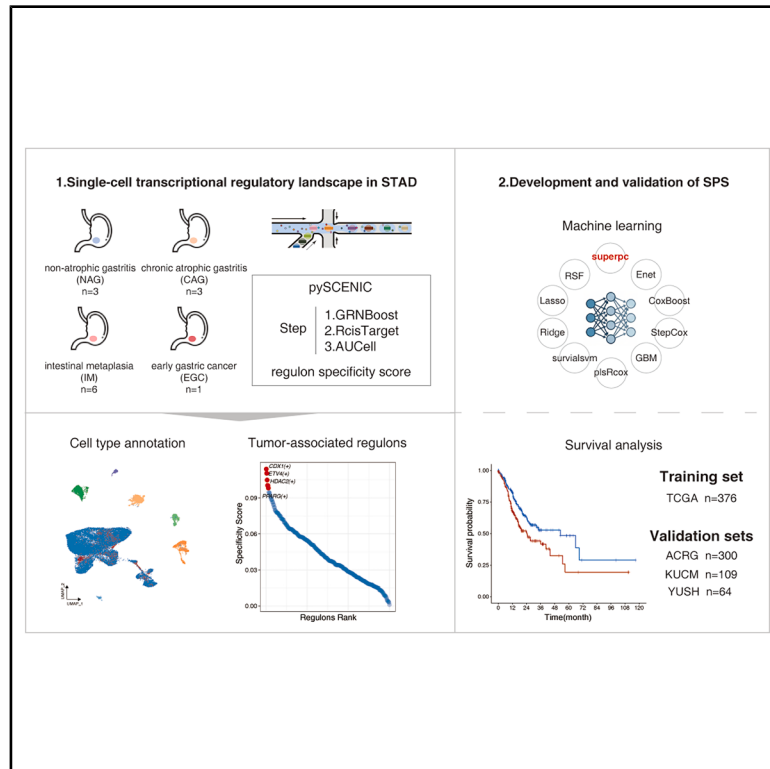


Integration of single-cell regulon atlas and bulk RNA-seq for individualized prognostic prediction in stomach adenocarcinoma

Graphical abstract



Authors

Aiping Zhang, Zhe Li, Xin Jiang, Yaoyao Li, Min Zhang, Yanbing Ding

Correspondence

aipingzhang2021@163.com (A.Z.),
ybding@yzu.edu.cn (Y.D.)

In brief

Health sciences; Medicine; Oncology

Highlights

- Single-cell transcriptional regulatory landscape in STAD
- Malignant epithelial cell-specific regulators-derived SPS predicts STAD prognosis
- SPS shows high accuracy and robustness, serves as an independent predictor for STAD



Article

Integration of single-cell regulon atlas and bulk RNA-seq for individualized prognostic prediction in stomach adenocarcinoma

Aiping Zhang,^{1,3,*} Zhe Li,^{2,3} Xin Jiang,¹ Yaoyao Li,¹ Min Zhang,¹ and Yanbing Ding^{1,4,*}¹Department of Gastroenterology, The Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, Jiangsu, China²Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China³These authors contributed equally⁴Lead contact*Correspondence: aipingzhang2021@163.com (A.Z.), ybding@yzu.edu.cn (Y.D.)<https://doi.org/10.1016/j.isci.2026.116100>

SUMMARY

Transcriptional regulators reflect cellular heterogeneity and are key for prognostic modeling. Given the poor prognosis of stomach adenocarcinoma (STAD), regulator-derived signatures are vital for risk stratification. Using multi-stage scRNA-seq data, we delineated the transcriptional regulatory landscape of STAD and identified *Helicobacter pylori*-associated epithelial heterogeneity in intestinal metaplasia. We then developed a 23-regulator machine learning-based STAD prognostic signature (SPS) from malignant epithelial cells to predict overall survival (OS). Patients with high-SPS exhibited significantly worse OS than patients with low-SPS (hazard ratio [HR] = 1.50, 95% CI: 1.09–2.09, log rank $p = 7.11 \times 10^{-3}$). Notably, SPS outperformed other established STAD prognosis models across various independent datasets and remained an independent prognostic factor after adjusting for clinical and pathologic factors. Moreover, integrating SPS with tumor stage and age showed superior accuracy to stage alone. Collectively, our study establishes a robust regulator-based prognostic signature, holding potential to facilitate precision prognostication in STAD.

INTRODUCTION

Gastric cancer (GC) has become the fifth most common cause of cancer incidence and mortality worldwide.¹ The tumor's aggressive progression and the common delay in diagnosis are major contributors to its high case-fatality rate.² Despite the improved early diagnosis and treatment levels facilitated by the application of the endoscopic examination, as well as advancements in surgical methods and systemic therapies, the five-year survival rate for patients with advanced GC was still dismal.³ Consequently, the development of robust prognostic signatures is critical for guiding the individualized clinical management of stomach adenocarcinoma (STAD).

Transcriptional regulation plays a central role in controlling gene expression programs underlying both physiological and pathological processes. Dysregulation of transcriptional regulators has been shown to profoundly influence tumor progression, including the remodeling of the tumor microenvironment and adaptive resistance to therapy.^{4,5} Coordinated activity of regulatory factors and their associated gene expression programs defines cellular phenotypes, forming complex transcriptional networks commonly referred to as gene regulatory networks (GRNs) or regulons.⁶ Disruptions in GRNs can lead to aberrant expression of the associated genes and contribute to the development of diseases, particularly cancers.⁷

Single-cell RNA sequencing (scRNA-seq) technology enables the analysis of gene expression at the cellular level, thereby facilitating the investigation of tumor heterogeneity and the tracking of disease progression.⁸ Recently, scRNA-seq technology has been leveraged to infer transcriptional regulatory networks, allowing the characterization of transcriptional programs underlying diverse cell states within complex tissues.⁹ Accordingly, characterizing cell-type-specific transcriptional regulatory networks provides a framework for dissecting transcriptional heterogeneity in GC.

In this study, we inferred transcriptional regulatory features in GC using scRNA-seq. We further identified regulatory features specific to malignant gastric epithelial cells and developed a prognostic risk model for patients with STAD. This work provides additional insights into tumor heterogeneity and supports improved prognostic assessment and personalized patient management.

