Supplementary Materials

- **Supplementary Methods**
- **Supplementary Results**
- **Supplementary References**
- **Supplementary Figures**
- **Supplementary Tables**

Supplementary Methods

1. Patients and enrolled criteria

The overall study design is shown in Figure 1, and the workflow of enrolled patients is shown in Supplementary Figure 1. We retrospectively collected data for 1,704 patients with gastric cancer (GC) in seven cancer centers. The inclusion criteria were: histologically confirmed gastric adenocarcinoma; no preoperative chemotherapy; and follow-up data available. We excluded those patients who had other synchronous malignant neoplasms or had received previous anticancer treatment or with unqualified H&E images.

2. Data processing

We conducted a systematic search for GC gene expression dataset, which were publicly accessible and had clinical annotations in the The Cancer Genome Atlas (TCGA; https://portal.gdc.cancer.gov/) and the Gene-Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) databases. Totally, we achieved six cohorts of 1,636 patients with GC for this study: TCGA-STAD, GSE62254/ACRG, GSE13861/YUSH, GSE26253/SMC, and GSE84437/KRIBB. Raw data for the microarray datasets generated by Affymetrix or Illumina platform were screened from the GEO, and processed for background adjustment, quantile normalization, and final summarization by Perl software and limma packages. The corresponding clinical information was downloaded or manually registered from the item page in the GEO dataset website. For some series whose clinical data could not be obtained through the aforementioned methods, we retrieved the exact clinical information from the

supplementary materials of relevant published papers [1]. Level 3 gene expression profile (FPKM normalized) and corresponding clinical data of the TCGA-STAD were downloaded from the TCGA database. Missing or updated clinical–genomic data was replenished from the UCSC Xena browser (GDC hub: https://gdc.xenahubs.net) and cBio Cancer Genomics Portal (cBioPortal: https://www.cbioportal.org/).

Fresh tumor tissue of GC was achieved from the postoperative specimen within 30 minutes and stored in RNA protective solution (Invitrogen, the USA). The following sequencing task was conducted by the BGI institute. In general, after passing standard procedure quality control, followed by RNA extraction, separation, and interruption, the extracted RNA is reverse transcribed, cDNA is synthesized, and adapter sequences are added to it to form an RNA-seq library. Next, RNA-seq libraries are sequenced at high throughput using Illumina sequencing platforms. Quality control of the raw data obtained by sequencing, removal of low-quality sequences and adapter sequences, and then comparison of sequencing reads to the reference genome or transcriptome to determine the expression level of each gene. Finally, calculate the expression of a gene by statistically comparing the number of reads to each gene using FPKM (base logarithm per million reads) as the expression measurement. Thus, we obtained a transcriptome matrix from cohort of the SMU hospital.

The original paired gene sequence and corresponding clinical information of the PRJEB25780 cohort (pembrolizumab treatment) were downloaded from the European Bioinformatics Institute (EMBL-EBI) database (https://www.ebi.ac.uk/). The adapter and low-quality sequences were removed from the raw data using Trim Galore software.

The quality of the samples after filtration was checked and adjusted by the FastQC software. Clean reads were compared with the human genome (HG38 version) using HISAT2 software, and then a read count of gene expression was generated using FeatureCounts software. Finally, the gene expression profile was normalized by the limma package.

Finally, all raw data was normalized using the standardize algorithm by R package to eliminate the batch effects.

3. TLSs-related gene features

In the present study, 39 TLSs-related genes to construct the generic gene expression profiling included CCL2/3/4/5/8/18/19/21, CXCL9/10/11/13 were chemokine signature genes; CXCL13, CD200, FBLN7, ICOS, SGPP2, SH2D1A, TIGIT, PDCD1 were T follicular helper cell (TFH cell) signature genes; CD4, CCR5, CXCR3, CSF2, IGSF6, IL2RA, CD38, CD40, CD5, MS4A1, SDC1, GFI1, IL1R1, IL1R2, IL10, CCL20, IRF4, TRAF6, STAT5A were T helper 1 cell (TH1 cell) and B cell signature genes; TNFRSF17 was plasma cell signature gene.

4. Hematoxylin-eosin (H&E) staining

In the present study, tumor tertiary lymphoid structures (TLSs) was stained and calculated in whole slide images (WSIs) using NanoZoomer Digital Pathology (NDP) platform. Firstly, formalin-fixed paraffin-embedded (FFPE) GC samples were processed for H&E staining as previously described [2]. Each tumor was made into paraffin sections with 3um thick. After dewaxing and hydration, the slides were stained with hematoxylin for 5 minutes, followed by running water flushing for 3 minutes,

hydrochloric acid alcohol differentiation for 30 seconds, running water flushing for 3 minutes, 1% ammonia reverting blue, running water flushing for 3 minutes, and finally eosin staining for 1 minute. Thereafter, the tissue slides were stored as digital pathology scans. To evaluate the existence of TLSs, two pathologists who were blinded to clinical outcomes independently classified all samples. A third pathologist was consulted when a difference of opinion arose between the two primary pathologists. The existence of TLSs was assessed morphologically on the whole slide image (WSIs) of H&E staining, using a previously published scale [3, 4]. Briefly, GC cases were classified according to TLSs status in the WSIs using NanoZoomer Digital Pathology (NDP) platform (Hamamatsu, Japan).

We then determined the patient-level TLSs status based on the following rule. 1) TLSs absence GC (TLS−): tumors without any TLSs; 2) TLSs presence (TLS+) GC: tumors with at least one TLSs.

5. Multi-omics analysis

Dataset from TCIA/TCGA-STAD was used to decode the associations between multi-omics characterization and TLSs status. Briefly, we achieved multi-omics data from TCGA-STAD cohort. We next decoded the associations between multi-omics characterization and TLSs status. We further evaluated the prognostic value of the H&E-based TLSs in patents with GC. Multi-omics data contained publicly available processed genomic data, level 3 normalized RNAseq data, level 3 microRNA data, level 3 reverse phase protein array (RPPA) data, and H&E slides. Heatmap and waterfall plot were generated to show a visually interpretable overview of copy number variations and

somatic mutation stratified by TLSs status from the genomic level using R package "GenVisR" and "pheatmap", respectively. Differential expression analysis was performed in the mRNA, miRNA, and proteomic expression profiles to identify differentially expressed molecules (Fold-change ≥ 2, $p < 0.05$) associated with TLSs status using R package "edgeR". Gene set enrichment analysis (GSEA) analysis to elucidate the underlying molecular signaling of TLSs status was conducted in the GSEA software. All parameters were set to their default values, and an adjusted p-value of < 0.05 was considered statistically significant.

6. Tumor microenvironment infiltrating cells and TME score dissecting

To quantify the composition of tumor-associated infiltrating cells in the GC samples, the proposed computational algorithms Microenvironment Cell Populations-counter (MCPcounter) was conducted [5]. Based on the gene expression profiles, the absolute abundance of ten kinds of immune-stromal associated cells, including two stromal cells (tumor-associated fibroblasts and endothelial cells) and eight immune cells (CD3 T cells, CD8 T cells, cytotoxic lymphocytes, B cell lineage, NK cells, monocytic lineage, myeloid dendritic cells, and neutrophils), was estimated by the MCPcounter algorithm.

The tumor-associated immune and stromal scores representing tumor microenvironment (TME) characterization were calculated based on the normalized gene-expression matrix using the ESTIMATE algorithm for each GC sample [6].

7. SHAP analysis

The Shapley value is a concept from coalitional game theory designed to fairly distribute the total surplus or reward attained by a coalition of players to each player in that coalition [7]. For an arbitrary coalitional game, $v(S)$: $\mathcal{P}(S) \mapsto \mathbb{R}$ (where S is the

set of players and $\mathcal{P}$ indicates the powerset), the Shapley value for a player i is defined as the marginal contribution of that player averaged over the set of all $d!$ possible orderings R of the d players in S:

$$\phi(i) = \frac{1}{d!}\sum_{R} v(S_i^R \cup i) - v(S_i^R)$$

where S_i^R indicates the set of players in S preceding player i in order R.

To use this value to allocate credit to features in a ML model, the model must first be defined as a coalitional game. Deciding exactly how to define a model as a game is non-trivial, and a variety of different approaches have been suggested [8-10]. The most popular, SHAP (SHapley Additive exPlanations) [8], defines the game as the conditional expectation of the output of a model f for a particular input sample $x \in \mathbb{R}^d$ given that the features in S have been observed:

$$v(S) = \mathbb{E}[f(x)|x_S]$$

Because modelling an exponential number of arbitrary conditional distributions is often intractable, in practice the simplifying assumption that input features are independent is often made, allowing the expected value to be calculated over the marginal distributions of the features not in each given set, rather than the conditional distributions [8].

8. Statistical analysis

Continuous variables were compared using the t-test or Mann-Whitney test. Categorical variables were compared with the chi-squared or Fisher's exact test. Survival curves were generated according to the Kaplan-Meier method and compared by the log-rank test. Univariate and multivariate analyses were performed using the Cox

proportional hazards model. Predictive features were selected using the Least Absolute Shrinkage and Selector Operation (LASSO) logistic regression algorithm by R package "glmnet" with 5-fold cross-validation. ROC curves were plotted by using the R packages "pROC". Nomogram were constructed by using the R package "rms". Decision curve analysis (DCA) was performed using the R package "ggDCA". Heatmap and waterfall plot were generated by R package "GenVisR" and "pheatmap", respectively. Differential expression analysis was performed using R package "edgeR". Volcano plot was generated by R package "ggplot2". SHAP was performed by "shap" library in python software. The gene set enrichment analysis (GSEA) was conducted in the GSEA software.

Supplementary Results

1. Multi-omics analysis

To identify the genetic events associated with TLSs, we compared the mutational profile of TLS+ and TLS−. We did not find any significant differences in copy number variations between the two groups, and the incidence of most somatic mutations was similar, although there was a higher incidence of ARID1A mutations (35.8% vs 25.2%), CDH1 mutations (12.8% vs 6.5%), CHD3 mutations (9.7% vs 4.7%), RHOA mutations (7.5% vs 4.7%), and MYC mutations (2.7% vs 0.9%) in the TLS+ group compared with the TLS− group (Figure 2C).

We further analyzed the transcriptional profile to identify dysregulated genes associated with TLSs (Figure 2D). We identified 145 differentially expressed genes of which most were correlated with immunoregulation and tumor activation or inhibition,

including FCRL1, CLNK, and APOH. We also identified 27 microRNAs (Figure 2E) and 27 proteins (Figure 2F) to be differentially expressed between TLS+ and TLS− groups. Of note, LCK (involves in the maturation of T lymphocytes) protein was overexpressed in the TLS+ group. All other dysregulated proteins and microRNAs exhibited key functions in tumor development, including CDK1, BAK1, ERBB2, hsa-mir-302c, and hsa-mir-4458. Furthermore, pathway analyses at transcriptomic (Figure 2D) and proteomic (Figure 2F) level found that compared with TLS− group, the TLS+ was positively correlated with tumor immune activation and apoptosis signaling, such as inflammatory response signaling, chemokine signaling, and apoptosis signaling, while displaying a negative correlation with tumor proliferation and metabolism signaling, such as MYC signaling and oxidative phosphorylation signaling (all FDR<0.05). Thus, our analyses did not find a clear correlation of TLSs with genomics, but revealed a strong link to the transcriptomic and proteomic landscapes.

