

# Characteristics of auto-quantified tumor-infiltrating lymphocytes and the prognostic value in adenocarcinoma of the esophagogastric junction, gastric adenocarcinoma, and esophageal squamous cell carcinoma

Hao Dong<sup>1,\*</sup>, Longqing Yao<sup>2,\*</sup>, Jiahui Fan<sup>3</sup>, Peipei Gao<sup>2,4</sup>, Xiaorong Yang<sup>4,5</sup>, Ziyu Yuan<sup>4</sup>, Tiejun Zhang<sup>1,4</sup>, Ming Lu<sup>4,5</sup>, Xingdong Chen<sup>2,4,6</sup>, Chen Suo<sup>1,4,7</sup>

<sup>1</sup>Department of Epidemiology and Ministry of Education Key Laboratory of Public Health Safety, School of Public Health, Fudan University, Shanghai, China

<sup>2</sup>State Key Laboratory of Genetic Engineering, Human Phenome Institute, School of Life Sciences, Fudan University, Shanghai, China

<sup>3</sup>Department of Clinical Laboratory, Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine, Shanghai, China

<sup>4</sup>Fudan University Taizhou Institute of Health Sciences, Taizhou, China

<sup>5</sup>Clinical Epidemiology Unit, Qilu Hospital of Shandong University, Jinan, China

<sup>6</sup>Yiwu Research Institute of Fudan University, Yiwu, Zhejiang, China

<sup>7</sup>Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, China

\*Co-first authors

**Correspondence to:** Chen Suo; email: [suchoen@fudan.edu.cn](mailto:suchoen@fudan.edu.cn)

**Keywords:** adenocarcinoma of the esophagogastric junction, gastric adenocarcinoma, esophageal squamous cell carcinoma, tumor-infiltrating lymphocytes, survival

**Received:** October 4, 2023

**Accepted:** June 10, 2024

**Published:** July 5, 2024

**Copyright:** © 2024 Dong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Background:** Adenocarcinoma of the esophagogastric junction (AEGJ) with a specific pathological profile and poor prognosis has limited therapeutic options. Previous studies have found that TILs exhibit distinct characteristics in different tumors and are correlated with tumor prognosis. We established cellular training sets to obtain auto-quantified TILs in pathological images. And we compared the characteristics of TILs in AEGJ with those in esophageal squamous cell carcinoma (ESCC) and gastric adenocarcinoma (GAC) to gain insight into the unique immune environments of these three tumors and investigate the prognostic value of TILs in these three tumors.

**Methods:** Utilizing a case-control study design, we analyzed 214 AEGJ, 256 GAC, and 752 ESCC cases. The TCGA dataset was used to validate prognostic value of auto-quantified TILs. The specific cellular training sets were established by experienced pathologists to determine TILs counts. Kruskal-Wallis test and multi-variable linear regression were conducted to explore TILs characteristics. Survival analyses were performed with Kaplan-Meier method and Cox proportional hazards model.

**Results:** The three cellular training sets of these cancers achieved a classification accuracy of 87.55 at least. The median auto-quantified TILs of AEGJ, GAC, and ESCC cases were 4.82%, 1.92%, and 0.12%, respectively. The TILs demonstrated varied characteristics under distinctive clinicopathological traits. The higher TILs proportion was

associated with better prognosis in esophagogastric cancers (all  $P < 0.05$ ) and was an independent prognostic biomarker on AEGJ in both datasets (Taixing: HR = 0.965, 95% CI = 0.938–0.994; TCGA: HR = 0.811, 95% CI = 0.712–0.925).

**Conclusions:** We found variations in TILs across ESCC, GAC, and AEGJ, as assessed by image processing algorithms. Additionally, TILs in these three cancers are an independent prognostic factor. This enhances our understanding of the unique immune characteristics of the three tumors, improving patient outcomes.

## INTRODUCTION

Globally, gastric and esophageal cancers were ranked as the fourth and sixth leading causes of cancer-related deaths, and were responsible for 769,000 and 544,000 deaths in 2020, respectively [1]. Given the advanced or metastatic nature of many gastroesophageal cases at diagnosis, the overall 5-year survival rate remains less than 20% in developing countries [2, 3]. The main factors affecting gastroesophageal cancers prognosis include tumor staging and grading, treatment method, living condition, and genetic marker [4, 5]. However, as a transitional region tumor from esophageal squamous epithelium to the gastric adenoid epithelium, it is possible that the oncological principles for esophageal and gastric cancer are not directly applicable to junctional cancer [6].

Accumulating studies reveal that the molecular characteristics, pathological course, and clinical behavior of junctional cancer differ from that of gastric and esophageal cancer [7]. Junctional cancer primarily refers to adenocarcinoma of the esophagogastric junction (AEGJ). It is based on Siewert's anatomical classification criteria and includes distal esophageal adenocarcinoma (EAC), cardiac cancer, and proximal gastric adenocarcinoma (GAC) [8]. The AEGJ incidence has risen rapidly in East Asia, North America, and Europe over the last few decades [9]. As the early symptoms are not obvious with a rapid progression, AEGJ is usually diagnosed in the late stages and has a 5-year survival rate of ~6% in the developing world [10]. Therefore, there is an urgent need to identify potential molecular markers to predict and improve the prognosis.

A specific component of the tumor immune micro-environment (TME), tumor-infiltrating lymphocytes (TILs) are reflective of host-tumor immune interactions and are predictive of patient prognosis [11, 12]. TILs primarily include T and B cells, and natural killer (NK) cells, which cooperate with tumor cells by releasing chemokines and cytokines that act as important tumorigenic and prognostic factors and determine tumor progression and aggressiveness [13]. Different cancer types have distinct TME, where numerous clinical studies that evaluated the TIL content in breast carcinoma, colorectal carcinoma, and non-small cell lung carcinoma

reported that higher TIL infiltration conferred a significant survival benefit [14–16]. Besides, the observation of the variations in TILs levels can recognize the population or cancer types with a high likelihood of reacting to immunotherapy [13]. However, there are few studies focusing on the characteristics of tumor-infiltrating lymphocytes (TILs) in esophagogastric tumors and their potential as prognostic markers to predict and improve survival in AEGJ patients. Additionally, the association between TILs and survival in esophageal squamous cell carcinoma (ESCC) and GAC remains controversial [17, 18].

Although the gold standard for evaluating TILs is based on routine haematoxylin–eosin (H&E) staining using a semi-quantitative scoring method, it may be subject to interobserver variability and costly [19]. Computational pathology has currently displayed promise in recognizing the biomarkers in tissues, and overcomes limitations related with manual grading and human bias [20]. Therefore, we establish cellular training sets for AEGJ, GAC, and ESCC based on the assessment of experienced pathologists. Then, quantification of the TILs on H&E staining sections using an open-source image processing tool that operates with minimal user intervention. We compared the AEGJ TILs characteristics with that of GAC and ESCC under demographic factors and clinical traits. We also examined the association of auto-assessed TILs as a quantitative variable with overall survival in both large datasets.

## RESULTS

### Demographic information

Table 1 displays the demographic information of the 214 AEGJ, 256 GAC, and 752 ESCC cases included in the analysis. There were significant differences between AEGJ, GAC, and ESCC for age, sex, tea drinking, wealth score, first-line treatment method, TNM staging, tumor differentiation grade, *Helicobacter pylori* (HP) infection status, and gastric atrophy (all  $P < 0.05$ ). Compared with ESCC and GAC, AEGJ cases were more likely to be older (mean age: 69.23 years), drink less tea (77.10%), have positive HP status (78.50%), receive combination therapy (24.30%), have advanced TNM stage (28.97%), and have gastric atrophy (25.23%).

**Table 1. Demographic information of 214 AEGJ, 256 GAC, and 752 ESCC cases in Taixing (2010–2014).**

Variables	AEGJ (n = 214), N (%)	GAC (n = 256), N (%)	ESCC (n = 752), N (%)	P-value
<b>Age, mean ± SD, years</b>	69.23 (7.61)	67.52 (9.56)	66.96 (8.50)	<b>0.002</b>
<b>Age (years)</b>				
<60	22 (10.28)	57 (22.27)	140 (18.62)	<b>0.002</b>
≥60	192 (89.72)	199 (77.73)	612 (81.38)	
<b>Sex</b>				<b>0.027</b>
Man	152 (71.03)	189 (73.83)	492 (65.43)	
Woman	62 (28.97)	67 (26.17)	260 (34.57)	
<b>Marriage</b>				<b>0.089</b>
Unmarried	12 (5.61)	9 (3.52)	25 (3.32)	
Married	153 (71.50)	186 (72.66)	591 (78.59)	
Divorce/widow	49 (22.90)	61 (23.83)	136 (18.09)	
<b>Educational level</b>				0.302
Illiteracy	74 (35.58)	80 (31.25)	276 (36.70)	
Primary or Secondary school	129 (60.28)	158 (61.72)	429 (57.05)	
High school and above	11 (5.14)	18 (7.03)	47 (6.25)	
<b>Cigarette smoking</b>				0.820
Never	86 (40.19)	103 (40.23)	313 (41.62)	
Ever or still	120 (56.07)	144 (56.25)	405 (53.86)	
Missing	8 (3.74)	9 (3.52)	34 (4.52)	
<b>Alcohol drinking</b>				<b>0.075</b>
Never	121 (56.54)	125 (48.83)	351 (46.68)	
Ever or still	88 (41.12)	122 (47.66)	366 (48.67)	
Missing	5 (2.34)	9 (3.52)	35 (4.65)	
<b>Tea drinking</b>				<b>0.013</b>
Never	165 (77.10)	170 (66.41)	493 (65.56)	
Ever	44 (20.56)	77 (30.08)	225 (29.92)	
Missing	5 (2.34)	9 (3.52)	34 (4.52)	
<b>Fruit intake (g/d)</b>				0.513
<25	110 (51.40)	136 (53.13)	417 (55.16)	
≥25	91 (42.52)	102 (39.84)	288 (38.10)	
Missing	13 (6.07)	18 (7.03)	47 (6.22)	
<b>Pickles intake (g/d)</b>				0.914
<10	124 (57.94)	141 (55.08)	419 (55.72)	
≥10	80 (37.38)	98 (38.28)	288 (38.30)	
Missing	10 (4.67)	17 (6.64)	45 (5.98)	
<b>BMI</b>				0.350
<18.5	27 (12.62)	32 (12.50)	85 (11.30)	
18.5–24	140 (65.42)	158 (61.72)	466 (61.97)	
≥24	46 (21.50)	66 (25.78)	201 (26.73)	
Missing	1 (0.47)	0 (0.00)	0 (0.00)	
<b>Wealth scores</b>				<b>0.003</b>
Q1	62 (28.97)	52 (20.31)	232 (30.85)	
Q2	45 (21.03)	55 (21.48)	139 (18.48)	
Q3	38 (17.76)	62 (24.42)	161 (21.41)	
Q4	43 (20.09)	47 (18.36)	138 (18.35)	
Q5	26 (12.15)	40 (15.63)	62 (8.24)	
<b>First-line treatment method</b>				<b>&lt;0.001</b>
Radiotherapy	9 (4.21)	9 (3.52)	88 (11.70)	
Chemotherapy	21 (9.81)	27 (10.55)	93 (12.37)	
Surgery	116 (54.21)	151 (58.98)	377 (50.13)	
Combination therapy	52 (24.30)	57 (22.27)	170 (22.61)	
Untreated	14 (6.54)	11 (4.30)	21 (2.79)	
Missing	2 (0.93)	1 (0.39)	3 (0.40)	

<b>TNM staging</b>				<b>0.004</b>
0+I+II	55 (25.70)	98 (38.28)	256 (30.04)	
III+IV	62 (28.97)	50 (19.53)	162 (21.54)	
Missing	97 (45.33)	108 (42.19)	334 (44.41)	
<b>Grade of differentiation</b>				<b>0.036</b>
Gx Grading cannot be evaluated	36 (16.82)	47 (18.36)	40 (5.32)	
G1 Highly differentiated	4 (0.53)	7 (2.73)	60 (7.98)	
G2 Medium differentiation	63 (29.44)	63 (24.61)	452 (60.11)	
G3 Poorly differentiated	29 (13.55)	48 (18.75)	134 (17.82)	
G4 Undifferentiated	11 (5.14)	9 (3.52)	45 (5.98)	
Missing	71 (33.18)	82 (32.03)	21 (2.79)	
<b><i>Helicobacter pylori</i></b>				<b>0.001</b>
HP+	168 (78.50)	181 (70.70)	484 (64.36)	
HP-	42 (19.63)	68 (26.56)	234 (31.12)	
Missing	4 (1.87)	7 (2.73)	34 (4.52)	
<b>Gastric atrophy</b>				<b>&lt;0.001</b>
Yes	54 (25.23)	51 (19.92)	98 (13.03)	
No	131 (61.21)	157 (61.33)	520 (69.15)	
Missing	29 (13.55)	157 (61.33)	134 (17.82)	

### Automated cellular recognition accuracy

We obtained the matched H&E-stained images of 214 AEGJ, 256 GAC, and 752 ESCC cases, each of which was from a solid tumor cross-section. The image processing approach automatically segmented the images and classified the cellular components into cancer cells, lymphocytes, and stromal cells. The classification was based on cellular training sets using an SVM classifier that pathologists had trained according to the cell features (Figure 1A). Cross-validation within the cellular training sets of the three cancers yielded overall classification accuracy of >87.55% (Supplementary Table 1). Furthermore, the overall correlation coefficients between automated recognition and the pathologists' quantitative assessment of AEGJ, GAC, and ESCC were 0.92, 0.93, and 0.93, respectively, and the TILs correlation coefficients were 0.94, 0.95, and 0.95, respectively. The correlation coefficients of the cancer cells and stromal cells were all >0.86 (Figure 1B; Supplementary Figure 1A, 1B). Furthermore, the automated recognition of the TILs proportion in AEGJ, GAC, and ESCC was consistent with the manual grading, and all Jonckheere-Terpstra tests yielded  $P = 0.001$  (Figure 1C; Supplementary Figure 1C, 1D). The AEGJ, GAC, and ESCC cellular training sets are provided in the Supplementary Datasets 1–3.

### The auto-quantified TILs characteristics in AEGJ, GAC, and ESCC

Based on the cellular training sets, we extracted the TILs proportions of 214 AEGJ, 256 GAC, and 752 ESCC cases. The examples of low, medium, and high TILs intensity circled by blue in ESCC, AEGJ, and

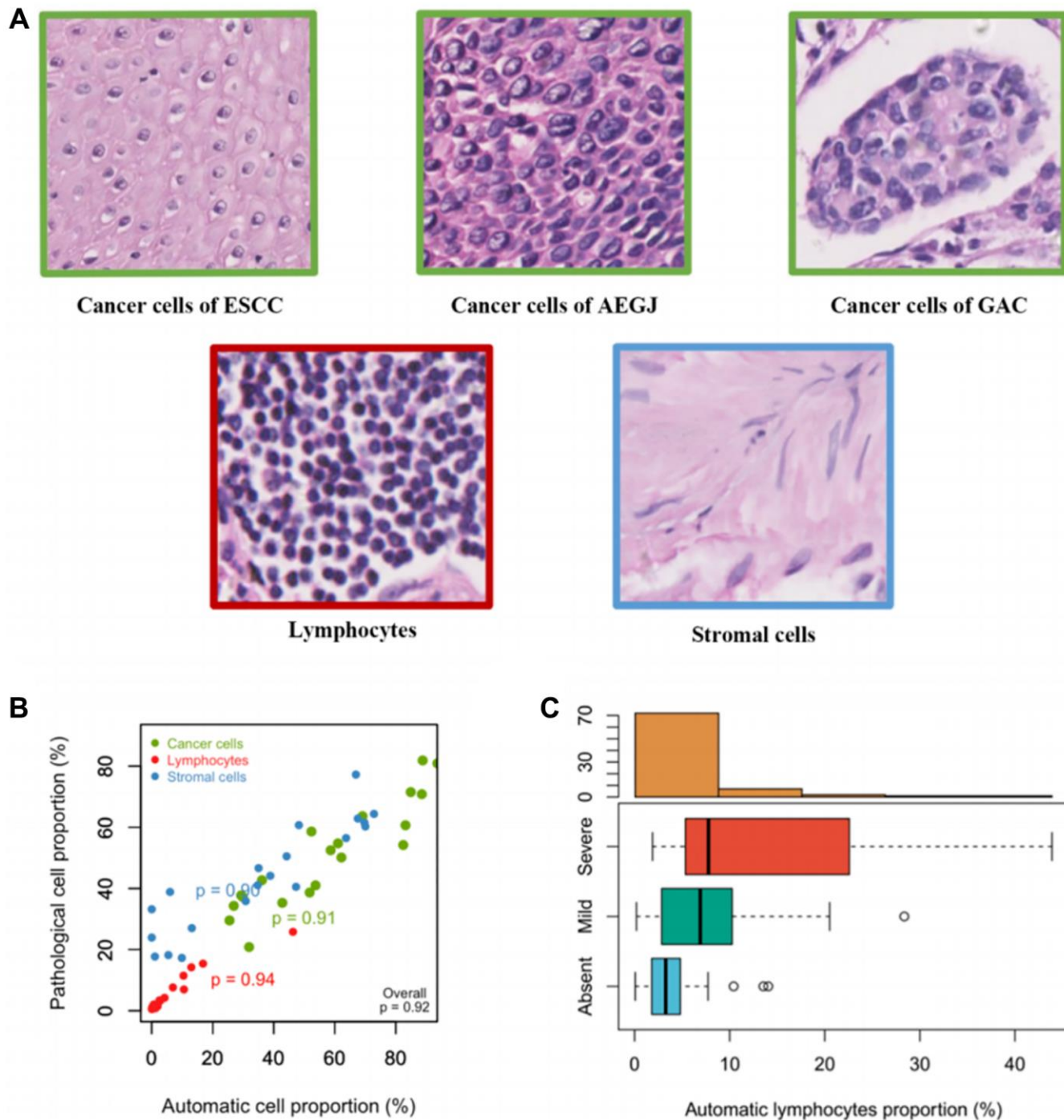
GAC H&E-stained tissue sections and their raw images were displayed (Figure 2A–2I and Supplementary Figure 2A–2I). The association between TILs proportion and demographic information of the above three gastroesophageal cancers was listed in Table 2. In the AEGJ cases, the TILs percentage was associated with the first-line treatment method ( $P = 0.030$ ), and the patients with the combination therapy (median TILs: 5.72%) had the highest TILs proportion compared to radiotherapy (median TILs: 1.83%), chemotherapy (median TILs: 4.10%), and surgery (median TILs: 2.77%). A similar association between TILs and the first-line treatment method had also discovered in GAC patients ( $P = 0.031$ ). Besides, the GAC patients drinking more tea (median TILs: 2.74%) were more likely to have a higher TILs proportion ( $P = 0.028$ ). In the ESCC cases, the patients eating fewer pickles (median TILs: 0.15%) were more likely to have a higher TILs level ( $P = 0.002$ ), and the trend tests demonstrated the TILs proportion increased with the BMI ranks ( $P = 0.008$ ).