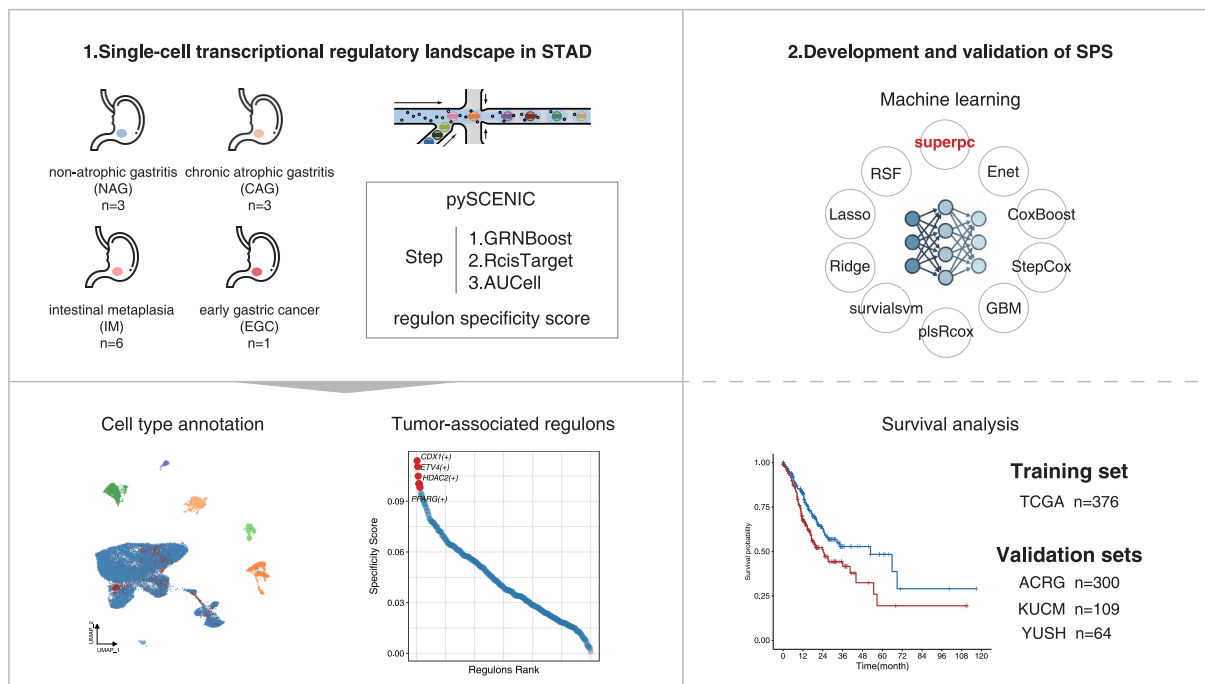
RESULTS

Construction of a single-cell-specific transcriptional regulatory network in STAD

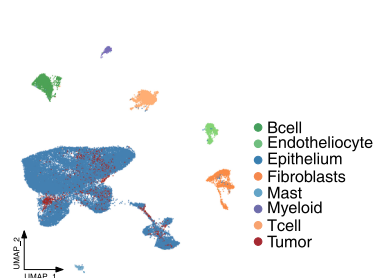
A graphical summary of the research methodology is illustrated in Figure 1A. We mapped a single-cell transcriptional regulatory landscape in STAD using an scRNA dataset (GSE134520). This dataset includes a total of 38,525 cells from 13 samples



A



B



C

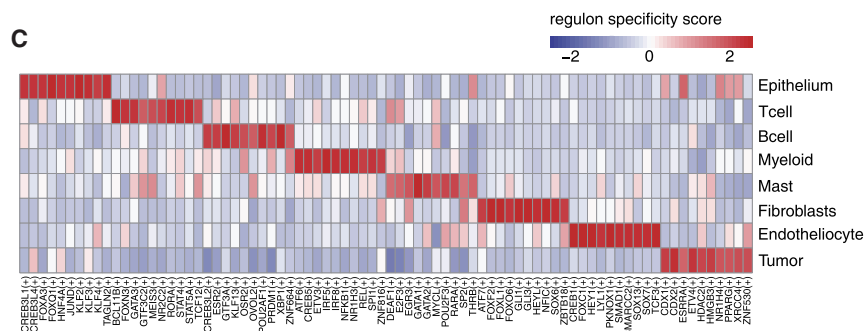


Figure 1. SCENIC analysis of human stomach adenocarcinoma single cell dataset

(A) The workflow of this study.

(B) The UMAP plot of 38,525 cells to visualize clusters colored by cell type.

(C) The heatmap demonstrates specific regulons across cell types.

The transcription factor (TF) regulatory activity (AUC cell score) was estimated by using pySCENIC. The color intensity was proportional to the scaled AUC cell score for each cell type, and the top 10 regulons of each cell type were shown.

representing the progressive histologic stages, i.e., 3 cases of non-atrophic gastritis (NAG), 3 cases of chronic atrophic gastritis (CAG), 6 cases of intestinal metaplasia (IM), and 1 case of early gastric cancer (EGC) (Figure 1A). The resulting 8 cell clusters were divided into 3 major cell lineages, including epithelial cells ($n = 30,687$), stromal cells ($n = 2,690$), and immune cells ($n = 5,148$) (Figure 1B). Subsequently, we employed the SCENIC algorithm to infer regulons for each cell subpopulation. By identifying co-expression modules between transcriptional regulators and candidate target genes, the algorithm constructs the underlying regulatory circuitry, defined as regulons (Figure 1C). We observed that some regulons were shared across cell types. Notably, many cell subpopula-

tions displayed subset-specific regulons. For instance, findings consistent with ours demonstrate that *FOXQ1* mRNA is localized to the gastric epithelium.¹⁰ Meanwhile, *ESR2* exhibits marked expression in B cells.¹¹ *POU2AF1* serves as a pivotal transcriptional coactivator for the POU2 family, modulating B cell development and malignant progression in B cells.¹² Our results indicate that regulons could capture the heterogeneous GRNs underlying cellular diversity in STAD.

Potential regulatory patterns in gastric epithelial cells associated with *Helicobacter pylori* infection

Helicobacter pylori colonizes the gastric epithelium and is a major risk factor for GC and precancerous lesions.^{13–15} Given

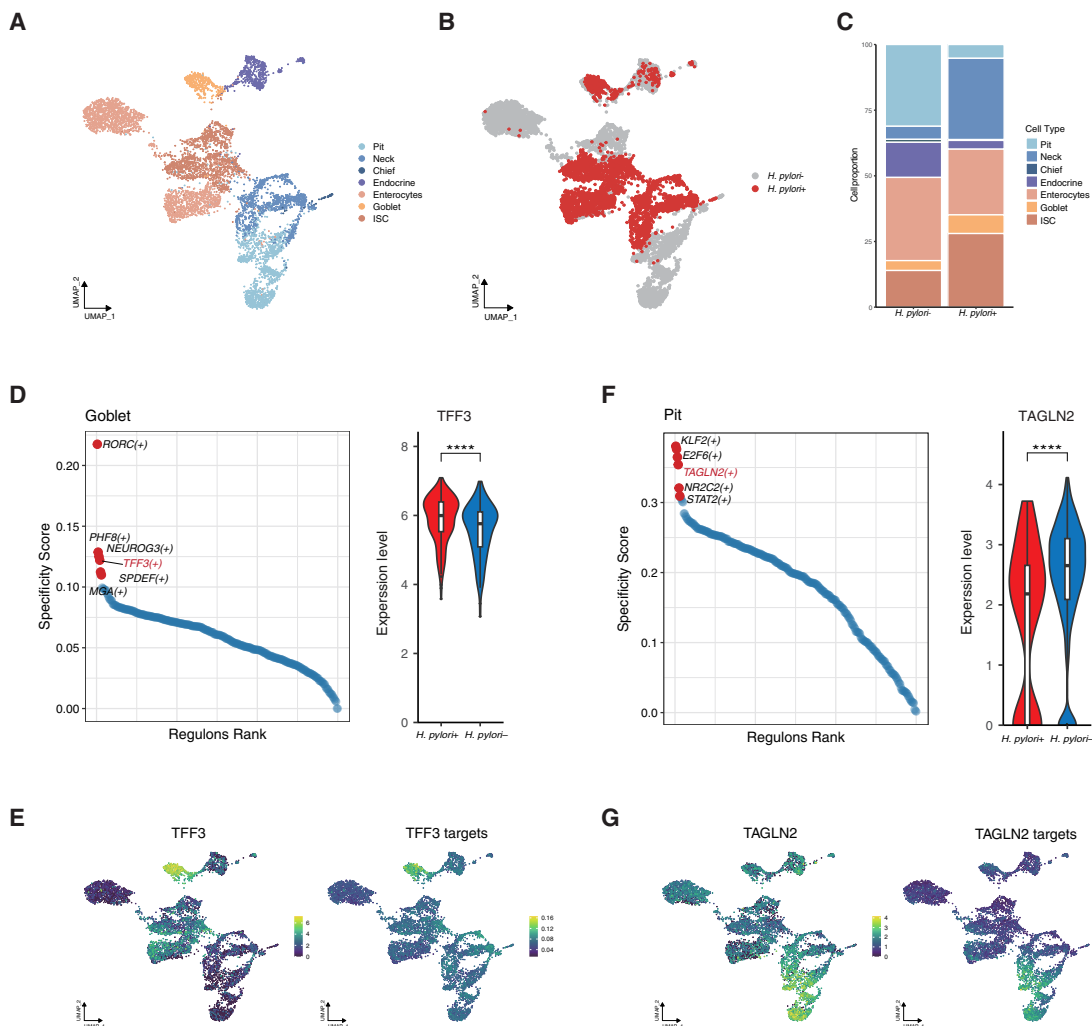


Figure 2. Epithelial-specific gene regulatory network in intestinal metaplasia

(A) UMAP plot of the major clusters of epithelial cells from 6 IM samples.