2. Workflow of model construction

The least absolute shrinkage and selection operator method (LASSO) is a popular method for regression of high-dimensional data. The method uses an L1 penalty to shrink some regression coefficients to exactly zero. We plotted the AUC versus log (λ), where λ is the tuning parameter for the LASSO logistic regression model. A value of λmin= 0.03142564 with log (λ) = -3.460131 was selected by maximizing AUC values in 5-fold cross validation. The optimal tuning parameter resulted in nine non-zero coefficients in the final logistic regression model. The following nine features were selected in the LASSO logistic regression model: MS4A1, CSF2, CXCL13, FBLN7,

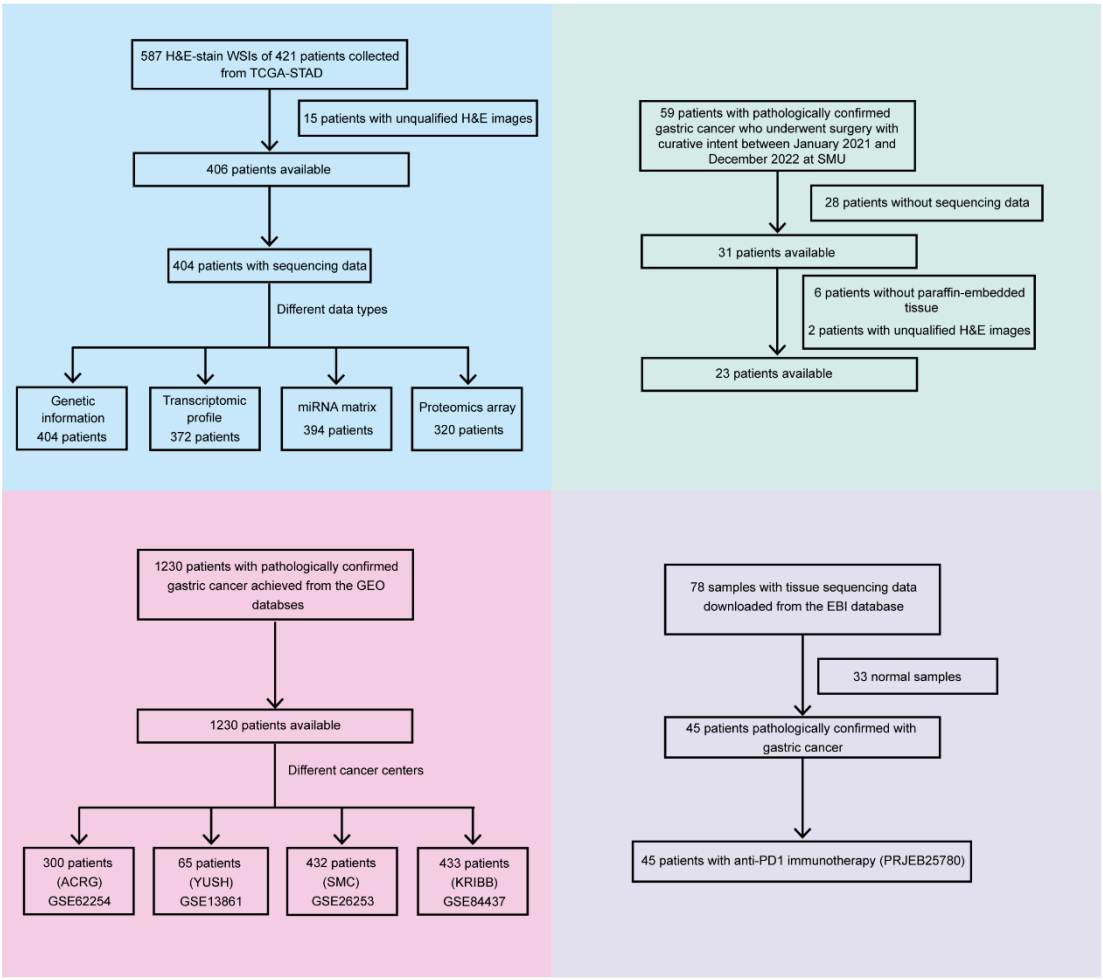
CCR5, TIGIT, CCL21, IRF4, and CD200.

3. Calculation formula

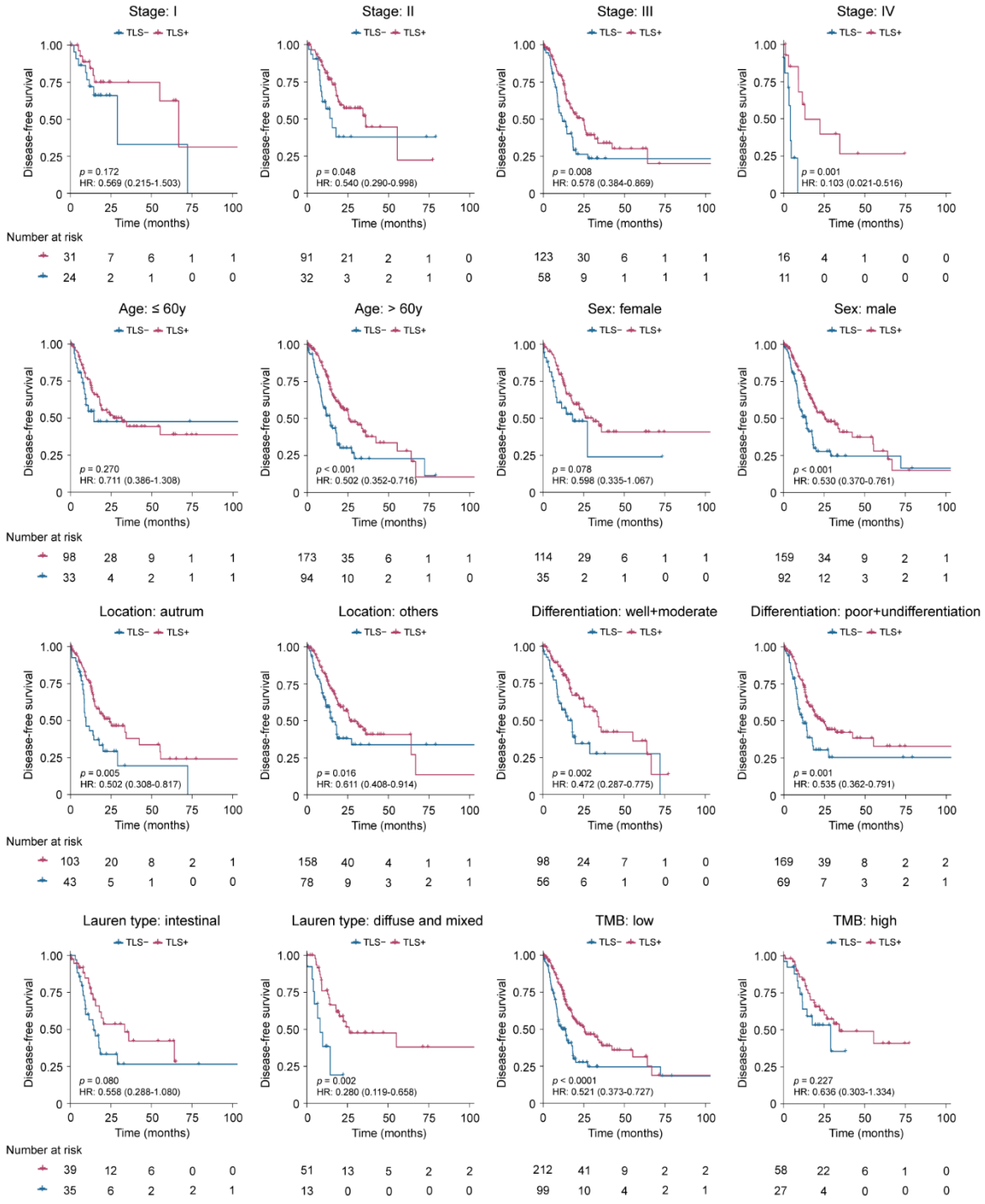
$$gsTLS = 0.3706 * MS4A1 - 0.2788 * CSF2 + 0.2308 * FBLN7 + 0.184 * CCR5 + 0.1185 * TIGIT + 0.0905 * CCL21 + 0.0826 * IRF4 + 0.0412 * CXCL13 + 0.0385 * CD200$$

Supplementary References

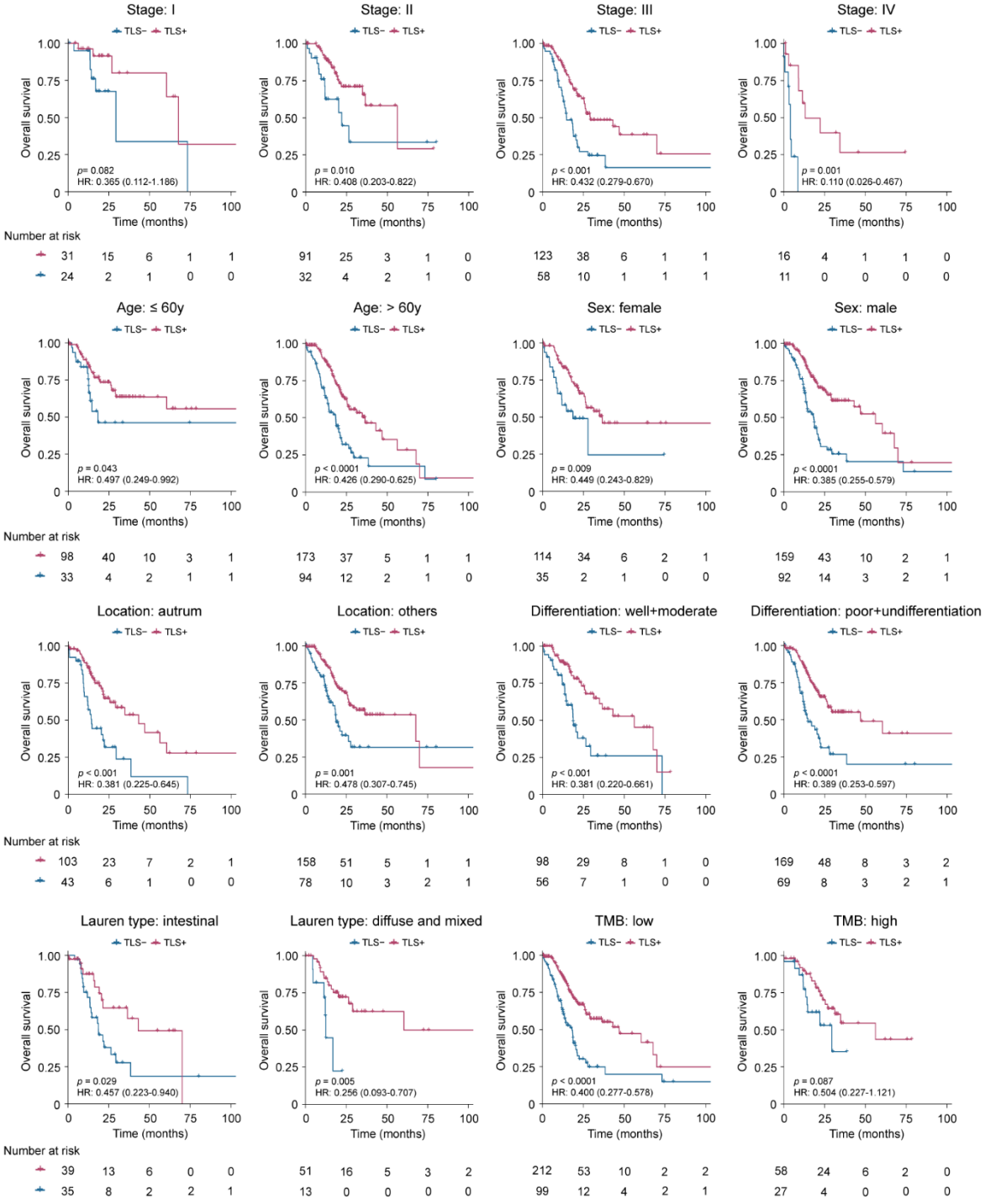
1. Oh SC, Sohn BH, Cheong JH, Kim SB, Lee JE, Park KC, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun. 2018;9(1):1777. Epub 2018/05/05. doi: 10.1038/s41467-018-04179-8. PubMed PMID: 29725014; PubMed Central PMCID: PMCPMC5934392.
2. Zhang C, Wang XY, Zuo JL, Wang XF, Feng XW, Zhang B, et al. Localization and density of tertiary lymphoid structures associate with molecular subtype and clinical outcome in colorectal cancer liver metastases. J Immunother Cancer. 2023;11(2). Epub 2023/02/10. doi: 10.1136/jitc-2022-006425. PubMed PMID: 36759015.
3. Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the era of cancer immunotherapy. Nat Rev Cancer. 2019;19(6):307-25. Epub 2019/05/17. doi: 10.1038/s41568-019-0144-6. PubMed PMID: 31092904.
4. Calderaro J, Petitprez F, Becht E, Laurent A, Hirsch TZ, Rousseau B, et al. Intra-tumoral tertiary lymphoid structures are associated with a low risk of early recurrence of hepatocellular carcinoma. Journal of hepatology. 2019;70(1):58-65. Epub 2018/09/15. doi: 10.1016/j.jhep.2018.09.003. PubMed PMID: 30213589.
5. 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. Epub 2016/10/22. doi: 10.1186/s13059-016-1070-5. PubMed PMID: 27765066; PubMed Central PMCID: PMCPMC5073889.
6. 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. Nat Commun. 2013;4:2612. Epub 2013/10/12. doi: 10.1038/ncomms3612. PubMed PMID: 24113773; PubMed Central PMCID: PMCPMC3826632.
7. Shapley LS. A value for n-person games. Class. game theory 69 (1997).
8. Lundberg, S. M. & Lee, S.-I. in Advances in Neural Information Processing Systems (eds Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., & Garnett, R.) 4765–4774 (Curran Associates, Inc., 2017).
9. Covert, I., Lundberg, S. & Lee, S.-I. Explaining by removing: a unified framework for model explanation. J. Mach. Learn. Res. 22, 1–90 (2021).
10. Štrumbelj, E. & Kononenko, I. Explaining prediction models and individual predictions with feature contributions. Knowl. Inf. Syst. 41, 647–665 (2014).