### The distribution of auto-quantified TILs proportion

We performed a crude comparison of the AEGJ TILs proportion with that of ESCC and GAC by The Kruskal-Wallis H test. The differences in the TILs proportions were statistically significant ( $P < 0.001$ ). The multiple comparisons corrected by the Bonferroni method found that the AEGJ cases had the highest TILs proportion (median, 4.82%), followed by GAC (median, 1.92%), and that of ESCC was the lowest (median, 0.12%) (Supplementary Figure 3). However, lymphocyte infiltration is associated with age, sex, BMI, TNM staging, tumor differentiation grade, and first-line treatment method [21–23]. Hence, based on

the above factors and the demographic information that differed among the three cancers, we compared the TILs proportion between these three cancers within each factor. The result revealed that significantly different among the above cancers still existed (Supplementary Table 2; all  $P < 0.001$ ). Moreover, the distribution of auto-quantified TILs proportion was tested with

Spearman correlation analysis and multi-variable linear regression. After adding the cancer type as a variable to the analysis-adjusted covariates, cancer type remained associated with TILs proportion (Table 3;  $\rho = 0.49$ ;  $P < 0.001$ ). The standardised effects of AEGJ and GAC were 0.36 and 0.28, respectively (Table 3;  $R^2 = 0.204$ ; adjusted  $R^2 = 0.179$ ;  $P < 0.001$ ). This indicated that the

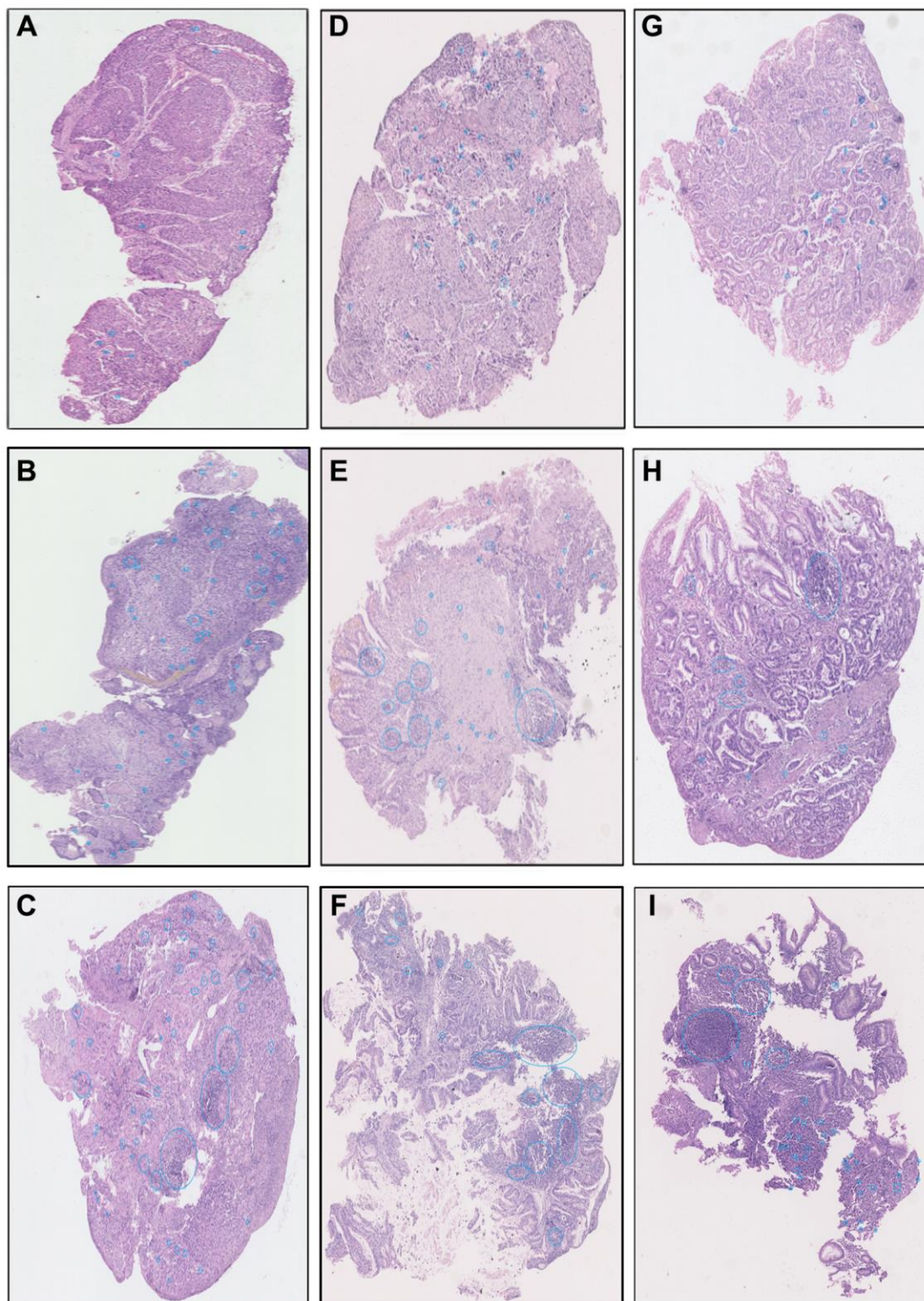


**Figure 1. Establishment and verifications of the cell training set in AEGJ.** (A) Example images of the three classes used in the classifier: cancer cells, lymphocytes, and stromal cells. (B) Cell proportions obtained by automated image analysis were compared to pathologists' counts for a total of 10,000 single cells in a representative set of 20 tissue samples within AEGJ. (C) TILs proportions versus manual grading for AEGJ TIL infiltration in random one-third samples.

cancer type contributed considerably to explaining variations in the distribution of auto-assessed TILs proportion. We also determined that the cancer type is AEGJ had the most influence on TILs proportion compared with GAC and ESCC.

### The prognostic value of auto-quantified TILs proportion in AEGJ, GAC, and ESCC

To explore whether there was a prognostic value of the auto-assessed TILs proportion in AEGJ, GAC, and



**Figure 2.** The examples images of low, medium, and high TILs intensity in ESCC (A–C), AEGJ (D–F), and GAC (G–I) H&E-stained tissue sections (H&E×200).

**Table 2. The association between TILs proportion and demographic information of AEGJ, GAC, and ESCC cases in Taixing.**

Variables	TILs (%) in AEGJ (n = 214)		TILs (%) in GAC (n = 256)		TILs (%) in ESCC (n = 752)	
	Median (IQR)	P-value	Median (IQR)	P-value	Median (IQR)	P-value
<b>Age (years)</b>		<b>0.775</b>		<b>0.711</b>		<b>0.090</b>
<60	3.82 (1.57, 11.05)		1.86 (0.16, 9.16)		<b>0.15 (0.03, 0.92)</b>	
≥60	4.88 (2.01, 11.21)		1.93 (0.47, 8.16)		<b>0.12 (0.02, 0.73)</b>	
<b>Sex</b>		<b>0.479</b>		<b>0.096</b>		<b>0.437</b>
Man	4.55 (1.71, 10.80)		2.38 (0.52, 9.14)		0.12 (0.02, 0.67)	
Woman	5.19 (2.41, 12.45)		1.27 (0.30, 6.54)		0.13 (0.02, 1.10)	
<b>Marriage</b>		<b>0.779</b>		<b>0.702</b>		<b>0.295</b>
Unmarried	8.02 (1.91, 11.32)		1.14 (0.43, 1.81)		0.11 (0.03, 0.43)	
Married	4.60 (1.86, 10.42)		2.36 (0.41, 9.30)		0.12 (0.02, 0.75)	
Divorce/widow	5.19 (2.52, 13.93)		2.18 (0.52, 3.55)		0.17 (0.03, 0.76)	
<b>Educational level</b>		<b>0.455</b>		<b>0.108</b>		<b>0.868</b>
Illiteracy	5.19 (2.58, 11.06)		1.16 (0.35, 4.48)		0.11 (0.02, 0.79)	
Primary or Secondary school	4.65 (1.66, 11.30)		2.50 (0.70, 10.43)		0.13 (0.02, 0.74)	
High school and above	3.23 (1.09, 7.31)		3.09 (0.11, 10.14)		0.12 (0.02, 0.92)	
<b>Cigarette smoking</b>		<b>0.683</b>		<b>0.025</b>		<b>0.173</b>
Never	4.34 (1.35, 13.89)		1.16 (0.21, 7.54)		0.13 (0.02, 1.06)	
Ever or still	4.95 (2.38, 10.50)		2.49 (0.72, 9.43)		0.11 (0.02, 0.63)	
Missing	NA		NA		NA	
<b>Alcohol drinking</b>		<b>0.357</b>		<b>0.432</b>		<b>0.252</b>
Never	5.42 (1.86, 13.35)		1.59 (0.36, 7.99)		0.13 (0.02, 0.91)	
Ever or still	4.17 (2.20, 9.41)		2.38 (0.51, 8.99)		0.11 (0.02, 0.64)	
Missing	NA		NA		NA	
<b>Tea drinking</b>		<b>0.271</b>		<b>0.028</b>		<b>0.997</b>
Never	4.65 (1.56, 11.01)		1.51 (0.29, 7.93)		0.12 (0.02, 0.72)	
Ever	5.36 (3.02, 11.11)		2.74 (1.14, 10.47)		0.12 (0.02, 0.75)	
Missing	NA		NA		NA	
<b>Fruit intake (g/d)</b>		<b>0.613</b>		<b>0.788</b>		<b>0.623</b>
<25	4.30 (2.07, 10.41)		2.23 (0.47, 7.80)		0.13 (0.02, 0.79)	
≥25	5.23 (1.51, 12.34)		1.90 (0.37, 9.58)		0.10 (0.02, 0.67)	
Missing	NA		NA		NA	
<b>Pickles intake (g/d)</b>		<b>0.091</b>		<b>0.164</b>		<b>0.002</b>
<10	5.12 (1.63, 16.54)		2.40 (0.51, 9.14)		0.15 (0.03, 0.96)	
≥10	3.95 (2.29, 9.66)		1.39 (0.36, 6.21)		0.08 (0.02, 0.47)	
Missing	NA		NA		NA	
<b>BMI</b>		<b>0.737</b>		<b>0.700</b>		<b>0.039</b>
<18.5	5.02 (1.32, 8.99)		1.83 (0.33, 6.88)		0.08 (0.02, 0.56)	
18.5–24	4.56 (1.99, 11.21)		2.34 (0.51, 8.08)		0.11 (0.02, 0.62)	
≥24	4.97 (2.09, 10.82)		1.86 (0.43, 9.15)		0.23 (0.02, 1.20)	
Missing	NA		NA		NA	
<b>Wealth scores</b>		<b>0.124</b>		<b>0.707</b>		<b>0.085</b>
Q1	4.99 (2.47, 11.56)		2.13 (0.37, 8.91)		0.12 (0.02, 0.69)	
Q2	2.94 (1.36, 5.60)		1.27 (0.36, 4.56)		0.13 (0.02, 0.96)	
Q3	5.22 (2.24, 13.93)		2.55 (0.38, 12.08)		0.09 (0.01, 0.45)	
Q4	4.26 (1.23, 10.69)		2.38 (0.67, 10.67)		0.13 (0.02, 0.51)	
Q5	5.71 (3.41, 13.55)		2.41 (0.93, 4.74)		0.25 (0.03, 2.05)	
<b>First-line treatment method</b>		<b>0.030</b>		<b>0.031</b>		<b>0.262</b>
Radiotherapy	1.83 (1.05, 3.49)		0.66 (0.18, 1.61)		0.03 (0.02, 0.43)	
Chemotherapy	4.10 (1.56, 5.30)		1.32 (0.30, 13.71)		0.10 (0.02, 0.48)	
Surgery	2.77 (0.40, 7.72)		0.82 (0.27, 2.74)		0.09 (0.02, 0.68)	
Combination therapy	5.72 (3.14, 11.94)		2.78 (0.73, 11.12)		0.15 (0.03, 0.80)	
Untreated	4.64 (1.81, 10.42)		1.32 (0.35, 3.21)		0.10 (0.02, 0.92)	
Missing	NA		NA		NA	

<b>TNM staging</b>		<b>0.750</b>		<b>0.841</b>		<b>0.996</b>
0+I+II	6.92 (3.12, 13.39)		2.64 (0.49, 10.91)		0.14 (0.02, 0.76)	
III+IV	5.34 (2.99, 11.56)		2.47 (0.74, 11.12)		0.15 (0.02, 0.92)	
Missing	NA		NA		NA	
<b>Grade of differentiation</b>		<b>0.207</b>		<b>0.930</b>		<b>0.954</b>
Gx Grading cannot be evaluated	4.04 (1.77, 12.42)		2.38 (0.82, 10.20)		0.16 (0.02, 1.30)	
G1 Highly differentiated	8.67 (2.55, 14.28)		1.60 (0.68, 10.19)		0.13 (0.03, 0.69)	
G2 Medium differentiation	7.72 (3.89, 13.39)		2.78 (1.13, 9.15)		0.13 (0.02, 0.84)	
G3 Poorly differentiated	7.17 (1.97, 12.81)		2.10 (0.26, 11.47)		0.11 (0.02, 0.56)	
G4 Undifferentiated	4.34 (2.42, 5.17)		0.85 (0.35, 3.50)		0.13 (0.05, 0.35)	
Missing	NA		NA		NA	
<b>Helicobacter pylori</b>		<b>0.599</b>		<b>0.721</b>		<b>0.707</b>
HP+	5.24 (2.99, 11.10)		1.83 (0.39, 6.87)		0.12 (0.02, 0.60)	
HP-	4.62 (1.84, 11.24)		2.28 (0.45, 9.66)		0.13 (0.02, 0.77)	
Missing	NA		NA		NA	
<b>Gastric atrophy</b>		<b>0.475</b>		<b>0.167</b>		<b>0.589</b>
Yes	4.96 (1.92, 10.58)		2.40 (0.73, 10.68)		0.12 (0.02, 0.73)	
No	4.34 (1.46, 11.06)		1.31 (0.36, 6.32)		0.09 (0.02, 0.42)	
Missing	NA		NA		NA	

**Table 3. Multi-variable linear regression analysis of the association between factors and TILs proportion.**

Variables	Correlation coefficient	Unstandardized coefficient (B)		Standardized coefficient (Beta)	P-value
		B (95% CI)	SE		
<b>Age</b>					
<60	0.01	Ref.			
≥60		-0.03 (-1.62, 1.57)	0.81	-0.01	0.975
<b>Sex</b>					
Woman	-0.03	Ref.			
Man		-0.06 (-1.54, 1.43)	0.75	0.00	0.942
<b>BMI</b>					
<18.5	0.04	Ref.			
18.5–24		0.36 (-1.72, 2.44)	1.06	0.02	0.734
≥24		0.50 (-1.77, 2.78)	1.16	0.03	0.664
<b>Helicobacter pylori</b>					
HP-	0.06	Ref.			
HP+		0.21 (-1.22, 1.63)	0.73	0.01	0.776
<b>Pickles intake</b>					
<10	0.04	Ref.			
≥10		-1.20 (-2.51, 0.11)	0.67	-0.07	0.07
<b>Tea drinking</b>					
Never	0.00	Ref.			
Ever		0.28 (-1.26, 1.83)	0.79	0.01	0.721
<b>First-line treatment method</b>					
Untreated	0.06	Ref.			
Radiotherapy		-13.94 (-22.20, -5.67)	4.21	-0.39	<b>0.001</b>
Chemotherapy		-14.42 (-22.58, -6.26)	4.15	-0.46	<b>0.001</b>
Surgery		-13.52 (-21.40, -5.63)	4.01	-0.74	<b>0.001</b>
Combination therapy		-13.85 (-21.81, -5.88)	4.06	-0.64	<b>0.001</b>



<b>Grade of differentiation</b>					
G1 Highly differentiated		Ref.			
G2 Medium differentiation		-0.03 (-2.53, 2.47)	1.27	-0.00	0.981
G3 Poorly differentiated	-0.06	0.51 (-2.35, 3.38)	1.46	0.02	0.725
G4 Undifferentiated		-1.46 (-4.87, 1.95)	1.74	-0.05	0.401
Gx Grading cannot be evaluated		1.21 (-2.39, 4.81)	1.83	0.03	0.510
<b>TNM staging</b>					
0+I+II		Ref.			
III+IV	0.03	0.21 (-1.24, 1.66)	0.74	0.01	0.776
<b>Cancer type</b>					
ESCC		Ref.			
GAC	0.49	6.07 (4.32, 7.82)	0.89	0.28	<0.001
AEGJ		8.28 (6.45, 10.12)	0.94	0.36	<0.001
Summary		R <sup>2</sup> = 0.204		Adjusted R <sup>2</sup> = 0.179	<0.001

ESCC, we utilized the cases from Taixing as discovery (Taixing: 214 AEGJ, 256 GAC, and 752 ESCC cases) and these of TCGA as validation (TCGA: 169 AEGJ, 222 GAC, and 70 ESCC cases). The AEGJ, GAC, and ESCC cases from the Taixing dataset were divided into high- and low-TIL groups using median TILs proportions of 4.82%, 1.92%, and 0.12% as cut-offs, respectively. The AEGJ, GAC, and ESCC cases from the TCGA dataset were divided using median TILs proportions of 1.99%, 4.14%, and 32.40%, respectively, as cut-offs, and evaluated the prognostic value in both datasets. Kaplan-Meier plots were generated to compare the OS based on high and low TILs proportions. There were statistically significant associations between better OS and higher TILs proportion in the three cancers of the Taixing dataset (Figure 3A, 3C, 3E; all  $P < 0.001$ ). As validation, we also identified statistically significant differences in OS between the high- and low-groups of the TCGA dataset (Figure 3B, 3D, 3F; all  $P < 0.05$ ).

