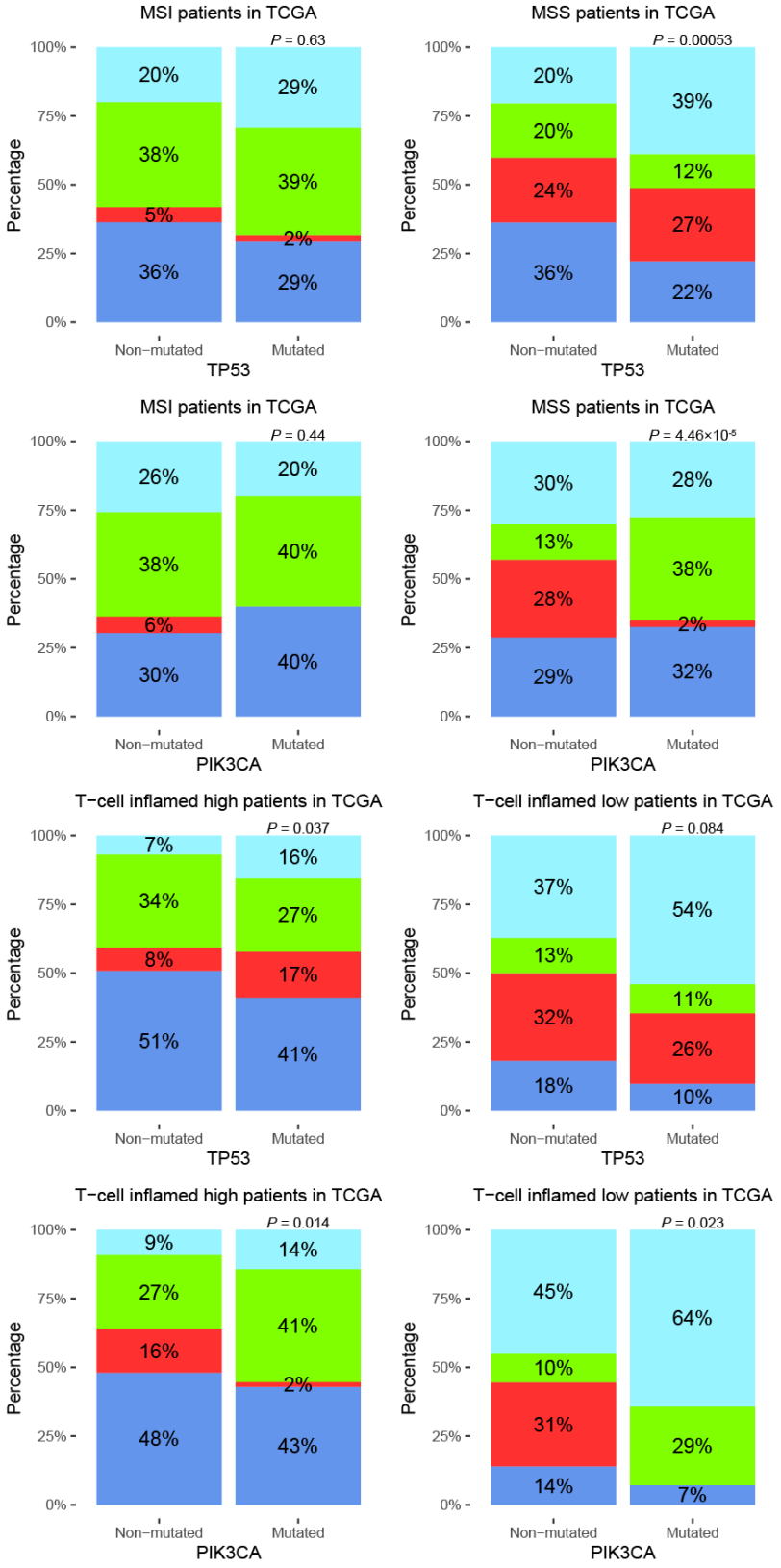


Figure S12. Genomic features significantly associated with Stroma status.