(B) UMAP plot of the epithelial cells from 3 *H. pylori*-infected IM samples and 3 uninfected IM samples.

(C) The proportion of epithelial cell types, with or without *H. pylori* infection.

(D and F) Left: Rank for regulons in goblet (D) and pit (F) cells based on the RSS. RSS, regulon specificity score. Right: Violin plots display the distribution of expression of *TFF3* (D) and *TAGLN2* (F), with or without *H. pylori* infection. Statistical significance was determined by the Wilcoxon rank-sum test. **** $p < 0.0001$.

(E and G) UMAP plots of the expression levels of transcriptional factor *TFF3* (left, E) and *TAGLN2* (left, G) and their target genes (right) in epithelial cells.

its impact on gastric epithelial cells, we investigated changes in transcriptional regulatory features of gastric epithelial cells associated with *H. pylori* infection. Based on *H. pylori* infection status from 6 IM samples, we categorized epithelial cells into *H. pylori*-positive and *H. pylori*-negative groups (Figures 2A and 2B). We observed apparent differences in gastric epithelial cell composition and global gene expression patterns between *H. pylori*-positive and *H. pylori*-negative samples (Figures 2C and S1, and Table S1). In particular, neck cells (4.98%–30.93%), goblet cells (3.77%–7.11%), and ISC cells (13.95%–28.03%) appeared more frequent in the *H. pylori*-infected biopsies, whereas chief cells (1.12%–0.34%) and pit cells (31.13%–5.25%) appeared less frequent (Figure 2C). We subsequently focused on the cell type-specific regulons

(Figure S2). Specifically, the *TFF3*-associated regulon exhibited elevated regulatory activity in goblet cells in *H. pylori*-infected biopsies (Figure 2D, 2E, and S1). Genes within this regulon, including *THBS2*, which has been previously linked to *H. pylori*-associated GC development and metastasis,¹⁶ showed goblet cell-enriched expression patterns (Figure 2E). Of note, among these regulon-associated genes, *FAM3D* and *ZG16B* displayed higher expression levels in *H. pylori*-infected samples (Figure S1). Meanwhile, the *TAGLN2*-associated regulon showed lower regulatory activity in pit cells (Figure 2F, 2G, and S1). To a certain extent, our study suggested an association between *H. pylori* infection and differences in cell type-specific regulatory states within gastric epithelial cells.

Development and validation of the STAD prognostic signature (SPS)

Moreover, we identified 175 transcriptional regulators that were preferentially associated with malignant epithelial cells. For example, *ETV4*, *ESRRA*, and other transcriptional regulators are significantly activated in malignant epithelial cells. Their overexpression in epithelial cells is implicated in the malignant progression of tumors (Figure 1C).^{17,18} Additionally, Gene Ontology (GO) enrichment analysis revealed that these transcriptional regulators were significantly enriched in gene sets related to transcription factor activation/inhibition (Figure S3A). We also used the MSigDB Hallmark gene sets and found the TNFA signaling via the NFKB gene set to be most significantly affected (Figure S3B).

Performing univariate Cox regression analysis on the 175 transcriptional regulators, setting $p < 0.05$ as the significance threshold, we identified 23 of them that were significantly associated with the prognosis of patients with GC (Figure 3A). Based on these 23 malignant epithelial-specific transcriptional regulators, we developed a prognostic signature for patients with STAD, defined as the STAD prognostic signature (SPS), using a machine learning-based integrative approach (Figure S4A). The TCGA dataset was designated as the training set, whereas the ACRG, KUCM, and YUSH datasets were employed as validation sets. By evaluating 117 combinations of 10 machine learning algorithms, we ultimately selected SuperPC to balance model complexity and predictive accuracy. This model achieved a mean C-index of 0.61 across all cohorts. The C-indices for the test cohorts ACRG, KUCM, and YUSH were 0.60, 0.62, and 0.64, respectively, all exceeding that of the discovery cohort (0.58), indicating robust performance (Figure S4B). By combining the risk score with the follow-up data, we stratified patients into high-risk or low-risk groups using the median value as the cutoff. Among the 376 patients with STAD with complete survival data in the TCGA, 188 were predicted to be high-risk, and 188 were predicted to be low-risk. The high-risk group demonstrated significantly worse overall survival (OS) than the low-risk group (Figure 3B, 5-year survival rate 19.5% vs. 48.4%, HR = 1.50, 95% CI: 1.09–2.09, log rank $p = 7.11 \times 10^{-3}$). To verify the robustness of the prognostic power of SPS, we employed a multivariate Cox regression model. Even after further adjusting for clinical and pathological factors such as tumor stage, gender, and age, the prognostic signature retained its significance as an independent prognostic factor ($p < 0.05$, Figure 3C). Then, we validated the prognostic potency of SPS in 3 independent validation datasets. In the Asian Cancer Research Group (ACRG) cohort with 300 patients with STAD, individuals in the high-risk group exhibited significantly worse OS compared to those in the low-risk group (Figure 3D, 5-year survival rate 43.2% vs. 61.3%, HR = 1.88, 95% CI: 1.35–2.63, log rank $p = 6.40 \times 10^{-4}$). Similar results were observed in the Kosin University College of Medicine (KUCM) cohort with 109 patients (Figure 3E, 5-year survival rate 40.3% vs. 60.7%, HR = 1.34, 95% CI: 0.77–2.33, log rank $p = 0.024$) and in the Yonsei University Severance Hospital (YUSH) cohort with 64 patients (Figure 3F, 5-year survival rate 40.6% vs. 75.0%, HR = 3.46, 95% CI: 1.46–8.21, log rank $p = 6.44 \times 10^{-3}$). Similarly, multivariate analysis with adjustment for clinical and pathologic factors in

validation cohorts further confirmed the independent prognostic value of SPS (Figure S5). Furthermore, across both the training and validation cohorts, the SPS_stage yielded a modest but consistent improvement in predictive performance compared to the robust clinical staging system. This enhancement was evidenced by an incremental gain in the C-index (mean $\Delta_{C\text{-index}}$ was 0.017) (Table S2), as well as relative advantages observed in time-dependent ROC (Figure S6), calibration (Figure S7), and decision curve analysis (DCA) (Figure S8), driven primarily by improved long-term prognostic performance.

Biological characteristics and preliminary validation of the SPS

We investigated the association between SPS and immune cell infiltration by CIBERSORTx and ssGSEA algorithms. As shown in Figure S9, we observed significant differences in immune cell infiltration patterns between high- and low- SPS groups. Specifically, the high SPS group exhibited reduced levels of immune-enhancing immune cells, including activated CD4⁺ T cells, and significantly elevated levels of immune-suppressive immune cells, such as mast cells, macrophages, and Treg cells. In addition, significant differences in eight cancer hallmark features were identified between the high- and low-SPS groups, with the high group showing higher activity (Figure S10A). Notably, the activities of the 23 SPS regulators were significantly correlated with these cancer hallmark pathways (Figure S10B).

