



OPEN

## Integrated bioinformatics analysis of noncoding RNAs with tumor immune microenvironment in gastric cancer

Jun Xu<sup>1</sup>, Shengnan Hu<sup>1</sup>, Qiuli Chen<sup>2</sup>, Lili Shu<sup>2</sup>, Peter Wang<sup>2</sup>✉ & Jianjiang Wang<sup>1</sup>✉

In recent years, molecular and genetic research hotspots of gastric cancer have been investigated, including microRNAs, long noncoding RNAs (lncRNAs) and messenger RNA (mRNAs). The study on the role of lncRNAs may help to develop personalized treatment and identify potential prognostic biomarkers in gastric cancer. The RNA-seq and miRNA-seq data of gastric cancer were downloaded from the TCGA database. Differential analysis of RNA expression between gastric cancer samples and normal samples was performed using the edgeR package. The ceRNA regulatory network was visualized using Cytoscape. KEGG pathway analysis of mRNAs in the ceRNA network was performed using the clusterProfiler package. CIBERSORT was used to distinguish 22 immune cell types and the prognosis-related genes and immune cells were determined using Kaplan-Meier and Cox proportional hazard analyses. To estimate these nomograms, we used receiver operating characteristic and calibration curve studies. The ceRNA regulation network of gastric cancer was built in this study, and the genes in the network were analyzed for prognosis. A total of 980 lncRNAs were differentially expressed, of which 774 were upregulated and 206 were downregulated. A survival study identified 15 genes associated with gastric cancer prognosis, including *VCAN-AS1*, *SERPINE1*, *AL139002.1*, *LINC00326*, *AC018781.1*, *C15orf54*, *hsa-miR-145*. Monocytes and Neutrophils were associated with the survival rate of gastric cancer. Our research uncovers new ceRNA network for the detection, treatment, and monitoring of gastric cancer.

### Abbreviations

ceRNA	Competing endogenous RNA
DElncRNA	Differentially expressed lncRNA
DEmRNA	Differentially expressed mRNA
DemiRNA	Differentially expressed miRNA
DERNA	Differentially expressed RNA
FENDRR	FOXF1 adjacent noncoding developmental regulatory RNA
FLOT	5-Fluorouracil, leucovorin, oxaliplatin and docetaxel
HAVCR1	Hepatitis A virus cellular receptor 1
GO	Gene ontology
KEGG	The kyoto encyclopedia of genes and genomes
lncRNA	Long noncoding RNA
miRNA	Microrna
mRNA	Messenger RNA
NKILA	Nuclear factor-κB-interacting lncRNA
OSCC	Oral squamous cell carcinoma
ROC	Receiver operating characteristic
SERPINE1	Serine protease inhibitor clade E member 1
TIME	Tumor immune microenvironment
TCGA	The cancer genome atlas
VCAN-AS1	VCAN antisense RNA 2

<sup>1</sup>First People's Hospital of Hangzhou Lin'an District, Affiliated Lin'an People's Hospital, Hangzhou Medical College, Hangzhou, China. <sup>2</sup>Department of Research and Development, Zhejiang Zhongwei Medical Research Center, Hangzhou 310018, Zhejiang, China. ✉email: wangpeter2@hotmail.com; 13615713000@163.com

Gastric cancer is one of the most common gastrointestinal malignancies in the world<sup>1</sup>. There were an estimated 26,380 new cases and 11,090 deaths of gastric cancer in the United States in 2022<sup>2</sup>. The incidence of gastric cancer varies significantly by region: highest in East Asia, Eastern Europe and South America, lowest in North and South Africa<sup>3</sup>. *Helicobacter Pylori* (HP) infection is the most important risk factor for gastric tumorigenesis<sup>4</sup>. Approximately 2 billion people worldwide are infected with HP, of which about 1 million will develop gastric cancer<sup>5</sup>. Hence, HP infection significantly increases the incidence of gastric cancer<sup>6</sup>. In addition to HP infection, gender, heredity, severe lack of sleep, irregular diet, excessive work and psychological stress can all contribute to the increased incidence of gastric cancer<sup>7</sup>.

Gastric cancer has low early diagnosis rate and 5 years survival rate<sup>8</sup>. Advanced gastric cancer can be metastasized to liver, pancreas, omentum, esophagus, bile duct and lymph node, and the treatment effect is poor<sup>9</sup>. Therefore, it is very important to find effective early diagnosis biomarkers and therapeutic targets for gastric cancer. Studies have found that the formation of gastric cancer is a complex process involving multiple factors and is usually associated with abnormal gene expressions<sup>10</sup>. At the same time, some studies have proved that noncoding RNAs (ncRNAs) are related to the development of cancer, and these studies have been continuously reported in various cancers<sup>11,12</sup>.

Long non-coding RNAs (lncRNAs) belong to noncoding RNAs with a length of more than 200 nucleotides<sup>13–15</sup>. Recently, some researchers found that many lncRNAs are new regulators of gene expression, and aberrant expression of some lncRNAs is involved in the pathogenesis and progression of malignant tumors<sup>16–18</sup>. These new lncRNAs may be developed into new clinical biomarkers or new therapeutic targets<sup>19–21</sup>. A large number of lncRNAs are associated with various cancers and have antitumor or oncogenic functions<sup>22–24</sup>. Altered lncRNA expression and its mutations are associated with tumorigenesis and metastasis<sup>25</sup>. Studies have confirmed that both lncRNAs and mRNAs can bind to miRNAs and regulate each other.

lncRNAs have been implicated in facilitating the evasion of immune surveillance by tumor cells through multiple mechanisms, including the promotion of an immunosuppressive microenvironment<sup>26–28</sup>. Multiple studies used comprehensive analysis to determine the functions of various lncRNAs in tumor immune microenvironment (TIME) in various cancers, such as colon cancer<sup>29</sup>, clear cell renal cell carcinoma<sup>30</sup>, breast cancer<sup>31,32</sup>, lung cancer<sup>33,34</sup>, thyroid cancer<sup>35</sup>, adrenocortical adenocarcinoma<sup>36</sup>, bladder cancer<sup>37</sup>, and head and neck squamous cell carcinoma<sup>38</sup>. lncRNA NKILA has demonstrated the ability to initiate apoptosis in tumor-specific T cells, thereby impeding their capacity to infiltrate the tumor, leading to tumor immune evasion<sup>39</sup>. lncRNA FENDRR mediated TIME and tumor suppression in non-small cell lung cancer (NSCLC)<sup>40</sup>. lncRNA RP5-881L22.5 was reported to play a crucial role in the TIME of tumors and in the colorectal cancer progression<sup>41</sup>. lncRNA WDFY3-AS2 has been identified to correlate with an immunosuppressive phenotype in the TIME in oral squamous cell carcinoma (OSCC)<sup>42</sup>. WDFY3-AS2 regulated Wnt/ $\beta$ -catenin pathway and promoted proliferation and metastasis in OSCC<sup>42</sup>. However, the role of lncRNAs in TIME of gastric cancer has not been elucidated. Therefore, to further investigate the molecular mechanisms underlying the development of gastric cancer, our study investigated the process of lncRNAs binding to miRNAs, mRNAs and proteins in gastric cancer to regulate gene expression. Moreover, we tried to comprehensively evaluate the correlation between lncRNAs and the prognosis and immune cell infiltration level of gastric cancer patients.

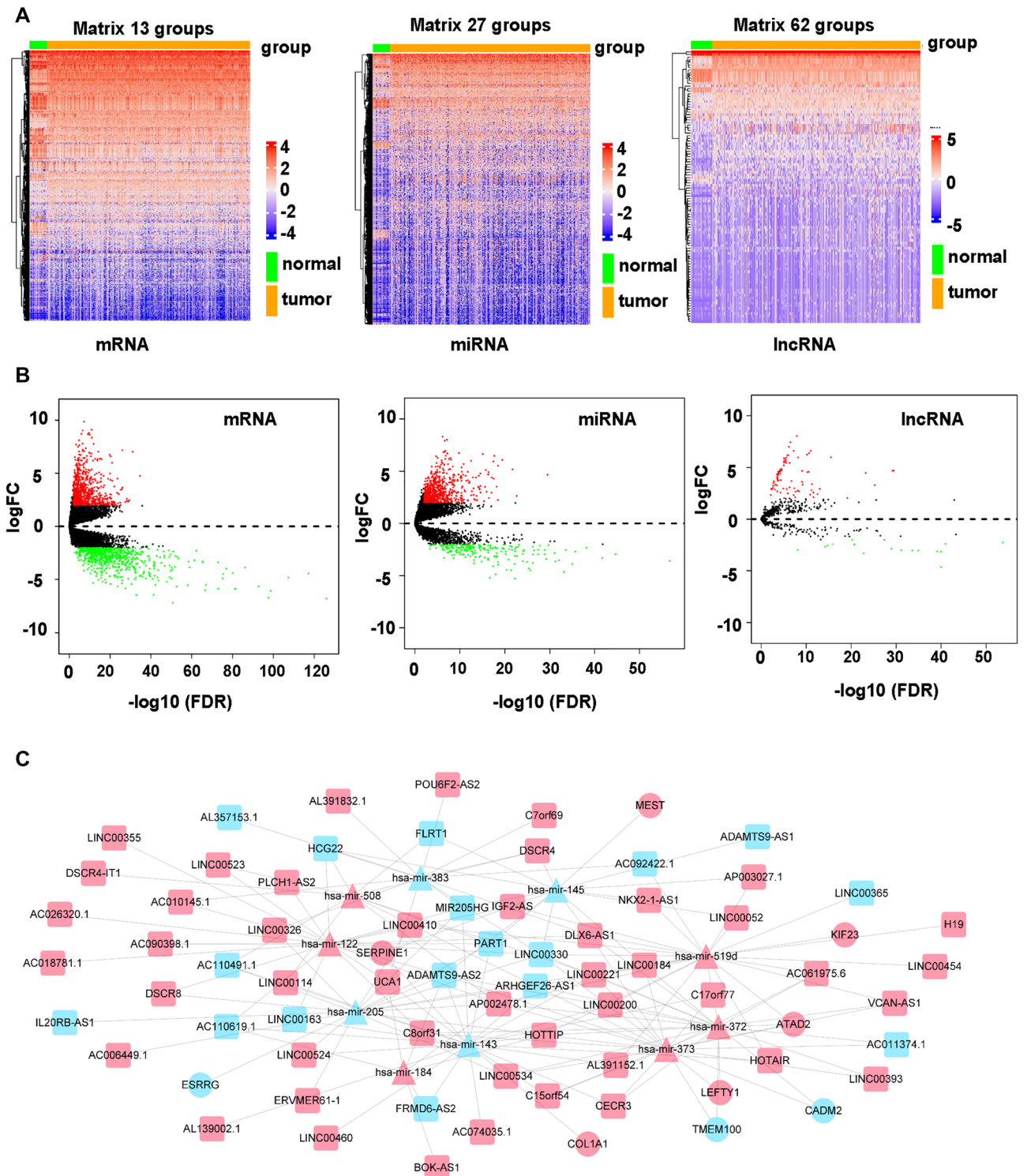
## Results

**Differentially expressed RNAs.** In this study, 980 differentially expressed lncRNAs, 104 differentially expressed miRNAs, and 1639 differentially expressed mRNAs were obtained by “DESeq2” analysis with cut-off ( $|\log_{2}FC| \geq 2$  and  $p < 0.01$ ). We found that among 980 DELncRNAs, 774 lncRNAs were upregulated, while 206 lncRNAs were downregulated. In addition, 87 miRNAs were elevated and 17 miRNAs were reduced. 870 mRNAs were upregulated, whereas 769 mRNAs were downregulated. Heatmaps of mRNAs, miRNAs, and lncRNAs were shown in Fig. 1A. The volcano plot of mRNAs, miRNAs, and lncRNAs was shown in Fig. 1B.