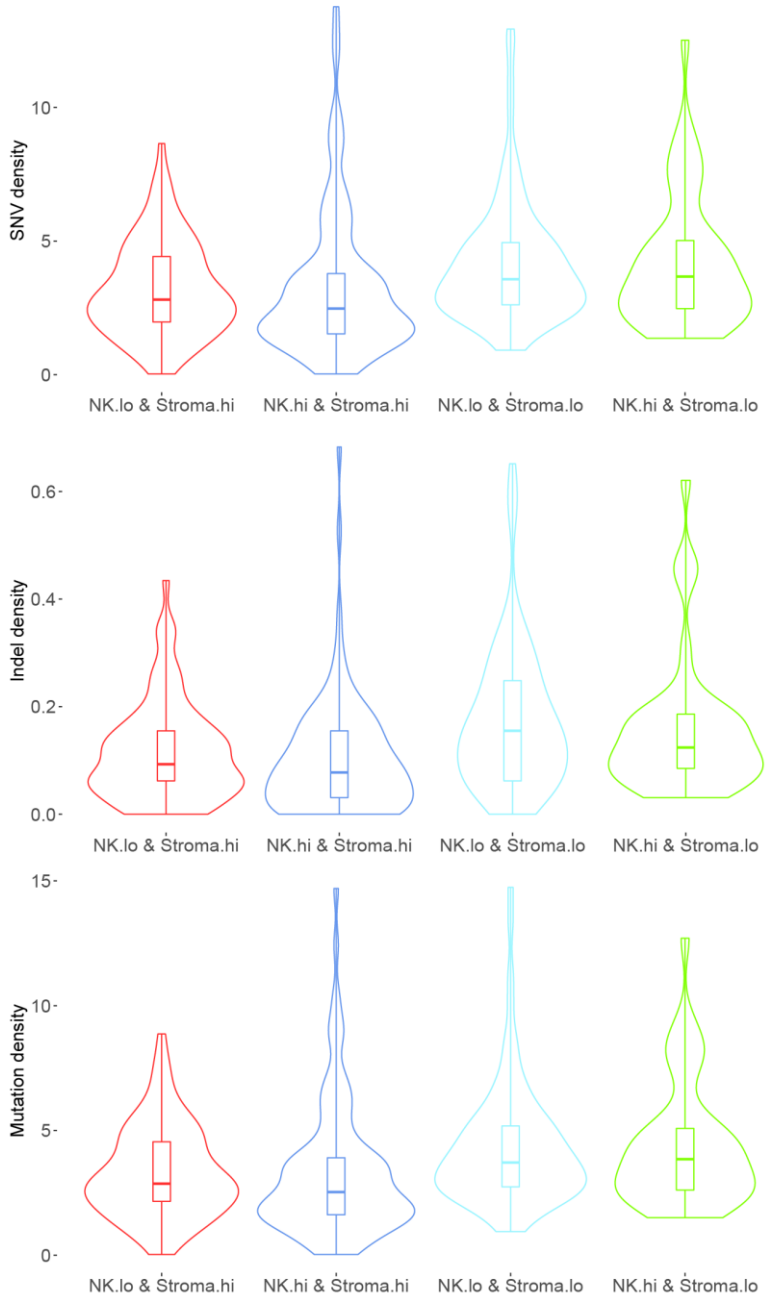


Figure S13. Genomic features correlated with both NK and stroma status

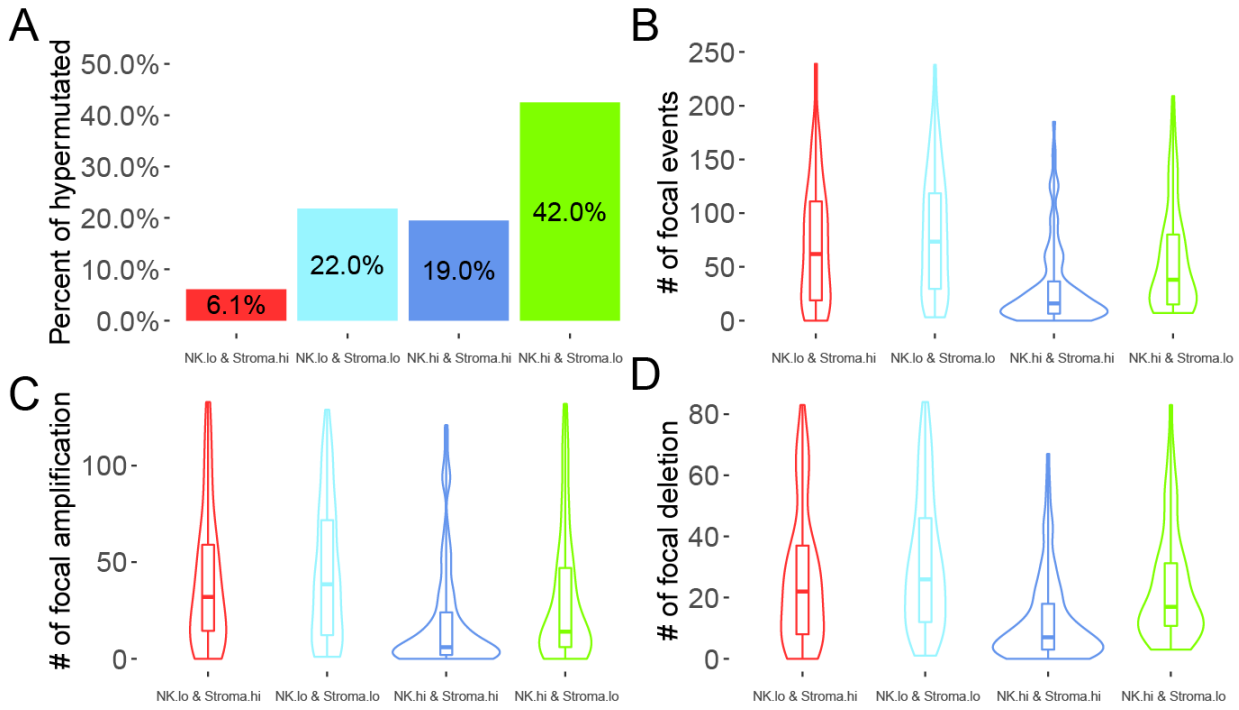


Figure S14. Genomic features correlated with NK or stroma status in the patients with CIN from the TCGA cohort.