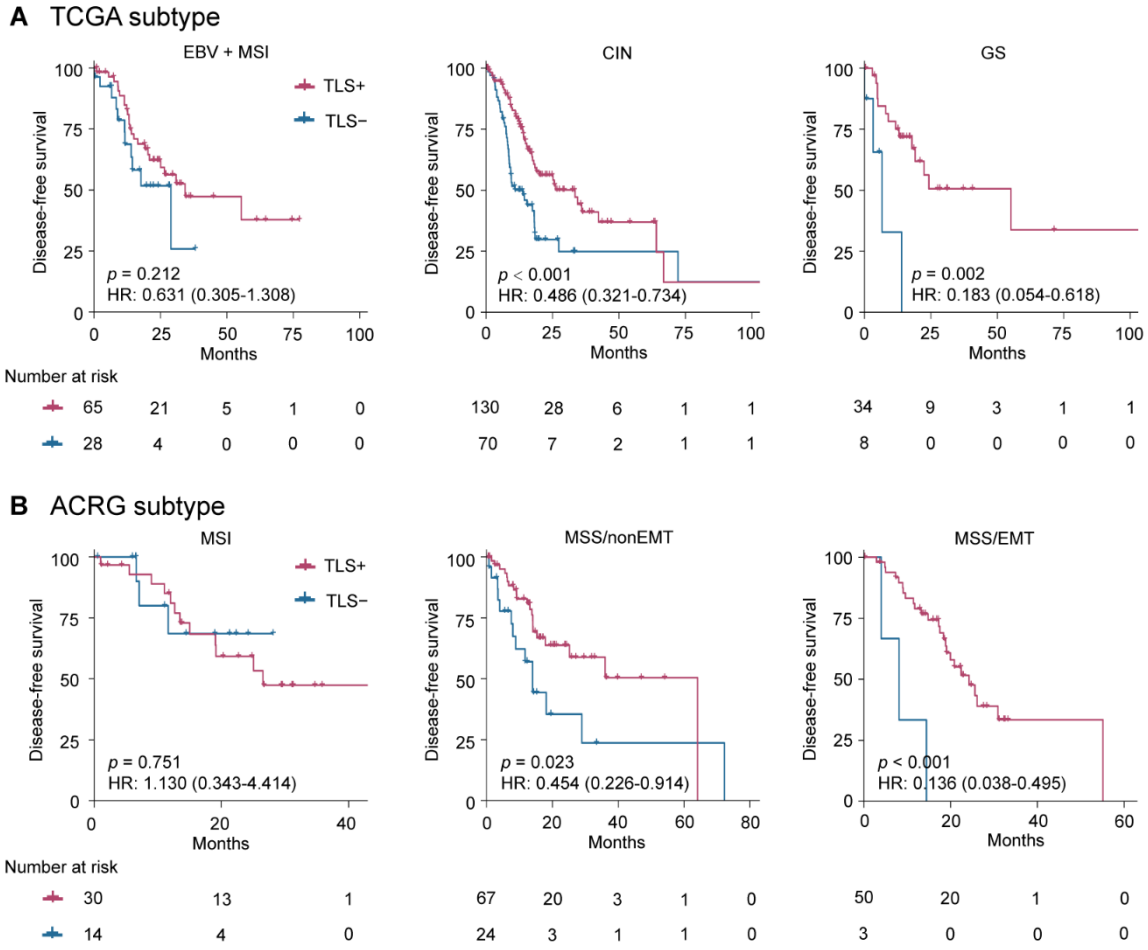
Supplementary Figure 1. Workflow of enrolled patients. SMU cohort: patients from SMU hospital; TCGA-STAD cohort: patients from the Cancer Genome Atlas database; ACRG cohort: patients from the GSE62254 profile; YUSH cohort: patients from the GSE13861 profile; SMC cohort: patients from the GSE26253 profile; KRIBB cohort: patients from the GSE84437 profile; PRJEB25780 cohort: patients from the EBI database.



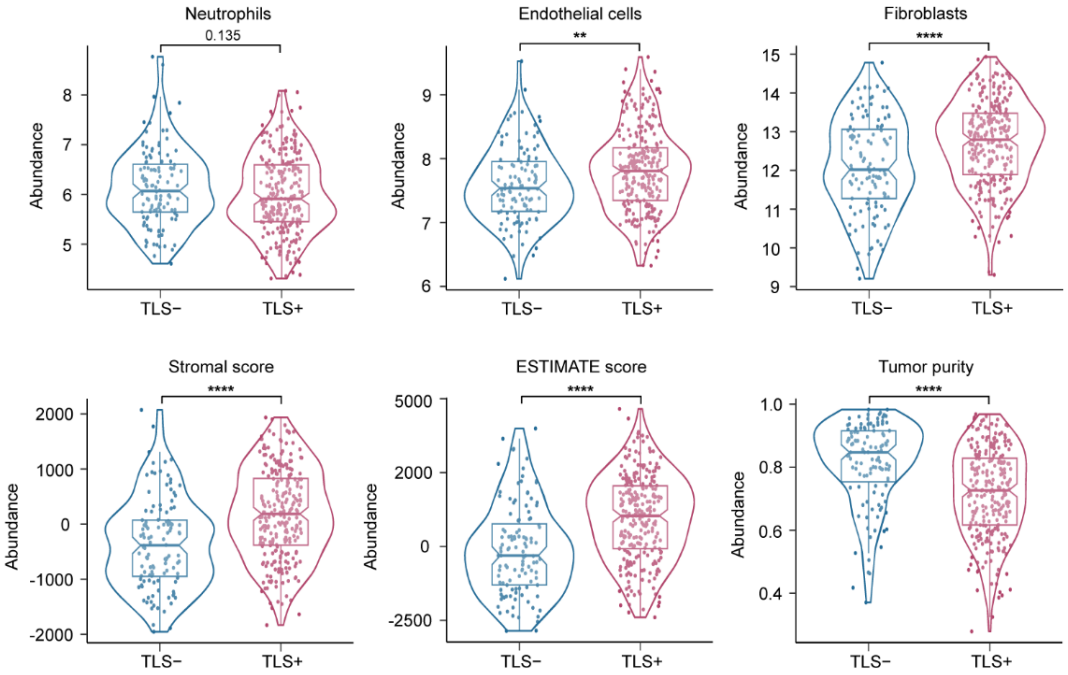
Supplementary Figure 2. Kaplan-Meier analyses of disease-free survival (DFS) according to H&E-based TLSs status stratified by clinicopathological risk factors in patients with gastric cancer from the TCGA-STAD cohort. TLSs, tertiary lymphoid structures. *P*-values were calculated by log-rank test.



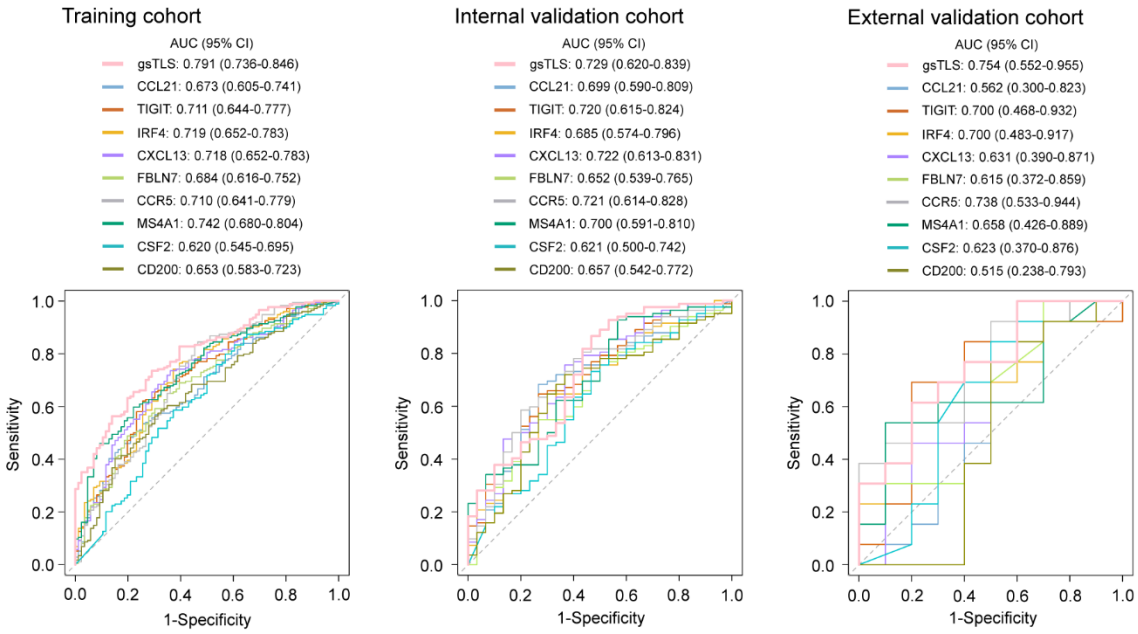
Supplementary Figure 3. Kaplan-Meier analyses of overall survival (OS) according to H&E-based TLSs status stratified by clinicopathological risk factors in patients with gastric cancer from the TCGA-STAD cohort. TLSs, tertiary lymphoid structures. *P*-values were calculated by log-rank test.



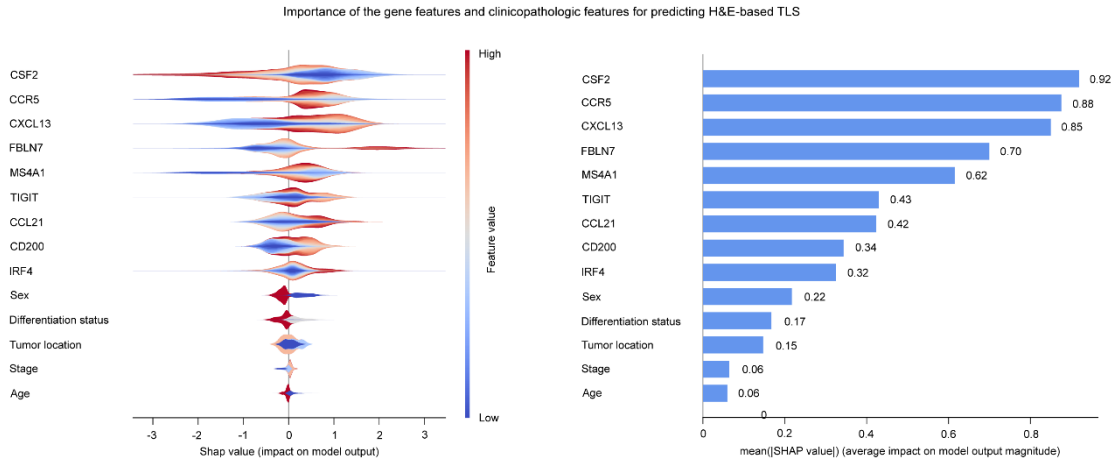
Supplementary Figure 4. Kaplan-Meier analyses of disease-free survival (DFS) according to H&E-based TLSs status stratified by molecular subtype in patients with gastric cancer from the TCGA-STAD cohort. (A) TCGA subtype. (B) ACRG subtype. TLSs, tertiary lymphoid structures. *P*-values were calculated by log-rank test.



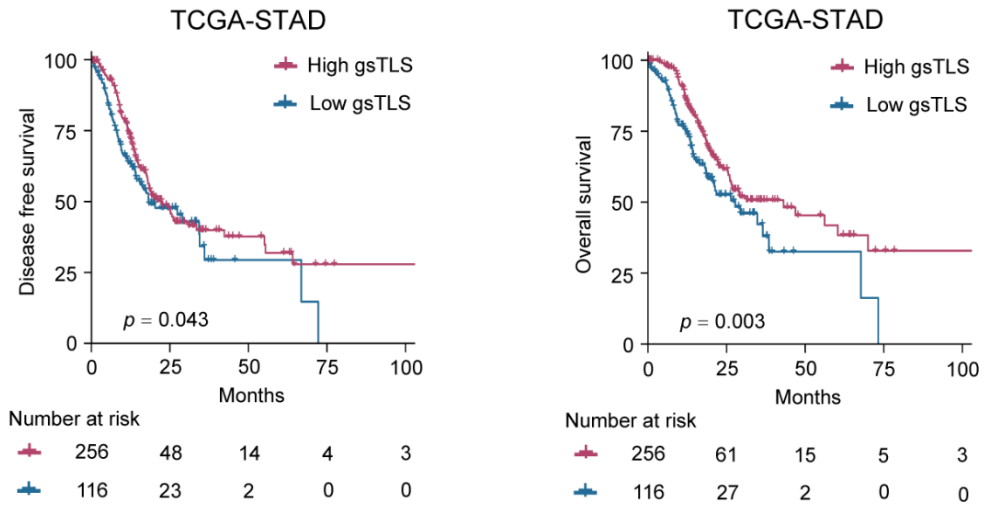
Supplementary Figure 5. Association of H&E-based TLSs status with tumor infiltrating immune-stromal cells and TME score. TLSs, tertiary lymphoid structures. *P*-values were calculated by unpaired t test.