To further substantiate the reliability of the SPS, we investigated the role of *GLIS2*, a key transcriptional regulator in our model. First, *GLIS2* exhibited significantly higher expression in GC samples compared to normal samples in both TCGA and GSE54129 cohorts (Figure S11A). We then silenced *GLIS2* expression in AGS GC cells using siRNAs targeting *GLIS2*. We performed RT-qPCR to confirm the knockdown efficiency at the mRNA level (Figure S11B). Further cell proliferation assays, including CCK-8 and EdU assays, demonstrated that the knockdown of *GLIS2* significantly inhibited cell proliferation (Figures S11C and S11D). These functional findings suggest that *GLIS2* acts as an oncogene in GC progression. Moreover, we investigated the downstream target genes of *GLIS2*. Based on SCENIC output and predictions from other databases (e.g., GTRD and ChIP-Atlas), *SYNPO* and *SH2B3* were suggested to be downstream targets of *GLIS2*. Consistent with this, we found a significant positive correlation between the expression of these genes and *GLIS2* (Figures S11E and S11F). Furthermore, RT-qPCR results showed that the expression levels of *SYNPO* and *SH2B3* were downregulated following *GLIS2* silencing (Figures S11E and S11F).

Comparison of the predictive performance of SPS with other prognostic signatures

We compared the predictive performance of SPS with 3 previously published prognostic signatures, including 6-gene signature,¹⁹ 5-gene signature,²⁰ and GPSGC.²¹ The results were summarized in Table 1 and Figure S5. Although the other 3 signatures exhibited distinct predictive performance for the OS across several datasets, the predictive effect of SPS was robust across all datasets, even after adjusting for other clinical and pathological factors.

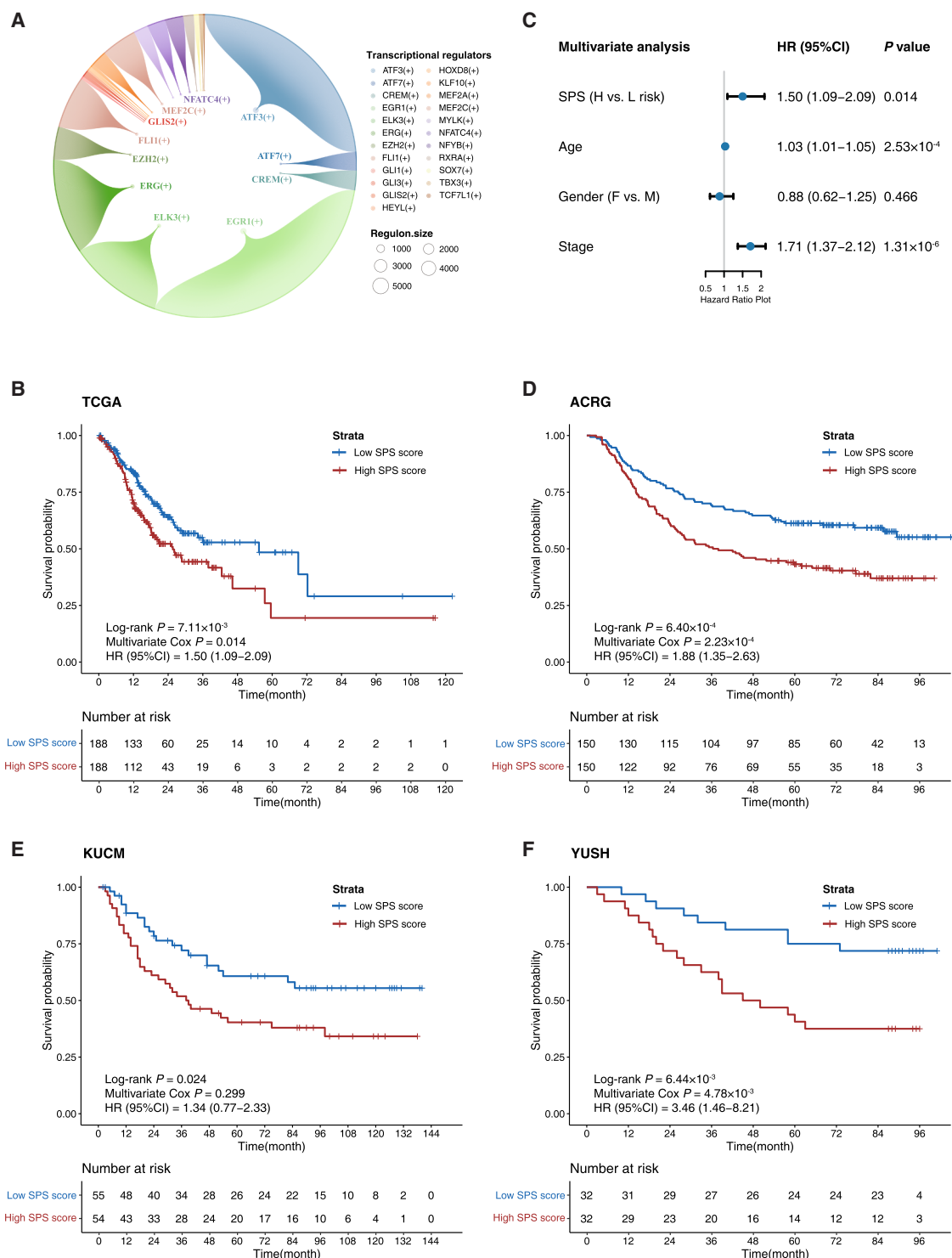


Figure 3. Development and validation of the SPS

(A) Identification of prognostic regulons. Large bubbles represent the transcriptional regulators, while the peripheral areas indicate their targets. Bubble size is proportional to the number of target genes within each regulon.

(B) Kaplan-Meier (KM) curve of the overall survival (OS) for patients between high- and low-risk groups in the TCGA discovery dataset.

(C) Multivariate analysis results assessing the prognostic efficiency of SPS. Solid dots represent the HR, and the closed-ended horizontal lines represent the 95% CIs. (D–F) Kaplan-Meier (KM) curves of OS for patients between high- and low-risk groups in three validation datasets: ACRG (D), KUCM (E), and YUSH (F). CI, confidence interval; HR, hazard ratio.

Table 1. Survival analyses of existing prognostic signatures in four datasets

Datasets	Signatures	Log rank <i>P</i>	C-index ^a	HR ^b	Cox <i>P</i> ^b
TCGA	SPS	7.11×10^{-3}	0.577	1.50	0.014
	6-gene signature	0.954	0.526	0.93	0.661
	5-gene signature	0.059	0.565	1.47	0.021
	GPSGC	0.012	0.599	1.62	0.004
ACRG	SPS	6.40×10^{-4}	0.603	1.88	2.23×10^{-4}
	6-gene signature	0.319	0.524	0.92	0.595
	5-gene signature	1.02×10^{-6}	0.647	1.97	9.77×10^{-5}
	GPSGC	1.58×10^{-3}	0.622	1.50	0.016
KUCM	SPS	0.024	0.614	1.34	0.299
	6-gene signature	0.197	0.599	1.07	0.814
	5-gene signature	0.015	0.650	1.24	0.457
	GPSGC	0.103	0.629	1.10	0.728
YUSH	SPS	6.44×10^{-3}	0.642	3.46	0.005
	6-gene signature	0.787	0.524	0.72	0.398
	5-gene signature	1.46×10^{-3}	0.733	4.44	9.90×10^{-4}
	GPSGC	0.541	0.599	1.59	0.220

C-index^a were generated by the univariate Cox proportional hazards model, HR^b and Cox *P*^b were generated by the multivariate Cox proportional hazards model.