We further examined the independent prognostic value of the auto-quantified TILs proportion. A multivariable Cox regression analysis was performed on the discovery and validation. In discovery, the marriage and TILs proportion were statistically related to the outcome in AEGJ cases (Table 4). Education level, first-line treatment method, gastric atrophy, and TILs proportion were significantly associated with prognosis in GAC cases (Supplementary Table 3). The age, sex, marriage, first-line treatment method, differentiation grade, and TILs proportion were significantly associated with OS in the ESCC cases (Supplementary Table 4). Specifically, a higher TIL percentage was associated with a better prognosis in AEGJ (adjusted HR (aHR) = 0.965; 95% CI = 0.938–0.994;  $P = 0.017$ ), GAC (aHR = 0.972; 95% CI = 0.949–0.995;  $P = 0.016$ ), and ESCC (aHR = 0.967; 95% CI = 0.938–0.997;  $P = 0.032$ ) cases (Table 4 and Supplementary Tables 3 and 4). Notably, after further adjusting for TNM staging and other

covariates, the associations between TILs proportion and OS were still significant in the AEGJ (aHR = 0.946; 95% CI = 0.907–0.986;  $P = 0.009$ ), GAC (aHR = 0.961; 95% CI = 0.931–0.993;  $P = 0.016$ ), and ESCC (aHR = 0.954; 95% CI = 0.911–0.998;  $P = 0.041$ ) cases (Supplementary Tables 5–7). In validation, except the ESCC (aHR = 0.984; 95% CI = 0.965–1.005;  $P = 0.133$ ) cases, the auto-quantified TILs also were an independent prognostic biomarker in AEGJ (aHR = 0.812; 95% CI = 0.712–0.925;  $P = 0.002$ ), GAC (aHR = 0.969; 95% CI = 0.942–0.996;  $P = 0.025$ ) cases (Supplementary Tables 8–10).

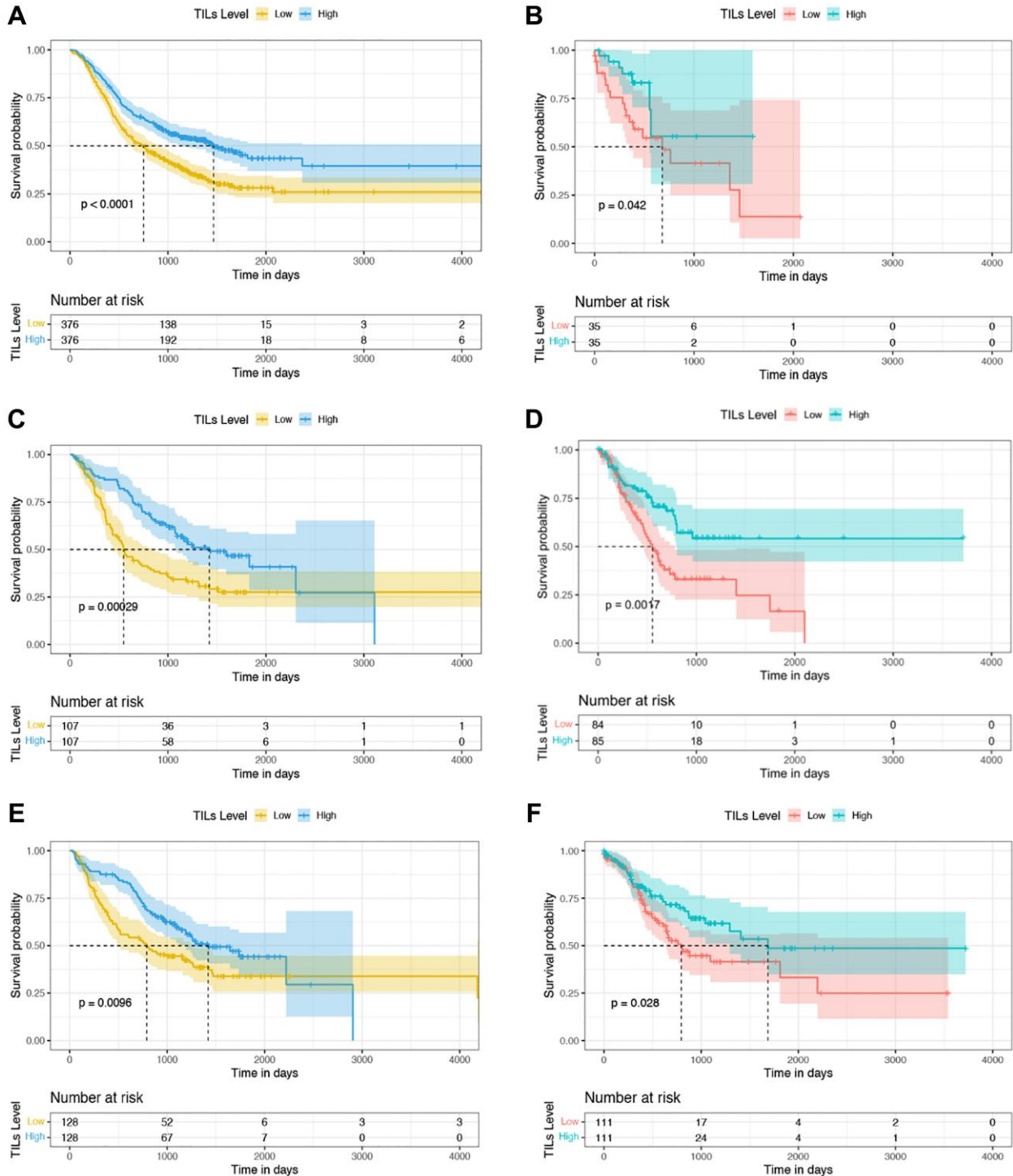
## DISCUSSION

Although a standardized methodology for manual TILs assessment exists, it has several limitations due to requiring professional pathologists, interobserver variability, and higher costs. To address these problems, our study established the cellular training sets, respectively, explored the characteristics of auto-quantified TILs in AEGJ, GAC, and ESCC. The prognostic value of auto-assessed TILs was investigated in the above esophagogastric tumors. The TILs proportion was distinctive between different demographic and clinical traits and was the highest in AEGJ compared with GAC and ESCC in Taixing. The auto-quantified TILs were an independent prognostic biomarker for AEGJ, GAC, and ESCC.

The characteristics of TILs infiltration are distinctive in different body mass, eating habits, and cancer treatment methods. The tumor microenvironment (TME) is a specific metabolic niche composed of various cellular components as well as the contents of the tumor interstitial space. Recent research data have revealed that high-fat diet-induced obesity contributes to the tumor cell fat uptake, whereas the CD8+ T cell intaking the energy was suppressed [24]. These distinctive

adaptations impaired the lymphocyte infiltration degree. However, our study showed the TILs proportion tends to enrich with increasing BMI ranks. This may be related to the types of TILs contained. The TILs include diverse immune cells, e.g., T cells, B cells, and NK cells [25]. In this research, we regard TILs as a major

category that might weaken the above association. Our study also discovered that patients drinking more tea and intaking fewer pickles were more likely to enrich the TILs. These results were consistent with previous experiments. Mantena et al. [26] proved that the tea polyphenols that originated from tea contributed to the



**Figure 3. Kaplan-Meier curves of OS based on TILs proportion in the discovery and validation datasets. (A, C, E)** Survival analysis between the high- and low-TILs groups in 752 ESCC, 214 AEGJ, and 256 GAC cases in Taixing, China, 2010–2014. **(B, D, F)** Survival analysis between the high- and low-TILs groups in 70 ESCC, 169 AEGJ, and 222 GAC cases of the TCGA dataset.

**Table 4. Univariable and multivariable Cox regression analyses of basic characteristics with OS in AEGJ of Taixing dataset (Discovery, N = 214).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			P-value	HR (95% CI)	P-value	aHR <sup>a</sup> (95% CI)
<b>TILs, Median (IQR)</b>	4.82 (1.90, 11.18)		<b>&lt;0.001</b>	0.964 (0.943, 0.985)	<b>0.017</b>	0.965 (0.938, 0.994)
<b>Age, mean (range)</b>	69.23 (49–84)		0.488	1.008 (0.986, 1.031)	0.849	1.003 (0.970, 1.037)
<b>Sex</b>						
Woman	62 (28.97)	2.44		Ref.		Ref.
Man	152 (71.03)	2.32	0.640	1.097 (0.743, 1.621)	0.852	0.950 (0.552, 1.634)
<b>Marriage</b>						
Divorce/widow	49 (22.90)	2.51		Ref.		Ref.
Unmarried	12 (5.61)	2.71	0.286	0.597 (0.231, 1.540)	<b>0.019</b>	0.125 (0.022, 0.708)
Married	153 (71.50)	2.26	0.819	0.953 (0.630, 1.441)	0.383	0.749 (0.391, 1.434)
<b>Educational level</b>						
High school and above	11 (5.14)	1.95		Ref.		Ref.
Primary or Secondary school	129 (60.28)	2.39	0.175	0.633 (0.327, 1.226)	0.253	1.881 (0.637, 5.550)
Illiteracy	74 (35.58)	2.52	0.130	0.586 (0.293, 1.171)	0.496	1.258 (0.650, 2.435)
<b>Cigarette smoking</b>						
Never	313 (41.62)	2.48		Ref.		Ref.
Ever or still	120 (56.07)	2.45	0.813	0.957 (0.666, 1.375)	0.927	1.023 (0.622, 1.683)
Missing	8 (3.74)					
<b>Alcohol drinking</b>						
Never	121 (56.54)	2.39		Ref.		Ref.
Ever or still	88 (41.12)	2.44	0.979	1.005 (0.701, 1.441)	0.767	0.930 (0.574, 1.505)
Missing	5 (2.34)					
<b>Tea drinking</b>						
Never	165 (77.10)	2.51		Ref.		Ref.
Ever	44 (20.56)	2.14	0.930	1.020 (0.656, 1.585)	0.909	1.033 (0.590, 1.809)
Missing	5 (2.34)					
<b>BMI</b>						
≥24	46 (21.50)	2.57		Ref.		Ref.
18.5–24	140 (65.42)	2.45	0.587	0.890 (0.585, 1.354)	0.706	0.893 (0.497, 1.607)
<18.5	27 (12.62)	1.68	0.611	1.171 (0.638, 2.151)	0.800	1.116 (0.477, 2.608)
Missing	1 (0.47)					
<b>Wealth scores</b>						
Q5	26 (12.15)	2.35		Ref.		Ref.
Q4	43 (20.09)	2.35	0.682	1.145 (0.600, 2.187)	0.647	0.817 (0.345, 1.939)
Q3	38 (17.76)	2.45	0.936	0.973 (0.494, 1.914)	0.916	0.952 (0.379, 2.390)
Q2	45 (21.03)	1.98	0.566	1.206 (0.636, 2.288)	0.457	0.718 (0.300, 1.719)
Q1	62 (28.97)	2.53	0.862	1.056 (0.570, 1.959)	0.430	0.714 (0.309, 1.647)
<b>First-line treatment method</b>						
Untreated	14 (6.54)	1.44		Ref.		Ref.
Radiotherapy	9 (4.21)	0.99	0.294	0.598 (0.229, 1.563)	0.424	0.473 (0.075, 2.967)
Chemotherapy	21 (9.81)	1.19	0.256	0.664 (0.327, 1.347)	0.610	0.723 (0.208, 2.516)
Surgery	116 (54.21)	2.94	<b>&lt;0.001</b>	0.235 (0.129, 0.426)	0.303	0.542 (0.169, 1.738)
Combination therapy	52 (24.30)	2.05	<b>0.002</b>	0.369 (0.196, 0.695)	0.469	0.637 (0.188, 2.162)
Missing	2 (0.93)					

<b>Grade of differentiation</b>							
G1 Highly differentiated	4 (0.53)	2.19		Ref.			Ref.
G2 Medium differentiation	63 (29.44)	3.00	0.573	0.661 (0.156, 2.790)	0.625	0.695 (0.161, 2.994)	
G3 Poorly differentiated	29 (13.55)	3.13	0.572	0.894 (0.204, 3.919)	0.926	0.931 (0.206, 4.203)	
G4 Undifferentiated	11 (5.14)	0.68	<b>0.048</b>	4.665 (1.015, 21.443)	0.058	4.655 (0.947, 22.879)	
Gx Grading cannot be evaluated	36 (16.82)	1.48	0.348	1.993 (0.473, 8.405)	0.396	1.886 (0.436, 8.156)	
Missing	71 (33.18)						
<b><i>Helicobacter pylori</i></b>							
HP-	42 (19.63)	2.36		Ref.			Ref.
HP+	168 (78.50)	2.36	0.959	1.012 (0.654, 1.564)	0.794	1.083 (0.597, 1.962)	
Missing	4 (1.87)						
<b>Gastric atrophy</b>							
No	131 (61.21)	2.26		Ref.			Ref.
Yes	54 (25.23)	2.13	0.116	1.362 (0.927, 2.001)	0.164	1.466 (0.855, 2.514)	
Missing	29 (13.55)						

<sup>a</sup>aHR with adjustment for TILs proportion, age, sex, differentiation grade, first-line treatment method, and BMI.

increasing recruitment of TILs in TME. Besides, eating excessive pickles leads to more nitrite intake, which easily oxidizes hemoglobin to methemoglobin, resulting in the lower oxygen-carrying capacity of blood and facilitating the formation of an immunosuppressive environment [27]. Additionally, the patients with combination therapy had the most TILs levels compared to radiotherapy, chemotherapy, and surgery in our cases. These results support that combination therapy produces fewer side effects on immune cells than applying the above three methods alone [28]. Hence, keeping a better diet and body quality, and choosing a suitable cancer therapy in the clinical field will help to improve immune infiltration to resist tumor growth.

The TILs infiltration showed a cancer-specificity in esophagogastric cancers. Quantification of TILs is growing in significance as evidence emerges of a reliable biomarker to reflect the better response to immunotherapeutic agents [29]. Characterizing the TIL proportion between different solid tumors would provide clues into the varied effectiveness of immunotherapy. In the general clinical field, the AEGJ, located between the esophagus and stomach is more likely to group with GAC. Nevertheless, increasing evidence demonstrated the AEGJ displayed a significant difference in immune molecular characteristics [30]. In this study, we revealed that the TILs proportion varied between esophagogastric cancer, where AEGJ had the highest TIL proportion. The absolute difference in the TILs proportions between AEGJ and GAC was smaller than that between ESCC and AEGJ or between ESCC and GAC. Our results were similar to previous studies. Mohamed et al. [31] studied 215 ESCC, 1176 EAC, and 1951 GAC cases, including gastric and gastroesophageal junction cancers, and reported that

ESCC exhibited a unique molecular profile, whereas GAC and AEGJ shared similarities, supporting the idea that squamous cell carcinomas and adenocarcinomas are entirely different diseases at the molecular level. In another study of 4125 tumor specimens from patients with 14 different gastrointestinal cancer types, Alberto et al. demonstrated that AEGJ had different immune characteristics from GAC and EAC [32]. Our results supported and complemented these findings, indicating the specificity of AEGJ in lymphocyte infiltration degree compared to GAC and ESCC. In addition, variations of TILs infiltration in cancers also indirectly reflect the different immunotherapy effects. Our results might present evidence for the specific selection of immunotherapy for esophagogastric cancers. However, some trial examinations of PD-1/PD-L1 blockade in upper gastrointestinal cancers enrolled patients with gastric cancers and AEGJ without distinction [33]. Therefore, our results also provide clues for future clinical immunotherapy in esophagogastric cancers and enhance precision therapy.

The auto-assessed TIL proportion is an independent prognostic biomarker in AEGJ, GAC, and ESCC patients. The cumulative studies have focused on the association between semiquantitative scoring of TILs levels and prognosis in esophagogastric cancers, and high TILs scores have been reported as a positive prognosis marker [34–37]. Despite the standardized efforts, the subjective nature and higher costs have limited its translational adoption into clinical practice [38]. Besides, the prognostic biomarkers for AEGJ are still under-explored. For this reason, we used the automatic algorithm to quantify the TILs percentage and investigate its prognostic value. We performed a survival analysis of AEGJ and determined that the auto-

quantified TILs proportion was an independent prognostic biomarker in Taixing and TCGA datasets. This finding was consistent with the results of two previous studies [39, 40], in which the TILs proportion was estimated by pathologists. However, several researchers demonstrated that the prognostic value of TILs in GAC and ESCC has not been defined [41–43]. Nevertheless, some studies also reported positive results [44–47]. In the present study, we also identified an association between higher auto-assessed TILs proportion and better overall survival in GAC and ESCC cases in both datasets. This discovery supported the idea of the prognostic value of TILs proportion in GAC and ESCC. Hence, we can predict the overall survival of AEGJ, GAC, and ESCC by auto-quantified TILs infiltration degree objectively and it has the potential for translation to the routine clinical and pathological application at minimal additional cost.