**ceRNA regulatory network.** To further understand the role of RNAs in gastric cancer, the interactions between miRNA-mRNA and lncRNA-miRNA served as the foundation for the development of the ceRNA network. The Cytoscape visualization of the ceRNA regulatory network was shown (Fig. 1C). The association between the prognosis and biomarkers in the STAD-linked ceRNA network was examined by the K-M survival evaluation. These analyses showed 15 genes were implicated in the prognosis of gastric cancer, including AC011374.1, AC018781.1, ADAMTS9-AS1, AL139002.1, AL391152.1, HOTTIP, FLRT1, NKX2-1-AS1, ADAMTS9-AS2, LINC00326, VCAN-AS1, SERPINE1, POU6F2-AS2, IGF2-AS, and miR-145 (Fig. 2).

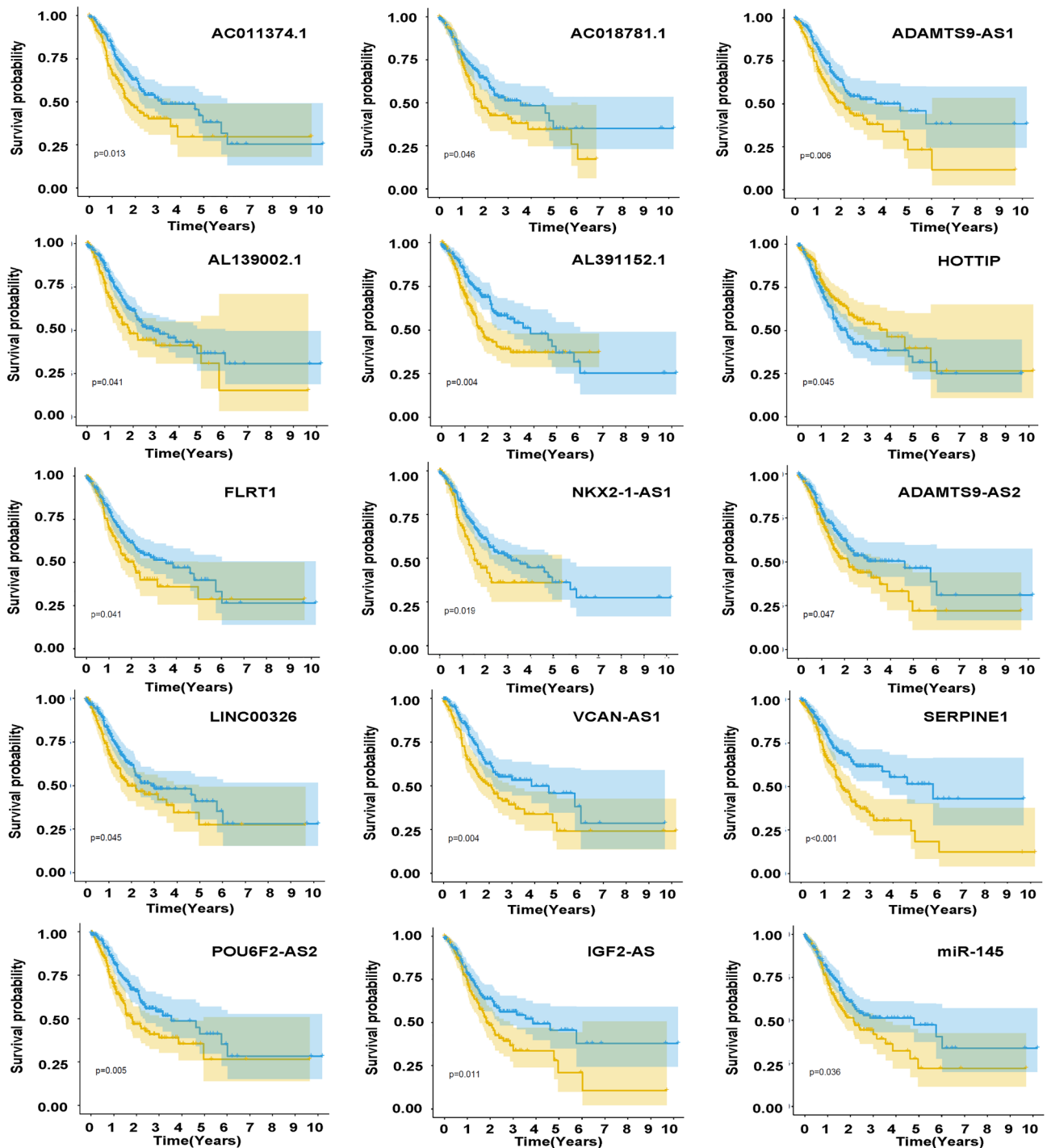
**Prognosis of related genes.** Lasso regression analysis was performed, and two plots were generated. The variation of coefficients at different log Lambda values was depicted (Fig. 3A). The horizontal axis represented log Lambda, while the vertical axis represented the values of the coefficients. The changes in partial likelihood deviance at different log Lambda values were showcased (Fig. 3B). The horizontal axis represented log Lambda, and the vertical axis represented the values of the partial likelihood deviance. VCAN-AS1, SERPINE1, AL139002.1, LINC00326, AC018781.1, C15orf54, hsa-miR-145 were obtained by Multivariate Cox regression analysis (Fig. 3C). The scores were then calculated, and the patients were classified into high-risk and low-risk groups based on the median risk score in the training set (Fig. 3D). The relationships between survival status and survival times of gastric cancer patients, ordered by their respective risk scores, were displayed (Fig. 3E).

The high -risk group had a worse chance of survival than the low-risk group (Fig. 4A). The AUC at 1 year, 3 years, 5 years were 0.665, 0.674 and 0.789 (Fig. 4B). The gene SERPINE1 was highly expressed in both high and low-risk groups (Fig. 4C). The Nomogram was established based on the independent prognostic indicators, which predicts survival rates for the first, 3 to 5 years. The genes in this nomogram model were VCAN-AS1,



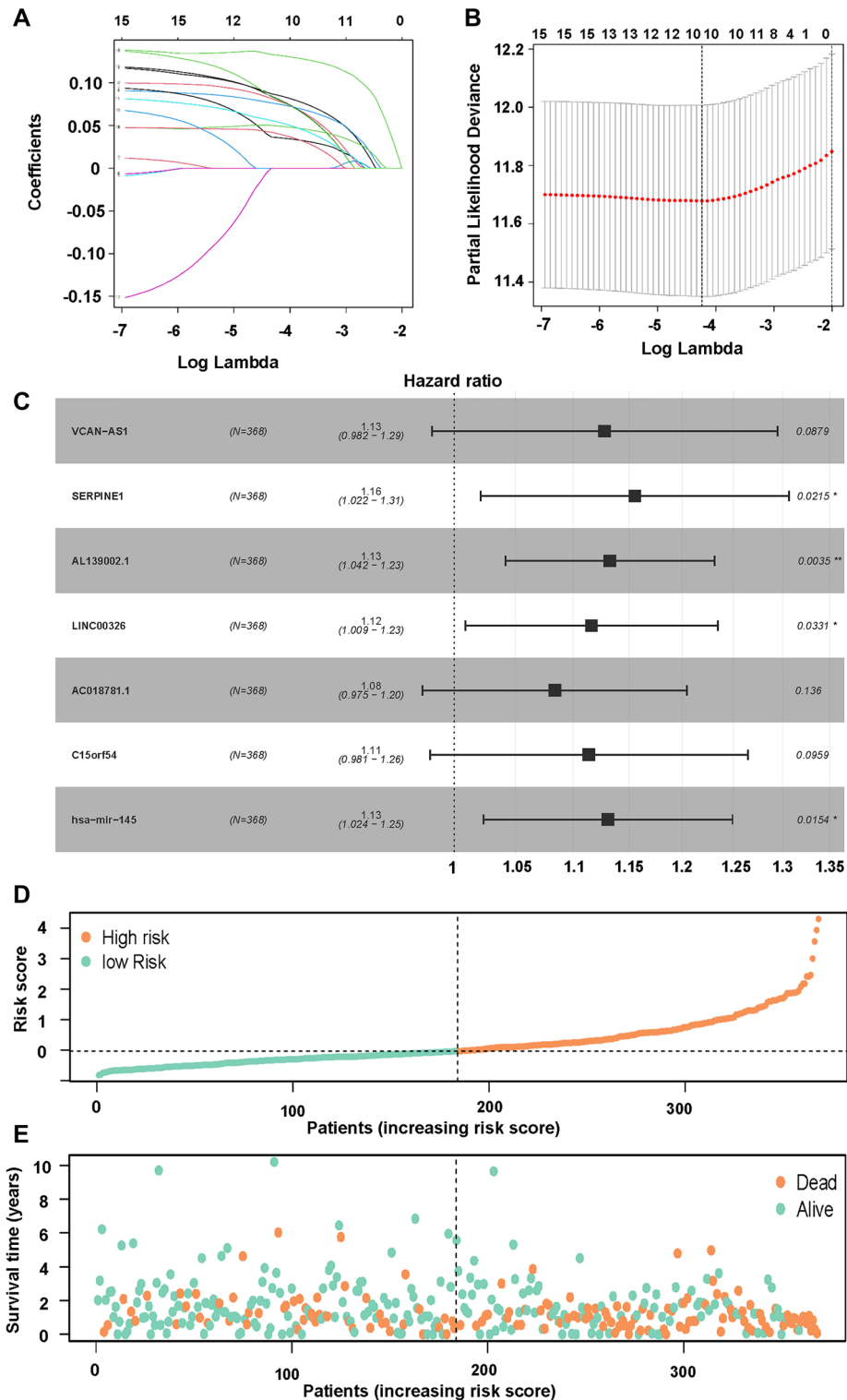
**Figure 1.** Volcano plots and heatmaps of differentially expressed mRNAs, miRNAs, and lncRNAs. (A) Volcano plot of differentially expressed mRNAs, miRNAs and lncRNAs. (B) Heatmap of differentially expressed mRNAs, miRNAs and lncRNAs. (C) lncRNA-miRNA-mRNA ceRNA network generated by Cytoscape. Red: upregulated RNAs. Green: downregulated RNAs. Triangle: miRNA. Diamond: lncRNA. Circular: mRNA.