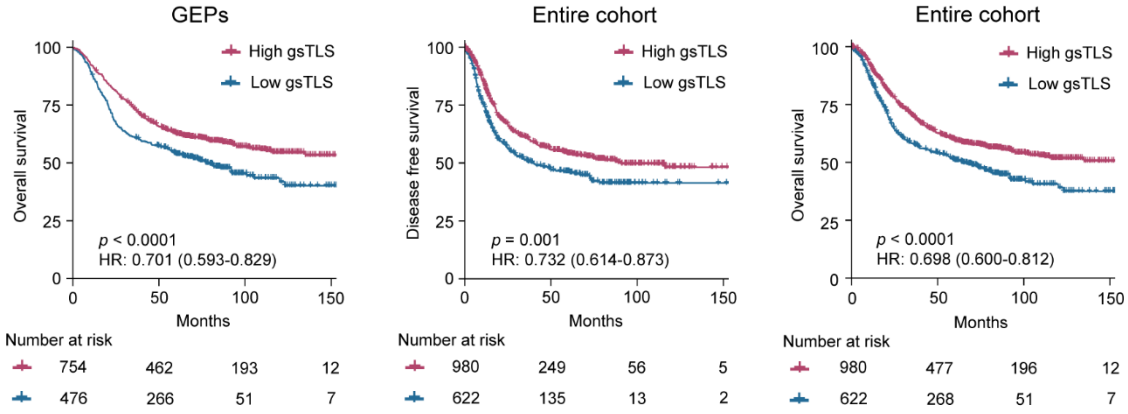
Supplementary Figure 6. ROC curves of the gsTLS and the single gene feature (9 genes) for predicting H&E-based TLSs in the training cohort, internal validation cohort, and external validation cohort. gsTLS, gene signature of tertiary lymphoid structures.



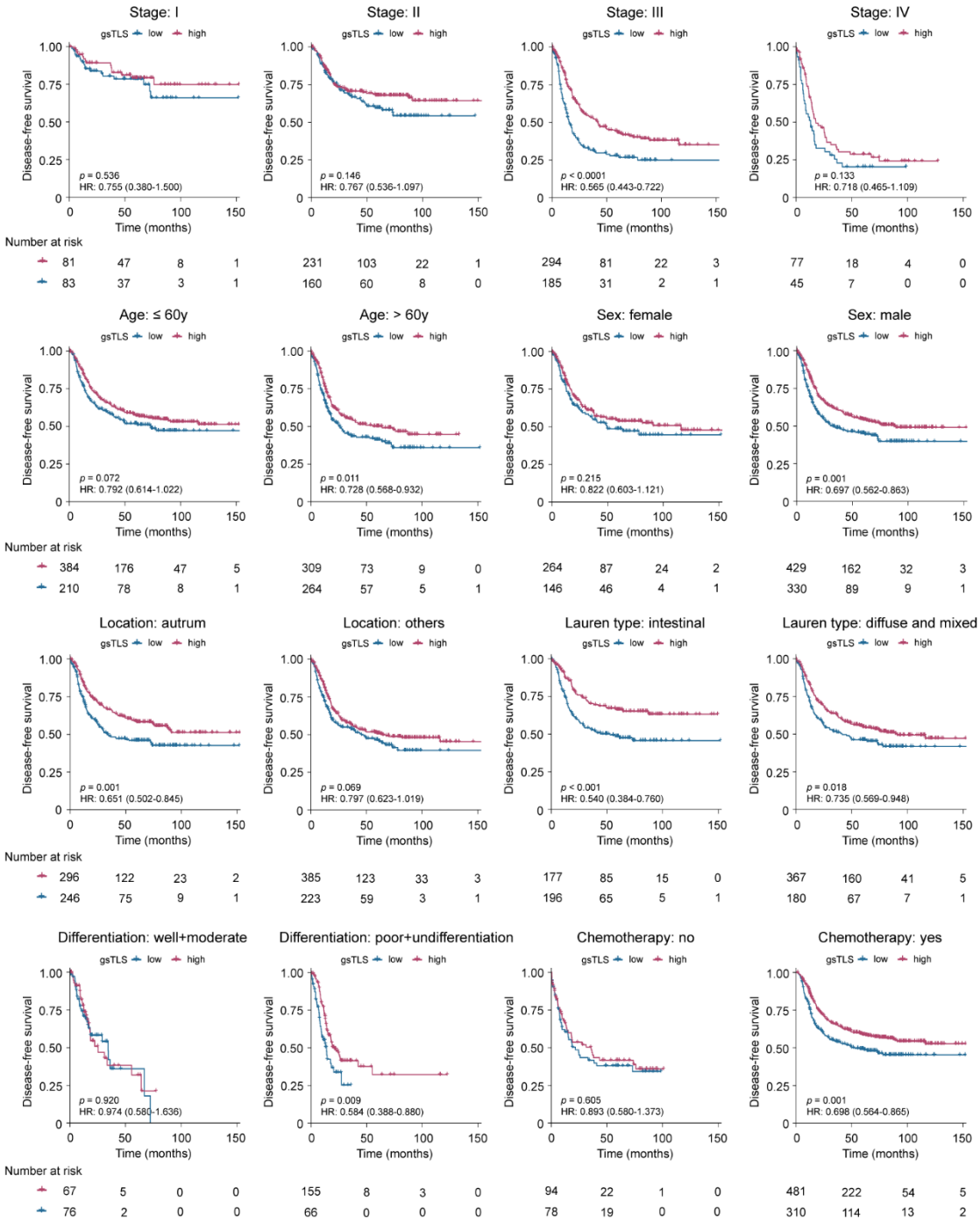
Supplementary Figure 7. SHAP interpretations on the TLS-associated gene features and clinicopathologic factors to predict the H&E-based TLSs. On the X-axis, the contribution of each feature is shown. The Shapley value is positively correlated with the importance. Moreover, a feature with a positive Shapley value will favorably impact the prediction (increase the possibility of TLSs presence). The influence of the value of the feature itself is shown on the Y-axis, for example, for CXCL13, a high value (in red) is associated with a positive Shapley value that will increase the possible of TLSs presence, while a low value (in blue) will decrease the Shapley value and the possible of TLSs presence.



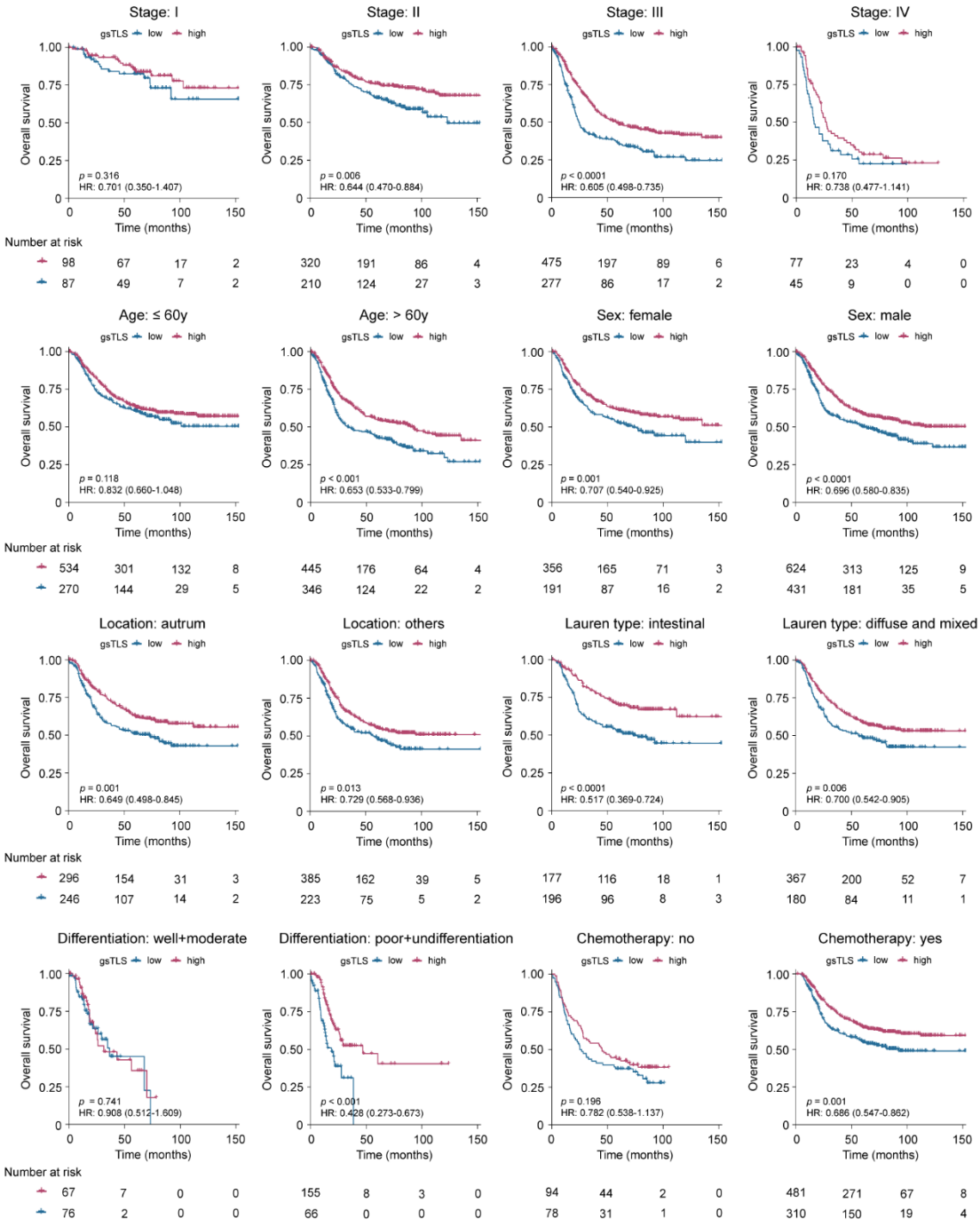
Supplementary Figure 8. Kaplan-Meier plots of DFS and OS according to the gsTLS in patients with gastric cancer in the TCGA-STAD cohort. gsTLS, gene signature of tertiary lymphoid structures. *P*-values were calculated by Breslow test.