As far as we know, this is the first relatively comprehensive establishment of cellular training sets for esophagogastric tumors to automatically quantify TILs infiltration in AEGJ, GAC, and ESCC. We obtained relatively full demographic characteristics and clinical information to explore the TILs characteristics. The findings contributed to more accurate tumor classification and immunotherapy outcome prediction. As an independent prognostic factor common to AEGJ, the auto-quantified TILs provided more evidence for its predictive value in upper gastrointestinal tumors. The findings established the foundation for further exploration of TME differences at specific immune cell level, providing crucial insights into immunotherapy and supporting the prognostic value of the auto-quantified TILs proportion in esophagogastric tumors, particularly in the AEGJ patients.

This study has some limitations. Our study aims to provide clues for immunotherapy in patients with upper gastrointestinal tumors by comparing the TILs level as the immune characteristic. Although the cases in our study did not receive immunotherapy to directly draw relevant conclusions, our results can provide data support for immunotherapy in AEGJ to some extent. Besides, multivariable survival analysis in the validation set showed that auto-quantified TILs were not an independent prognostic factor for ESCC, which may be related to the small sample size of ESCC in TCGA. However, considering the large sample size of ESCC in the Taixing dataset, we can still consider the independent prognostic value of auto-quantified TILs in ESCC.

Future research should investigate the association between the auto-quantified TILs proportion and clinical outcomes in patients received immunotherapy.

Moreover, incorporating additional independent datasets to deeply validate the independent prognostic value of TILs in patients with AEGJ, GAC and ESCC will enhance the value of the clinical application of this biomarker.

## CONCLUSIONS

The findings of this study suggested that TILs levels determined by CRImage based on three different cell training sets are showing distinctive characteristics between various demographic information, clinical traits, and cancer types. The auto-quantified TILs are an independent prognostic factor in AEGJ, GAC, and ESCC patients, and are associated with a favorable prognosis. It is a cost-effective biomarker to predict and improve prognosis in clinical and pathological research.

## MATERIALS AND METHODS

### Patients and sample selection

The analysis was based on a 2010–2014 population-based case-control study in Taixing, Jiangsu, China, where there is a high incidence of upper gastrointestinal cancer. Patients were mainly recruited from the endoscopy units of the four most prominent hospitals in Taixing, the People's Hospital of Taixing, the Second People's Hospital of Taixing, the Third People's Hospital of Taixing, and the Hospital of Traditional Chinese Medicine of Taixing. Patients with suspected ESCC, AEGJ, and GAC on endoscopy were invited to participate. The demographic information of patients was obtained using questionnaires and pathological sections of formalin-fixed, paraffin-embedded tissue blocks were obtained from the pathology department. The following inclusion criteria of the participants were applied: (1) age 40–85 years old and living in Taixing for >5 years; (2) suspected ESCC, AEGJ, and GAC by endoscopy that was subsequently pathologically confirmed; (3) preserved H&E staining tumor section images. The detailed research designs were described previously [48–50]. Based on the above criteria, the 1005 ESCC, 292 AEGJ, and 340 GAC cases were included. We also excluded 233 ESCC (23.18%), 71 AEGJ (24.32%), and 80 GAC (23.53%) cases with low-quality images due to uneven H&E-stained, unflattening, and blurred images. And the 20 ESCC (1.99%), 7 AEGJ (2.40%), and 4 GAC (1.18%) patients who missed the survival time were excluded. Finally, the 752 ESCC, 214 AEGJ, and 256 GAC cases were analyzed. Figure 4 illustrates the case selection flow. In the TCGA dataset, we excluded 34 AEGJ (16.75%), 51 GAC (18.68%), and 19 ESCC (21.35%) cases with low-quality images from the 203 AEGJ, 273 GAC, and 89 ESCC cases, respectively. The 169 AEGJ, 222 GAC,

and 70 ESCC cases from the TCGA dataset were analyzed to validate the prognostic value of auto-quantified TILs.

### Pathological image processing pipeline

We utilized the pathological image processing pipeline published in our previous study [51]. The images of the H&E-stained tumor sections were processed

using the R package CRImage developed by Yuan et al. [52]. Based on watershed segmentation and Otsu thresholding for haematoxylin-positive nuclei, this tool was embedded with the EImage R package and a support vector machine (SVM) [52], to achieve color transformation and segmentation of the nuclei, then analyzed each morphological feature of each nuclei detected, such as shape, intensity, and texture features. The resulting morphological and textural features were

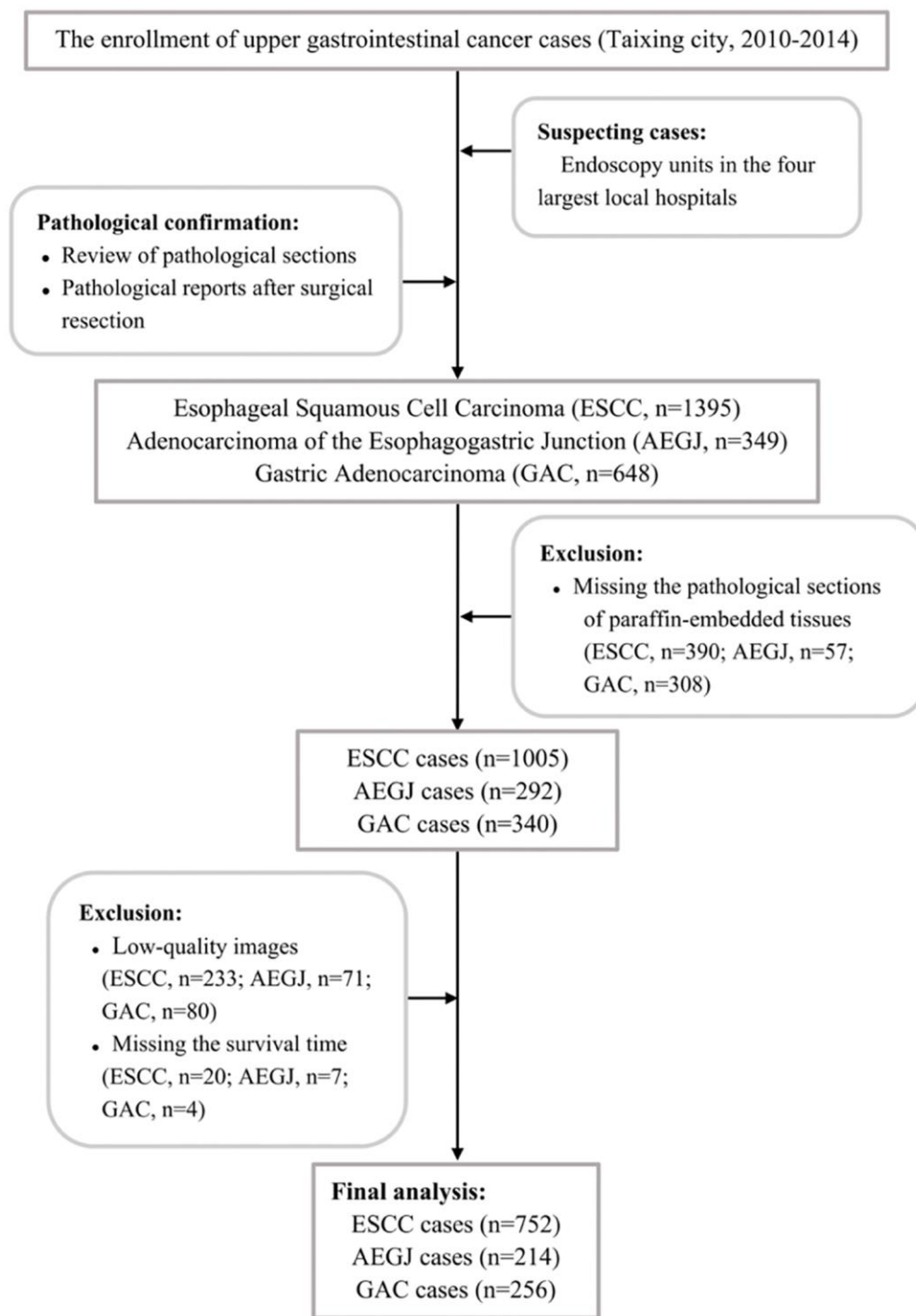


Figure 4. The flowchart of inclusion and exclusion of ESCC, AEGJ, and GAC cases in Taixing, Jiangsu (2010–2014).

input into the SVM for the supervised classification of cancer cells, lymphocytes, and stromal cells. The cancer cells exhibited large nuclei and variable texture and shape, and lymphocytes were small, round, and contained basophilic nuclei. Therefore, cancer cells and lymphocytes could be reliably differentiated from stromal cells that contained the elongated nuclei of fibroblasts and endothelial cells.

We selected the regions containing tumor cells, lymphocytes, and stromal cells from the tissue images and imported them into EBImage for conversion to the LAB color space. The mean and standard deviation of each channel were computed to convert the image to grayscale for further segmentation and cell recognition. Subsequently, an Otsu threshold to partition the image into foreground and background was constructed by the algorithm of maximization of the between-class variance method and morphological opening. Leveraging both the image grayscale and the threshold, the algorithm can eliminate noise and refine the cell edges. Eventually, the watershed segmentation was performed to separate cell clusters, automatically outlining recognized cells in the image.

The senior pathologists were invited to discern the circled cell types, identifying them as cancer cells, lymphocytes, and stromal cells. Subsequently, the EBImage toolkit integrated within CRImage was employed to extract 43 cellular features, encompassing nucleus perimeter, major axis, eccentricity, and the count of neighboring cells, among additional metrics. These features were then exported to construct training sets comprising cellular characteristics for each cancer type. Considering the cytomorphological differences between AEGJ, GAC, and ESCC, we established cell training sets for each cancer. We performed three verifications to test the accuracy of the image analysis tool based on our training sets: (1) cross-validation within each training set; (2) correlation between 10,000 single-cell annotations by the pathologists and automated recognition; (3) comparison of auto-quantified TILs proportion with manual infiltration grade of TILs evaluated by expert pathologists in random one-third samples of each cancer type.

We enhanced the computational efficiency of the algorithm by dividing each entire H&E-stained tumor section image into 100 equal parts. The established AEGJ, GAC, and ESCC cell feature training sets were loaded into the SVM embedded in CRImage to create cell classifiers, respectively. Utilizing the corresponding classifier to read the target H&E-stained tumor tissue sections, we obtained the cell counts and measured lymphocyte infiltration of each sample by calculating the TILs proportion which refers to lymphocyte counts divided by the total cell counts.

## Statistical analysis

The distributions of the patient's demographic characteristics were summarized and presented as counts and percentages. Differences in ordered categorical or continuous variables between cancer types were assessed with the Kruskal-Wallis H test and Wilcoxon rank sum tests. The Pearson chi-square test was used for disordered categorical variables. The multiple comparisons were corrected by the Bonferroni method. The correlations between the auto-assessed TILs percentage and demographic information were analyzed using a Kruskal-Wallis H test since the proportion of TILs did not meet the prerequisites for the parametric testing method. Spearman correlation analysis and multi-variable linear regression were performed to explore the distribution of TILs proportion between cancer types with adjustment for age, sex, body mass index (BMI), first-line treatment method, tumor differentiation grade, tumor-node-metastasis (TNM) staging, i.e., factors correlated with TILs levels. Besides, we also adjusted for *Helicobacter pylori* infection status, pickles intake, and tea drinking which were significant in univariable correlation analyses.

Overall survival (OS) was defined as the time from diagnosis to the time of any-cause death or the last follow-up. The OS was evaluated with Kaplan-Meier curves and compared using log-rank tests. The Cox proportional hazard model was conducted for univariable and multivariable association analyses between the auto-quantified TILs proportion and OS, adjusted for age, sex, BMI, first-line treatment method, and tumor differentiation grade, i.e., factors associated with esophagogastric cancers survival. The multivariable model adjusting for TNM staging and other demographic characteristics was used for the 117 AEGJ, 148 GAC, and 418 ESCC cases with TNM staging information. The hazard ratio (HR) and 95% confidence interval (CI) were estimated. The statistical analyses were performed using R version 4.1.2 (<http://cran.r-project.org>). Two-sided  $P < 0.05$  was considered statistically significant.

## Availability of data and materials

The datasets generated during and analyzed during the current study are available from the corresponding author upon reasonable request.

## Abbreviations

AEGJ: Adenocarcinoma of the Esophagogastric Junction; aHR: adjusted HR; ESCC: Esophageal Squamous Cell Carcinoma; GAC: Gastric Adenocarcinoma; OS: Overall Survival; SVM: Support Vector Machine; TILs: Tumor-Infiltrating Lymphocytes.

## AUTHOR CONTRIBUTIONS

HD designed the study and made substantial contributions to establishing the AEGJ and GAC cellular training sets, and participated in data analysis and writing the manuscript. LY contributed to establishing the ESCC cellular training set and the conception. JF, PG, and XY participated in the design of the study and in the critical revision of the manuscript draft. ZY, TZ, ML, and LJ contributed to the design of the study and conceived data recording. CS and XC conceived the study, performed experiments, essentially revised the manuscript, and contributed equally to this article. All authors provided comments on the draft manuscript and approved the final manuscript.

## ACKNOWLEDGMENTS

We would like to thank the interviewers and technicians at Fudan University Taizhou Institute of Health Sciences for their contribution to data collection and sample preparation; the staff of Taixing Center for Disease Control and Prevention for organizing the fieldwork; and the staff of the study hospitals for their assistance in sample collection.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

## ETHICAL STATEMENT AND CONSENT

This study was performed in line with the principles of the declaration of Helsinki. Approval was granted by the institutional review board at the School of Life Sciences, Fudan University, and the institutional review board at Qilu Hospital, Shandong University (Ethical Approval Number: KYLL-2018(KS)-204). The authors affirm that all human research participants provided informed consent for publication of the individual details in this article.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Numbers: 82122060, 82073637, 81973116), the National Key Research and Development Program of China (Grant Number: 2019YFC1315804), the Innovation Grant from Science and Technology Commission of Shanghai Municipality, China (Grant Number: 20ZR1405600).

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:[33538338](https://pubmed.ncbi.nlm.nih.gov/33538338/)
2. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev.* 2014; 23:700–13. <https://doi.org/10.1158/1055-9965.EPI-13-1057> PMID:[24618998](https://pubmed.ncbi.nlm.nih.gov/24618998/)
3. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, Cunningham D. Oesophageal cancer. *Nat Rev Dis Primers.* 2017; 3:17048. <https://doi.org/10.1038/nrdp.2017.48> PMID:[28748917](https://pubmed.ncbi.nlm.nih.gov/28748917/)
4. Wang D, Dai J, Suo C, Wang S, Zhang Y, Chen X. Molecular subtyping of esophageal squamous cell carcinoma by large-scale transcriptional profiling: Characterization, therapeutic targets, and prognostic value. *Front Genet.* 2022; 13:1033214. <https://doi.org/10.3389/fgene.2022.1033214> PMID:[36425064](https://pubmed.ncbi.nlm.nih.gov/36425064/)
5. Sung H, Hu N, Yang HH, Giffen CA, Zhu B, Song L, Su H, Wang C, Parisi DM, Goldstein AM, Taylor PR, Hyland PL. Association of high-evidence gastric cancer susceptibility loci and somatic gene expression levels with survival. *Carcinogenesis.* 2017; 38:1119–28. <https://doi.org/10.1093/carcin/bgx090> PMID:[29028942](https://pubmed.ncbi.nlm.nih.gov/29028942/)
6. Chevallay M, Bollschweiler E, Chandramohan SM, Schmidt T, Koch O, Demanzoni G, Mönig S, Allum W. Cancer of the gastroesophageal junction: a diagnosis, classification, and management review. *Ann N Y Acad Sci.* 2018; 1434:132–8. <https://doi.org/10.1111/nyas.13954> PMID:[30138540](https://pubmed.ncbi.nlm.nih.gov/30138540/)
7. Hui Z, Xianglin M. Association of HOTAIR expression with PI3K/Akt pathway activation in adenocarcinoma of esophagogastric junction. *Open Med (Wars).* 2016; 11:36–40. <https://doi.org/10.1515/med-2016-0008> PMID:[28352764](https://pubmed.ncbi.nlm.nih.gov/28352764/)
8. Ichihara S, Uedo N, Gotoda T. Considering the esophagogastric junction as a 'zone'. *Dig Endosc.* 2017 (Suppl 2); 29:3–10. <https://doi.org/10.1111/den.12792>



PMID:[28425656](#)