SERPINE1, AL139002.1, LINC00326, AC018781.1, C15orf54, hsa-miR-145 (Figs. 4D). Moreover, the calibration curve for nomogram models showed a good agreement between the actual survival rate and the predicted 3 years overall survival rate, suggesting that this model was accurate in its predictions (Fig. 4E).

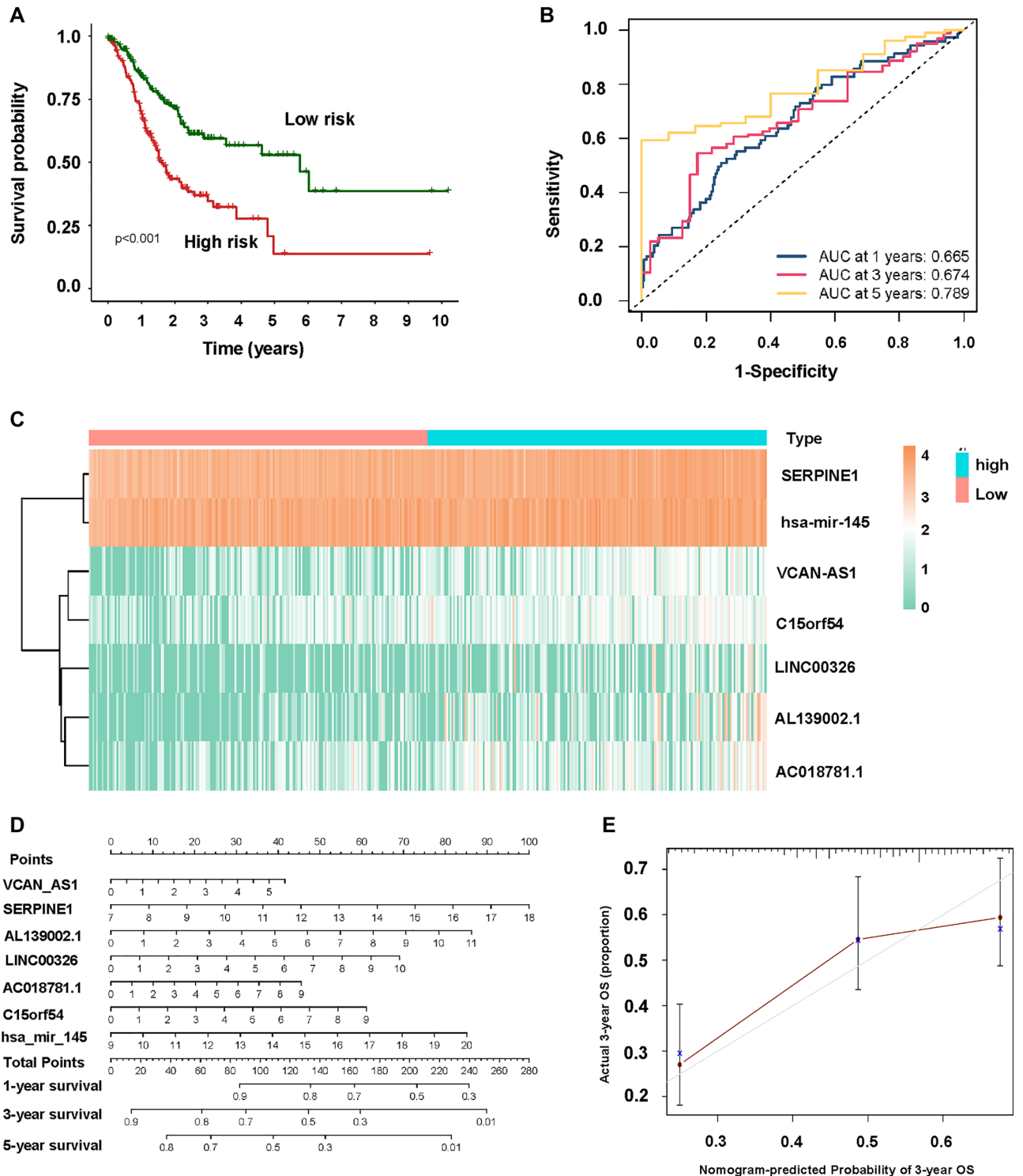


**Figure 2.** Prognosis-related genes were illustrated by the K-M survival evaluation.

**Investigation of immune infiltration and gastric cancer.** The immune cell content in the samples was shown in Fig. 5A. It was depicted that B cells naïve, T cells CD4 memory resting, Macrophages M0, M1, M2 were higher in tumor cells, while B cells memory, plasma cells, T cells CD8, T cells CD4 memory activated and Monocytes were lower in tumor cells (Fig. 5B). The heatmap was also demonstrated that various immune cells were differently expressed in tumor cells (Fig. 5C). CD4 memory activated T cells was associated with survival probability. High number of CD4 memory activated T cells had a good survival probability (Fig. 6A). The association among 22 different types of immune cells was presented (Fig. 6B). The expression of B cells naïve, Mast cells resting and NK cells activated was higher in younger than in older age, while the expression of Mast cells activated, Neutrophils, NK cells resting was lower than in older age (Fig. 7). The expression of Macrophage M0, Mast cells activated, and Plasma cells was higher in G1/2 than in G3 grade, while the expression of Macrophage



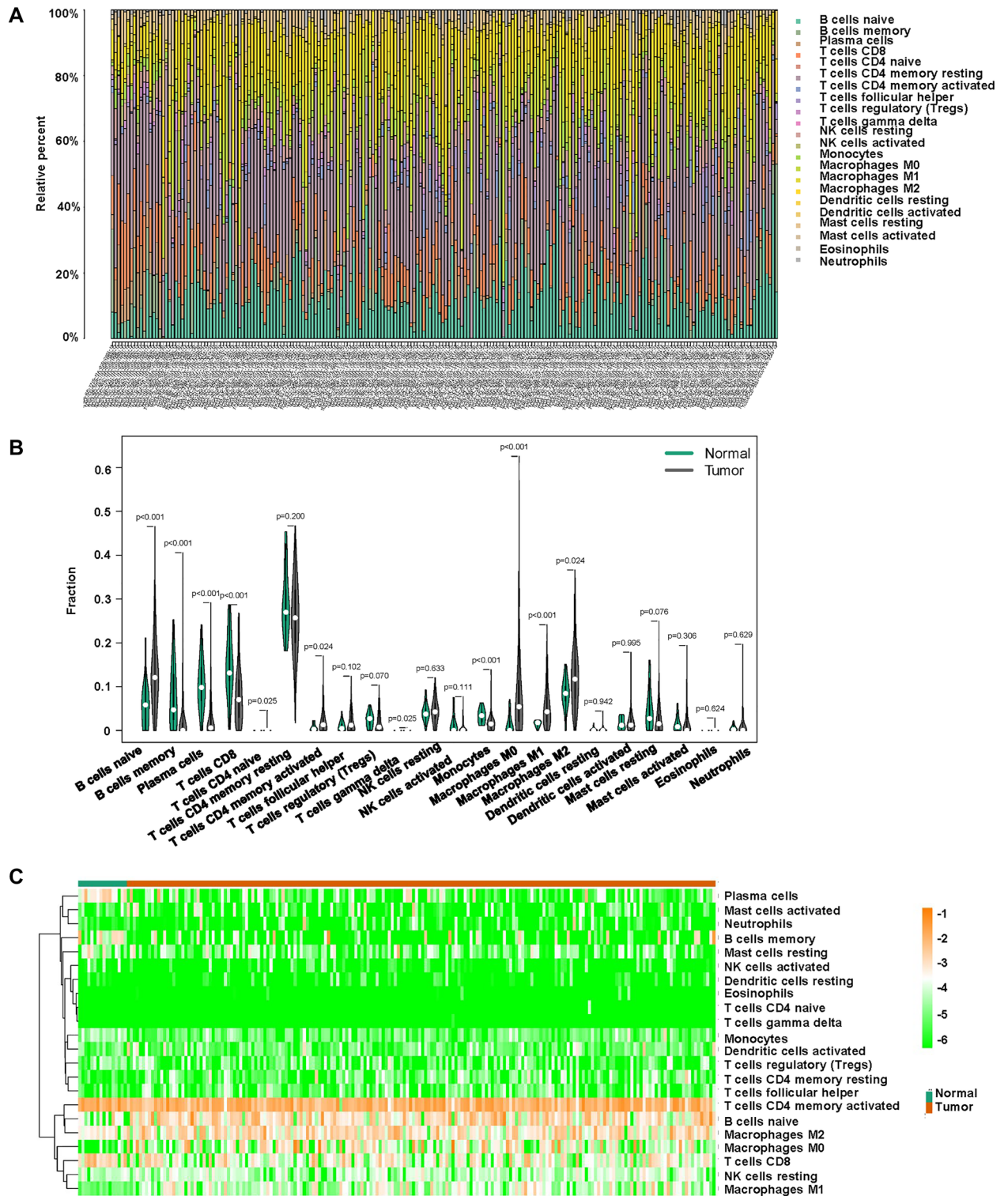
**Figure 3.** LASSO and multivariate regression models screening genes. **(A)** LASSO coefficient profiles of genes in STAD; dotted vertical lines were drawn at the optimal values by using the minimum criteria. **(B)** LASSO coefficient profiles of the candidate OS-related GRGs with nonzero coefficients determined by the optimal lambda. A coefficient profile plot was produced against the  $\log\lambda$  sequence. **(C)** Forest plot of several genes was involved in the ceRNA network. **(D)** Risk score analyses of GC patients in the training cohort based on the seven-GRGs signature. Distribution of risk scores per patient. **(E)** Relationships between survival status and survival times of GC patients ranked by risk score. The black dotted line represents the optimum cut-off point dividing patients into low and high-risk groups.