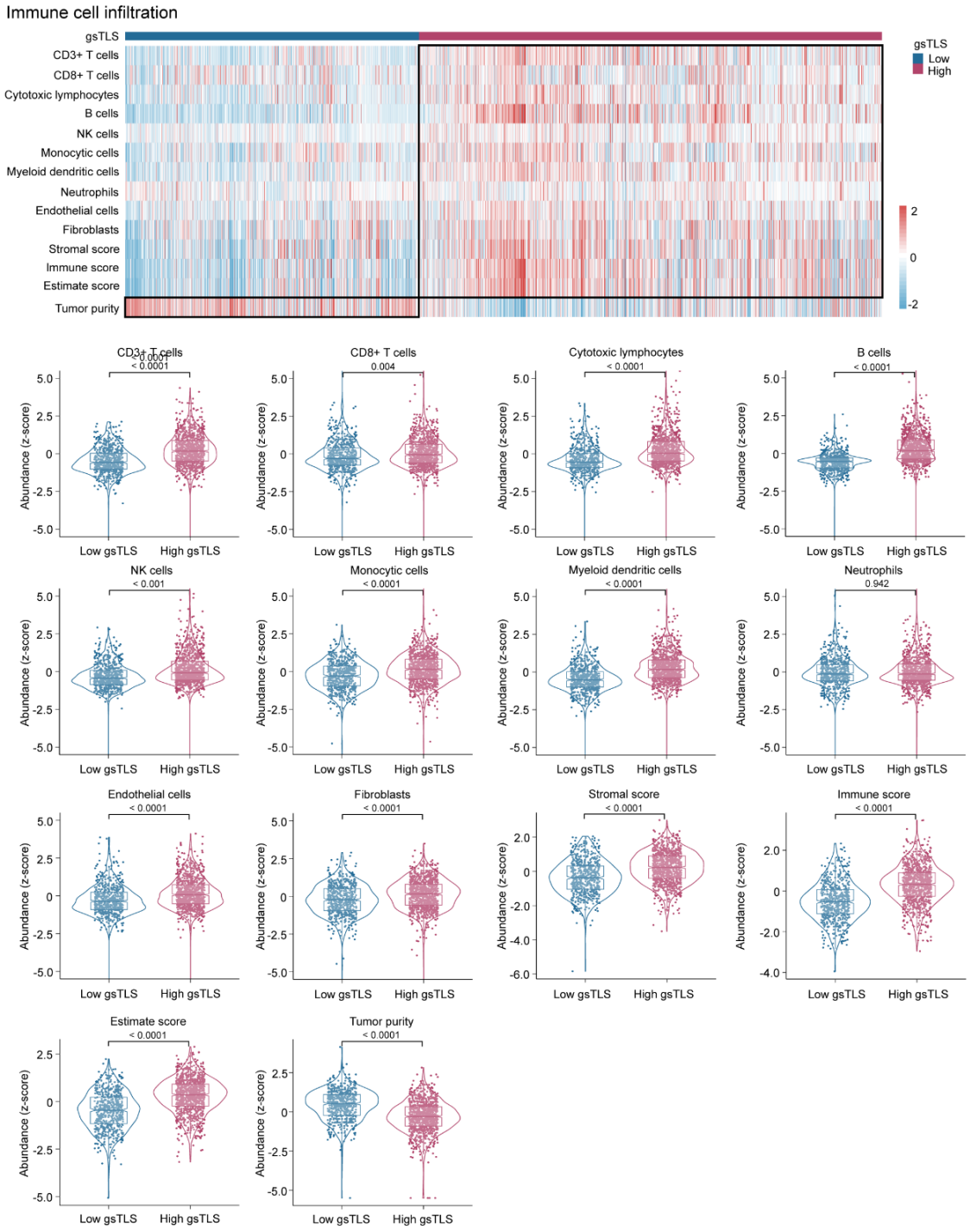
Supplementary Figure 9. Kaplan-Meier plots of DFS and OS according to the gsTLS in patients with gastric cancer in the GEPs cohort and entire cohort. gsTLS, gene signature of tertiary lymphoid structures. *P*-values were calculated by log-rank test.



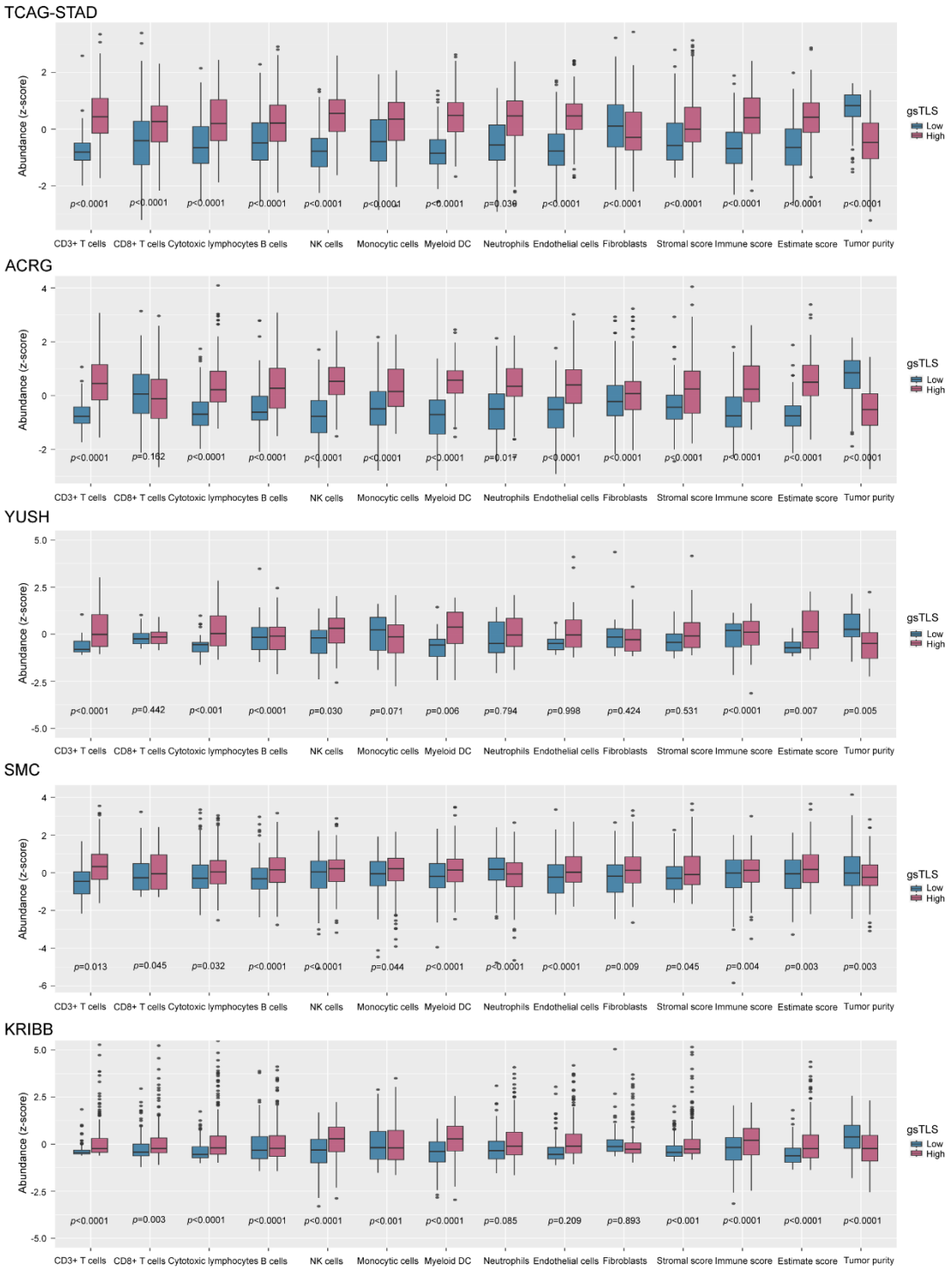
Supplementary Figure 10. Kaplan-Meier analyses of disease-free survival (DFS) according to gsTLS status stratified by clinicopathological risk factors in patients with gastric cancer from the TCGA-STAD, ACRG, YUSH, SMC, and KRIBB cohorts. gsTLS, gene signature of tertiary lymphoid structures. *P*-values were calculated by log-rank test.



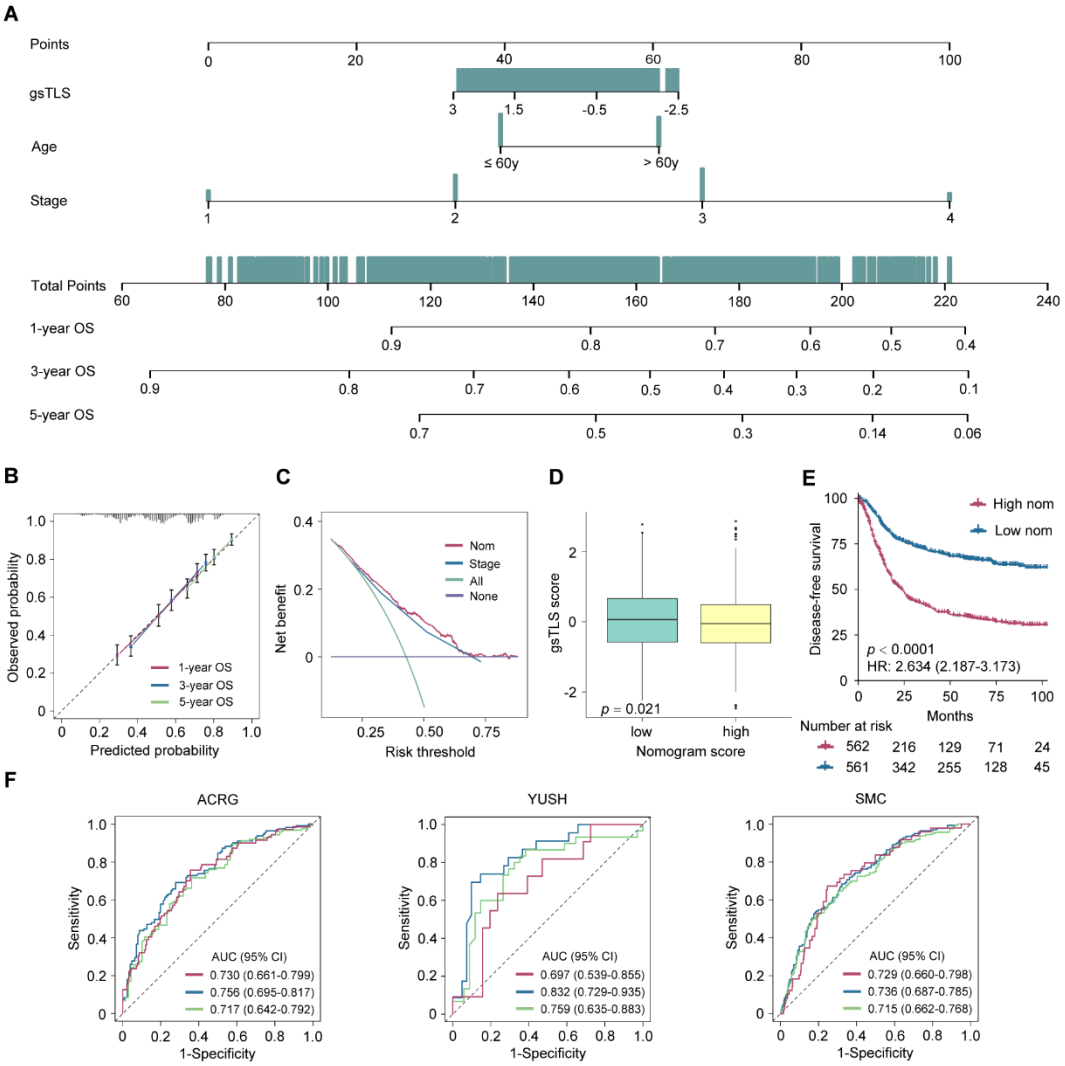
Supplementary Figure 11. Kaplan-Meier analyses of overall survival (OS) according to gsTLS status stratified by clinicopathological risk factors in patients with gastric cancer from the TCGA-STAD, ACRG, YUSH, SMC, and KRIBB cohorts. gsTLS, gene signature of tertiary lymphoid structures. *P*-values were calculated by log-rank test.



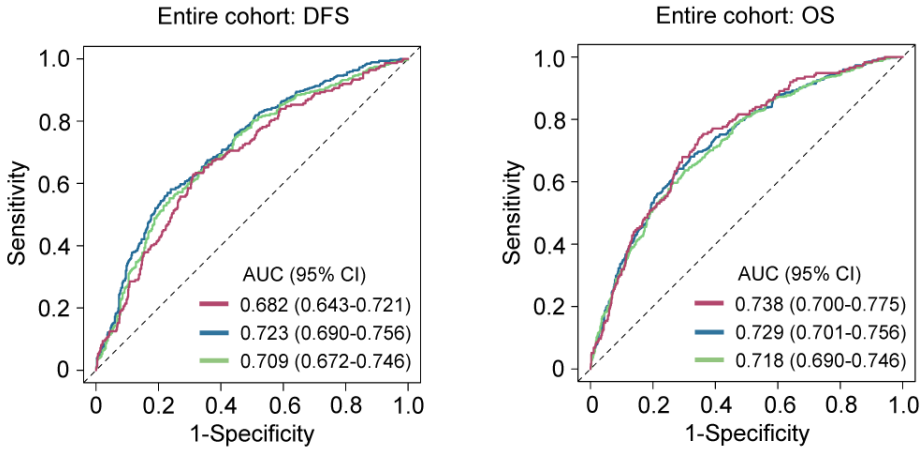
Supplementary Figure 12. Association of gsTLS status with tumor infiltrating immune-stromal cells and TME score in the entire cohort. gsTLS, gene signature of tertiary lymphoid structures. *P*-values were calculated by unpaired t test.