9. Tang W, Chen S, Chen Y, Lin J, Lin J, Wang Y, Liu C, Kang M. Programmed death-1 polymorphisms is associated with risk of esophagogastric junction adenocarcinoma in the Chinese Han population: A case-control study involving 2,740 subjects. *Oncotarget*. 2017; 8:39198–208. <https://doi.org/10.18632/oncotarget.17338> PMID:[28487496](#)
10. Alsina M, Diez M, Tabernero J. Emerging biological drugs for the treatment of gastroesophageal adenocarcinoma. *Expert Opin Emerg Drugs*. 2021; 26:385–400. <https://doi.org/10.1080/14728214.2021.2010705> PMID:[34814781](#)
11. Eismann L, Rodler S, Buchner A, Schulz GB, Volz Y, Bischoff R, Ebner B, Westhofen T, Casuscelli J, Waidelich R, Stief C, Schlenker B, Ledderose S. Identification of the Tumor Infiltrating Lymphocytes (TILs) Landscape in Pure Squamous Cell Carcinoma of the Bladder. *Cancers (Basel)*. 2022; 14:3999. <https://doi.org/10.3390/cancers14163999> PMID:[36010989](#)
12. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, Christie M, van de Vijver K, Estrada MV, Gonzalez-Ericsson PI, Sanders M, Solomon B, Solinas C, et al. Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinomas, Non-Small Cell Lung Carcinoma and Mesothelioma, Endometrial and Ovarian Carcinomas, Squamous Cell Carcinoma of the Head and Neck, Genitourinary Carcinomas, and Primary Brain Tumors. *Adv Anat Pathol*. 2017; 24:311–35. <https://doi.org/10.1097/PAP.000000000000161> PMID:[28777143](#)
13. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, Christie M, van de Vijver K, Estrada MV, Gonzalez-Ericsson PI, Sanders M, Solomon B, Solinas C, et al. Assessing Tumor-infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group: Part 1: Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma In Situ, Metastatic Tumor Deposits and Areas for Further Research. *Adv Anat Pathol*. 2017; 24:235–51. <https://doi.org/10.1097/PAP.000000000000162> PMID:[28777142](#)
14. Rakaee M, Adib E, Ricciuti B, Sholl LM, Shi W, Alessi JV, Cortellini A, Fulgenzi CAM, Viola P, Pinato DJ, Hashemi S, Bahce I, Houda I, et al. Association of Machine Learning-Based Assessment of Tumor-Infiltrating Lymphocytes on Standard Histologic Images With Outcomes of Immunotherapy in Patients With NSCLC. *JAMA Oncol*. 2023; 9:51–60. <https://doi.org/10.1001/jamaoncol.2022.4933> PMID:[36394839](#)
15. Yoo SY, Park HE, Kim JH, Wen X, Jeong S, Cho NY, Gwon HG, Kim K, Lee HS, Jeong SY, Park KJ, Han SW, Kim TY, et al. Whole-Slide Image Analysis Reveals Quantitative Landscape of Tumor-Immune Microenvironment in Colorectal Cancers. *Clin Cancer Res*. 2020; 26:870–81. <https://doi.org/10.1158/1078-0432.CCR-19-1159> PMID:[31757879](#)
16. Bai Y, Cole K, Martinez-Morilla S, Ahmed FS, Zugazagoitia J, Staaf J, Bosch A, Ehinger A, Nimeus E, Hartman J, Acs B, Rimm DL. An Open-Source, Automated Tumor-Infiltrating Lymphocyte Algorithm for Prognosis in Triple-Negative Breast Cancer. *Clin Cancer Res*. 2021; 27:5557–65. <https://doi.org/10.1158/1078-0432.CCR-21-0325> PMID:[34088723](#)
17. Yoshioka T, Miyamoto M, Cho Y, Ishikawa K, Tsuchikawa T, Kadoya M, Li L, Mishra R, Ichinokawa K, Shoji Y, Matsumura Y, Shichinohe T, Hirano S, et al. Infiltrating regulatory T cell numbers is not a factor to predict patient's survival in oesophageal squamous cell carcinoma. *Br J Cancer*. 2008; 98:1258–63. <https://doi.org/10.1038/sj.bjc.6604294> PMID:[18349839](#)
18. Zheng X, Song X, Shao Y, Xu B, Chen L, Zhou Q, Hu W, Zhang D, Wu C, Tao M, Zhu Y, Jiang J. Prognostic role of tumor-infiltrating lymphocytes in gastric cancer: a meta-analysis. *Oncotarget*. 2017; 8:57386–98. <https://doi.org/10.18632/oncotarget.18065> PMID:[28915679](#)
19. Kos Z, Roblin E, Kim RS, Michiels S, Gallas BD, Chen W, van de Vijver KK, Goel S, Adams S, Demaria S, Viale G, Nielsen TO, Badve SS, et al, and International Immuno-Oncology Biomarker Working Group. Pitfalls in assessing stromal tumor infiltrating lymphocytes (sTILs) in breast cancer. *NPJ Breast Cancer*. 2020; 6:17. <https://doi.org/10.1038/s41523-020-0156-0> PMID:[32411819](#)
20. Lan L, Feng K, Wu Y, Zhang W, Wei L, Che H, Xue L, Gao Y, Tao J, Qian S, Cao W, Zhang J, Wang C, Tian M. Phenomic Imaging. *Phenomics*. 2023; 3:597–612. <https://doi.org/10.1007/s43657-023-00128-8> PMID:[38223684](#)
21. Dodd KC, Menon M. Sex bias in lymphocytes: Implications for autoimmune diseases. *Front*

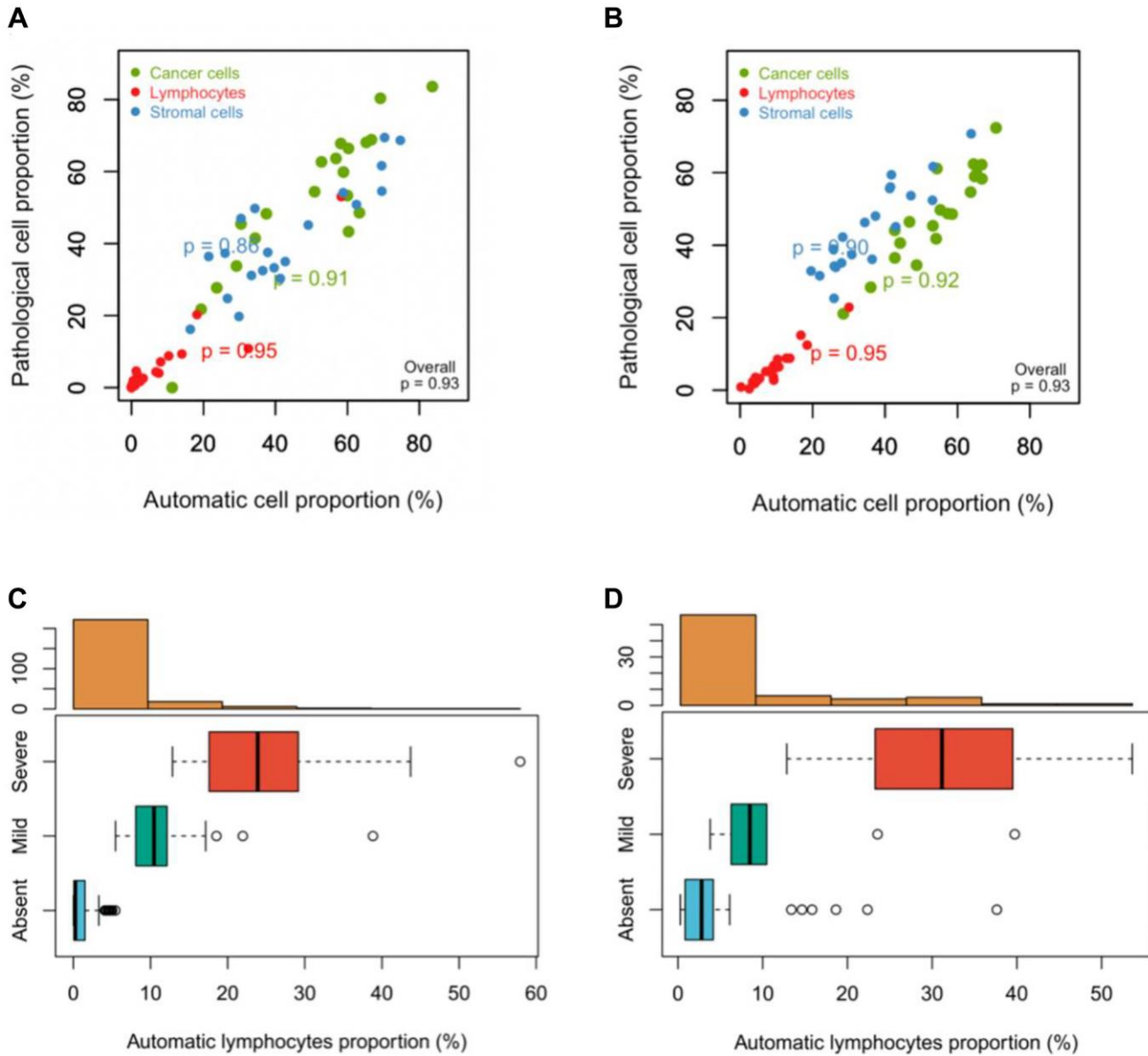
- Immunol. 2022; 13:945762.  
<https://doi.org/10.3389/fimmu.2022.945762>  
PMID:[36505451](https://pubmed.ncbi.nlm.nih.gov/36505451/)
22. Murad LD, Silva TQ, Schilithz AOC, Fernandes PV, Monteiro M, Murad LB, Fialho E. Low body mass index is associated with reduced intratumoral CD4<sup>+</sup> T-lymphocyte infiltration in laryngeal squamous cell carcinoma patients. *Nutr Res.* 2022; 102:1–12.  
<https://doi.org/10.1016/j.nutres.2022.02.005>  
PMID:[35366454](https://pubmed.ncbi.nlm.nih.gov/35366454/)
23. Takada K, Kashiwagi S, Asano Y, Goto W, Morisaki T, Shibutani M, Tanaka H, Hirakawa K, Ohira M. Differences in tumor-infiltrating lymphocyte density and prognostic factors for breast cancer by patient age. *World J Surg Oncol.* 2022; 20:38.  
<https://doi.org/10.1186/s12957-022-02513-5>  
PMID:[35177074](https://pubmed.ncbi.nlm.nih.gov/35177074/)
24. Ringel AE, Drijvers JM, Baker GJ, Catozzi A, García-Cañaveras JC, Gassaway BM, Miller BC, Juneja VR, Nguyen TH, Joshi S, Yao CH, Yoon H, Sage PT, et al. Obesity Shapes Metabolism in the Tumor Microenvironment to Suppress Anti-Tumor Immunity. *Cell.* 2020; 183:1848–66.e26.  
<https://doi.org/10.1016/j.cell.2020.11.009>  
PMID:[33301708](https://pubmed.ncbi.nlm.nih.gov/33301708/)
25. Zhu Y, Yu X, Thamphiwatana SD, Zheng Y, Pang Z. Nanomedicines modulating tumor immunosuppressive cells to enhance cancer immunotherapy. *Acta Pharm Sin B.* 2020; 10:2054–74.  
<https://doi.org/10.1016/j.apsb.2020.08.010>  
PMID:[33304779](https://pubmed.ncbi.nlm.nih.gov/33304779/)
26. Zheng YY, Viswanathan B, Kesarwani P, Mehrotra S. Dietary agents in cancer prevention: an immunological perspective. *Photochem Photobiol.* 2012; 88:1083–98.  
<https://doi.org/10.1111/j.1751-1097.2012.01128.x>  
PMID:[22372381](https://pubmed.ncbi.nlm.nih.gov/22372381/)
27. Kohn MC, Melnick RL, Ye F, Portier CJ. Pharmacokinetics of sodium nitrite-induced methemoglobinemia in the rat. *Drug Metab Dispos.* 2002; 30:676–83.  
<https://doi.org/10.1124/dmd.30.6.676>  
PMID:[12019195](https://pubmed.ncbi.nlm.nih.gov/12019195/)
28. Hassanian H, Asadzadeh Z, Baghbanzadeh A, Derakhshani A, Dufour A, Rostami Khosroshahi N, Najafi S, Brunetti O, Silvestris N, Baradaran B. The expression pattern of Immune checkpoints after chemo/radiotherapy in the tumor microenvironment. *Front Immunol.* 2022; 13:938063.  
<https://doi.org/10.3389/fimmu.2022.938063>  
PMID:[35967381](https://pubmed.ncbi.nlm.nih.gov/35967381/)
29. Lee N, Zakka LR, Mihm MC Jr, Schatton T. Tumour-infiltrating lymphocytes in melanoma prognosis and cancer immunotherapy. *Pathology.* 2016; 48:177–87.  
<https://doi.org/10.1016/j.pathol.2015.12.006>  
PMID:[27020390](https://pubmed.ncbi.nlm.nih.gov/27020390/)
30. Yu X, Xiao H, Zhao B, Zhang X, Wang G. DNA repair gene ERCC1 C118T polymorphism predicts sensitivity of recurrent esophageal cancer to radiochemotherapy in a Chinese population. *Thorac Cancer.* 2015; 6:741–8.  
<https://doi.org/10.1111/1759-7714.12251>  
PMID:[26557912](https://pubmed.ncbi.nlm.nih.gov/26557912/)
31. Salem ME, Puccini A, Xiu J, Raghavan D, Lenz HJ, Korn WM, Shields AF, Philip PA, Marshall JL, Goldberg RM. Comparative Molecular Analyses of Esophageal Squamous Cell Carcinoma, Esophageal Adenocarcinoma, and Gastric Adenocarcinoma. *Oncologist.* 2018; 23:1319–27.  
<https://doi.org/10.1634/theoncologist.2018-0143>  
PMID:[29866946](https://pubmed.ncbi.nlm.nih.gov/29866946/)
32. Salem ME, Puccini A, Grothey A, Raghavan D, Goldberg RM, Xiu J, Korn WM, Weinberg BA, Hwang JJ, Shields AF, Marshall JL, Philip PA, Lenz HJ. Landscape of Tumor Mutation Load, Mismatch Repair Deficiency, and PD-L1 Expression in a Large Patient Cohort of Gastrointestinal Cancers. *Mol Cancer Res.* 2018; 16:805–12.  
<https://doi.org/10.1158/1541-7786.MCR-17-0735>  
PMID:[29523759](https://pubmed.ncbi.nlm.nih.gov/29523759/)
33. Jin Z, Yoon HH. The promise of PD-1 inhibitors in gastro-esophageal cancers: microsatellite instability vs. PD-L1. *J Gastrointest Oncol.* 2016; 7:771–88.  
<https://doi.org/10.21037/jgo.2016.08.06>  
PMID:[27747091](https://pubmed.ncbi.nlm.nih.gov/27747091/)
34. Gao Y, Guo W, Geng X, Zhang Y, Zhang G, Qiu B, Tan F, Xue Q, Gao S, He J. Prognostic value of tumor-infiltrating lymphocytes in esophageal cancer: an updated meta-analysis of 30 studies with 5,122 patients. *Ann Transl Med.* 2020; 8:822.  
<https://doi.org/10.21037/atm-20-151>  
PMID:[32793667](https://pubmed.ncbi.nlm.nih.gov/32793667/)
35. Jiang D, Song Q, Wei X, Yu Z, Liu Y, Wang H, Wang X, Huang J, Su J, Hong Y, Xu Y, Xu C, Hou Y. PMS2 Expression With Combination of PD-L1 and TILs for Predicting Survival of Esophageal Squamous Cell Carcinoma. *Front Oncol.* 2022; 12:897527.  
<https://doi.org/10.3389/fonc.2022.897527>  
PMID:[35865481](https://pubmed.ncbi.nlm.nih.gov/35865481/)
36. Kim HI, Kim SY, Yu JE, Shin SJ, Roh YH, Cheong JH, Hyung WJ, Noh SH, Park CG, Lee HJ. Contrasting Prognostic Effects of Tumor-Infiltrating Lymphocyte Density in Cardia and Non-cardia Gastric Adenocarcinomas. *J Gastric Cancer.* 2020; 20:190–201.  
<https://doi.org/10.5230/jgc.2020.20.e21>  
PMID:[32596002](https://pubmed.ncbi.nlm.nih.gov/32596002/)