**Figure 4.** The association of ceRNAs and survival. (A) KM analysis of the high-risk and low-risk groups. (B) Time-dependent receiver operating characteristic (ROC) analysis for OS prediction of prediction model. (C) Heatmap of seven differentially expressed genes. (D) Nomogram based on multiple Cox regression. (E) Calibration curve for 3 years survival.

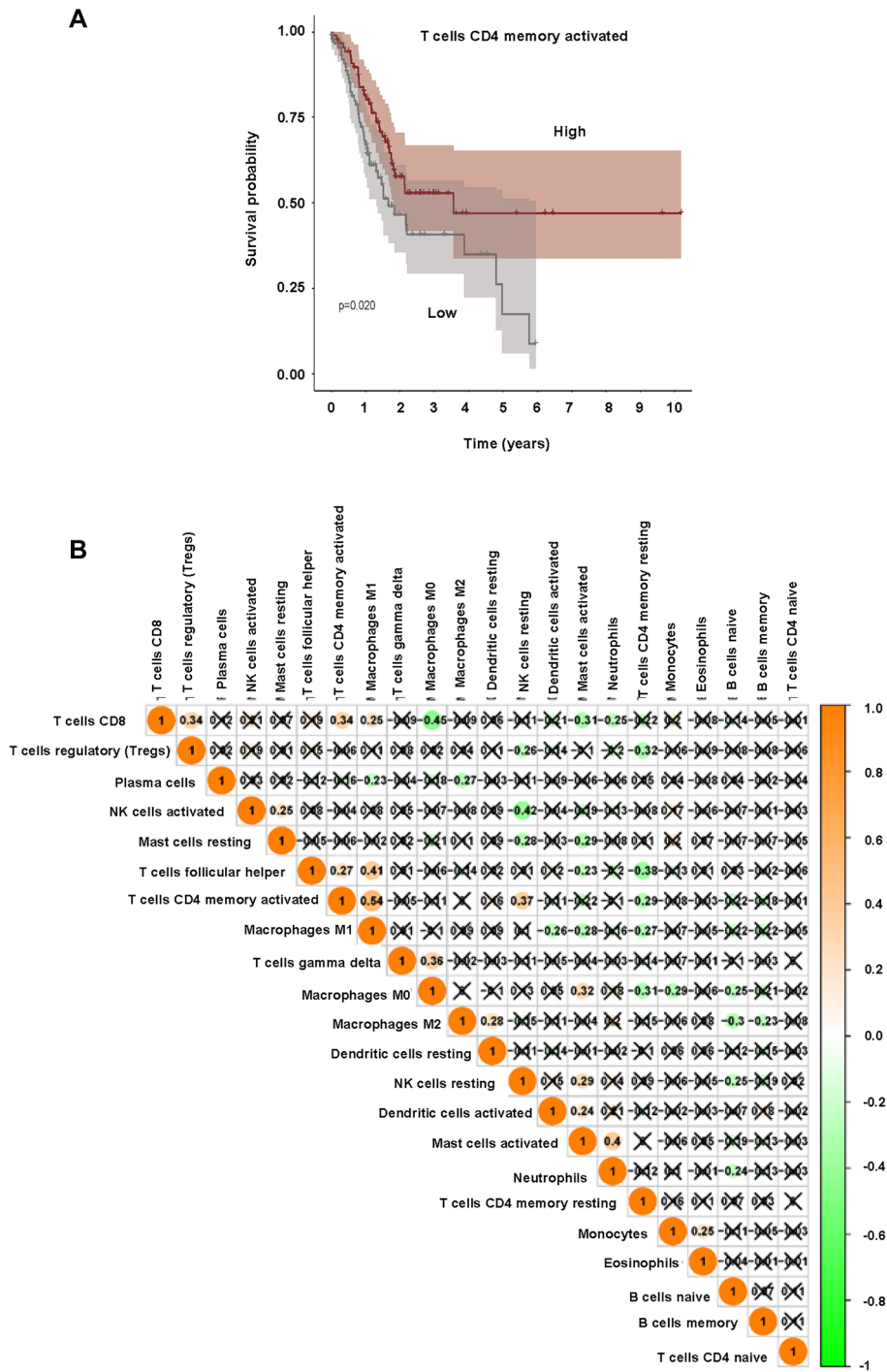
M1, Mast cells resting, Monocytes, T cells CD4 memory activated, T cells CD8, and T cells follicular helper was lower in G1/2 than in G3 grade (Fig. 7).

**Monocytes and neutrophils in gastric cancer.** Next, we found no significant difference for survival probability between the high-risk group and the low-risk group (Fig. 8A). The data from AUC curve showed that the nomogram survival prediction model had a reasonable accuracy. The AUC values at 1 year, 3 years, and



**Figure 5.** The association of immune cells and STAD. (A) The percentage of 22 immune cell subpopulations in STAD patients. (B) The fraction of immune cells in STAD patients. (C) Heatmap of immune infiltration between normal and tumor groups.

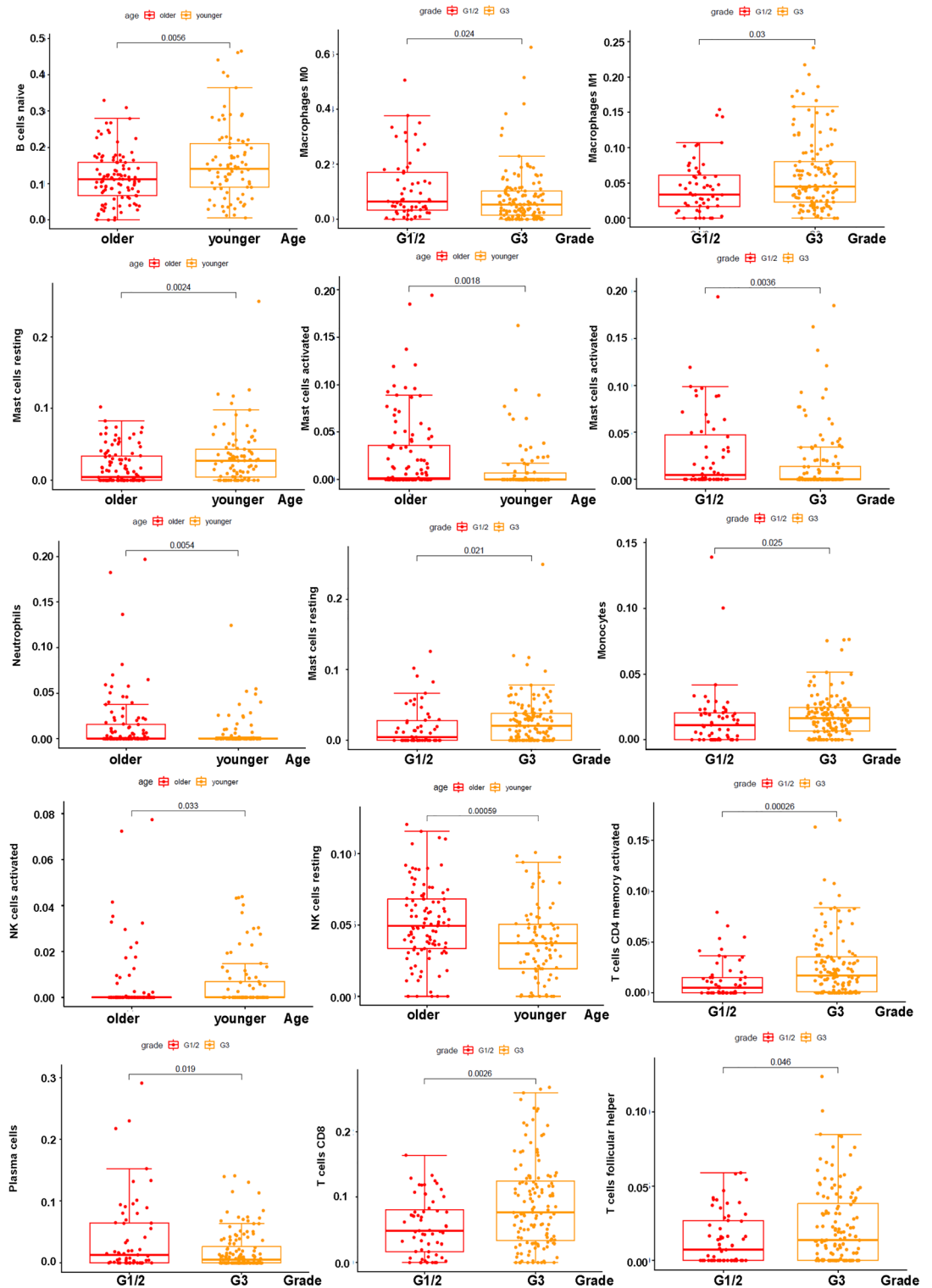
5 years were 0.569, 0.525, and 0.605, respectively (Fig. 8A). Cox regression analysis obtained two predictable related immune cells, Monocytes and Neutrophils. The Nomogram was established based on the immune cells, which predicts survival rates for the first, 3 to 5 years. The immune cells in this nomogram model were Monocytes and Neutrophils (Fig. 8B). Moreover, the calibration curve for nomogram models demonstrated a consistent agreement between the actual survival rate and the predicted 3 years overall survival rate, implying that this model was accurate in its predictions (Fig. 8B). In addition, hsa-miR-145 and AC018781.1 were associated



**Figure 6.** The correlation of immune cells and STAD. (A) T cells CD4 memory activated and survival in STAD. (B) Co-expression patterns among fractions of immune cells.

with Monocytes, while AC018781.1, AL139002.1, SERPINE1, VCAV-AS1 and C15orf54 were associated with Neutrophils (Fig. 8C-D). In the high-risk group, Monocytes and neutrophils were highly expressed (Fig. 9A). Pearson's correlation assessed the association between biomarkers and immune cells. C15orf54 had the strongest relationship with Neutrophils with a coefficient of 0.32. VCAN-AS1 had a good association with Neutrophils with a coefficient of 0.28. In addition, hsa-miR-145 had a stronger relationship with Monocytes with a coefficient of 0.31 (Fig. 9B).