Integrated prognostic signature by combining the SPS with clinical factors

In the multivariate Cox regression analyses, the tumor stage, age, and SPS were identified as independent prognostic factors in all datasets, indicating their complementary value (Figure S5). To further improve the predictive performance, we derived a new signature in the TCGA dataset by combining tumor stage, age, and SPS, which was calculated as $(0.41 \times \text{SPS}) + (0.53 \times \text{stage}) + (0.03 \times \text{age})$ and named as SPS_stage_age. In both discovery and validation datasets, we observed that the SPS_stage_age composite index yielded more pronounced prognostic distinctions compared to the binary SPS alone (Figures 4A–4D). Compared to individual factors such as tumor stage, or SPS alone, the integrated SPS_stage_age index exhibited enhanced accuracy in predicting survival across all datasets (Figure 4E).

DISCUSSION

Characterizing GRNs is essential in cancer research for elucidating both cancer prognosis and gene interactions.²² In this study, we employed scRNA-seq data from GSE134520 to construct a single-cell regulatory network atlas for STAD using the SCENIC algorithm and explored differences associated with *H. pylori* infection status within this framework. We further

developed a novel prognostic risk model termed the SPS and validated it in both training and validation datasets. Notably, this model enhanced the accuracy of OS estimation in STAD by integrating molecular and clinical features, thereby supporting more informed clinical decision-making.

In our study, we noted heterogeneity in cell-type-specific transcriptional regulation within the gastric epithelium associated with *H. pylori* infection. Notably, the inferred regulatory differences appear to be cell-type specific, with distinct regulatory programs observed across goblet cells, pit cells, and other epithelial lineages. Within this framework, the *TFF3*- and *TAGLN2*-associated regulons served as representative examples of infection-associated regulatory state shifts in different epithelial lineages, consistent with previous studies.^{23,24} While our analyses are inferential and do not establish direct mechanistic regulation, these observations provide insights into the regulatory patterns associated with *H. pylori* infection.

In addition, we established the SPS, a scoring index comprising 23 transcriptional regulators with prognostic significance. To explore the biological significance of this transcriptional regulatory signature, we examined the core constituent genes. Notably, several candidate transcriptional regulators have been implicated in GC tumorigenesis. For instance, existing studies indicate that *EGR1* promotes GC progression and metastasis by activating various downstream genes through distinct mechanisms, such as promoting cell cycle progression or reprogramming fatty acid metabolism.^{25,26} The E26 transformation-specific (ETS) transcription factor *ELK3* is required for the migration and invasion of GC cells.²⁷ *GLI1* acts as a direct transcriptional target and a reliable indicator of Hedgehog (Hh) signaling pathway,²⁸ a signaling cascade that is critical for tumorigenesis and sustaining cancer stemness. Previous research has indicated that *GLI1* can be activated by the AKT-mTOR pathway. The co-expression of *GLI1* and *p-AKT* is related with aggressive cellular functions such as proliferation, migration, and drug-resistance, and could serve as an indicator of poor prognosis in GC.²⁹ A study investigating the genetic susceptibility to GC revealed that *GLI3* promotes GC progression and correlates with adverse outcomes in patients.³⁰ Findings from these molecular mechanism studies could provide evidence for the reliability of our prognostic model.

In the TCGA training dataset, the SPS reveals that the five-year survival rate in the high-risk group was significantly lower than that of the low-risk group, and this result was consistent in the validation sets. Importantly, compared to the risk models proposed by Cho et al. (the 6-gene signature),¹⁹ Zhang et al. (the 5-gene signature),²⁰ and Cai et al. (gene expression-based prognostic signature (GPSGC)),²¹ our risk score (SPS) demonstrates superior accuracy in predicting patient survival. This highlights the robust predictive power of SPS for patient outcomes. Moreover, our analysis revealed that prognostic models integrating gene expression profiling demonstrated better predictive accuracy than the conventional clinical model utilizing only tumor stage information. By constructing malignant epithelial cell-specific transcriptional regulatory networks, we uncovered the regulatory essence of tumor progression and its association with clinical outcomes, while also eliminating the influence of tumor heterogeneity and varying tumor cell proportions.

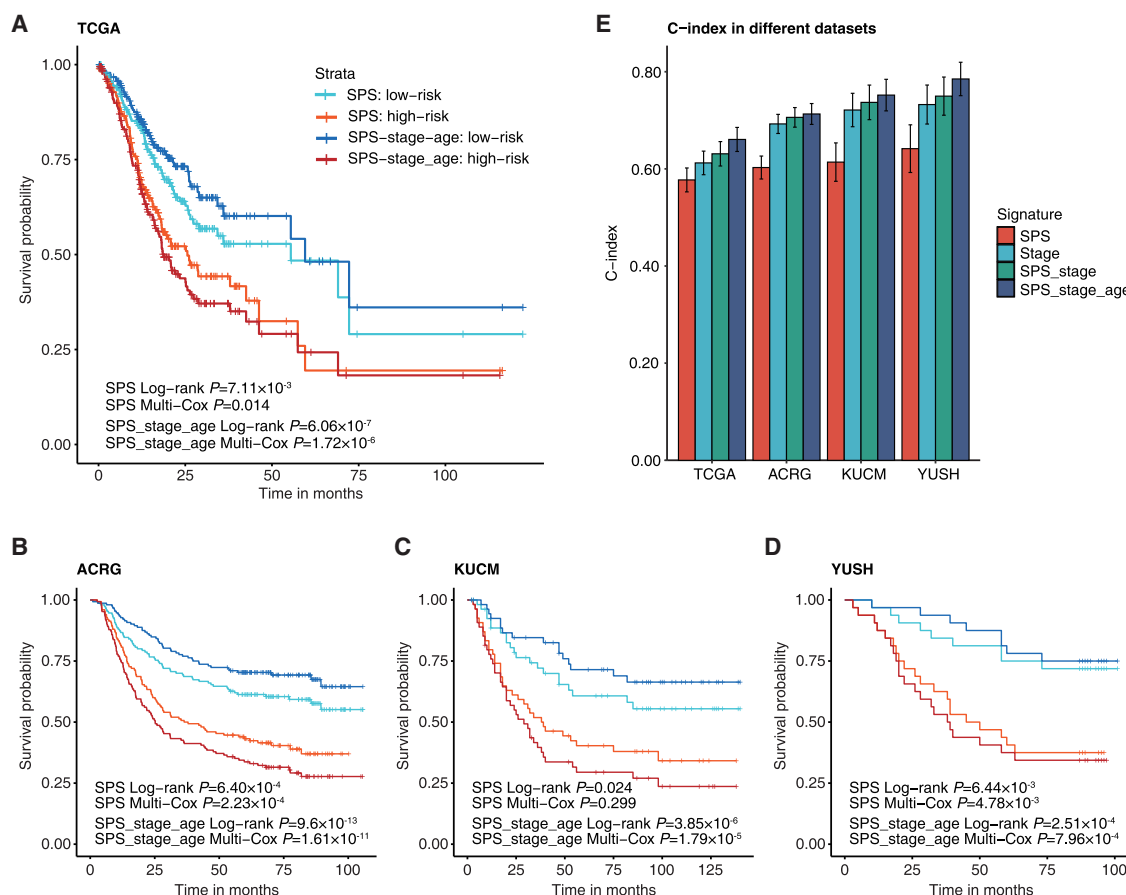


Figure 4. Performance of SPS_stage_age in discovery and validation datasets

(A–D) The KM curves of OS for patients between high- and low-risk groups determined by SPS and SPS_stage_age signatures in the TCGA discovery dataset (A), and independent validation datasets ACRG (B), KUCM (C), and YUSH (D).