37. Zou W, Zhou ML, Zhang LY, Yang JN, Yang W, Wang YQ, Yi YX, Li GC, Zhang Z. Immune Score Predicts Outcomes of Gastric Cancer Patients Treated with Adjuvant Chemoradiotherapy. *J Oncol*. 2021; 2021:9344124.  
<https://doi.org/10.1155/2021/9344124>  
PMID:[34987582](https://pubmed.ncbi.nlm.nih.gov/34987582/)
38. Van Bockstal MR, François A, Altinay S, Arnould L, Balkenhol M, Broeckx G, Burguès O, Colpaert C, Dedeurwaerdere F, Dessauvage B, Duwel V, Floris G, Fox S, et al. Interobserver variability in the assessment of stromal tumor-infiltrating lymphocytes (sTILs) in triple-negative invasive breast carcinoma influences the association with pathological complete response: the IVITA study. *Mod Pathol*. 2021; 34:2130–40.  
<https://doi.org/10.1038/s41379-021-00865-z>  
PMID:[34218258](https://pubmed.ncbi.nlm.nih.gov/34218258/)
39. Knief J, Lazar-Karsten P, Wellner U, Hummel R, Thorns C. Density of CD8-positive tumor-infiltrating T-lymphocytes is an independent prognostic factor in adenocarcinoma of the esophagogastric junction. *Histol Histopathol*. 2019; 34:1121–9.  
<https://doi.org/10.14670/HH-18-109>  
PMID:[30938449](https://pubmed.ncbi.nlm.nih.gov/30938449/)
40. Knief J, Reddemann K, Petrova E, Herhahn T, Wellner U, Thorns C. High Density of Tumor-infiltrating B-Lymphocytes and Plasma Cells Signifies Prolonged Overall Survival in Adenocarcinoma of the Esophagogastric Junction. *Anticancer Res*. 2016; 36:5339–45.  
<https://doi.org/10.21873/anticanres.11107>  
PMID:[27798897](https://pubmed.ncbi.nlm.nih.gov/27798897/)
41. Fukuoka E, Yamashita K, Tanaka T, Sawada R, Sugita Y, Arimoto A, Fujita M, Takiguchi G, Matsuda T, Oshikiri T, Nakamura T, Suzuki S, Kakeji Y. Neoadjuvant Chemotherapy Increases PD-L1 Expression and CD8<sup>+</sup> Tumor-infiltrating Lymphocytes in Esophageal Squamous Cell Carcinoma. *Anticancer Res*. 2019; 39:4539–48.  
<https://doi.org/10.21873/anticanres.13631>  
PMID:[31366557](https://pubmed.ncbi.nlm.nih.gov/31366557/)
42. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst*. 2005; 97:142–6.  
<https://doi.org/10.1093/jnci/dji024>  
PMID:[15657344](https://pubmed.ncbi.nlm.nih.gov/15657344/)
43. Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. *Ann Oncol*. 2012; 23:3155–62.  
<https://doi.org/10.1093/annonc/mds181>  
PMID:[22847812](https://pubmed.ncbi.nlm.nih.gov/22847812/)
44. Jiang D, Liu Y, Wang H, Wang H, Song Q, Sujie A, Huang J, Xu Y, Zeng H, Tan L, Hou Y, Xu C. Tumour infiltrating lymphocytes correlate with improved survival in patients with esophageal squamous cell carcinoma. *Sci Rep*. 2017; 7:44823.  
<https://doi.org/10.1038/srep44823>  
PMID:[28322245](https://pubmed.ncbi.nlm.nih.gov/28322245/)
45. Liu K, Yang K, Wu B, Chen H, Chen X, Chen X, Jiang L, Ye F, He D, Lu Z, Xue L, Zhang W, Li Q, et al. Tumor-Infiltrating Immune Cells Are Associated With Prognosis of Gastric Cancer. *Medicine (Baltimore)*. 2015; 94:e1631.  
<https://doi.org/10.1097/MD.0000000000001631>  
PMID:[26426650](https://pubmed.ncbi.nlm.nih.gov/26426650/)
46. Suo C, Chen H, Binczyk F, Zhao R, Fan J, Yang X, Yuan Z, Kreil D, Łabaj P, Zhang T, Lu M, Jin L, Polańska J, et al. Tumor infiltrating lymphocyte signature is associated with single nucleotide polymorphisms and predicts survival in esophageal squamous cell carcinoma patients. *Aging (Albany NY)*. 2021; 13:10369–86.  
<https://doi.org/10.18632/aging.202798>  
PMID:[33819921](https://pubmed.ncbi.nlm.nih.gov/33819921/)
47. Wu ZD, Wang JC, Zhang HJ, Lin JQ, Wu ZZ, Zheng XW, Chen G. [Prognostic values of stromal tumor-infiltrating lymphocytes and lymphocyte-activation gene-3 in advanced gastric adenocarcinoma]. *Zhonghua Bing Li Xue Za Zhi*. 2022; 51:1007–12.  
<https://doi.org/10.3760/cma.j.cn112151-20220412-00273>  
PMID:[36207914](https://pubmed.ncbi.nlm.nih.gov/36207914/)
48. Gao P, Cai N, Yang X, Yuan Z, Zhang T, Lu M, Jin L, Ye W, Suo C, Chen X. Association of Helicobacter pylori and gastric atrophy with adenocarcinoma of the esophagogastric junction in Taixing, China. *Int J Cancer*. 2022; 150:243–52.  
<https://doi.org/10.1002/ijc.33801>  
PMID:[34498732](https://pubmed.ncbi.nlm.nih.gov/34498732/)
49. Yang X, Chen X, Zhuang M, Yuan Z, Nie S, Lu M, Jin L, Ye W. Smoking and alcohol drinking in relation to the risk of esophageal squamous cell carcinoma: A population-based case-control study in China. *Sci Rep*. 2017; 7:17249.  
<https://doi.org/10.1038/s41598-017-17617-2>  
PMID:[29222520](https://pubmed.ncbi.nlm.nih.gov/29222520/)
50. Zhang T, Yang X, Yin X, Yuan Z, Chen H, Jin L, Chen X, Lu M, Ye W. Poor oral hygiene behavior is associated with an increased risk of gastric cancer: A population-based case-control study in China. *J Periodontol*. 2022; 93:988–1002.  
<https://doi.org/10.1002/JPER.21-0301>  
PMID:[34599526](https://pubmed.ncbi.nlm.nih.gov/34599526/)

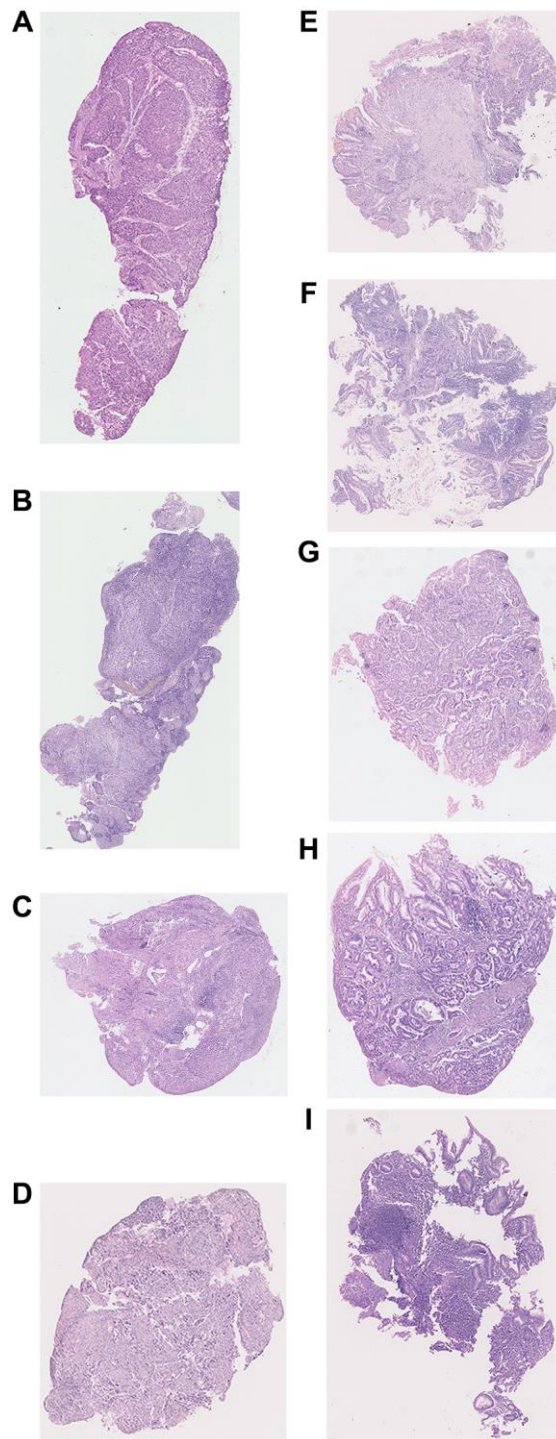
51. Dong H, Yao L, Fan J, Yang X, Yuan Z, Zhang T, Lu M, Chen X, Suo C. Development and application of a tool for automatically evaluating tumor lymphocytes infiltration in esophagogastric cancers. *J Mod Oncol*. 2024; 32:870–7.
52. Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin SF, Schwarz RF, Curtis C, Dunning MJ, Bardwell H, Johnson N, Doyle S, Turashvili G, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med*. 2012; 4:157ra143.  
<https://doi.org/10.1126/scitranslmed.3004330>  
PMID:[23100629](https://pubmed.ncbi.nlm.nih.gov/23100629/)

# SUPPLEMENTARY MATERIALS

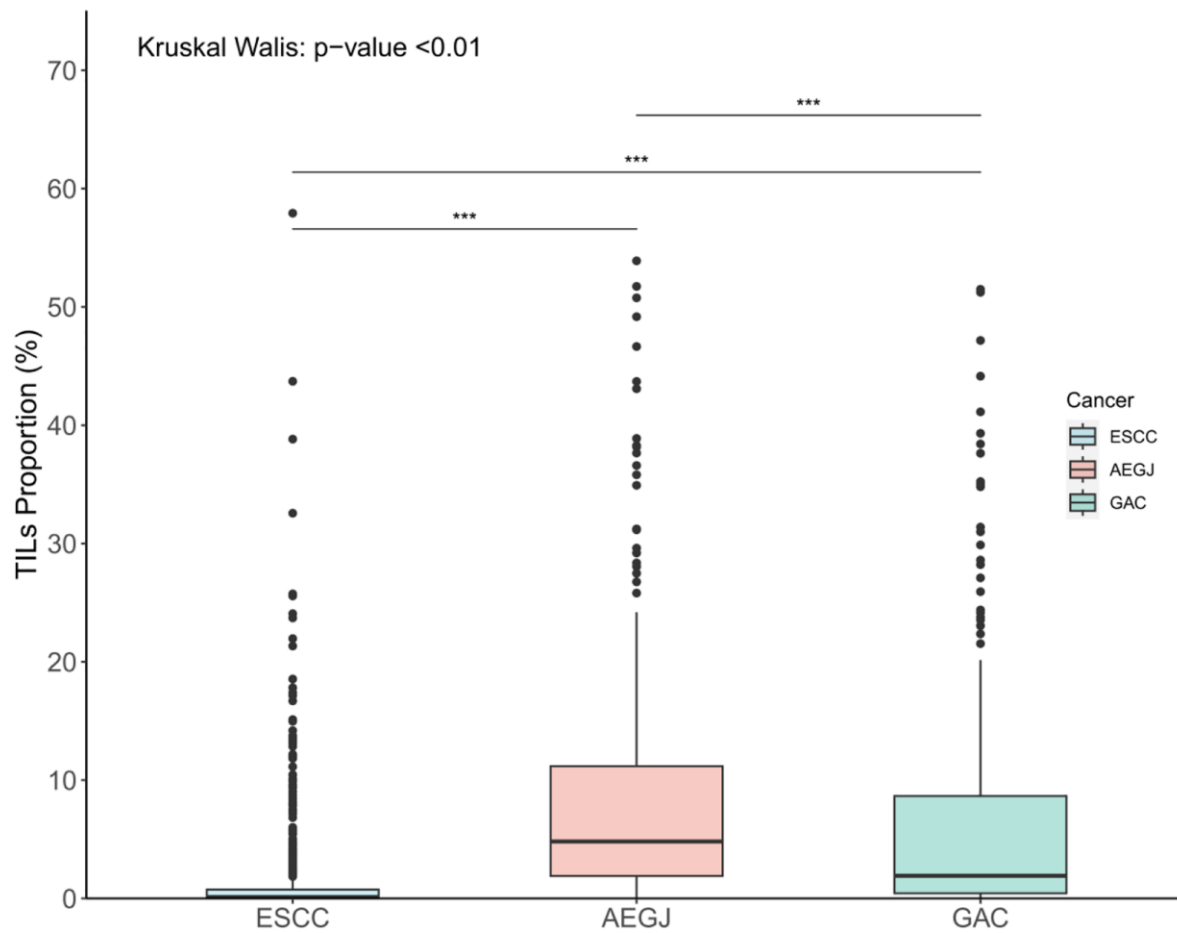
## Supplementary Figures



**Supplementary Figure 1. Verifications of the cell training sets.** (A, B) Cell proportions obtained by automated image analysis were compared to a pathologist's counts for a total of 10,000 single cells in the two representative sets of 20 ESCC and 20 GAC tissue samples. (C, D) TIL proportions versus manual grading of TIL infiltration in random one-third samples of ESCC and GAC.



**Supplementary Figure 2. The raw images of TILs intensity examples in ESCC, AEGJ, and GAC H&E-stained tissue sections (H&E×200).** (A–C) The H&E-stained tumor tissue sections with low TILs grade (0.00%, 0.02%), medium TILs grade (0.02%, 0.75%), and high TILs grade (0.75%, 57.92%) infiltration in ESCC. (D–F) The H&E-stained tumor tissue sections with low TILs grade (0.01%, 1.87%), medium TILs grade (1.87%, 11.14%), and high TILs grade (11.14%, 53.89%) infiltration in AEGJ. (G–I) The H&E-stained tumor tissue sections with low TILs grade (0.00%-0.43%), medium TILs grade (0.43%, 8.54%), and high TILs grade (8.54%, 75.05%) infiltration in GAC.



Supplementary Figure 3. Comparison of the TILs proportions in 752 ESCC, 214 AEGJ, and 256 GAC cases without stratification.

## Supplementary Tables

**Supplementary Table 1. Classification accuracy by 2-fold cross-validation for cancer cells, lymphocytes, and stromal cells based on the AEGJ, GAC, and ESCC training and testing sets.**

Classes	SVM Prediction (Training set)				SVM Prediction (Testing set)			
	Cancer	Lymphocyte	Stromal	Recall (%)	Cancer	Lymphocyte	Stromal	Recall (%)
<b>AEGJ</b>								
Cancer	426	28	34	87.30	463	14	35	90.43
Lymphocyte	23	425	7	93.41	18	408	3	95.78
Stromal	58	4	187	75.10	38	4	209	83.27
Precision (%)	84.02	93.00	82.02		89.21	95.77	84.62	
<b>GAC</b>								
Cancer	530	1	4	99.07	523	3	5	98.40
Lymphocyte	9	748	1	98.68	4	729	0	99.45
Stromal	12	0	189	94.03	10	0	220	95.65
Precision (%)	96.19	99.87	97.42		97.39	99.59	97.78	
<b>ESCC</b>								
Cancer	410	4	98	80.08	438	4	91	82.18
Lymphocyte	3	166	48	76.50	4	167	34	81.46
Stromal	11	7	594	97.06	19	15	569	94.36
Precision (%)	96.70	93.79	80.27		95.01	89.78	81.99	

The confusion matrix of the real classes of the cells (columns) and the predicted classes (rows) are shown, where Precision was calculated by true positives/(true positives + false positives) and Recall was calculated by true positives/(true positives + false negatives).

**Supplementary Table 2. The comparison of TILs proportion in the 214 AEGJ, 256 GAC, and 752 ESCC cases.**

Variables	TILs proportion, Median (IQR)			P-value
	AEGJ (%)	GAC (%)	ESCC (%)	
<b>All patients</b>	4.82 (1.90, 11.18)	1.92 (0.44, 8.66)	0.12 (0.02, 0.75)	<0.001
<b>Age</b>				
<60	3.82 (1.57, 11.05)	1.86 (0.16, 9.16)	0.15 (0.03, 0.92)	<0.001
≥60	4.88 (2.01, 11.21)	1.93 (0.47, 8.16)	0.12 (0.02, 0.73)	<0.001
<b>Sex</b>				
Man	4.55 (1.71, 10.80)	2.38 (0.52, 9.14)	0.12 (0.02, 0.67)	<0.001
Woman	5.19 (2.41, 12.45)	1.27 (0.30, 6.54)	0.13 (0.02, 1.10)	<0.001
<b>BMI</b>				
<18.5	5.02 (1.32, 8.99)	1.83 (0.33, 6.88)	0.08 (0.02, 0.56)	<0.001
18.5-24	4.56 (1.99, 11.21)	2.34 (0.51, 8.08)	0.11 (0.02, 0.62)	<0.001
≥24	4.97 (2.09, 10.82)	1.86 (0.43, 9.15)	0.23 (0.02, 1.20)	<0.001
<b>Tea drinking</b>				
Never	4.65 (1.56, 11.01)	1.51 (0.29, 7.93)	0.12 (0.02, 0.72)	<0.001
Ever	5.36 (3.02, 11.11)	2.74 (1.14, 10.47)	0.12 (0.02, 0.75)	<0.001
<b>Times of tooth brushing daily</b>				
<2	4.94 (2.03, 11.21)	2.18 (0.46, 9.09)	0.12 (0.02, 0.75)	<0.001
≥2	3.98 (1.29, 6.62)	1.79 (0.33, 8.08)	0.12 (0.02, 0.75)	<0.001



<b>Wealth scores</b>					
Q1	4.99 (2.47, 11.56)	2.13 (0.37, 8.91)	0.12 (0.02, 0.69)	<0.001	
Q2	2.94 (1.36, 5.60)	1.27 (0.36, 4.56)	0.13 (0.02, 0.96)	<0.001	
Q3	5.22 (2.24, 13.93)	2.55 (0.38, 2.55)	0.09 (0.01, 0.45)	<0.001	
Q4	4.26 (1.23, 10.69)	2.38 (0.67, 2.38)	0.13 (0.02, 0.51)	<0.001	
Q5	5.71 (3.41, 13.55)	2.41 (0.93, 4.74)	0.25 (0.03, 2.05)	<0.001	
<b><i>Helicobacter pylori</i></b>					
HP+	4.62 (1.84, 11.24)	2.28 (0.45, 9.66)	0.13 (0.02, 0.77)	<0.001	
HP-	5.24 (2.99, 11.10)	1.83 (0.39, 6.87)	0.12 (0.02, 0.60)	<0.001	
<b>Gastric atrophy</b>					
Yes	4.62 (1.84, 11.24)	2.28 (0.45, 9.66)	0.13 (0.02, 0.77)	<0.001	
No	5.24 (2.99, 11.10)	1.83 (0.39, 6.87)	0.12 (0.02, 0.60)	<0.001	
<b>First-line treatment method</b>					
Radiotherapy	2.77 (0.40, 7.72)	0.82 (0.27, 2.74)	0.09 (0.02, 0.68)	<0.001	
Chemotherapy	4.10 (1.56, 5.30)	1.32 (0.30, 13.71)	0.10 (0.02, 0.48)	<0.001	
Surgery	5.72 (3.14, 11.94)	2.78 (0.73, 11.12)	0.15 (0.03, 0.80)	<0.001	
Combination therapy	4.64 (1.81, 10.42)	1.32 (0.35, 3.21)	0.10 (0.02, 0.92)	<0.001	
Untreated	1.83 (1.05, 3.49)	0.66 (0.18, 1.61)	0.03 (0.02, 0.43)	0.002	
<b>TNM staging</b>					
0+I+II	6.92 (3.12, 13.39)	2.64 (0.49, 10.91)	0.14 (0.02, 0.76)	<0.001	
III+IV	5.34 (2.99, 11.56)	2.47 (0.74, 11.12)	0.15 (0.02, 0.92)	<0.001	
<b>Grade of differentiation</b>					
Gx Grading cannot be evaluated	4.04 (1.77, 12.42)	2.38 (0.82, 10.20)	0.16 (0.02, 1.30)	<0.001	
G1 Highly differentiated	8.67 (2.55, 14.28)	1.60 (0.68, 10.19)	0.13 (0.03, 0.69)	0.008	
G2 Medium differentiation	7.72 (3.89, 13.39)	2.78 (1.13, 9.15)	0.13 (0.02, 0.84)	<0.001	
G3 Poorly differentiated	7.17 (1.97, 12.81)	2.10 (0.26, 11.47)	0.11 (0.02, 0.56)	<0.001	
G4 Undifferentiated	4.34 (2.42, 5.17)	0.85 (0.35, 3.50)	0.13 (0.05, 0.35)	<0.001	