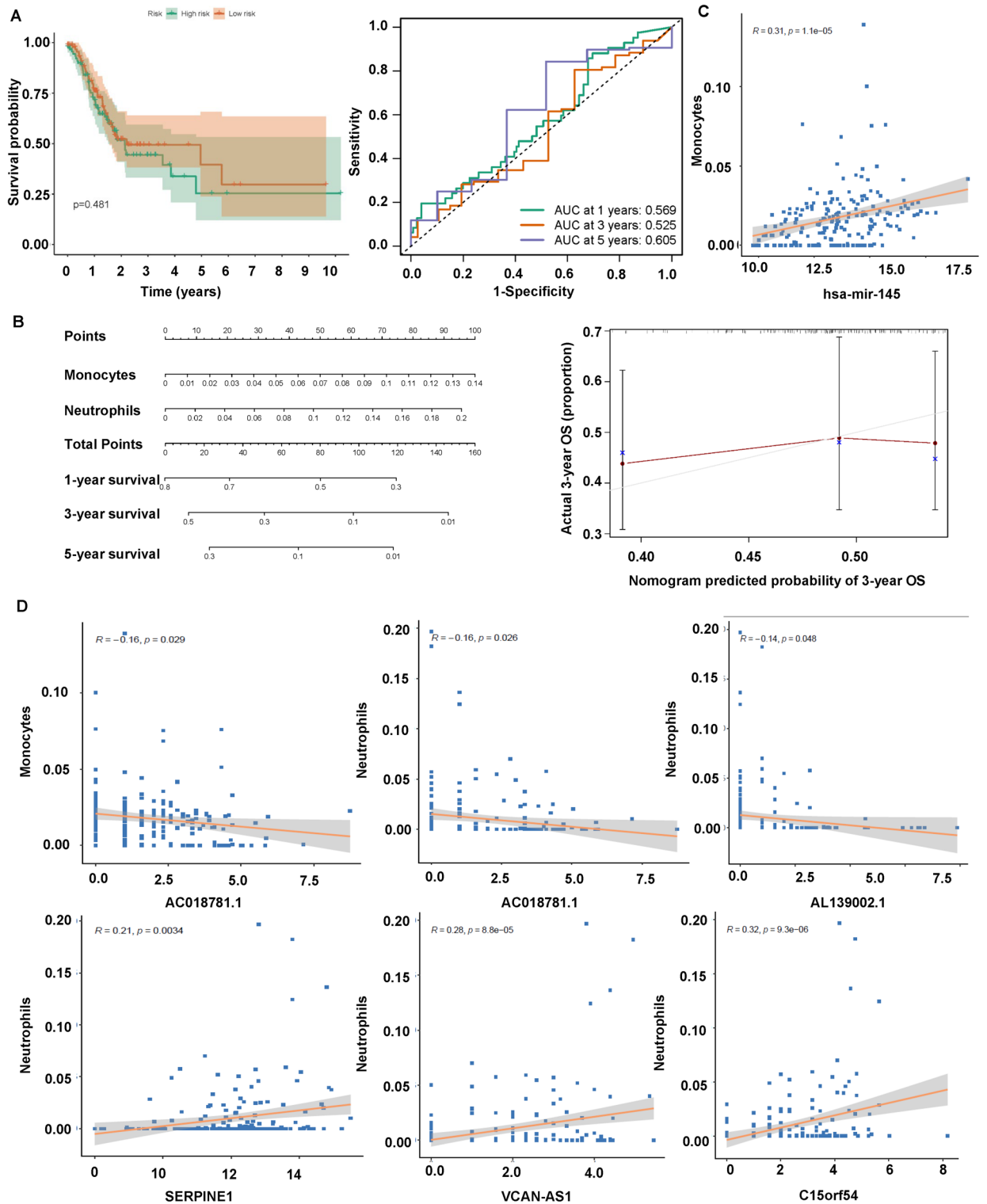




**Figure 7.** Correlation between immune cells and clinical features.

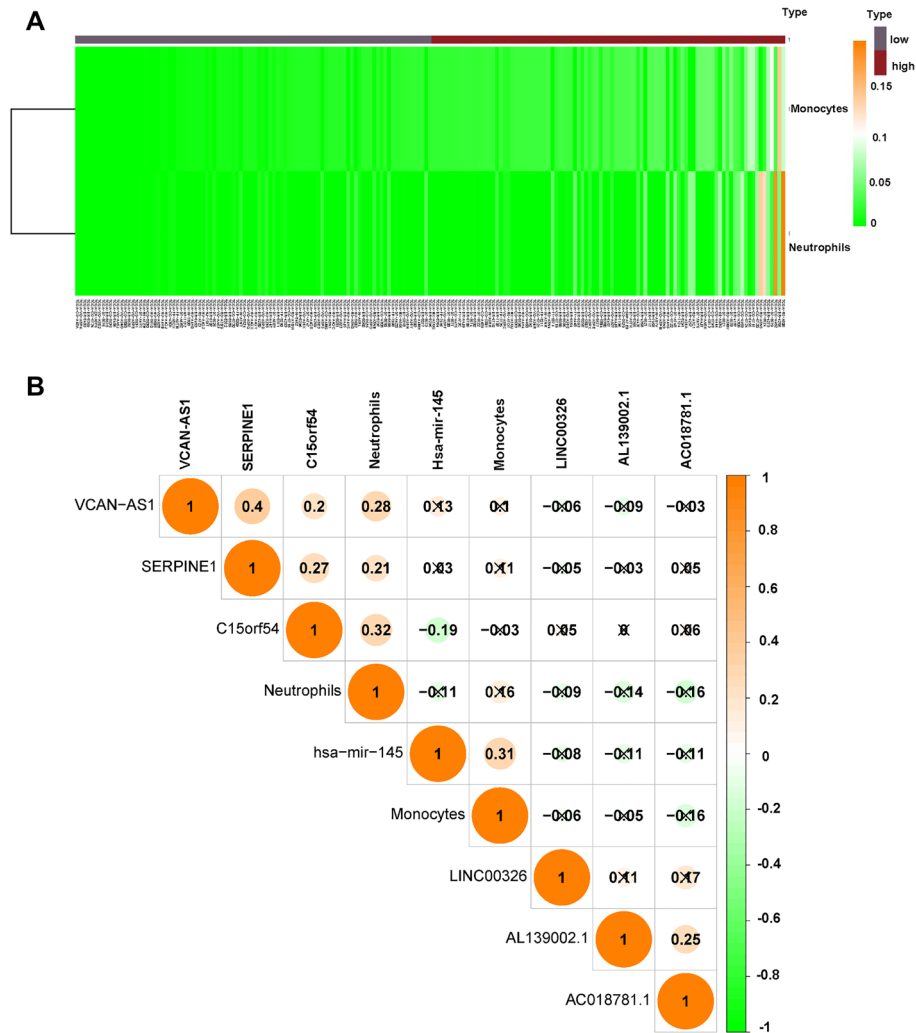
### Discussion

Although the incidence and mortality of gastric cancer have decreased in recent years, gastric cancer is still one of the most common malignant tumors in the world and causes a heavy medical burden globally<sup>3</sup>. To identify potential biomarkers associated with the diagnosis, treatment and prognosis of gastric cancer, comprehensive bioinformatics analysis may help to achieve this goal and develop personalized treatment for gastric cancer patients. Based on RNA expression profiles from TCGA, 980 lncRNAs, 104 miRNAs and 1639 mRNAs were identified, which were differentially expressed between tumors and normal tissues. The current work used



**Figure 8.** The association of immune cells and survival in STAD. (A) Survival probability of high and low-risk groups (left). AUC curve at 1, 3, 5 years (right). (B) Nomogram for predicting patients' outcome based on immune cells (left). Calibration curves for assessing the discrimination and accuracy of the nomogram (right). (C–D) The association between immune cells and prognostic genes.

lncRNA-associated ceRNA to identify key biomarkers that were related to the prognosis of gastric cancer. The lncRNAs VCAN-AS1, AL139002.1, LINC00326, AC018781.1, and C15orf54 served as hub nodes in the ceRNA network, which targeted other miRNAs and mRNAs.



**Figure 9.** The association of immune cells and survival in STAD. **(A)** Heatmap of the immune cells (monocytes and neutrophils) in high-risk and low-risk groups. **(B)** correlation matrix of immune cells and genes.

LncRNAs have also been considered as an underappreciated novel therapeutic target for gastric cancer<sup>43</sup>. Their diverse roles in cancer progression offer new opportunities to disrupt metastasis in clinical settings<sup>44</sup>. LncRNAs play important roles in gene regulation, including regulation of gene activation and silencing, X chromosome inactivation, alternative splicing, and post-translational regulation. mRNAs, miRNAs, and lncRNAs are linked in an intricate network of crosstalk by competing endogenous RNAs<sup>45</sup>. LncRNA VCAN-AS1 has been reported to promote the malignant tumor behaviors via modulating the miR-106a-5p-involved STAT3/HIF-1 $\alpha$  axis in breast cancer<sup>46</sup>. Two studies also validated that lncRNA VCAN-AS1 was abnormally expressed in gastric cancer by constructing the ceRNA regulatory network<sup>47,48</sup>. Feng et al.<sup>49</sup> reported that lncRNA VCAN-AS1 contributed to gastric cancer progression via regulation of p53 expression. Hence, lncRNA VCAN-AS1 might be a potential factor to influence gastric cancer development and progression. Due to the unclear role of lncRNA VCAN-AS1 in TIME, further investigation is required to determine the functions of VCAN-AS1 in regulating TIME in gastric cancer.

One study has confirmed that AL139002.1 sponged miR-490-3p and regulated the expression of Hepatitis A virus cellular receptor 1 (HAVCR1) in gastric cancer, contributing to the development of gastric cancer<sup>50</sup>. In addition, LINC00326 has been found to regulate hepatocarcinogenic lipid metabolism<sup>51</sup>. In non-small cell lung carcinoma, overexpression of LINC00326 attenuated tumor progression via blockade of Wnt/ $\beta$ -catenin pathway through miR-657/DKK2 axis<sup>52</sup>. Liu et al.<sup>53</sup> used a competing endogenous RNA network and found that 3 lncRNAs, C15orf54, ADAMTS9-AS1 and AL391152.1, were involved in survival rate of gastric cancer patients. LncRNA AC018781.1 was observed to be an independent risk factors for gastric cancer by a bioinformatic and integrated analysis<sup>48</sup>.

In this study, we found the important role of miR-145 in gastric cancer. In line with this finding, numerous studies have validated the functions of miR-145 in gastric cancer progression<sup>54</sup>. For example, miR-145 inhibited cell proliferation, invasion and migration as well as cell cycle progression via suppression of transcription factor Sp1 in gastric cancer<sup>55</sup>. Xing et al. reported that miR-145 targeted the expression of catenin- $\delta$ 1, MYO6 and

inhibited cell invasion in gastric cancer<sup>56,57</sup>. Chang et al.<sup>58</sup> found that miR-145 induced the antiproliferative effects via targeting E2F3 in gastric cancer. Low expression of miR-145-5p was associated with poor prognosis in gastric cancer<sup>59</sup>. It is clear that the role of miR-145 has been well studied in gastric cancer.