(E) Comparison of the SPS_stage_age signature with SPS and tumor stage. Error bars in the bar graph represent the standard deviation of the C-index.

In conclusion, our findings depict a comprehensive landscape of the single-cell regulon atlas, uncovering the potential heterogeneity of GRNs within STAD tissues, and offer preliminary insights into *H. pylori*-associated epithelial alterations. We further developed an individualized prognostic signature using the malignant epithelial-specific transcriptional regulators for the survival stratification of patients with STAD. This signature holds promise as a valuable predictive tool for STAD survival, suggesting potential for further personalized clinical management of patients with STAD. Its ability to potentially inform treatment decisions highlights its translational significance.

Limitations of the study

Although we constructed a robust signature and validated its predictive efficiency, this study has several limitations. Firstly, refining a well-established prognostic standard such as the clinical stage is inherently challenging; however, the modest yet steady gains provided by the SPS_stage represent a meaningful marginal improvement in risk stratification. The sustained performance across multiple cohorts and metrics validates the poten-

tial of the SPS_stage as a reliable supplementary tool to enhance conventional staging, arguing against overfitting. Nevertheless, the prognostic efficacy of SPS was tested in retrospective cohorts, and its clinical utility requires independent validation in larger, prospective cohorts of patients with STAD. Additionally, the generalizability of the findings might be constrained by the sample size in this study, highlighting the need for future research involving larger-scale validation to confirm the model's robustness. Secondly, patient prognosis may also be influenced by other clinicopathological variables, such as treatment regimens, tumor burden, and comorbidities. Unfortunately, similar to most large-scale databases, detailed information on these factors is unavailable in the datasets used in our study. Further research incorporating these variables is warranted to validate and expand upon our findings. Thirdly, regarding regulon inference, while SCENIC is a powerful tool, the regulators it identifies remain computational predictions. To bolster the reliability of our inferences, we provided supporting evidence by incorporating validation from multiple datasets as well as performing some functional experiments on the key regulator *GLIS2*. However, further experimental validation is warranted to confirm the

underlying mechanisms in future studies. Regarding *H. pylori* infection-related stratification, we acknowledge the current absence of single-cell validation cohorts and the inherent challenges in performing functional validation at the single-cell level. Consequently, while our data highlight significant alterations in *TFF3* and *TAGLN2*, their clinical implications warrant cautious interpretation. To further solidify these insights, future efforts should prioritize epithelial cell subclassification through both advanced data analytics and organoid experiments. Fourthly, the immediate clinical application of the 23-gene signature may face challenges. However, this model represents an optimal trade-off between complexity and prognostic accuracy, as reducing the gene count would compromise performance. While current transcriptomic platforms make such panels increasingly feasible, cost and implementation remain critical barriers. Therefore, developing simpler, clinically tractable alternatives, such as IHC panels focusing on key regulators, is a crucial direction for future research to enhance translational potential. Finally, despite the fact that the scRNA-seq regulatory network prognostic model outperforms conventional transcriptomic features, the inherent limitations in information content and dimensionality may compromise the accuracy and comprehensiveness of inferred gene regulatory interactions compared to single-cell multi-omics. Specifically, during GC development, the transcriptional dysregulation of specific genes may also be influenced by epigenetic mechanisms such as histone modifications, DNA methylation, and non-coding RNAs.³¹

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Yanbing Ding (ybding@yzu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The datasets employed in this work can be accessed via the TCGA (<https://portal.gdc.cancer.gov/>) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>). Detailed information about datasets is provided in Table S3. The code for our main analysis is available at https://github.com/AP-Zhang/Regulon-based_Survival_Pipeline. Further details and additional supporting data are available upon request from the corresponding authors.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (no. 82273084), the NSFC Young Scientists Fund (no. 82404358), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_3569), and the Health Key Project of the Basic Research Program of Yangzhou (Joint Special Project, No. 2024-1-05).

AUTHOR CONTRIBUTIONS

A.Z.: conceptualization, data curation, formal analysis, methodology, visualization, writing - original draft, and writing - review and editing. Z.L.: data curation, formal analysis, methodology, visualization, and writing - review and editing. X.J.: validation. Y.L.: funding acquisition and supervision. M.Z.: funding acquisition and supervision. Y.D.: funding acquisition, project administration, and supervision. All authors have read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used Zhipu Qingyan (powered by ChatGLM) in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Data collection
 - Cells
- METHOD DETAILS
 - Pre-processing of single-cell RNA sequencing data
 - Identification of malignant epithelium
 - Transcriptional regulatory networks analysis
 - Gene set enrichment analysis
 - Construction of a gene signature from prognostic regulons
 - Evaluation of model predictive performance and clinical utility
 - Assessment of immune cell infiltration and tumor-related features enrichment
 - Transfections
 - Real-time quantitative PCR (RT-qPCR)
 - Cell proliferation assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2026.116100>.

Received: September 9, 2025

Revised: January 26, 2026

Accepted: May 8, 2026

REFERENCES

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 74, 229–263. <https://doi.org/10.3322/caac.21834>.
2. Karimi, P., Islami, F., Anandasabapathy, S., Freedman, N.D., and Kamangar, F. (2014). Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol. Biomarkers Prev.* 23, 700–713. <https://doi.org/10.1158/1055-9965.Epi-13-1057>.
3. Ilic, M., and Ilic, I. (2022). Epidemiology of stomach cancer. *World J. Gastroenterol.* 28, 1187–1203. <https://doi.org/10.3748/wjg.v28.i12.1187>.
4. Hayakawa, Y., Ariyama, H., Stancikova, J., Sakitani, K., Asfaha, S., Renz, B.W., Dubeykovskaya, Z.A., Shibata, W., Wang, H., Westphalen, C.B., et al. (2015). Mist1 Expressing Gastric Stem Cells Maintain the Normal and Neoplastic Gastric Epithelium and Are Supported by a Perivascular Stem Cell Niche. *Cancer Cell* 28, 800–814. <https://doi.org/10.1016/j.ccell.2015.10.003>.
5. Cao, L., Weng, K., Li, L., Lin, G., Zhao, Y., Gao, Y., Huang, X., Chen, Q., Wang, J., Zheng, C., et al. (2024). BATF2 inhibits the stem cell-like properties and chemoresistance of gastric cancer cells through