**Supplementary Table 3. Univariable and multivariable Cox regression analyses of basic characteristics with OS in GAC of Taixing dataset (Discovery, *N* = 256).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			<i>P</i> -value	HR (95% CI)	<i>P</i> -value	aHR <sup>a</sup> (95% CI)
<b>TILs, Median (IQR)</b>	1.92 (0.44, 8.66)		<b>0.030</b>	0.979 (0.961, 0.998)	<b>0.016</b>	0.972 (0.949, 0.995)
<b>Age, mean (range)</b>	67.52 (41–85)		<b>0.023</b>	1.022 (1.003, 1.041)	<b>0.919</b>	1.001 (0.977, 1.026)
<b>Sex</b>						
Woman	67 (26.17)	2.16		Ref.		Ref.
Man	189 (73.83)	2.65	0.061	0.711 (0.500, 1.016)	0.574	0.854 (0.493, 1.479)
<b>Marriage</b>						
Divorce/widow	61 (23.83)	2.56		Ref.		Ref.
Unmarried	9 (3.52)	0.98	0.103	1.955 (0.874, 4.375)	0.390	1.790 (0.475, 6.746)
Married	186 (72.66)	2.71	0.108	0.740 (0.512, 1.068)	0.665	1.135 (0.640, 2.011)
<b>Educational level</b>						
High school and above	18 (7.03)	3.47		Ref.		Ref.
Primary or Secondary school	158 (61.72)	2.64	0.095	2.038 (0.883, 4.705)	<b>0.027</b>	3.871 (1.163, 12.881)
Illiteracy	80 (31.25)	2.16	<b>0.027</b>	2.602 (1.113, 6.079)	0.077	3.272 (0.881, 12.148)

<b>Cigarette smoking</b>							
Never	103 (40.23)	2.16		Ref.			Ref.
Ever or still	144 (56.25)	2.73	0.077	0.737 (0.525, 1.034)	0.497		0.776 (0.373, 1.614)
Missing	9 (3.52)						
<b>Alcohol drinking</b>							
Never	125 (48.83)	2.64		Ref.			Ref.
Ever or still	122 (47.66)	2.62	0.536	0.899 (0.642, 1.259)	0.552		1.189 (0.673, 2.101)
Missing	9 (3.52)						
<b>Tea drinking</b>							
Never	170 (66.41)	2.64		Ref.			Ref.
Ever	77 (30.08)	2.48	0.856	0.967 (0.672, 1.391)	0.070		1.671 (0.960, 2.909)
Missing	9 (3.52)						
<b>BMI</b>							
≥24	66 (25.78)	3.01		Ref.			Ref.
18.5–24	158 (61.72)	2.55	0.434	1.176 (0.783, 1.766)	0.489		1.230 (0.684, 2.211)
<18.5	32 (12.50)	2.25	0.107	1.564 (0.908, 2.695)	0.103		1.938 (0.875, 4.294)
Missing	0 (0.00)						
<b>Wealth scores</b>							
Q5	40 (15.63)	2.64		Ref.			Ref.
Q4	47 (18.36)	2.68	0.448	0.792 (0.434, 1.447)	0.903		1.056 (0.438, 2.548)
Q3	62 (24.42)	3.14	0.453	0.810 (0.468, 1.403)	0.850		1.081 (0.484, 2.415)
Q2	55 (21.48)	1.99	0.508	1.202 (0.697, 2.070)	0.532		1.296 (0.574, 2.926)
Q1	52 (20.31)	2.40	0.321	1.309 (0.769, 2.227)	0.990		1.005 (0.461, 2.188)
<b>First-line treatment method</b>							
Untreated	11 (4.30)	0.81		Ref.			Ref.
Combination therapy	57 (22.27)	2.32	<b>0.003</b>	0.344 (0.169, 0.698)	<b>0.007</b>		0.175 (0.049, 0.620)
Surgery	151 (58.98)	3.00	<b>&lt;0.001</b>	0.193 (0.098, 0.380)	<b>&lt;0.001</b>		0.140 (0.043, 0.451)
Chemotherapy	27 (10.55)	0.86	0.658	0.844 (0.399, 1.786)	0.241		0.472 (0.135, 1.656)
Radiotherapy	9 (3.52)	1.21	0.413	0.681 (0.271, 1.709)	0.778		0.817 (0.202, 3.314)
Missing	1 (0.39)						
<b>Grade of differentiation</b>							
G1 Highly differentiated	7 (2.73)	3.59		Ref.			Ref.
G2 Medium differentiation	63 (24.61)	3.14	0.762	0.795 (0.181, 3.503)	0.799		0.822 (0.183, 3.694)
G3 Poorly differentiated	48 (18.75)	2.40	0.162	2.786 (0.663, 11.711)	0.162		2.835 (0.657, 12.237)
G4 Undifferentiated	9 (3.52)	3.48	0.320	2.300 (0.446, 11.868)	0.435		1.979 (0.357, 10.977)
Gx Grading cannot be evaluated	47 (18.36)	1.99	0.071	3.733 (0.893, 15.603)	0.242		2.467 (0.543, 11.209)
Missing	82 (32.03)						
<b>Helicobacter pylori</b>							
HP–	68 (26.56)	2.27		Ref.			Ref.
HP+	181 (70.70)	2.65	0.138	0.761 (0.531, 1.092)	0.543		0.836 (0.470, 1.488)
Missing	7 (2.73)						
<b>Gastric atrophy</b>							
No	157 (61.33)	2.71		Ref.			Ref.
Yes	51 (19.92)	2.10	<b>0.017</b>	1.623(1.091, 2.414)	<b>0.006</b>		2.183 (1.246, 3.826)
Missing	48 (18.75)						

<sup>a</sup>aHR with adjustment for TILs, age, sex, grade of differentiation, first-line treatment method, BMI.

**Supplementary Table 4. Univariable and multivariable Cox regression analyses of basic characteristics with OS in ESCC of Taixing dataset (Discovery, *N* = 752).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			<i>P</i> -value	HR (95% CI)	<i>P</i> -value	aHR <sup>a</sup> (95% CI)
<b>TILs, Median (IQR)</b>	0.12 (0.02, 0.75)		<b>0.008</b>	0.960 (0.931, 0.990)	<b>0.032</b>	0.967 (0.938, 0.997)
<b>Age, mean (range)</b>	66.96 (42–85)		<b>&lt;0.001</b>	1.033 (1.021, 1.045)	<b>&lt;0.001</b>	1.022 (1.010, 1.035)
<b>Sex</b>						
Woman	260 (34.57)	2.62		Ref.		Ref.
Man	492 (65.43)	2.35	<b>0.107</b>	1.180 (0.965, 1.444)	<b>0.020</b>	1.289 (1.040, 1.599)
<b>Marriage</b>						
Divorce/widow	136 (18.09)	2.54		Ref.		Ref.
Unmarried	25 (3.32)	1.31	<b>0.003</b>	2.105 (1.286, 3.445)	<b>0.037</b>	1.798 (1.036, 3.121)
Married	591 (78.59)	2.46	0.938	0.990 (0.773, 1.269)	0.346	1.137 (0.870, 1.486)
<b>Educational level</b>						
High school and above	47 (6.25)	2.43		Ref.		Ref.
Primary or Secondary school	429 (57.05)	2.61	0.796	1.057 (0.694, 1.609)	0.880	1.034 (0.669, 1.598)
Illiteracy	276 (36.70)	2.06	0.135	1.385 (0.903, 2.122)	0.190	1.374 (0.854, 2.210)
<b>Cigarette smoking</b>						
Never	313 (41.62)	2.59		Ref.		Ref.
Ever or still	405 (53.86)	2.40	0.600	1.053 (0.867, 1.280)	0.423	0.884 (0.653, 1.120)
Missing	34 (4.52)					
<b>Alcohol drinking</b>						
Never	351 (46.68)	2.43		Ref.		Ref.
Ever or still	366 (48.67)	2.45	0.523	0.939 (0.775, 1.139)	0.208	0.849 (0.658, 1.096)
Missing	35 (4.65)					
<b>Tea drinking</b>						
Never	493 (65.56)	2.51		Ref.		Ref.
Ever	225 (29.92)	2.33	0.483	1.076 (0.876, 1.322)	0.181	1.180 (0.926, 1.504)
Missing	34 (4.52)					
<b>BMI</b>						
≥24	201 (26.73)	2.77		Ref.		Ref.
18.5-24	466 (61.97)	2.22	<b>0.007</b>	1.377 (1.093, 1.735)	0.058	1.259 (0.993, 1.598)
<18.5	85 (11.30)	1.75	<b>0.001</b>	1.771 (1.281, 2.449)	0.063	1.378 (0.983, 1.932)
Missing	0 (0.00)					
<b>Wealth scores</b>						
Q5	62 (8.24)	2.67		Ref.		Ref.
Q4	138 (18.35)	2.63	0.232	1.262 (0.862, 1.848)	0.287	1.235 (0.837, 1.822)
Q3	161 (21.41)	2.49	0.240	1.253 (0.860, 1.826)	0.580	1.114 (0.760, 1.635)
Q2	139 (18.48)	2.31	0.158	1.318 (0.898, 1.935)	0.112	1.373 (0.928, 2.029)
Q1	232 (30.85)	2.18	<b>0.030</b>	1.480 (1.039, 2.109)	0.254	1.235 (0.859, 1.776)
<b>First-line treatment method</b>						
Untreated	21 (2.79)	0.89		Ref.		Ref.
Radiotherapy	88 (11.70)	1.68	<b>&lt;0.001</b>	0.328 (0.199, 0.540)	<b>0.001</b>	0.391 (0.224, 0.683)
Chemotherapy	93 (12.37)	1.49	<b>&lt;0.001</b>	0.368 (0.225, 0.602)	<b>0.004</b>	0.448 (0.259, 0.775)
Surgery	377 (50.13)	2.88	<b>&lt;0.001</b>	0.166 (0.105, 0.263)	<b>&lt;0.001</b>	0.236 (0.140, 0.399)
Combination therapy	170 (22.61)	1.84	<b>&lt;0.001</b>	0.291 (0.182, 0.466)	<b>0.001</b>	0.400 (0.236, 0.678)
Missing	3 (0.40)					

<b>Grade of differentiation</b>							
G1 Highly differentiated	60 (7.98)	2.85		Ref.			Ref.
G2 Medium differentiation	452 (60.11)	2.61	0.091	1.426 (0.945, 2.152)	0.142		1.366 (0.9005, 2.073)
G3 Poorly differentiated	134 (17.82)	2.05	<b>0.003</b>	1.966 (1.260, 3.069)	0.101		1.463 (0.929, 2.303)
G4 Undifferentiated	45 (5.98)	1.51	<b>0.003</b>	2.215 (1.302, 3.770)	0.087		1.607 (0.933, 2.768)
Gx Grading cannot be evaluated	40 (5.32)	1.68	<b>0.003</b>	2.250 (1.317, 3.845)	<b>0.031</b>		1.814 (1.055, 3.117)
Missing	21 (2.79)						
<b>Helicobacter pylori</b>							
HP-	234 (31.12)	2.45		Ref.			Ref.
HP+	484 (64.36)	2.38	0.271	1.124 (0.913, 1.383)	0.355		1.108 (0.892, 1.376)
Missing	34 (4.52)						
<b>Gastric atrophy</b>							
No	520 (69.15)	2.47		Ref.			Ref.
Yes	98 (13.03)	2.31	0.635	1.070 (0.808, 1.418)	0.468		1.114 (0.832, 1.490)
Missing	134 (17.82)						

<sup>a</sup>aHR with adjustment for TILs, age, sex, grade of differentiation, first-line treatment method, BMI.

**Supplementary Table 5. Univariable and multivariable Cox regression analyses of basic characteristics with OS in AEGJ of Taixing dataset (Discovery, N = 117).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			P-value	HR (95% CI)	P-value	aHR <sup>a</sup> (95% CI)
<b>TILs, Median (IQR)</b>	5.42 (3.04, 12.86)		<b>0.009</b>	0.957 (0.926, 0.989)	<b>0.009</b>	0.946 (0.907, 0.986)
<b>Age, mean (range)</b>	69 (49–84)		0.546	1.010 (0.978, 1.044)	0.553	1.012 (0.973, 1.053)
<b>Sex</b>						
Woman	35 (29.91)	2.56		Ref.		Ref.
Man	82 (70.09)	2.82	0.628	0.874 (0.507, 1.506)	0.605	0.849 (0.458, 1.577)
<b>Marriage</b>						
Divorce/widow	28 (23.93)	3.05		Ref.		Ref.
Unmarried	9 (7.69)	2.56	0.579	0.703 (0.202, 2.448)	0.073	0.149 (0.019, 1.194)
Married	80 (68.38)	2.71	0.600	1.175 (0.643, 2.147)	0.712	0.857 (0.379, 1.938)
<b>Educational level</b>						
High school and above	6 (5.13)	2.59		Ref.		Ref.
Primary or Secondary school	66 (56.41)	2.83	0.285	0.597 (0.232, 1.537)	0.407	0.637 (0.219, 1.851)
Illiteracy	45 (38.46)	2.70	0.520	0.729 (0.278, 1.909)	0.608	0.720 (0.205, 2.524)
<b>Cigarette smoking</b>						
Never	45 (38.46)	3.26		Ref.		Ref.
Ever or still	70 (59.83)	2.53	0.331	1.298 (0.767, 2.198)	0.322	1.364 (0.738, 2.522)
Missing	2 (1.71)					
<b>Alcohol drinking</b>						
Never	70 (59.83)	3.11		Ref.		Ref.
Ever or still	45 (38.46)	2.53	0.790	1.073 (0.637, 1.809)	0.755	0.912 (0.511, 1.628)
Missing	2 (1.71)					
<b>Tea drinking</b>						
Never	88 (75.21)	2.84		Ref.		Ref.
Ever	27 (23.08)	2.17	0.480	1.235 (0.688, 2.218)	0.822	1.079 (0.555, 2.101)
Missing	2 (1.71)					

<b>BMI</b>							
≥24	27 (23.08)	2.93		Ref.		Ref.	
18.5-24	78 (66.67)	2.75	0.792	0.925 (0.519, 1.649)	0.434	0.841 (0.423, 1.674)	
<18.5	11 (9.40)	2.70	0.493	0.681 (0.228, 2.040)	0.435	0.625 (0.193, 2.030)	
Missing	1 (0.85)						
<b>Wealth scores</b>							
Q5	14 (11.97)	2.98		Ref.		Ref.	
Q4	24 (20.51)	2.25	0.391	1.521 (0.584, 3.963)	0.636	0.769 (0.260, 2.279)	
Q3	24 (20.51)	2.60	0.572	1.322 (0.502, 3.481)	0.959	0.972 (0.327, 2.894)	
Q2	22 (18.80)	3.67	0.621	1.277 (0.485, 3.365)	0.893	0.931 (0.328, 2.642)	
Q1	33 (28.21)	2.88	0.987	1.008 (0.391, 2.600)	0.661	0.792 (0.279, 2.247)	
<b>First-line treatment method</b>							
Untreated	1 (0.85)	1.58		Ref.		Ref.	
Radiotherapy	4 (3.42)	1.16	0.744	0.684 (0.070, 6.676)	0.969	0.946 (0.059, 15.110)	
Chemotherapy	4 (3.42)	2.08	0.859	0.819 (0.091, 7.400)	0.990	1.015 (0.108, 9.568)	
Surgery	82 (70.09)	2.82	0.226	0.291 (0.039, 2.150)	0.900	1.153 (0.125, 10.65)	
Combination therapy	26 (22.22)	3.03	0.102	0.176 (0.022, 1.415)	0.731	0.663 (0.064, 6.916)	
Missing	0 (0.00)						
<b>TNM staging</b>							
0+I+II	55 (47.01)	2.94		Ref.		Ref.	
III+IV	62 (52.99)	2.14	<b>0.027</b>	1.795 (1.068, 3.014)	0.200	1.462 (0.813, 2.611)	
<b>Grade of differentiation</b>							
G1 Highly differentiated	4 (3.42)	2.19		Ref.		Ref.	
G2 Medium differentiation	61 (52.14)	3.00	0.574	0.661 (0.156, 2.797)	0.771	0.804 (0.184, 3.512)	
G3 Poorly differentiated	26 (22.22)	2.73	0.979	0.980 (0.223, 4.319)	0.919	1.082 (0.234, 4.998)	
G4 Undifferentiated	8 (6.84)	0.68	0.087	3.971 (0.819, 19.250)	0.100	4.242 (0.757, 23.762)	
Gx Grading cannot be evaluated	14 (11.97)	1.70	0.378	1.971 (0.436, 8.907)	0.365	2.128 (0.416, 10.891)	
Missing	4 (3.42)						
<b>Helicobacter pylori</b>							
HP-	19 (16.24)	2.70		Ref.		Ref.	
HP+	95 (81.20)	2.80	0.987	0.994 (0.503, 1.964)	0.968	0.983 (0.419, 2.251)	
Missing	3 (2.56)						
<b>Gastric atrophy</b>							
No	70 (59.83)	2.66		Ref.		Ref.	
Yes	29 (24.79)	2.76	0.461	1.234 (0.706, 2.159)	0.786	0.914 (0.475, 1.755)	
Missing	18 (15.38)						

<sup>a</sup>aHR with adjustment for TILs, age, sex, grade of differentiation, TNM staging, first-line treatment method, BMI.