SERPINE1 gene has been studied in a variety of human cancers, including gastric cancer. One group reported a SERPINE1-based immune gene signature, which can predict immunotherapy response and patient prognosis in gastric cancer<sup>60</sup>. Another group uncovered that SERPINE1 enhanced malignant progression and associated with poor prognosis in gastric cancer patients<sup>61</sup>. Several analyses support SERPINE1 as a potential prognostic biomarker in gastric cancer<sup>62–64</sup>. One study used data mining in combination with bioinformatics dissected that SERPINE1 was associated with immunoinfiltration and might be a diagnosis and prognosis biomarker in stomach adenocarcinoma<sup>65</sup>. High SERPINE1 expression group presented higher immune cells' expression, including CD4+ T cells, neutrophils, CD8+ T cells, macrophages and B cells. Moreover, SERPINE1 was associated with immune cells in the TIME of stomach adenocarcinoma<sup>65</sup>. Another study discovered that SERPINE1 expression was correlated with several types of immune cells, such as neutrophils, macrophages, CD4+ T cells, CD8+ T cells, B cells, and dendritic cells in clear cell renal cell carcinoma<sup>66</sup>.

Recently, evidence suggested that SERPINE1 was associated with immune infiltrates in gastric cancer<sup>67</sup>. SERPINE1 enhanced the inhibitory TIME, which was positively associated with the infiltration level of neutrophils, macrophage M2, resting NK cells, and activated mast cells. SERPINE1 was negatively associated with plasma cells and B cell memory in gastric cancer<sup>67</sup>. Consistently, in our study, we identified that SERPINE1 expression was positively correlated with neutrophils in gastric cancer.

CD4+ regulatory T cells were identified to be associated with survival in gastric cancer<sup>68</sup>. In gastric cancer patients, CD4+ T cells and regulatory T cells were enriched, and lower expression of miR-128-3p was correlated with overall survival. Moreover, miR-128-3p targeted the expression of IL16 in gastric cancer<sup>68</sup>. One group reported that high infiltration of CD4+ T cells was linked to worse overall survival in gastric cancer<sup>69</sup>. Another group revealed that CD4+ memory T cell-related genes were correlated with clinical overall survival in patients with gastric cancer<sup>70</sup>. Additionally, lncRNAs, including A2M-AS1, C2orf27A, and ZNF667-AS1, impaired activation of CD4+ T cells and affected prognosis of gastric cancer patients<sup>71</sup>. In gastric cancer patients with FLOT chemotherapy (5-Fluorouracil, Leucovorin, Oxaliplatin and Docetaxel), CD4+/CD8+ lymphocyte ratio was elevated and predicted favourable therapy response<sup>72</sup>. In our study, CD4 memory activated T cells were linked to survival of gastric cancer patients. Without a doubt, more exploration is necessary to determine the mechanism of CD4 T cells-mediated survival in gastric cancer.

In summary, we successfully identified lncRNAs closely related to the occurrence and development of gastric cancer based on the TCGA database. Based on these lncRNAs, a ceRNA network based on the lncRNA-miRNA-mRNA regulatory mechanism was constructed. Several lncRNAs, miRNAs and mRNAs associated with gastric cancer prognosis were screened in the ceRNA network by survival analysis. These indicators provide new targets for the prognosis evaluation of gastric cancer patients. At the same time, our study also confirmed the relationship between immunity and lncRNA-based ceRNA regulatory network. This study will enable us to identify more useful targets to develop effective treatment strategies for gastric cancer. It has several limitations in this study. In TCGA database, there are lower numbers of control groups compared with the tumor groups. The discrepancy between the control group and the experimental group could lead to potential bias in the analysis results. In addition, it is important to mention that in vitro and in vivo experiments are necessary to validate our findings in the future.

## Materials and methods

**Data collecting from TCGA.** RNA-sequencing data, miRNA profiling and clinical information of STAD were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>). Illumina HiSeqRNASeq platforms were used to obtain mRNA and lncRNA data. HiSeqmiRNASeq platforms were used to collect miRNA data. Data were collected for gastric cancer samples and normal samples. Data were further organized by ID conversion, filtering, merging, correction and clinical information. Demography, histology, tumor stage and TNM of the STAD were obtained (Supplementary Table 1 and supplementary file 1). Pre-filtered low count genes ( $\text{rowMeans}(\text{data}) > 1$ ) were used to preprocess RNA sequencing profiles. We converted the downloaded data into count format through R, so as to analyze these samples and screen out differentially expressed RNAs. EdgeR (v.3.28.0) is an R package dedicated to analyzing DEGs (differentially expressed genes). Standard settings for DEGs were  $\text{FDR} < 0.01$  and  $\log \text{fold change (FC)} > 2$ . DElncRNA (differentially expressed lncRNA), DEMRNA (differentially expressed mRNA) and DemRNA (differentially expressed miRNA) were visualized using the ggplot package. The Ensembl database was used to annotate genes in RNAseq, and we excluded RNAs not included in the Ensembl database. Afterwards, we extracted lncRNA expression profiles and mRNA expression profiles from the RNAseq expression matrix.

**Construction of CeRNA regulatory network.** To further understand the role of DERNA (differentially expressed RNA) in gastric cancer, we constructed a ceRNA regulatory network between lncRNA-miRNA-mRNA to explore the relationship between them. Then, a ceRNA network was built using the hypergeometric test and correlation analysis to identify the differentially expressed miRNAs that can control lncRNAs and mRNAs ( $p < 0.05$  as the filter threshold criteria). DElncRNAs, DEMiRNAs and DEMRNAs were incorporated into ceRNA regulatory network. We visualized the ceRNA regulatory network with Cytoscape. PPI network of mRNAs involved in the ceRNA network were generated by STRING<sup>73</sup>.

**GO analysis and KEGG pathway analysis.** Functional analysis of genes from different organisms, which includes three gene ontology of cellular component (CC), molecular function (MF) and biological process (BP),

was further conducted through the Gene ontology (GO) database<sup>74</sup>. The clusterProfiler (v.3.14.3) package was adopted to perform GO functional annotation and KEGG pathway analysis of mRNAs in the ceRNA network<sup>75</sup>.

**Prognosis analysis of related genes.** The clinical information of gastric cancer patients was downloaded from TCGA database and then we extracted the survival information of patients (survival time and survival status of patients)<sup>76</sup>. Subsequently, survival information with the expression matrix of RNAs in the ceRNA network was combined. To determine whether there is a relationship between biomarker expression and STAD survival, Kaplan-Meier (K-M) survival analysis was performed<sup>77</sup>. The next step was to find significant factors in the original Cox model that had prognostic significance and to incorporate those into the reduced Cox proportional hazard model<sup>78</sup>. In order to test the accuracy of the created multifactorial model, LASSO regression uses contraction to lower the data value to a particular point<sup>79</sup>. Ultimately, a nomogram was developed using the multivariate model to forecast the prognosis of STAD. The nomogram might make it possible for doctors to assign each biomarker a prognosis value based on how it expresses<sup>80</sup>. Indicators of prognosis and total 3- and 5 years survival may be found in the sum of the individual values. The accuracy of the nomograms was evaluated using receiver operating characteristic curve (ROC) analysis and calibration<sup>81</sup>.

**Immune infiltration in STAD.** Gene expression data were used by CIBERSORT to estimate the proportion and abundance of different types of immune cells in tumor and normal groups<sup>82</sup>. To assess the percentages of 22 immune cell types in STAD samples, we used CIBERSORT in this instance. Only instances with CIBERSORT results of  $P \leq 0.05$  were considered for further analysis. Important immune cells in tumor samples were distinguished from non-malignant samples using Wilcoxon rank-sum<sup>83</sup>. Next, K-M survival analysis was used to see whether the number of particular immune cells was related to overall STAD survival<sup>84</sup>. Different immune cells were incorporated into a Cox proportional hazard model following LASSO regression analysis, and a nomogram for STAD prognosis prediction was created<sup>85</sup>. The bias and precision of the nomogram were evaluated using the concordance index of the Cox mode.

**Association between selected RNAs and immune cells.** To examine the immune cells that are connected to survival, univariate Cox regression and Kaplan-Meier survival analysis were employed<sup>77</sup>. In parallel, multivariate Cox regression analysis and Lasso regression were used to create the final immune cell model<sup>86,87</sup>. An equation that forecasts medical outcomes was graphically represented as a nomogram. With nomograms, patients accumulate points based on the severity of their risk factors<sup>85</sup>. The fold difference in gene expression between tumor tissues and normal tissues served as the basis for the risk factors of two nomograms that we constructed for the study. Then, we forecast the patients' 1-, 3-, or 5 years survival rate.

## Data availability

The data of this study are available from the corresponding author upon reasonable request.