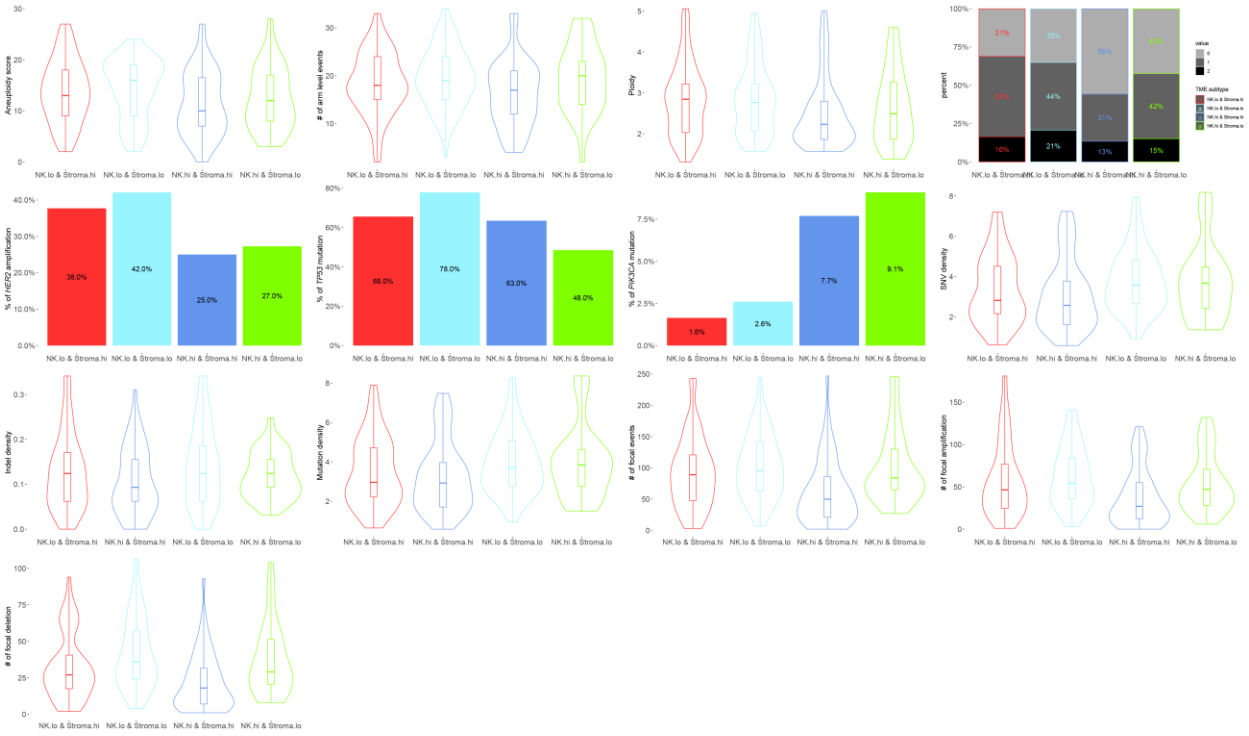


Figure S15. Top hallmark pathways correlated with NK cell abundance