**Supplementary Table 6. Univariable and multivariable Cox regression analyses of basic characteristics with OS in GAC of Taixing dataset (Discovery, *N* = 148).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			<i>P</i> -value	HR (95% CI)	<i>P</i> -value	aHR <sup>a</sup> (95% CI)
TILs, Median (IQR)	2.55 (0.51, 11.06)		0.192	0.983 (0.958, 1.009)	<b>0.016</b>	0.961 (0.931, 0.993)
Age, mean (range)	65.22 (44–85)		0.853	1.003 (0.974, 1.032)	0.600	0.990 (0.955, 1.027)

<b>Sex</b>							
Woman	39 (26.35)	2.88		Ref.			Ref.
Man	109 (73.65)	2.75	0.982	0.993 (0.568, 1.736)	0.896	0.955 (0.483, 1.891)	
<b>Marriage</b>							
Divorce/widow	31 (20.95)	2.88		Ref.			Ref.
Unmarried	3 (2.03)	2.01	0.364	1.987 (0.452, 8.740)	0.321	0.319 (0.034, 3.046)	
Married	114 (77.03)	2.87	0.608	0.858 (0.478, 1.540)	0.766	1.131 (0.503, 2.546)	
<b>Educational level</b>							
High school and above	14 (9.46)	3.42		Ref.			Ref.
Primary or Secondary school	95 (64.19)	2.81	0.262	1.962 (0.605, 6.362)	<b>0.045</b>	4.728 (1.032, 21.662)	
Illiteracy	39 (26.35)	2.82	0.111	2.679 (0.798, 9.001)	<b>0.043</b>	5.534 (1.056, 28.992)	
<b>Cigarette smoking</b>							
Never	59 (39.86)	2.87		Ref.			Ref.
Ever or still	88 (59.46)	2.82	0.746	0.920 (0.555, 1.524)	0.101	0.496 (0.214, 1.148)	
Missing	1 (0.68)						
<b>Alcohol drinking</b>							
Never	75 (50.68)	2.98		Ref.			Ref.
Ever or still	72 (48.65)	2.64	0.492	1.191 (0.723, 1.961)	0.440	1.390 (0.602, 3.210)	
Missing	1 (0.68)						
<b>Tea drinking</b>							
Never	100 (67.57)	2.91		Ref.			Ref.
Ever	47 (31.76)	2.46	0.139	1.477 (0.882, 2.473)	0.299	1.475 (0.708, 3.070)	
Missing	1 (0.68)						
<b>BMI</b>							
≥24	43 (29.05)	3.14		Ref.			Ref.
18.5-24	90 (60.81)	2.64	0.513	1.215 (0.677, 2.181)	0.188	1.678 (0.777, 3.626)	
<18.5	15 (10.14)	2.75	0.369	1.477 (0.631, 3.454)	0.135	2.193 (0.783, 6.143)	
Missing	0 (0.00)						
<b>Wealth scores</b>							
Q5	22 (14.86)	2.62		Ref.			Ref.
Q4	30 (20.27)	2.97	0.479	0.733 (0.311, 1.731)	0.648	0.792 (0.291, 2.158)	
Q3	40 (27.03)	3.34	0.174	0.564 (0.246, 1.289)	0.187	0.518 (0.195, 1.376)	
Q2	30 (20.27)	2.44	0.744	0.869 (0.375, 2.015)	0.477	0.698 (0.259, 1.881)	
Q1	26 (17.57)	2.41	0.527	1.291 (0.585, 2.848)	0.416	0.668 (0.252, 1.768)	
<b>First-line treatment method</b>							
Untreated	2 (1.35)	1.19		Ref.			Ref.
Radiotherapy	3 (2.03)	0.98	0.997	1.004 (0.167, 6.039)	0.070	0.043 (0.001, 1.276)	
Chemotherapy	11 (7.43)	0.93	0.942	0.945 (0.206, 4.342)	<b>0.031</b>	0.026 (0.001, 0.723)	
Surgery	106 (7.16)	3.11	<b>0.004</b>	0.120 (0.028, 0.509)	<b>0.002</b>	0.007 (0.001, 0.161)	
Combination therapy	25 (16.89)	2.64	<b>0.028</b>	0.180 (0.039, 0.828)	<b>0.005</b>	0.010 (0.001, 0.253)	
Missing	1 (0.68)						
<b>TNM staging</b>							
0-I+II	98 (66.22)	3.27		Ref.			Ref.
III+IV	50 (33.78)	1.85	<b>&lt;0.001</b>	4.568 (2.733, 7.633)	<b>0.001</b>	3.115 (1.546, 6.278)	
<b>Grade of differentiation</b>							
G1 Highly differentiated	7 (4.73)	3.59		Ref.			Ref.
G2 Medium differentiation	60 (40.54)	3.14	0.755	0.789 (0.178, 3.501)	0.643	0.696 (0.151, 3.218)	

G3 Poorly differentiated	40 (27.03)	2.55	0.225	2.460 (0.575, 10.519)	0.673	1.407 (0.288, 6.878)
G4 Undifferentiated	7 (4.73)	3.48	0.315	2.460 (0.575, 10.519)	0.438	2.045 (0.335, 12.482)
Gx Grading cannot be evaluated	23 (15.54)	1.79	0.093	3.576 (0.809, 15.801)	0.922	1.089 (0.199, 5.957)
Missing	11 (7.43)					
<b><i>Helicobacter pylori</i></b>						
HP-	34 (22.97)	2.52		Ref.		Ref.
HP+	108 (72.97)	2.87	0.763	0.914 (0.510, 1.639)	0.721	1.141 (0.554, 2.349)
Missing	6 (4.05)					
<b>Gastric atrophy</b>						
No	90 (60.81)	3.14		Ref.		Ref.
Yes	30 (20.27)	2.10	<b>0.002</b>	2.547 (1.407, 4.609)	<b>0.047</b>	2.196 (1.011, 4.771)
Missing	28 (18.92)					

<sup>a</sup>aHR with adjustment for TILs, age, sex, grade of differentiation, TNM staging, first-line treatment method, BMI.

**Supplementary Table 7. Univariable and multivariable Cox regression analyses of basic characteristics with OS in ESCC of Taixing dataset (Discovery, N = 418).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			P-value	HR (95% CI)	P-value	aHR <sup>a</sup> (95% CI)
<b>TILs, Median (IQR)</b>	0.14 (0.02, 0.84)		<b>0.045</b>	0.958 (0.918, 0.999)	<b>0.041</b>	0.954 (0.911, 0.998)
<b>Age, mean (range)</b>	65.60 (42–85)		<b>0.003</b>	1.024 (1.008, 1.041)	<b>0.016</b>	1.021 (1.004, 1.039)
<b>Sex</b>						
Woman	152 (36.36)	2.89		Ref.		Ref.
Man	266 (63.64)	2.54	<b>0.023</b>	0.716 (0.537, 0.954)	<b>0.004</b>	1.574 (1.157, 2.140)
<b>Marriage</b>						
Divorce/widow	70 (16.75)	2.58		Ref.		Ref.
Unmarried	9 (2.15)	0.99	<b>0.019</b>	2.649 (1.172, 5.986)	<b>0.005</b>	3.765 (1.504, 9.424)
Married	339 (81.10)	2.74	0.697	1.076 (0.745, 1.553)	0.147	1.341 (0.902, 1.993)
<b>Educational level</b>						
High school and above	25 (5.98)	2.81		Ref.		Ref.
Primary or Secondary school	252 (60.29)	2.71	0.583	1.188 (0.642, 2.200)	0.787	1.091 (0.579, 2.055)
Illiteracy	141 (33.73)	2.61	0.416	1.300 (0.690, 2.449)	0.340	1.398 (0.702, 2.786)
<b>Cigarette smoking</b>						
Never	180 (43.06)	2.76		Ref.		Ref.
Ever or still	219 (52.39)	2.61	0.223	1.187 (0.901, 1.564)	0.179	0.745 (0.485, 1.144)
Missing	19 (4.55)					
<b>Alcohol drinking</b>						
Never	194 (46.41)	2.61		Ref.		Ref.
Ever or still	205 (49.04)	2.76	0.914	1.015 (0.773, 1.333)	0.094	0.735 (0.513, 1.053)
Missing	199 (4.55)					
<b>Tea drinking</b>						
Never	285 (68.18)	2.70		Ref.		Ref.
Ever	115 (27.51)	2.55	0.268	1.178 (0.881, 1.576)	0.749	1.058 (0.749, 1.493)
Missing	18 (4.31)					
<b>BMI</b>						
≥24	122 (29.19)	2.82		Ref.		Ref.
18.5–24	248 (59.33)	2.61	0.095	1.308 (0.955, 1.791)	0.564	0.564 (0.796, 1.521)

<18.5	48 (11.48)	2.09	<b>0.047</b>	1.579 (1.006, 2.479)	0.406	0.406 (0.765, 1.937)
Missing	0 (0.00)					
<b>Wealth scores</b>						
Q5	46 (11.00)	2.92		Ref.		Ref.
Q4	84 (20.10)	2.66	0.149	1.489 (0.867, 2.558)	0.332	1.316 (0.755, 2.295)
Q3	78 (18.66)	2.57	0.429	1.258 (0.712, 2.223)	0.918	1.031 (0.575, 1.849)
Q2	85 (20.33)	2.62	0.259	1.373 (0.791, 2.382)	0.201	1.445 (0.822, 2.542)
Q1	125 (29.90)	2.68	0.068	1.615 (0.964, 2.706)	0.245	1.373 (0.804, 2.343)
<b>First-line treatment method</b>						
Untreated	2 (0.47)	0.38		Ref.		Ref.
Radiotherapy	39 (9.33)	2.25	<b>&lt;0.001</b>	0.063 (0.014, 0.282)	<b>&lt;0.001</b>	0.059 (0.128, 0.274)
Chemotherapy	46 (11.00)	1.53	<b>&lt;0.001</b>	0.078 (0.018, 0.343)	<b>&lt;0.001</b>	0.063 (0.139, 0.283)
Surgery	250 (59.81)	2.89	<b>&lt;0.001</b>	0.040 (0.009, 0.171)	<b>&lt;0.001</b>	0.054 (0.125, 0.236)
Combination therapy	81 (19.38)	2.55	<b>&lt;0.001</b>	0.048 (0.011, 0.211)	<b>&lt;0.001</b>	0.059 (1.331, 0.263)
Missing	0 (0.00)					
<b>TNM staging</b>						
0+I+II	256 (61.24)	3.04		Ref.		Ref.
III+IV	162(38.76)	1.56	<b>&lt;0.001</b>	2.435 (1.863, 3.182)	<b>&lt;0.001</b>	2.273 (1.684, 3.070)
<b>Grade of differentiation</b>						
G1 Highly differentiated	39 (9.33)	3.13		Ref.		Ref.
G2 Medium differentiation	283 (67.70)	2.76	0.067	1.701 (0.963, 3.003)	0.115	1.596 (0.893, 2.852)
G3 Poorly differentiated	51 (12.20)	1.75	<b>0.003</b>	2.611 (1.373, 4.966)	0.131	1.671 (0.859, 3.251)
G4 Undifferentiated	34 (8.13)	2.32	<b>0.009</b>	2.505 (1.253, 5.010)	0.589	1.223 (0.589, 2.542)
Gx Grading cannot be evaluated	9 (2.15)	1.59	<b>0.011</b>	3.286 (1.309, 8.247)	0.121	2.145 (0.818, 5.625)
Missing	2 (0.48)					
<b>Helicobacter pylori</b>						
HP-	135 (32.30)	2.61		Ref.		Ref.
HP+	263 (62.92)	2.66	0.328	1.156 (0.864, 1.547)	0.519	1.103 (0.818, 1.488)
Missing	20 (4.78)					
<b>Gastric atrophy</b>						
No	289 (69.14)	2.68		Ref.		Ref.
Yes	54 (12.92)	2.50	0.849	1.040 (0.695, 1.556)	0.750	1.071 (0.701, 1.638)
Missing	75 (17.94)					

<sup>a</sup>aHR with adjustment for TILs, age, sex, grade of differentiation, TNM staging, first-line treatment method, BMI.

**Supplementary Table 8. Univariable and multivariable Cox regression analyses of basic characteristics with OS in AEGJ of TCGA dataset (Validation, N = 169).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis	
			P-value	HR (95% CI)	P-value	aHR <sup>a</sup> (95% CI)
<b>TILs, Median (IQR)</b>	1.99 (0.63, 6.38)		<b>0.003</b>	0.897 (0.835, 0.964)	<b>0.002</b>	0.812 (0.712, 0.925)
<b>Age, mean (range)</b>	65.43 (27–90)		0.806	1.003 (0.982, 1.024)	0.344	1.014 (0.985, 1.045)
<b>Sex</b>						
Woman	46 (27.22)	0.97		Ref.		Ref.
Man	123 (72.78)	1.36	0.530	1.207 (0.671, 2.169)	0.444	1.325 (0.646, 2.718)



<b>First-line treatment method</b>							
Pharmaceutical Therapy	89 (52.66)	1.25		Ref.			Ref.
Radiotherapy	80 (47.34)	1.26	0.613	0.887 (0.559, 1.409)	0.234		1.447 (0.788, 2.657)
<b>Grade of differentiation</b>							
Grade 1–2	46 (27.22)	1.45		Ref.			Ref.
Grade 3	72 (42.60)	1.08	0.073	1.787 (0.948, 3.366)	0.125		1.676 (0.866, 3.240)
Missing	51 (30.18)						
<b>TNM staging</b>							
I+II	71 (42.01)	1.28		Ref.			Ref.
III+IV	85 (50.30)	1.35	<b>0.014</b>	1.921 (1.140, 3.236)	0.644		1.167 (0.606, 2.249)
Missing	13 (7.69)						

<sup>a</sup>aHR with adjustment for TILs proportion, age, sex, first-line treatment method, grade of differentiation, and TNM staging.

**Supplementary Table 9. Univariable and multivariable Cox regression analyses of basic characteristics with OS in GAC of TCGA dataset (Validation, N = 222).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis		
			P-value	HR (95% CI)	P-value	aHR <sup>a</sup> (95% CI)	
<b>TILs, Median (IQR)</b>	4.14 (1.17, 12.99)		<b>0.004</b>	0.961 (0.935, 0.988)	<b>0.025</b>	0.969 (0.942, 0.996)	
<b>Age, mean (range)</b>	65.52 (30–90)		<b>0.019</b>	1.027 (1.004, 1.050)	<b>0.002</b>	1.039 (1.014, 1.064)	
<b>Sex</b>							
Woman	80 (27.22)	1.72		Ref.		Ref.	
Man	142 (72.78)	1.17	0.492	1.172 (0.746, 1.841)	0.797	1.064 (0.662, 1.711)	
<b>First-line treatment method</b>							
Pharmaceutical Therapy	109 (49.10)	1.30		Ref.		Ref.	
Radiotherapy	113 (50.90)	1.44	0.873	0.966 (0.627, 1.486)	0.912	1.026 (0.646, 1.631)	
<b>Grade of differentiation</b>							
Grade 1–2	83 (37.39)	1.44		Ref.		Ref.	
Grade 3	133 (59.91)	1.18	0.085	1.510 (0.944, 2.414)	0.119	1.486 (0.904, 2.442)	
Missing	6 (2.70)						
<b>TNM staging</b>							
I+II	95 (42.79)	1.66		Ref.		Ref.	
III+IV	123 (55.41)	1.10	<b>0.001</b>	2.780 (1.704, 4.533)	<b>0.001</b>	2.464 (1.468, 4.134)	
Missing	4 (1.80)						

<sup>a</sup>aHR with adjustment for TILs proportion, age, sex, first-line treatment method, grade of differentiation, and TNM staging.

**Supplementary Table 10. Univariable and multivariable Cox regression analyses of basic characteristics with OS in ESCC of TCGA dataset (Validation, N = 70).**

Characteristics	No. of patients (%)	Median survival (years)	Univariable analysis		Multivariable analysis		
			P-value	HR (95% CI)	P-value	aHR <sup>a</sup> (95% CI)	
TILs, Median (IQR)	33.00 (9.77, 62.62)		0.080	0.986 (0.970, 1.002)	0.133	0.985 (0.965, 1.005)	
Age, mean (range)	57.01 (36–90)		<b>0.031</b>	1.052 (1.005, 1.102)	0.058	1.063 (0.998, 1.132)	
<b>Sex</b>							
Woman	11 (27.22)	1.95		Ref.		Ref.	
Man	59 (72.78)	1.04	<b>0.045</b>	4.546 (1.032, 20.040)	0.070	6.690 (0.854, 52.415)	

<b>First-line treatment method</b>						
Pharmaceutical Therapy	36 (70.00)	1.06		Ref.		Ref.
Radiotherapy	34 (50.90)	1.05	0.542	0.777 (0.346, 1.746)	0.187	0.492 (0.171, 1.411)
<b>Grade of differentiation</b>						
Grade 1-2	49 (70.00)	1.05		Ref.		Ref.
Grade 3	15 (21.43)	1.10	0.290	0.516 (0.152, 1.758)	0.182	0.357 (0.079, 1.619)
Missing	6 (8.57)					
<b>TNM staging</b>						
I+II	43 (61.43)	1.10		Ref.		Ref.
III+IV	26 (31.17)	1.02	<b>0.048</b>	2.258 (1.006, 5.067)	0.150	2.125 (0.762, 5.922)
Missing	1 (1.80)					

<sup>a</sup>aHR with adjustment for TILs proportion, age, sex, first-line treatment method, grade of differentiation, and TNM staging.

## **Supplementary Datasets**

Please browse Full Text version to see the data of Supplementary Datasets 1–3.

**Supplementary Dataset 1. AEGJ cell feature training set.**

**Supplementary Dataset 2. ESCC cell feature training set.**

**Supplementary Dataset 3. GAC cell feature training set.**