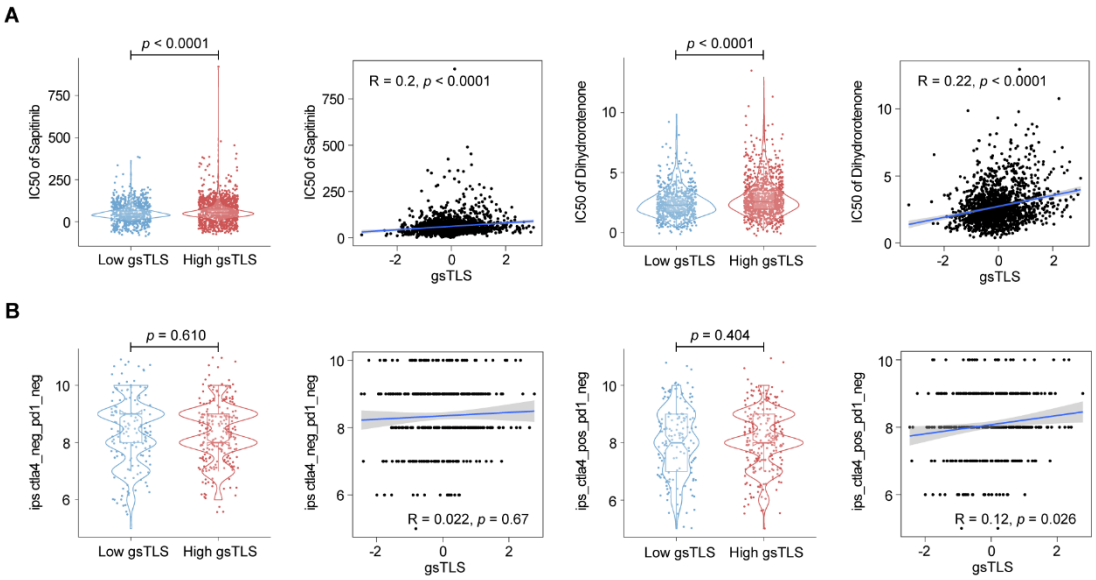
Supplementary Figure 13. Association of gsTLS status with tumor infiltrating immune-stromal cells and TME score in five independent cohorts of the TCGA-STAD, ACRG, YUSH, SMC, and KRIBB. gsTLS, gene signature of tertiary lymphoid structures. *P*-values were calculated by unpaired t test.



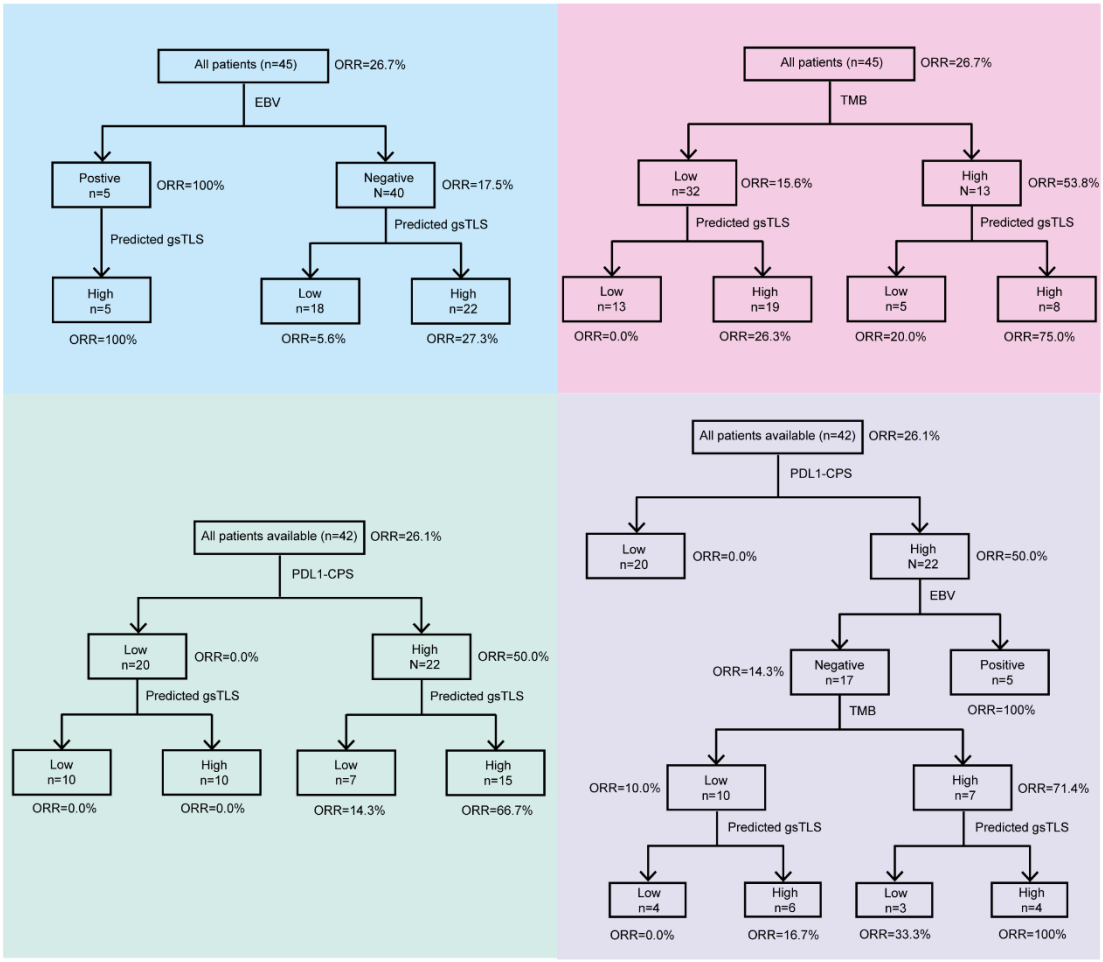
Supplementary Figure 14. Integrated nomograms and its performance to predict 1-, 3-, 5-year disease free survival (DFS) for patients with gastric cancer from the TCGA-STAD, ACRG, YUSH, SMC, and KRIBB cohorts. (A) To determine how many points toward the probability of DFS the patient receives for his or her gsTLS, locate the patient's gsTLSs on their axis, draw a line straight upward to the point axis, repeat this process for each variable, sum the points achieved for each of the risk factors, locate the final sum on the Total Point axis, and draw a line straight down to find the patient's probability of DFS. (B) Calibration of the nomogram in terms of agreement between predicted and observed 1-year, 3-year, and 5-year survival probability. (C) Decision curve analysis of DFS. (D) The gsTLS of different nomogram status. (E) Kaplan-Meier plots of DFS according to the output score of the nomogram. (F) Time-dependent ROC curves of the nomogram in the ACRG, YUSH, and SMC cohorts.



Supplementary Figure 15. 1-, 3-, 5-year time-dependent ROC curves of the nomogram in the entire cohort. DFS, disease-free survival; OS, overall survival.



Supplementary Figure 16. Therapeutic Response predicted by The Cancer Immunome Atlas (TCIA) database and oncoPredict database. (A) two representative drugs from oncoPredict database which are sensitive in low gsTLS group with a low IC50. (B) the predictive value of gsTLS for immunotherapy response showed by the TCIA database. gsTLS, gene signature of tertiary lymphoid structures.



Supplementary Figure 17. The predictive value of gsTLS for immunotherapy response was independent from EBV, TMB, and PDL1-CPS. gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 1. Clinicopathological characteristics of patients with gastric cancer.

Variables	TCGA-STAD	SMU	ACRG	YUSH	SMC	KRIBB	PRJEB25780
No. of patients	406	23	300	65	432	433	45
Median age (interquartile range)	67 (58-73)	58 (50-66)	64 (55-70)	63 (54-69)	53 (43-60)	62 (53-68)	–
Sex (%)							
Female	149 (36.7%)	12 (52.2)	101 (33.7)	19 (29.2)	152 (35.2)	137 (31.6)	13 (28.9)
Male	251 (61.8%)	11 (47.8)	199 (66.3)	46 (70.8)	280 (64.8)	296 (68.4)	32 (71.1)
Unknown	6 (1.5%)	0	0	0	0	0	0
Location (%)							
Cardia	149 (36.7)	4 (17.4)	30 (10.0)	5 (7.7)	54 (12.5)	–	–
Body	87 (21.4)	2 (8.7)	107 (35.7)	31 (47.7)	139 (32.2)	–	–
Antrum	146 (36.0)	15 (65.2)	150 (50.0)	26 (40.0)	226 (52.3)	–	–
Whole	–	2 (8.7)	12 (4.0)	1 (1.5)	13 (3.0)	–	–
Unknown	24 (5.9)	0	1 (0.3)	0	0	–	–
Lauren type (%)							
Intestinal	65 (16.0)	7 (30.4)	150 (50.0)	19 (29.2)	139 (32.2)	–	–
Diffuse or mixed	74 (18.2)	16 (69.6)	150 (50.0)	42 (64.6)	293 (67.8)	–	–
Unknown	267 (65.8)	0	0	4 (6.2)	0	–	–
Differentiation status (%)							
Well or moderate	154 (37.9)	6 (26.1)	–	–	–	–	13 (28.9)
Poor or undifferentiated	238 (58.6)	17 (73.9)	–	–	–	–	22 (48.9)
Unknown	14 (3.4)	0	–	–	–	–	10 (22.2)
Depth of invasion (%)							
T1	20 (4.9)	4 (17.4)	1 (0.3)	–	–	11 (2.5)	–
T2	85 (20.9)	4 (17.4)	187 (62.3)	–	–	39 (9.0)	–
T3	176 (43.3)	7 (30.4)	91 (30.3)	–	–	92 (21.2)	–
T4	111 (27.3)	8 (34.8)	21 (7.0)	–	–	291 (67.2)	–
Unknown	14 (3.4)	0	0	–	–	0	–
Lymph node metastasis (%)							
N0	124 (30.5)	3 (13.0)	38 (12.7)	–	–	80 (18.5)	–
N1	101 (24.9)	7 (30.4)	131 (43.7)	–	–	189 (43.6)	–
N2	82 (20.2)	4 (17.4)	80 (26.7)	–	–	131 (30.3)	–
N3	78 (19.2)	9 (39.1)	51 (17.0)	–	–	33 (7.6)	–
Unknown	21 (5.2)	0	0	–	–	0	–
Distant metastasis (%)							
M0	353 (86.9)	21 (91.3)	273 (91.0)	61 (93.8)	–	433 (100)	0
M1	27 (6.7)	2 (8.7)	27 (9.0)	4 (6.2)	–	0	45 (100)
Unknown	26 (6.4)	0	0	0	–	0	0
TNM stage (%)							
I	55 (13.5)	4 (17.4)	31 (10.3)	12 (18.5)	68 (15.7)	21 (4.8)	0
II	123 (30.3)	7 (30.4)	96 (32.0)	12 (18.5)	167 (38.7)	139 (32.1)	0
III	181 (44.6)	10 (43.5)	146 (48.7)	37 (56.9)	130 (30.1)	273 (63.0)	0
IV	27 (6.7)	2 (8.7)	27 (9.0)	4 (6.2)	67 (15.5)	0	45 (100)
Unknown	20 (4.9)	0	0	0	0	0	0

Supplementary Table 2. Clinicopathological characteristics of patients according to the TLSs in the TCGA-STAD cohorts.

Variables	TCGA-STAD		
	TLSs		
	Absence	Presence	*P*
No. of patients		406	
Median age (interquartile range)	69 (60-74)	65 (57-72)	0.021
Sex (%)			0.008
Female	35 (27.3)	114 (41.0)	
Male	92 (71.9)	159 (57.2)	
Unknown	1 (0.8)	5 (1.8)	
Location (%)			0.555
Cardia	52 (40.6)	97 (34.9)	
Body	26 (20.3)	61 (21.9)	
Antrum	43 (33.6)	103 (37.1)	
Whole	0	0	
Unknown	7 (5.5)	17 (6.1)	
Lauren type (%)			0.001
Intestinal	35 (27.3)	39 (14.0)	
Diffuse or mixed	13 (10.2)	52 (18.7)	
Unknown	80 (62.5)	187 (67.3)	
Differentiation status (%)			0.269
Well or moderate	56 (43.7)	98 (35.3)	
Poor or undifferentiated	69 (53.9)	169 (60.8)	
Unknown	3 (2.3)	11 (4.0)	
Depth of invasion (%)			0.160
T1	11 (8.6)	9 (3.2)	
T2	25 (19.5)	60 (21.6)	
T3	55 (43.0)	121 (43.5)	
T4	35 (27.3)	76 (27.3)	
Unknown	2 (1.6)	12 (4.3)	
Lymph node metastasis (%)			0.590
N0	42 (32.8)	82 (29.5)	
N1	27 (21.1)	74 (26.6)	
N2	29 (22.7)	53 (9.1)	
N3	25 (19.5)	53 (19.1)	
Unknown	5 (3.9)	16 (5.8)	
Distant metastasis (%)			0.392
M0	111 (86.7)	242 (87.1)	
M1	11 (8.6)	16 (5.8)	
Unknown	6 (4.7)	20 (7.2)	
TNM stage (%)			0.097
I	24 (18.8)	31 (11.2)	
II	32 (25.0)	91 (32.7)	
III	58 (45.3)	123 (44.2)	
IV	11 (8.6)	16 (5.8)	
Unknown	3 (2.3)	17 (6.1)	

TLSs, tertiary lymphoid structures.