Received: 29 May 2023; Accepted: 26 August 2023

Published online: 11 September 2023

## References

- Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249. <https://doi.org/10.3322/caac.21660> (2021).
- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA A Cancer J. Clin.* **72**, 7–33. <https://doi.org/10.3322/caac.21708> (2022).
- Thrift, A. P. & El-Serag, H. B. Burden of gastric cancer. *Clin. Gastroenterol. Hepatol.* **18**, 534–542. <https://doi.org/10.1016/j.cgh.2019.07.045> (2020).
- Kim, Y. H. & Shin, S. W. Helicobacter pylori and prevention of gastric cancer. *N. Engl. J. Med.* **378**, 2244–2245. <https://doi.org/10.1056/NEJMc1805129> (2018).
- Kavitt, R. T. & Cifu, A. S. Management of helicobacter pylori infection. *JAMA* **317**, 1572–1573. <https://doi.org/10.1001/jama.2017.1949> (2017).
- Lee, Y.-C. *et al.* Association between helicobacter pylori eradication and gastric cancer incidence: A systematic review and meta-analysis. *Gastroenterology* **150**, 1113–1124.e1115. <https://doi.org/10.1053/j.gastro.2016.01.028> (2016).
- Choi, I. J. *et al.* Family history of gastric cancer and helicobacter pylori treatment. *N. Engl. J. Med.* **382**, 427–436. <https://doi.org/10.1056/NEJMoa1909666> (2020).
- Suzuki, H. & Matsuzaki, J. Gastric cancer: Evidence boosts Helicobacter pylori eradication. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 458–460. <https://doi.org/10.1038/s41575-018-0023-8> (2018).
- Smyth, E. Missing a GOLDen opportunity in gastric cancer. *Lancet Oncol.* **18**, 1561–1563. [https://doi.org/10.1016/S1470-2045\(17\)30719-2](https://doi.org/10.1016/S1470-2045(17)30719-2) (2017).
- Zhou, Z. *et al.* The long noncoding RNA D63785 regulates chemotherapy sensitivity in human gastric cancer by targeting miR-422a. *Mol. Ther. Nucl. Acids* **12**, 405–419. <https://doi.org/10.1016/j.omtn.2018.05.024> (2018).
- Liu, J. & Shang, G. The roles of noncoding RNAs in the development of osteosarcoma stem cells and potential therapeutic targets. *Front. Cell Dev. Biol.* **10**, 773038. <https://doi.org/10.3389/fcell.2022.773038> (2022).
- Winkle, M., El-Daly, S. M., Fabbri, M. & Calin, G. A. Noncoding RNA therapeutics - Challenges and potential solutions. *Nat. Rev. Drug Discov.* **20**, 629–651. <https://doi.org/10.1038/s41573-021-00219-z> (2021).
- Luo, Y., Zhou, R., Huang, N., Sun, L. & Liao, W. Effect of long non-coding RNA EIF3J-AS1 on multi-drug resistance and autophagy in gastric cancer. *J. Clin. Oncol.* **35**, e15581–e15581. [https://doi.org/10.1200/JCO.2017.35.15\\_suppl.e15581](https://doi.org/10.1200/JCO.2017.35.15_suppl.e15581) (2017).
- Chen, X., Liu, Y., Liu, H., Wang, Z. W. & Zhu, X. Unraveling diverse roles of noncoding RNAs in various human papillomavirus negative cancers. *Pharmacol. Ther.* **238**, 108188. <https://doi.org/10.1016/j.pharmthera.2022.108188> (2022).
- Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-022-00566-8> (2023).

16. Chen, T., Liu, J., Zhang, H., Li, J. & Shang, G. Long intergenic noncoding RNA00265 enhances cell viability and metastasis via targeting miR-485-5p/USP22 axis in osteosarcoma. *Front. Oncol.* **12**, 907472. <https://doi.org/10.3389/fonc.2022.907472> (2022).
17. Chen, W. *et al.* Comprehensive analysis of lncRNA-mediated ceRNA network for hepatocellular carcinoma. *Front. Oncol.* **12**, 1042928. <https://doi.org/10.3389/fonc.2022.1042928> (2022).
18. Jiang, W., Pan, S., Chen, X., Wang, Z. W. & Zhu, X. The role of lncRNAs and circRNAs in the PD-1/PD-L1 pathway in cancer immunotherapy. *Mol. Cancer* **20**, 116. <https://doi.org/10.1186/s12943-021-01406-7> (2021).
19. Yuan, L. *et al.* Long non-coding RNAs towards precision medicine in gastric cancer: Early diagnosis, treatment, and drug resistance. *Mol. Cancer* **19**, 96. <https://doi.org/10.1186/s12943-020-01219-0> (2020).
20. Xie, W. *et al.* Emerging roles of long noncoding RNAs in chemoresistance of pancreatic cancer. *Semin. Cancer Biol.* **33**, 303–318. <https://doi.org/10.1016/j.semcancer.2020.11.004> (2022).
21. Jiang, W. *et al.* Long non-coding RNAs as a determinant of cancer drug resistance: Towards the overcoming of chemoresistance via modulation of lncRNAs. *Drug Resist. Update* **50**, 100683. <https://doi.org/10.1016/j.drug.2020.100683> (2020).
22. Chen, D.-L. *et al.* Long non-coding RNA XIST regulates gastric cancer progression by acting as a molecular sponge of miR-101 to modulate EZH2 expression. *J. Exp. Clin. Cancer Res.* **35**, 142. <https://doi.org/10.1186/s13046-016-0420-1> (2016).
23. Begliarzade, S. *et al.* Long non-coding RNAs as promising biomarkers and therapeutic targets in cervical cancer. *Noncoding RNA Res.* **8**, 233–239. <https://doi.org/10.1016/j.ncrna.2023.02.006> (2023).
24. Fattahi, S. *et al.* LINC02688 and PP7080 as novel biomarkers in early diagnosis of gastric cancer. *Noncoding RNA Res.* **6**, 86–91. <https://doi.org/10.1016/j.ncrna.2021.04.002> (2021).
25. Suman, P., Chhichholiya, Y., Kaur, P., Ghosh, S. & Munshi, A. Long non-coding RNAs involved in different steps of cancer metastasis. *Clin. Transl. Oncol.* **24**, 997–1013. <https://doi.org/10.1007/s12094-021-02761-z> (2022).
26. Hu, Q. *et al.* Oncogenic lncRNA downregulates cancer cell antigen presentation and intrinsic tumor suppression. *Nat. Immunol.* **20**, 835–851. <https://doi.org/10.1038/s41590-019-0400-7> (2019).
27. Zhang, Y., Liu, Q. & Liao, Q. Long noncoding RNA: A dazzling dancer in tumor immune microenvironment. *J. Exp. Clin. Cancer Res.* **39**, 231. <https://doi.org/10.1186/s13046-020-01727-3> (2020).
28. Zhang, Z. *et al.* Pan-cancer characterization of lncRNA modifiers of immune microenvironment reveals clinically distinct de novo tumor subtypes. *NPJ Genom. Med.* **6**, 52. <https://doi.org/10.1038/s41525-021-00215-7> (2021).
29. Chen, W. *et al.* Comprehensive analysis of immune infiltrates of ferroptosis-related long noncoding RNA and prediction of colon cancer patient prognoses. *J. Immunol. Res.* **2022**, 9480628. <https://doi.org/10.1155/2022/9480628> (2022).
30. Chen, Y. *et al.* Development and validation of a novel 5 cuproptosis-related long noncoding RNA signature to predict diagnosis, prognosis, and drug therapy in clear cell renal cell carcinoma. *Transl. Androl. Urol.* **12**, 622–641. <https://doi.org/10.21037/tau-23-65> (2023).
31. Guo, Q., Qiu, P., Pan, K. & Lin, J. Comprehensive analysis of cuproptosis-related long non-coding RNA signature and personalized therapeutic strategy of breast cancer patients. *Front. Oncol.* **12**, 1081089. <https://doi.org/10.3389/fonc.2022.1081089> (2022).
32. Jiang, H., Sun, J., Liu, F., Wu, X. & Wen, Z. An immune-related long noncoding RNA pair as a new biomarker to predict the prognosis of patients in breast cancer. *Front. Genet.* **13**, 895200. <https://doi.org/10.3389/fgene.2022.895200> (2022).
33. Zeng, C., Yu, H., Liu, X., Liu, Q. & Jin, J. Identification and validation of a novel necroptosis-related long noncoding RNA prognostic signature for lung adenocarcinoma. *Biomed. Res. Int.* **2022**, 9710540. <https://doi.org/10.1155/2022/9710540> (2022).
34. Wan, R. *et al.* Discovery of tumor immune infiltration-related snoRNAs for predicting tumor immune microenvironment status and prognosis in lung adenocarcinoma. *Comput. Struct. Biotechnol. J.* **19**, 6386–6399. <https://doi.org/10.1016/j.csbj.2021.11.032> (2021).
35. Lin, Y., Gan, F., He, X., Deng, H. & Li, Y. Identification of ferroptosis-associated long noncoding RNA prognostic model and tumor immune microenvironment in thyroid cancer. *J. Immunol. Res.* **2022**, 5893998. <https://doi.org/10.1155/2022/5893998> (2022).
36. Ding, Y. *et al.* Integrated analyses reveal potential functional N6-methyladenosine-related long noncoding RNAs in adrenocortical adenocarcinoma. *Front. Cell Dev. Biol.* **10**, 851748. <https://doi.org/10.3389/fcell.2022.851748> (2022).
37. Zhang, L. *et al.* Identification of immune-related lncRNA signature to predict prognosis and immunotherapeutic efficiency in bladder cancer. *Front. Oncol.* **10**, 542140. <https://doi.org/10.3389/fonc.2020.542140> (2020).
38. Zhou, C. *et al.* Construction of an m6A-related lncRNA pair prognostic signature and prediction of the immune landscape in head and neck squamous cell carcinoma. *J. Clin. Lab Anal.* **36**, e24113. <https://doi.org/10.1002/jcla.24113> (2022).
39. Huang, D. *et al.* NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. *Nat. Immunol.* **19**, 1112–1125. <https://doi.org/10.1038/s41590-018-0207-y> (2018).
40. Pan, H. *et al.* lncRNA FENDRR-mediated tumor suppression and tumor-immune microenvironment changes in non-small cell lung cancer. *Transl. Cancer Res.* **9**, 3946–3959. <https://doi.org/10.21037/tcr-20-2147> (2020).
41. Zong, H., Zou, J. Q., Huang, J. P. & Huang, S. T. Potential role of long noncoding RNA RP5-881L22.5 as a novel biomarker and therapeutic target of colorectal cancer. *World J. Gastrointest. Oncol.* **14**, 2108–2121. <https://doi.org/10.4251/wjgo.v14.i11.2108> (2022).
42. Lin, X., Ding, J. M., Zheng, X. Z. & Chen, J. G. Immunity-related long noncoding RNA WDFY3-AS2 inhibited cell proliferation and metastasis through Wnt/beta-catenin signaling in oral squamous cell carcinoma. *Arch Oral. Biol.* **147**, 105625. <https://doi.org/10.1016/j.archoralbio.2023.105625> (2023).
43. Huang, Y. *et al.* lncRNA AK023391 promotes tumorigenesis and invasion of gastric cancer through activation of the PI3K/Akt signaling pathway. *J. Exp. Clin. Cancer Res.* **36**, 194. <https://doi.org/10.1186/s13046-017-0666-2> (2017).
44. Zhu, X. *et al.* A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol. Cancer* **15**, 60. <https://doi.org/10.1186/s12943-016-0544-0> (2016).
45. Li, Y. *et al.* The multifaceted role of long non-coding RNA in gastric cancer: Current status and future perspectives. *Int. J. Biol. Sci.* **17**, 2737–2755. <https://doi.org/10.7150/ijbs.61410> (2021).
46. Du, P. *et al.* Long non-coding RNA VCAN-AS1 promotes the malignant behaviors of breast cancer by regulating the miR-106a-5p-mediated STAT3/HIF-1 $\alpha$  pathway. *Bioengineered* **12**, 5028–5044. <https://doi.org/10.1080/21655979.2021.1960774> (2021).
47. Wang, J., Ding, Y., Wu, Y. & Wang, X. Identification of the complex regulatory relationships related to gastric cancer from lncRNA-miRNA-mRNA network. *J. Cell Biochem.* **121**, 876–887. <https://doi.org/10.1002/jcb.29332> (2020).
48. Liao, Y. *et al.* Bioinformatic and integrated analysis identifies an lncRNA-miRNA-mRNA interaction mechanism in gastric adenocarcinoma. *Genes Genom.* **43**, 613–622. <https://doi.org/10.1007/s13258-021-01086-z> (2021).
49. Feng, L. *et al.* Long noncoding RNA VCAN-AS1 contributes to the progression of gastric cancer via regulating p53 expression. *J. Cell Physiol.* **235**, 4388–4398. <https://doi.org/10.1002/jcp.29315> (2020).
50. Chen, Y. & Zhang, R. Long non-coding RNA AL139002.1 promotes gastric cancer development by sponging microRNA-490-3p to regulate hepatitis A virus cellular receptor 1 expression. *Bioengineered* **12**, 1927–1938. <https://doi.org/10.1080/21655979.2021.1922329> (2021).
51. Sondergaard, J. N. *et al.* CCT3-LINC00326 axis regulates hepatocarcinogenic lipid metabolism. *Gut* **71**, 2081–2092. <https://doi.org/10.1136/gutjnl-2021-325109> (2022).
52. Zhang, Y. *et al.* Upregulation of long intergenic non-coding RNA LINC00326 inhibits non-small cell lung carcinoma progression by blocking Wnt/beta-catenin pathway through modulating the miR-657/dickkopf WNT signaling pathway inhibitor 2 axis. *Biol. Direct* **18**, 3. <https://doi.org/10.1186/s13062-023-00359-9> (2023).