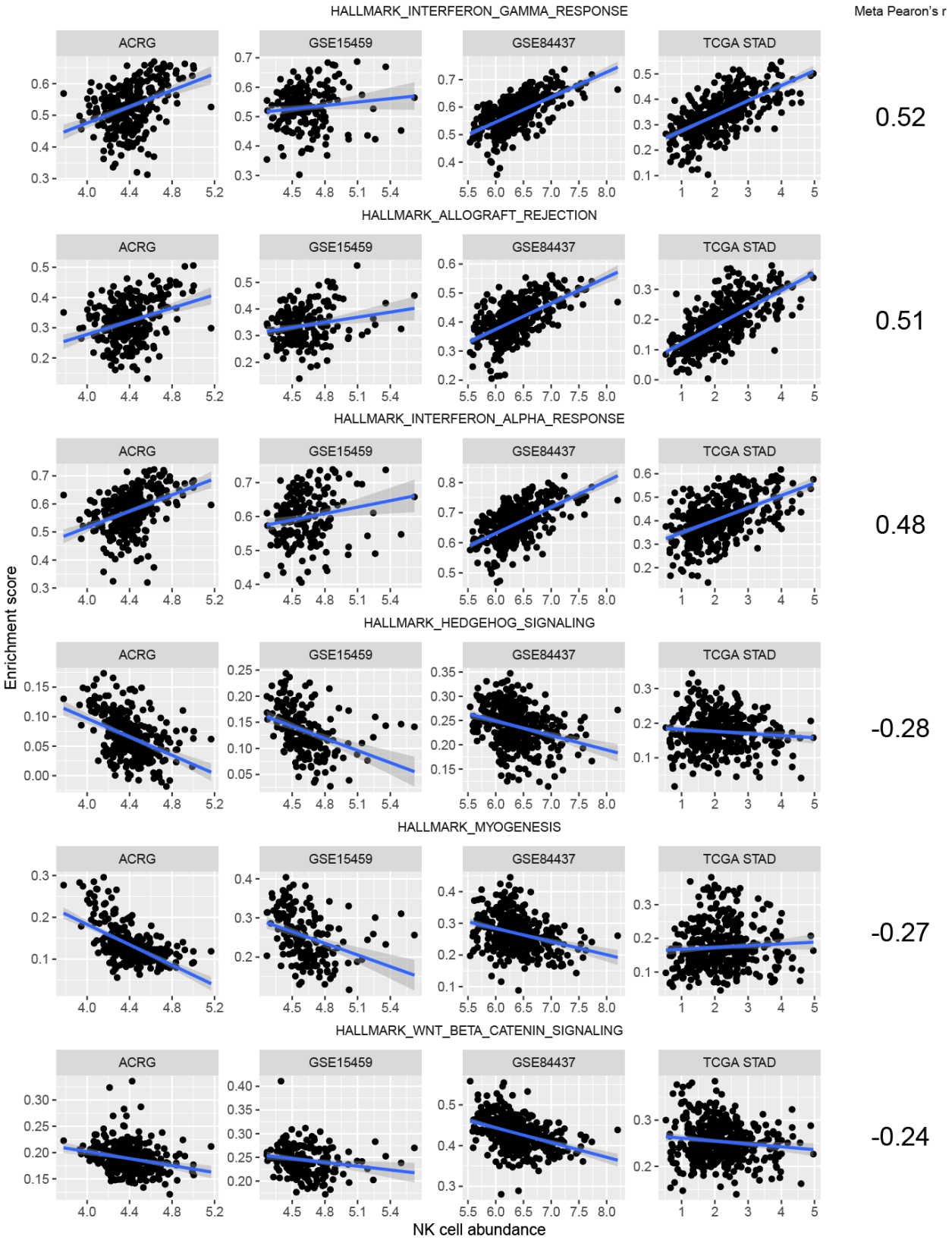


Figure S16. Top hallmark pathways correlated with Stroma score