Supplementary Table 3. Association of the TLSs, clinicopathological characteristics with disease-free and overall survival.

Variables	Disease-free survival		Overall survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
Univariate analysis				
TLSs (presence vs. absence)	0.540 (0.400-0.729)	<0.0001	0.413 (0.297-0.574)	<0.0001
Age (years) (>60 vs. ≤60)	1.233 (0.898-1.693)	0.195	1.679 (1.159-2.432)	0.006
Sex (male vs. female)	1.273 (0.934-1.735)	0.127	1.119 (0.795-1.574)	0.520
Tumor location	1.054 (0.885-1.255)	0.558	1.052 (0.865-1.278)	0.614
Differentiation	1.157 (0.870-1.540)	0.316	1.046 (0.762-1.436)	0.779
Lauren type	0.892 (0.551-1.445)	0.643	0.702 (0.406-1.215)	0.207
TNM stage (IV vs. III vs. II vs. I)	1.504 (1.236-1.830)	<0.0001	1.586 (1.268-1.984)	<0.0001
Multivariate analysis				
TLSs (presence vs. absence)	0.507 (0.374-0.687)	<0.0001	0.410 (0.292-0.575)	<0.0001
Age (years) (>60 vs. ≤60)	–	–	1.654 (1.127-2.428)	0.010
TNM stage (IV vs. III vs. II vs. I)	1.548 (1.270-1.887)	<0.0001	1.727 (1.368-2.181)	<0.0001

TLSs, tertiary lymphoid structures.

Supplementary Table 4. Clinicopathological characteristics of patients according to the TLSs in the training and validation cohorts.

Variables	Training cohort			Internal validation cohort			External validation cohort		
	TLSs			TLSs			TLSs		
	Absence	Presence	*P*	Absence	Presence	*P*	Absence	Presence	*P*
No. of patients		260			112			23	
Median age (interquartile range)	70 (63-76)	66 (58-72)	0.003	65 (53-71)	63 (54-72)	0.910	58 (48-67)	58 (53-66)	0.885
Sex (%)			0.004			0.672			0.414
Female	19 (22.1)	70 (40.2)		12 (40.0)	37 (45.1)		4 (40.0)	8 (61.5)	
Male	67 (77.9)	104 (59.8)		18 (60.0)	45 (54.9)		6 (60.0)	5 (38.5)	
Location (%)			0.246			0.135			0.636
Cardia	37 (43.0)	61 (35.1)		12 (40.0)	24 (29.3)		2 (20.0)	2 (15.4)	
Body	15 (17.4)	45 (25.9)		8 (26.7)	14 (17.1)		0	2 (15.4)	
Antrum	31 (36.0)	58 (33.3)		9 (30.0)	42 (51.2)		7 (70.0)	8 (61.5)	
Whole	0	0		0	0		1 (10.0)	1 (7.7)	
Unknown	3 (3.5)	10 (5.7)		1 (3.3)	2 (2.4)		0	0	
Lauren type (%)			0.002			0.159			0.968
Intestinal	25 (29.1)	20 (11.5)		7 (23.3)	13 (15.9)		3 (30.0)	4 (30.8)	
Diffuse or mixed	9 (10.5)	32 (18.4)		3 (10.0)	18 (22.0)		7 (70.0)	9 (69.2)	
Unknown	52 (60.5)	122 (70.1)		20 (66.7)	51 (62.2)		0	0	
Differentiation status (%)			0.997			0.056			0.660
Well or moderate	36 (41.8)	70 (40.2)		14 (46.7)	23 (28.0)		2 (20.0)	4 (30.8)	
Poor or undifferentiated	50 (58.1)	99 (56.9)		14 (46.7)	58 (70.7)		8 (80.0)	9 (69.2)	
Unknown	0	5 (2.9)		2 (6.7)	1 (1.2)		0	0	
Depth of invasion (%)			0.081			0.387			0.123
T1	8 (9.3)	5 (2.9)		2 (6.7)	4 (4.9)		3 (30.0)	1 (7.7)	
T2	22 (25.6)	46 (26.4)		3 (10.0)	9 (11.0)		3 (30.0)	1 (7.7)	
T3	29 (33.7)	75 (43.1)		18 (60.0)	37 (45.1)		1 (10.0)	6 (46.2)	
T4	27 (31.4)	44 (25.3)		6 (20.0)	30 (36.6)		3 (30.0)	5 (38.5)	
Unknown	0	4 (2.3)		1 (3.3)	2 (2.4)		0	0	
Lymph node metastasis (%)			0.919			0.596			0.860
N0	29 (33.7)	54 (31.0)		8 (26.7)	24 (29.3)		1 (10.0)	2 (15.4)	
N1	20 (23.3)	42 (24.1)		6 (20.0)	26 (31.7)		4 (40.0)	3 (23.1)	
N2	20 (23.3)	35 (20.1)		6 (20.0)	15 (18.3)		2 (20.0)	2 (15.4)	
N3	15 (17.4)	34 (19.5)		8 (26.7)	15 (18.3)		3 (30.0)	6 (46.2)	
Unknown	2 (2.3)	9 (5.2)		2 (6.7)	2 (2.4)		0	0	
Distant metastasis (%)			0.991			0.256			0.846
M0	76 (88.4)	153 (87.9)		26 (86.7)	74 (90.2)		9 (90.0)	12 (92.3)	
M1	5 (5.8)	10 (5.7)		4 (13.3)	5 (6.1)		1 (10.0)	1 (7.7)	
Unknown	5 (5.8)	11 (6.3)		0	3 (3.7)		0	0	
TNM stage (%)			0.033			0.580			0.495
I	22 (25.6)	21 (12.1)		2 (6.7)	8 (9.8)		3 (30.0)	1 (7.7)	
II	20 (23.3)	61 (35.1)		8 (26.7)	27 (32.9)		3 (30.0)	4 (30.8)	
III	39 (45.3)	73 (42.0)		14 (46.7)	40 (48.8)		3 (30.0)	7 (53.8)	
IV	5 (5.8)	10 (5.7)		4 (13.3)	5 (6.1)		1 (10.0)	1 (7.7)	
Unknown	0	9 (5.2)		2 (6.7)	2 (2.4)		0	0	

TLSs, tertiary lymphoid structures.

Supplementary Table 5. The optimal cut-off value for gsTLS was determined using Youden's index in the training cohort.

	Cutoff	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	Accuracy	PPV	NPV
gsTLS	−0.2794	0.791 (0.736-0.846)	73.6 (66.4-79.9)	70.9 (60.1-80.2)	72.7	83.7 (76.8-89.1)	57.0 (47.1-66.5)

gsTLS, gene signature of tertiary lymphoid structures; AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

Supplementary Table 6. Clinical characteristics of patients according to the gsTLS in the training and two validation cohorts.

Variables	Training cohort, n = 260			Internal validation cohort, n = 112			External validation cohort, n = 23		
	low gsTLS	high gsTLS	*P*	low gsTLS	high gsTLS	*P*	low gsTLS	high gsTLS	*P*
Median age (interquartile range)	69 (61-74)	66 (58-73)	0.041	66 (57-71)	62 (54-72)	0.638	66 (58-74)	53 (44-59)	0.005
Sex (%)			0.047			0.238			0.214
Female	29 (27.1)	60 (39.2)		14 (35.9)	35 (47.9)		3 (33.3)	9 (64.3)	
Male	78 (72.9)	93 (60.8)		25 (64.1)	38 (52.1)		6 (66.7)	5 (35.7)	
Tumor location (%)			0.477			0.824			0.662
Cardia	41 (38.3)	57 (37.3)		13 (33.3)	23 (31.5)		2 (22.2)	2 (14.3)	
Body	21 (19.6)	39 (25.5)		9 (23.1)	13 (17.8)		0	2 (14.3)	
Antrum	40 (37.4)	49 (32.0)		17 (43.6)	34 (46.6)		6 (66.7)	9 (64.3)	
Whole	0	0		0	0		1 (11.1)	1 (7.1)	
Unknown	5 (4.7)	8 (5.2)		0	3 (4.1)		0	0	
Differentiation status (%)			0.001			0.107			0.643
Well or moderate	8 (8.0)	19 (13.4)		17 (43.6)	20 (27.4)		3 (33.3)	3 (21.4)	
Poor or undifferentiated	23 (23.0)	33 (23.2)		19 (48.7)	53 (72.6)		6 (66.7)	11 (78.6)	
Unknown	69 (69.0)	90 (63.4)		3 (7.7)	0		0	0	
Lauren type (%)			0.002			0.181			0.363
Intestinal	59 (55.1)	47 (30.8)		8 (20.5)	12 (16.4)		4 (44.4)	3 (21.4)	
Diffuse or mixed	47 (43.9)	102 (66.7)		4 (10.3)	17 (23.3)		5 (55.6)	11 (78.6)	
Unknown	1 (0.9)	4 (2.6)		27 (69.2)	44 (60.3)		0	0	
Depth of invasion (%)			0.175			0.394			0.361
T1	7 (6.5)	6 (3.9)		4 (10.3)	2 (2.7)		0	4 (28.6)	
T2	35 (32.7)	33 (21.6)		4 (10.3)	8 (11.0)		2 (22.2)	2 (14.3)	
T3	39 (36.4)	65 (42.5)		19 (48.7)	36 (49.3)		3 (33.3)	4 (28.6)	
T4	26 (24.3)	45 (29.4)		11 (28.2)	25 (34.2)		4 (44.4)	4 (28.6)	
Unknown	0	4 (2.6)		1 (2.6)	2 (2.7)		0	0	
Lymph node metastasis (%)			0.043			0.099			0.188
N0	33 (30.8)	50 (32.7)		15 (38.5)	17 (23.3)		0	3 (21.4)	
N1	29 (27.1)	33 (21.6)		6 (15.4)	26 (35.6)		3 (33.3)	4 (28.6)	
N2	29 (27.1)	26 (17.0)		6 (15.4)	15 (20.5)		1 (11.1)	3 (21.4)	
N3	13 (12.1)	36 (23.5)		9 (23.1)	14 (19.2)		5 (55.5)	4 (28.6)	
Unknown	3 (2.8)	8 (5.2)		3 (7.7)	1 (1.4)		0	0	
Distant metastasis (%)			0.289			0.063			0.742-
M0	96 (89.7)	133 (86.9)		32 (82.1)	68 (93.2)		8 (88.9)	13 (92.9)	
M1	4 (3.7)	11 (7.2)		6 (15.4)	3 (4.1)		1 (11.1)	1 (7.1)	
Unknown	7 (6.5)	9 (5.9)		1 (2.6)	2 (2.7)		0	0	
Stage (%)			0.033			0.034			0.361
I	26 (24.3)	17 (11.1)		6 (15.4)	4 (5.5)		0	4 (28.6)	
II	29 (27.1)	52 (34.0)		10 (25.6)	25 (34.2)		3 (33.3)	4 (28.6)	
III	47 (43.9)	65 (42.5)		15 (38.5)	39 (53.4)		5 (55.6)	5 (35.7)	
IV	4 (3.7)	11 (7.2)		6 (15.4)	3 (4.1)		1 (11.1)	1 (7.1%)	
Unknown	1 (0.9)	8 (5.2)		2 (5.1)	2 (2.7)		0	0	

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 7. Clinical characteristics of patients according to the gsTLS in the ACRG and YUSH cohorts.