- PTEN/AKT/ β -catenin pathway. *Theranostics* 14, 7007–7022. <https://doi.org/10.7150/thno.98389>.
6. Iacono, G., Massoni-Badosa, R., and Heyn, H. (2019). Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* 20, 110. <https://doi.org/10.1186/s13059-019-1713-4>.
7. van Dam, S., Vösa, U., van der Graaf, A., Franke, L., and de Magalhães, J.P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. *Brief. Bioinform.* 19, 575–592. <https://doi.org/10.1093/bib/bbw139>.
8. Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50, 1–14. <https://doi.org/10.1038/s12276-018-0071-8>.
9. Dong, J., Wu, X., Zhou, X., Gao, Y., Wang, C., Wang, W., He, W., Li, J., Deng, W., Liao, J., et al. (2023). Spatially resolved expression landscape and gene-regulatory network of human gastric corpus epithelium. *Protein Cell* 14, 433–447. <https://doi.org/10.1093/procel/pwac059>.
10. Verzi, M.P., Khan, A.H., Ito, S., and Shivdasani, R.A. (2008). Transcription factor foxq1 controls mucin gene expression and granule content in mouse stomach surface mucous cells. *Gastroenterology* 135, 591–600. <https://doi.org/10.1053/j.gastro.2008.04.019>.
11. Peckham, H., Radziszewska, A., Sikora, J., de Gruijter, N.M., Restuadi, R., Kartawinata, M., Martin-Gutierrez, L., Robinson, G.A., Deakin, C.T., Wedderburn, L.R., et al. (2025). Estrogen influences class-switched memory B cell frequency only in humans with two X chromosomes. *J. Exp. Med.* 222, e20241253. <https://doi.org/10.1084/jem.20241253>.
12. He, T., Xiao, L., Qiao, Y., Klingbeil, O., Young, E., Wu, X.S., Mannan, R., Mahapatra, S., Redin, E., Cho, H., et al. (2024). Targeting the mSWI/SNF complex in POU2F-POU2AF transcription factor-driven malignancies. *Cancer Cell* 42, 1336–1351.e9. <https://doi.org/10.1016/j.ccell.2024.06.006>.
13. Lim, M.C.C., Jantaree, P., and Naumann, M. (2023). The conundrum of Helicobacter pylori-associated apoptosis in gastric cancer. *Trends Cancer* 9, 679–690. <https://doi.org/10.1016/j.trecan.2023.04.012>.
14. Lu, B., and Li, M. (2014). Helicobacter pylori eradication for preventing gastric cancer. *World J. Gastroenterol.* 20, 5660–5665. <https://doi.org/10.3748/wjg.v20.i19.5660>.
15. Gomaa, A., Maacha, S., Peng, D., Soutto, M., Genoula, M., Bhat, N., Cao, L., Zhu, S., Castells, A., Chen, Z., et al. (2024). SOX9 is regulated by AURKA in response to Helicobacter pylori infection via EIF4E-mediated cap-dependent translation. *Cancer Lett.* 593, 216939. <https://doi.org/10.1016/j.canlet.2024.216939>.
16. Qi, C., Shi, H., Fan, M., Chen, W., Yao, H., Jiang, C., Meng, L., Pang, S., and Lin, R. (2023). Microvesicles from bone marrow-derived mesenchymal stem cells promote Helicobacter pylori-associated gastric cancer progression by transferring thrombospondin-2. *Cell Commun. Signal.* 21, 274. <https://doi.org/10.1186/s12964-023-01127-y>.
17. Lu, Y.X., Ju, H.Q., Liu, Z.X., Chen, D.L., Wang, Y., Zhao, Q., Wu, Q.N., Zeng, Z.L., Qiu, H.B., Hu, P.S., et al. (2018). ME1 Regulates NADPH Homeostasis to Promote Gastric Cancer Growth and Metastasis. *Cancer Res.* 78, 1972–1985. <https://doi.org/10.1158/0008-5472.Can-17-3155>.
18. Li, F.N., Zhang, Q.Y., Li, O., Liu, S.L., Yang, Z.Y., Pan, L.J., Zhao, C., Gong, W., Shu, Y.J., and Dong, P. (2021). ESRRA promotes gastric cancer development by regulating the CDC25C/CDK1/CyclinB1 pathway via DSN1. *Int. J. Biol. Sci.* 17, 1909–1924. <https://doi.org/10.7150/ijbs.57623>.
19. Cho, J.Y., Lim, J.Y., Cheong, J.H., Park, Y.Y., Yoon, S.L., Kim, S.M., Kim, S.B., Kim, H., Hong, S.W., Park, Y.N., et al. (2011). Gene expression signature-based prognostic risk score in gastric cancer. *Clin. Cancer Res.* 17, 1850–1857. <https://doi.org/10.1158/1078-0432.Ccr-10-2180>.
20. Zhang, Y., Wu, X., Zhang, C., Wang, J., Fei, G., Di, X., Lu, X., Feng, L., Cheng, S., and Yang, A. (2020). Dissecting expression profiles of gastric precancerous lesions and early gastric cancer to explore crucial molecules in intestinal-type gastric cancer tumorigenesis. *J. Pathol.* 251, 135–146. <https://doi.org/10.1002/path.5434>.
21. Cai, W.Y., Dong, Z.N., Fu, X.T., Lin, L.Y., Wang, L., Ye, G.D., Luo, Q.C., and Chen, Y.C. (2020). Identification of a Tumor Microenvironment-relevant Gene set-based Prognostic Signature and Related Therapy Targets in Gastric Cancer. *Theranostics* 10, 8633–8647. <https://doi.org/10.7150/thno.47938>.
22. Xu, P., and Zhang, B. (2023). Multiscale network modeling reveals the gene regulatory landscape driving cancer prognosis in 32 cancer types. *Genome Res.* 33, 1806–1817. <https://doi.org/10.1101/gr.278063.123>.
23. Huang, K.K., Ma, H., Chong, R.H.H., Uchiyama, T., Lian, B.S.X., Zhu, F., Sheng, T., Srivastava, S., Tay, S.T., Sundar, R., et al. (2023). Spatiotemporal genomic profiling of intestinal metaplasia reveals clonal dynamics of gastric cancer progression. *Cancer Cell* 41, 2019–2037.e8. <https://doi.org/10.1016/j.ccell.2023.10.004>.
24. Kaise, M., Miwa, J., Tashiro, J., Ohmoto, Y., Morimoto, S., Kato, M., Ura-shima, M., Ikegami, M., and Tajiri, H. (2011). The combination of serum trefoil factor 3 and pepsinogen testing is a valid non-endoscopic biomarker for predicting the presence of gastric cancer: a new marker for gastric cancer risk. *J. Gastroenterol.* 46, 736–745. <https://doi.org/10.1007/s00535-011-0396-8>.
25. Liu, H.T., Liu, S., Liu, L., Ma, R.R., and Gao, P. (2018). EGR1-Mediated Transcription of lncRNA-HNF1A-AS1 Promotes Cell-Cycle Progression in Gastric Cancer. *Cancer Res.* 78, 5877–5890. <https://doi.org/10.1158/0008-5472.Can-18-1011>.
26. Tang, Y., Zhao, B., Wang, W., Chen, H., Zhang, J., Xie, Y., Chen, Y., Lin, F., Li, Y., Zhai, X., and Zhou, W. (2025). The EGR1-mediated lncRNA TENM3-AS1 potentiates gastric cancer metastasis via reprogramming fatty acid metabolism. *Mol. Cancer* 24, 165. <https://doi.org/10.1186/s12943-025-02341-7>.
27. Lee, M., Cho, H.J., Park, K.S., and Jung, H.Y. (2022). ELK3 Controls Gastric Cancer Cell Migration and Invasion by Regulating ECM Remodeling-Related Genes. *Int. J. Mol. Sci.* 23, 3709. <https://doi.org/10.3390/ijms23073709>.
28. Jiang, J. (2022). Hedgehog signaling mechanism and role in cancer. *Semin. Cancer Biol.* 85, 107–122. <https://doi.org/10.1016/j.semcancer.2021.04.003>.
29. Yao, Y., Zhou, D., Shi, D., Zhang, H., Zhan, S., Shao, X., Sun, K., Sun, L., Wu, G., Tian, K., et al. (2019). GLI1 overexpression promotes gastric cancer cell proliferation and migration and induces drug resistance by combining with the AKT-mTOR pathway. *Biomed. Pharmacother.* 111, 993–1004. <https://doi.org/10.1016/j.biopha.2019.01.018>.
30. Tao, B., Wang, Z., Wang, X., Song, A., Liu, J., Wang, J., Zhang, Q., Chen, Z., Wang, Z., Xu, W., et al. (2025). An inherited predisposition allele promotes gastric cancer via enhancing deubiquitination-mediated activation of epithelial-to-mesenchymal transition signaling. *J. Clin. Investig.* 135, e179617. <https://doi.org/10.1172/jci179617>.
31. Chen, H., Su, Y., Yang, L., Xi, L., Li, X., Lan, B., Liu, M., and Xuan, C. (2023). CBX8 promotes lung adenocarcinoma growth and metastasis through transcriptional repression of CDKN2C and SCEL. *J. Cell. Physiol.* 238, 2710–2723. <https://doi.org/10.1002/jcp.31124>.
32. Zhang, P., Yang, M., Zhang, Y., Xiao, S., Lai, X., Tan, A., Du, S., and Li, S. (2019). Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell Rep.* 27, 1934–1947.e5. <https://doi.org/10.1016/j.celrep.2019.04.052>.
33. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
34. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.
35. Bartfeld, S., Bayram, T., van de Wetering, M., Huch, M., Begthel, H., Kujala, P., Vries, R., Peters, P.J., and Clevers, H. (2015). In vitro expansion of human gastric epithelial stem cells and their responses to bacterial