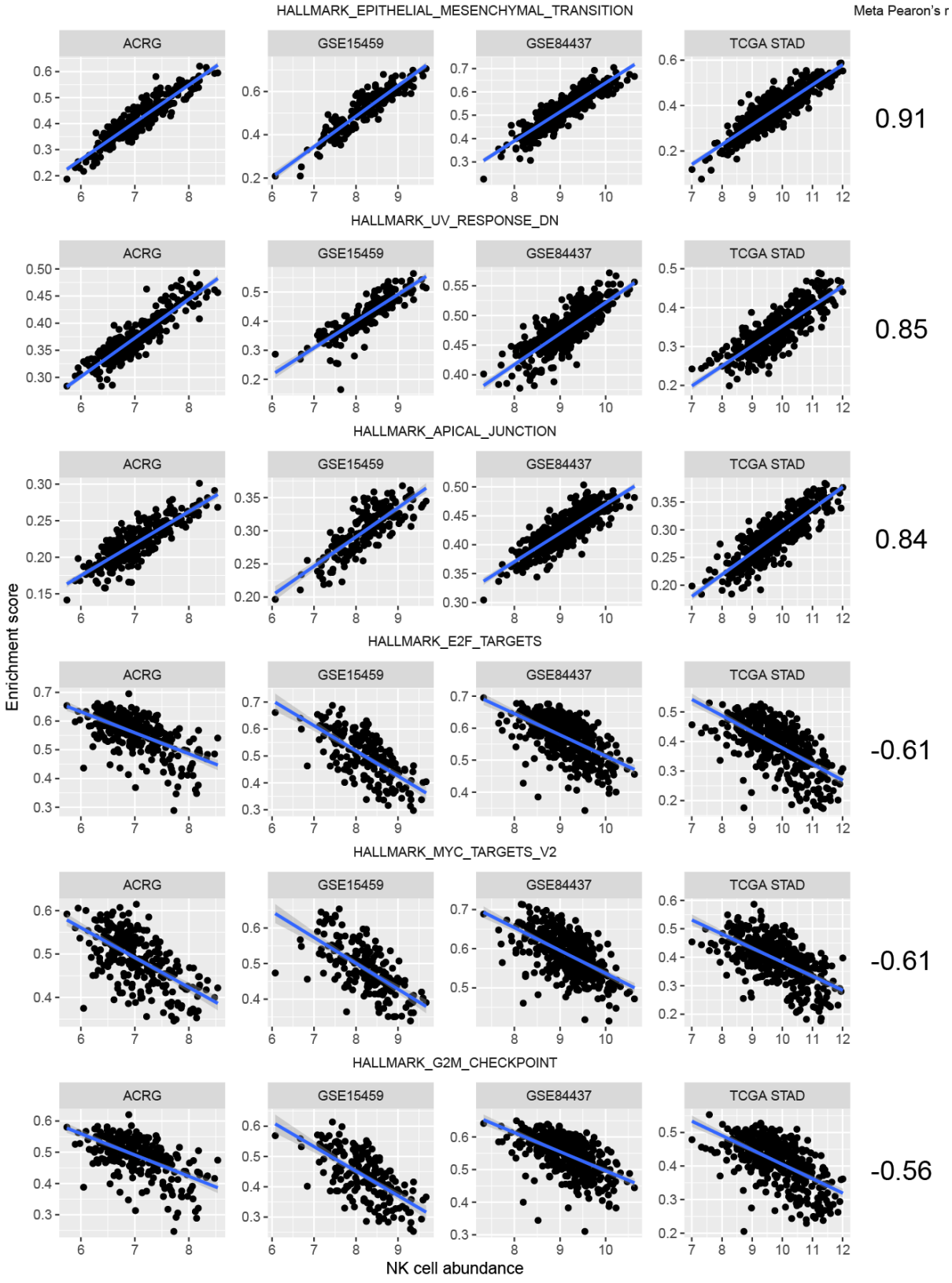


Figure S17. Differentially expressed cytokines in different NK (A) and Stroma (B) groups.