53. Liu, M., Li, J., Huang, Z. & Li, Y. Gastric cancer risk-scoring system based on analysis of a competing endogenous RNA network. *Transl. Cancer Res.* **9**, 3889–3902. <https://doi.org/10.21037/tcr-19-2977> (2020).
54. Wang, J., Sun, Z., Yan, S. & Gao, F. Effect of miR-145 on gastric cancer cells. *Mol. Med. Rep.* **19**, 3403–3410. <https://doi.org/10.3892/mmr.2019.10015> (2019).
55. Qiu, T. *et al.* MiR-145, miR-133a and miR-133b inhibit proliferation, migration, invasion and cell cycle progression via targeting transcription factor Sp1 in gastric cancer. *FEBS Lett.* **588**, 1168–1177. <https://doi.org/10.1016/j.febslet.2014.02.054> (2014).
56. Xing, A. Y. *et al.* Catenin-delta1, negatively regulated by miR-145, promotes tumour aggressiveness in gastric cancer. *J. Pathol.* **236**, 53–64. <https://doi.org/10.1002/path.4495> (2015).
57. Lei, C. *et al.* miR-143 and miR-145 inhibit gastric cancer cell migration and metastasis by suppressing MYO6. *Cell Death Dis.* **8**, e3101. <https://doi.org/10.1038/cddis.2017.493> (2017).
58. Chang, S. *et al.* miR-145 mediates the antiproliferative and gene regulatory effects of vitamin D3 by directly targeting E2F3 in gastric cancer cells. *Oncotarget* **6**, 7675–7685. <https://doi.org/10.18632/oncotarget.3048> (2015).
59. Zhang, Y. *et al.* Downregulation of miR-145-5p correlates with poor prognosis in gastric cancer. *Eur. Rev. Med. Pharmacol. Sci.* **20**, 3026–3030 (2016).
60. Xu, X. *et al.* A SERPINE1-based immune gene signature predicts prognosis and immunotherapy response in gastric cancer. *Pharmaceuticals (Basel)* <https://doi.org/10.3390/ph15111401> (2022).
61. Chen, S. *et al.* SERPINE1 overexpression promotes malignant progression and poor prognosis of gastric cancer. *J. Oncol.* **2022**, 2647825. <https://doi.org/10.1155/2022/2647825> (2022).
62. Xu, B., Bai, Z., Yin, J. & Zhang, Z. Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer. *PeerJ* **7**, e7091. <https://doi.org/10.7717/peerj.7091> (2019).
63. Liao, P. *et al.* Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer. *Onco Targets Ther.* **11**, 6969–6980. <https://doi.org/10.2147/OTT.S173934> (2018).
64. Zhu, Z. *et al.* Comprehensive analysis reveals CTHRC1, SERPINE1, VCAN and UPK1B as the novel prognostic markers in gastric cancer. *Transl. Cancer Res.* **9**, 4093–4110. <https://doi.org/10.21037/tcr-20-211> (2020).
65. Zhai, Y. *et al.* Data mining combines bioinformatics discover immunoinfiltration-related gene SERPINE1 as a biomarker for diagnosis and prognosis of stomach adenocarcinoma. *Sci. Rep.* **13**, 1373. <https://doi.org/10.1038/s41598-023-28234-7> (2023).
66. Guo, L., An, T., Wan, Z., Huang, Z. & Chong, T. SERPINE1 and its co-expressed genes are associated with the progression of clear cell renal cell carcinoma. *BMC Urol.* **23**, 43. <https://doi.org/10.1186/s12894-023-01217-6> (2023).
67. Feng, L., Li, G., Li, D., Duan, G. & Liu, J. Cuproptosis-related gene SERPINE1 is a prognostic biomarker and correlated with immune infiltrates in gastric cancer. *J. Cancer Res. Clin. Oncol.* <https://doi.org/10.1007/s00432-023-04900-1> (2023).
68. Fang, W., Shi, C., Wang, Y., Song, J. & Zhang, L. microRNA-128-3p inhibits CD4+ regulatory T cells enrichment by targeting interleukin 16 in gastric cancer. *Bioengineered* **13**, 1025–1038. <https://doi.org/10.1080/21655979.2021.2017566> (2022).
69. You, Q. *et al.* Serum CD4 is associated with the infiltration of CD4(+)T cells in the tumor microenvironment of gastric cancer. *J. Immunol. Res.* **2021**, 6539702. <https://doi.org/10.1155/2021/6539702> (2021).
70. Ning, Z. K. *et al.* Molecular subtypes and CD4(+) memory T cell-based signature associated with clinical outcomes in gastric cancer. *Front. Oncol.* **10**, 626912. <https://doi.org/10.3389/fonc.2020.626912> (2020).
71. Yao, F. *et al.* LncRNAs target ferroptosis-related genes and impair activation of CD4(+) T cell in gastric cancer. *Front. Cell. Dev. Biol.* **9**, 797339. <https://doi.org/10.3389/fcell.2021.797339> (2021).
72. Skubleny, D. *et al.* Increased CD4/CD8 lymphocyte ratio predicts favourable neoadjuvant treatment response in gastric cancer: A prospective pilot study. *World J. Gastrointest. Oncol.* **15**, 303–317. <https://doi.org/10.4251/wjgo.v15.i2.303> (2023).
73. Freilich, R., Arhar, T., Abrams, J. L. & Gestwicki, J. E. Protein–protein interactions in the molecular chaperone network. *Acc. Chem. Res.* **51**, 940–949. <https://doi.org/10.1021/acs.accounts.8b00036> (2018).
74. Blake, J. A. & Harris, M. A. The gene ontology (GO) project: Structured vocabularies for molecular biology and their application to genome and expression analysis. *Curr. Protoc. Bioinf.* <https://doi.org/10.1002/0471250953.bi0702s23> (2008).
75. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34. <https://doi.org/10.1093/nar/27.1.29> (1999).
76. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol. (Pozn)* **19**, A68–77. <https://doi.org/10.5114/wo.2014.47136> (2015).
77. Goel, M. K., Khanna, P. & Kishore, J. Understanding survival analysis: Kaplan–Meier estimate. *Int. J. Ayurveda Res.* **1**, 274–278. <https://doi.org/10.4103/0974-7788.76794> (2010).
78. Deo, S. V., Deo, V. & Sundaram, V. Survival analysis–part 2: Cox proportional hazards model. *Indian J. Thorac. Cardiovasc. Surg.* **37**, 229–233. <https://doi.org/10.1007/s12055-020-01108-7> (2021).
79. Kim, S. M., Kim, Y., Jeong, K., Jeong, H. & Kim, J. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography* **37**, 36–42. <https://doi.org/10.14366/usg.16045> (2018).
80. Barlin, J. N. *et al.* Nomogram for predicting 5-year disease-specific mortality after primary surgery for epithelial ovarian cancer. *Gynecol. Oncol.* **125**, 25–30. <https://doi.org/10.1016/j.ygyno.2011.12.423> (2012).
81. Park, S. H., Goo, J. M. & Jo, C. H. Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean J. Radiol.* **5**, 11–18. <https://doi.org/10.3348/kjr.2004.5.1.11> (2004).
82. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243–259. [https://doi.org/10.1007/978-1-4939-7493-1\\_12](https://doi.org/10.1007/978-1-4939-7493-1_12) (2018).
83. Dutta, S. & Datta, S. A rank-sum test for clustered data when the number of subjects in a group within a cluster is informative. *Biometrics* **72**, 432–440. <https://doi.org/10.1111/biom.12447> (2016).
84. Stel, V. S., Dekker, F. W., Tripepi, G., Zoccali, C. & Jager, K. J. Survival analysis I: The Kaplan–Meier method. *Nephron Clin. Pract.* **119**, c83–88. <https://doi.org/10.1159/000324758> (2011).
85. Mell, L. K. *et al.* Nomogram to predict the benefit of intensive treatment for locoregionally advanced head and neck cancer. *Clin. Cancer Res.* **25**, 7078–7088. <https://doi.org/10.1158/1078-0432.Ccr-19-1832> (2019).
86. Stel, V. S., Dekker, F. W., Tripepi, G., Zoccali, C. & Jager, K. J. Survival analysis II: Cox regression. *Nephron Clin. Pract.* **119**, c255–260. <https://doi.org/10.1159/000328916> (2011).
87. Chintalapudi, N. *et al.* LASSO regression modeling on prediction of medical terms among seafarers’ health documents using tidy text mining. *Bioengineering (Basel)* <https://doi.org/10.3390/bioengineering9030124> (2022).

## Author contributions

J.X. and S.H. wrote the manuscript and performed experiments. Q.C. and L.S. performed the statistical analyses and edited the manuscript. P.W. edited and revised the manuscript. J.W. designed and supervised the study, and revised the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41444-3>.

**Correspondence** and requests for materials should be addressed to P.W. or J.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023