Variables	ACRG, n = 300			YUSH, n = 65		
	low gsTLS	high gsTLS	*P*	low gsTLS	high gsTLS	*P*
Median age (interquartile range)	66 (60-71)	61 (52-68)	<0.001	64 (59-71)	61 (51-69)	0.064
Sex (%)			0.001			0.417
Female	29 (22.8)	72 (41.6)		6 (23.1)	13 (33.3)	
Male	98 (77.2)	101 (58.4)		20 (76.9)	26 (66.7)	
Tumor location (%)			0.058			0.392
Cardia	14 (11.0)	16 (9.2)		1 (3.8)	4 (10.3)	
Body	38 (29.9)	69 (39.9)		14 (53.8)	17 (43.6)	
Antrum	72 (56.7)	78 (45.1)		9 (34.6)	17 (43.6)	
Whole	2 (1.6)	10 (5.8)		1 (3.8)	0	
Unknown	5 (0.8)	0		1 (3.8)	1 (2.6)	
Lauren type (%)			<0.001			0.572
Intestinal	74 (41.8)	65 (25.5)		6 (23.1)	13 (33.3)	
Diffuse or mixed	103 (58.2)	190 (74.5)		18 (69.2)	24 (61.5)	
Unknown	0	0		2 (7.7)	2 (5.1)	
Depth of invasion (%)			0.040			–
T1	1 (0.8)	0		–	–	
T2	88 (69.3)	99 (57.2)		–	–	
T3	28 (22.0)	63 (36.4)		–	–	
T4	10 (7.9)	11 (6.4)		–	–	
Unknown	0	0		–	–	
Lymph node metastasis (%)			0.708			–
N0	17 (13.4)	21 (12.1)		–	–	
N1	53 (41.7)	78 (45.1)		–	–	
N2	32 (25.2)	48 (27.7)		–	–	
N3	25 (19.7)	26 (15.0)		–	–	
Unknown	0	0		–	–	
Distant metastasis (%)			0.415			0.644
M0	118 (92.9)	155 (89.6)		25 (96.2)	36 (92.3)	
M1	9 (7.1)	18 (10.4)		1 (3.8)	3 (7.7)	
Unknown	0	0		0	0	
Stage (%)			0.639			0.856
I	15 (11.8)	16 (9.2)		5 (19.2)	7 (17.9)	
II	43 (33.9)	53 (30.6)		4 (15.4)	8 (20.5)	
III	60 (47.2)	86 (49.7)		16 (61.5)	21 (53.8)	
IV	9 (7.1)	18 (10.4)		1 (3.8)	3 (7.7)	
Unknown	0	0		0	0	

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 8. Clinical characteristics of patients according to the gsTLS in the SMC and KRIBB cohorts.

Variables	SMC, n = 432			KRIBB, n = 433		
	low gsTLS	high gsTLS	*P*	low gsTLS	high gsTLS	*P*
Median age (interquartile range)	55 (45-61)	51 (42-60)	0.010	63 (56-71)	61 (50-67)	<0.001
Sex (%)			0.260			0.828
Female	68 (38.4)	84 (32.9)		45 (30.8)	92 (32.1)	
Male	109 (61.6)	171 (67.1)		101 (69.2)	195 (67.9)	
Tumor location (%)			0.015			–
Cardia	21 (11.9)	33 (12.9)		–	–	
Body	45 (25.4)	94 (36.9)		–	–	
Antrum	108 (61.0)	118 (46.3)		–	–	
Whole	3 (1.7)	10 (3.9)		–	–	
Lauren type (%)			0.001			–
Intestinal	82 (64.6)	68 (39.3)		–	–	
Diffuse or mixed	45 (35.4)	105 (60.6)		–	–	
Depth of invasion (%)			–			0.188
T1	–	–		1 (0.7)	10 (3.5)	
T2	–	–		10 (6.8)	29 (10.1)	
T3	–	–		37 (25.3)	55 (19.2)	
T4	–	–		98 (67.1)	193 (67.2)	
Lymph node metastasis (%)			–			0.061
N0	–	–		30 (20.5)	50 (17.4)	
N1	–	–		54 (37.0)	135 (47.0)	
N2	–	–		54 (37.0)	77 (26.8)	
N3	–	–		8 (5.5)	25 (8.7)	
Stage (%)			0.394			0.311
I	31 (17.5)	37 (14.5)		4 (2.7)	17 (5.9)	
II	74 (41.8)	93 (36.5)		50 (34.2)	89 (31.0)	
III	47 (26.6)	83 (32.5)		92 (63.0)	181 (63.1)	
IV	25 (14.1)	42 (16.5)		0	0	

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 9. Clinical characteristics of patients according to the gsTLS in the immunotherapy cohort (PRJEB25780).

Variables	PRJEB25780, n = 45		
	low gsTLS	high gsTLS	*P*
Sex (%)			0.188
Female	3 (16.7)	10 (37.0)	
Male	15 (83.3)	17 (63.0)	
Differentiation status (%)			0.134
Well or moderate	9 (50.0)	4 (14.8)	
Poor or undifferentiated	6 (33.3)	16 (59.3)	
Unknown	3 (16.7)	7 (25.9)	
Stage (%)			–
I	0	0	
II	0	0	
III	0	0	
IV	18 (100)	27 (100)	

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 10. Univariate association of the gsTLS, clinicopathological characteristics with disease-free and overall survival.

Variables	Disease-free survival		Overall survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
ACRG cohort				
gsTLS (high vs. low)	0.730 (0.514-1.038)	0.080	0.704 (0.512-0.968)	0.031
Age (years) (>60 vs. ≤60)	1.171 (0.819-1.673)	0.388	1.263 (0.906-1.761)	0.168
Sex (male vs. female)	0.967 (0.669-1.399)	0.860	0.905 (0.647-1.265)	0.559
Tumor location	0.897 (0.701-1.147)	0.385	0.866 (0.695-1.080)	0.202
Lauren type	1.653 (1.155-2.366)	0.006	1.743 (1.261-2.410)	0.001
TNM stage (IV vs. III vs. II vs. I)	2.729 (2.096-3.555)	<0.0001	2.491 (1.973-3.146)	<0.0001
YUSH cohort				
gsTLS (high vs. low)	0.518 (0.267-1.008)	0.053	0.427 (0.208-0.876)	0.020
Age (years) (>60 vs. ≤60)	1.452 (0.730-2.886)	0.288	2.126 (0.972-4.650)	0.059
Sex (male vs. female)	0.861 (0.421-1.759)	0.681	0.829 (0.387-1.773)	0.628
Tumor location	1.114 (0.652-1.902)	0.694	1.136 (0.630-2.049)	0.671
Lauren type	1.017 (0.493-2.099)	0.964	0.741 (0.347-1.584)	0.440
TNM stage (IV vs. III vs. II vs. I)	1.782 (1.173-2.708)	0.007	2.010 (1.246-3.241)	0.004
SMC cohort				
gsTLS (high vs. low)	0.690 (0.513-0.927)	0.014	0.733 (0.540-0.994)	0.046
Age (years) (>60 vs. ≤60)	0.957 (0.671-1.366)	0.810	1.037 (0.724-1.484)	0.843
Sex (male vs. female)	1.070 (0.785-1.459)	0.669	1.035 (0.753-1.424)	0.830
Tumor location	0.877 (0.724-1.063)	0.182	0.877 (0.720-1.068)	0.190
Lauren type	1.272 (0.913-1.772)	0.155	1.393 (0.987-1.965)	0.059
TNM stage (IV vs. III vs. II vs. I)	1.975 (1.676-2.327)	<0.0001	2.077 (1.750-2.465)	<0.0001
KRIBB cohort				
gsTLS (high vs. low)	–	–	0.716 (0.541-0.947)	0.019
Age (years) (>60 vs. ≤60)	–	–	1.531 (1.163-2.015)	0.002
Sex (male vs. female)	–	–	1.254 (0.926-1.697)	0.143
TNM stage (IV vs. III vs. II vs. I)	–	–	2.354 (1.761-3.147)	<0.0001
TCGA-STAD cohort				
gsTLS (high vs. low)	0.791 (0.579-1.080)	0.140	0.643 (0.455-0.907)	0.012
Age (years) (>60 vs. ≤60)	1.339 (0.957-1.873)	0.089	1.930 (1.296-2.872)	0.001
Sex (male vs. female)	1.402 (1.009-1.947)	0.044	1.288 (0.895-1.853)	0.173
Tumor location	1.076 (0.898-1.289)	0.427	1.042 (0.853-1.275)	0.685
Differentiation	1.247 (0.923-1.685)	0.151	1.130 (0.809-1.578)	0.474
Lauren type	0.804 (0.488-1.327)	0.394	0.621 (0.349-1.105)	0.105
TNM stage (IV vs. III vs. II vs. I)	1.403 (1.148-1.713)	0.001	1.485 (1.184-1.863)	0.001

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 11. Multivariate association of the gsTLS, clinicopathological characteristics with disease-free and overall survival.

Variables	Disease-free survival		Overall survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
ACRG cohort				
gsTLS (high vs. low)	0.562 (0.390-0.810)	0.002	0.543 (0.390-0.756)	<0.001
Lauren type	1.404 (0.964-2.046)	0.077	1.478 (1.050-2.082)	0.025
TNM stage (IV vs. III vs. II vs. I)	2.755 (2.093-3.626)	<0.0001	2.465 (1.935-3.140)	<0.0001
YUSH cohort				
gsTLS (high vs. low)	0.452 (0.230-0.889)	0.021	0.346 (0.165-0.722)	0.002
TNM stage (IV vs. III vs. II vs. I)	1.907 (1.232-2.950)	0.004	2.277 (1.362-3.807)	0.005
SMC cohort				
gsTLS (high vs. low)	0.594 (0.441-0.800)	0.001	0.627 (0.461-0.851)	0.003
TNM stage (IV vs. III vs. II vs. I)	2.029 (1.723-2.389)	<0.0001	2.129 (1.795-2.527)	<0.0001
KRIBB cohort				
gsTLS (high vs. low)	–	–	0.743 (0.560-0.984)	0.038
Age (years) (>60 vs. ≤60)	–	–	1.538 (1.167-2.028)	0.002
TNM stage (IV vs. III vs. II vs. I)	–	–	2.396 (1.789-3.210)	<0.0001
TCGA-STAD cohort				
gsTLS (high vs. low)	0.779 (0.567-1.070)	0.123	0.607 (0.428-0.862)	0.005
Age (years) (>60 vs. ≤60)	–	–	2.066 (1.377-3.099)	<0.001
Sex (male vs. female)	1.345 (0.959-1.884)	0.086	–	–
TNM stage (IV vs. III vs. II vs. I)	1.403 (1.151-1.709)	0.001	1.626 (1.292-2.047)	<0.0001

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 12. Association of the gsTLS, clinicopathological characteristics with disease-free and overall survival in the entire cohort including TCGA-STAD and GEPs.

Variables	Disease-free survival		Overall survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
Univariate analysis				
gsTLS (vs.)				
Low	1	Reference	1	Reference
High	0.732 (0.614-0.873)	0.001	0.698 (0.600-0.812)	<0.0001
Age (years)				
≤60	1	Reference	1	Reference
>60	1.438 (1.206-1.714)	<0.0001	1.588 (1.366-1.845)	<0.0001
Sex				
Female	1	Reference	1	Reference
Male	1.104 (0.917-1.329)	0..295	1.086 (0.926-1.274)	0.312
Differentiation				
Well or moderate	1	Reference	1	Reference
Poor or undifferentiated	1.247 (0.923-1.685)	0.151	1.130 (0.809-1.578)	0.474
Lauren type				
Intestinal	1	Reference	1	Reference
Diffuse or mixed	1.163 (0.948-1.428)	0.148	1.188 (0.969-1.457)	0.097
Tumor location				
Whole	1	Reference	1	Reference
Cardia	1.791 (0.904-3.548)	0.095	1.730 (0.900-3.327)	0.100
Body	1.445 (0.738-2.827)	0.283	1.339 (0.706-2.537)	0.371
Antrum	1.388 (0.713-2.703)	0.335	1.263 (0.670-2.381)	0.470
TNM stage				
I	1	Reference	1	Reference
II	1.652 (1.124-2.428)	0.011	1.667 (1.139-2.439)	0.009
III	3.700 (2.574-5.316)	<0.0001	4.001 (2.792-5.733)	<0.0001
IV	5.373 (3.590-8.043)	<0.0001	6.436 (4.281-9.676)	<0.0001
Multivariate analysis				
gsTLS				
Low	1	Reference	1	Reference
High	0.669 (0.560-0.799)	<0.0001	0.658 (0.565-0.766)	<0.0001
Age (years)				
≤60	1	Reference	1	Reference
>60	1.478 (1.235-1.767)	<0.0001	1.629 (1.399-1.898)	<0.0001
TNM stage (IV vs. III vs. II vs. I)				
I	1	Reference	1	Reference
II	1.705 (1.160-2.506)	0.007	1.653 (1.129-2.419)	0.010
III	3.961 (2.753-5.701)	<0.0001	4.148 (2.893-5.948)	<0.0001
IV	6.145 (4.095-9.221)	<0.0001	7.289 (4.843-10.970)	<0.0001

gsTLS, gene signature of tertiary lymphoid structures.

Supplementary Table 13. Comparing the prediction accuracy of the integrated nomogram with gsTLS and TNM stage and Age in the entire cohort including TCGA-STAD and GEPs.

Variable	Disease-free survival		Overall survival	
	C-index (95% CI)	*P*	C-index (95% CI)	*P*
Nomogram	0.667 (0.642-0.692)	Reference	0.679 (0.659-0.699)	Reference
gsTLS	0.524 (0.497-0.551)	<0.0001	0.533 (0.509-0.557)	<0.0001
TNM stage	0.641 (0.617-0.665)	<0.0001	0.646 (0.627-0.665)	<0.0001
Age	0.552 (0.529-0.575)	<0.0001	0.560 (0.541-0.579)	<0.0001