- infection. *Gastroenterology* 148, 126–136.e6. <https://doi.org/10.1053/j.gastro.2014.09.042>.
36. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* 15, 2247–2276. <https://doi.org/10.1038/s41596-020-0336-2>.
 37. Suo, S., Zhu, Q., Saadatpour, A., Fei, L., Guo, G., and Yuan, G.C. (2018). Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. *Cell Rep.* 25, 1436–1445.e3. <https://doi.org/10.1016/j.celrep.2018.10.045>.
 38. Liu, H., Zhang, W., Zhang, Y., Adegboro, A.A., Fasoranti, D.O., Dai, L., Pan, Z., Liu, H., Xiong, Y., Li, W., et al. (2024). Mime: A flexible machine-learning framework to construct and visualize models for clinical characteristics prediction and feature selection. *Comput. Struct. Biotechnol. J.* 23, 2798–2810. <https://doi.org/10.1016/j.csbj.2024.06.035>.
 39. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 14, 7. <https://doi.org/10.1186/1471-2105-14-7>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
RPMI 1640 medium	VivaCell	C3010-0500
Fetal Bovine Serum (FBS)	BioChannel	BC-SE-FBS01
Penicillin-Streptomycin	Gibco	15140122
jetPRIME	Polypius	101000046
Cell Counting Kit-8 (CCK-8)	Dojindo	CK04-500T
EdU assay kit	RiboBio	C10310-1
Critical commercial assays		
TRIzol reagent	Invitrogen	15596018
PrimeScript RT reagent Kit with gDNA Eraser	TaKaRa	RR047A
TB Green Premix Ex Taq	TaKaRa	RR420A
Deposited data		
scRNA-seq data	GEO database	GSE134520
TCGA-STAD bulk RNA-seq data	NIH GDC data portal	https://portal.gdc.cancer.gov
microarray dataset (ACRG)	GEO database	GSE62254
microarray dataset (KUCM)	GEO database	GSE26901
microarray dataset (YUSH)	GEO database	GSE13861
Experimental models: Cell lines		
human gastric cancer cell line	IMMOCELL, Xiamen	AGS
Oligonucleotides		
See Tables S4 and S5	N/A	N/A
Software and algorithms		
R software	https://www.r-project.org	v4.3.3
inferCNV	https://github.com/broadinstitute/infercnv	v1.20.0
pySCENIC	https://github.com/aertslab/pySCENIC	v0.12.1
clusterProfiler	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html	v4.14.6
CIBERSORTx	https://cibersortx.stanford.edu	N/A
Survival	https://cran.r-project.org/web/packages/survival/index.html	v3.8.3
Survminer	https://cran.r-project.org/web/packages/survminer/index.html	v0.4.9
Seurat	https://cran.r-project.org/web/packages/Seurat/index.html	v4.4.0
GraphPad Prism	https://www.graphpad.com/	V10

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Data collection

We analyzed gene expression profiles of 858 patients from 5 stomach adenocarcinoma (STAD) cohorts, including one scRNA-seq dataset, one bulk RNA-seq dataset and 3 microarray datasets, as detailed in [Table S3](#). The STAD cancer scRNA-seq data (GSE134520)³² were downloaded from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) database for analysis.

The bulk RNA-seq data and clinical information of 376 STAD patients were obtained from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov>). We retained genes with median TPM greater than 1. We also evaluated the predictive efficiency of our signature in three human GC microarray datasets (GSE62254 [Asian Cancer Research Group (ACRG) cohort], GSE26901 [Kosin

University College of Medicine (KUCM) cohort]), and GSE13861 [Yonsei University Severance Hospital (YUSH) cohort] from the GEO database. Our study was restricted to STAD patients who had available transcriptomic data along with full clinical details, including age, gender, stage, and overall survival time. All gene expression values were log2-transformed ($\log_2(x + 1)$) within each dataset prior to survival analyses.

Cells

The AGS gastric cancer cell line (Sex: Female) was obtained from IMMOCELL (Xiamen, Fujian, China). AGS cells were cultured in RPMI 1640 medium containing 10% fetal bovine serum (FBS) and 1% antibiotics at 37°C in a humidified incubator of 5% CO₂.

METHOD DETAILS

Pre-processing of single-cell RNA sequencing data

The single-cell RNA sequencing dataset was processed using Seurat (v4.4.0).³³ To remove low quality cells and likely multiple captures, we filtered out cells with the following criteria: UMI counts <500 or >20,000, gene counts <400 or >4,500, and mitochondrial content >20%. After filtering, we retained 38,525 cells for downstream analysis. Key QC metrics (UMI, nFeature, and mitochondrial content) for each sample were evaluated to ensure data quality. We applied the `NormalizeData` and `ScaleData` functions to normalize and scale the gene expression data. Subsequently, we identified the highly variable genes (HVGs) using `FindVariableFeatures` to perform principal component analysis. We selected the top 30 principal components and performed clustering using `FindClusters`. Batch effect correction was performed with `Harmony`,³⁴ and dimensionality reduction was achieved via UMAP, yielding a two-dimensional embedding for visualizing cellular.

We then assigned cell type annotation to each cluster based on the expression of specific marker genes.³⁵ Specifically, immunological cells were categorized into B cells (*CD79A* and *MZB1*), T cells (*CD3D* and *CD2*), mast cells (*CPA3* and *TPSB2*), and myeloid cells (*CSF1R* and *CD14*). Stromal cells were identified by the expression of *DCN* and *COL1A2*. Epithelial cells were also identified using a range of markers, including *EPCAM* and *KRT18*, and further subclassified into various types, including neck cells (*MUC6*), pit cells (*MUC5AC* and *TFF1*), chief cells (*PGA3* and *PGA4*), parietal cells (*ATP4A* and *ATP4B*), endocrine cells (*CHGA* and *CHGB*), enterocytes (*FABP1* and *APOA1*), goblet cells (*MUC2* and *ITLN1*) and intestinal stem cells (*OLFM4*).

Identification of malignant epithelium

In this study, we calculated the copy number variations (CNVs) score of each gastric cell using the “inferCNV” v1.20.0 R package. The inferCNV analysis was performed with default parameters. Epithelial cells from normal gastric samples served as the reference. We conducted hierarchical clustering on the CNV scores of all genes via the k-means algorithm. We identified non-malignant epithelial cells as those with low CNV scores and clustered with reference cells. Conversely, cells exhibiting elevated CNV levels were designated as malignant.

Transcriptional regulatory networks analysis

To identify transcriptional regulatory networks in STAD, SCENIC algorithm was implemented by pySCENIC (Single-Cell rEgulatory Network Inference and Clustering, v0.12.1).³⁶ First, GENIE3 was employed to infer co-expression based regulatory relationships linking transcriptional regulators to candidate target genes. Motif enrichment analysis was then performed using RcisTarget to refine these relationships and identify putative direct regulatory interactions. Subsequently, regulon activity was quantified at the single-cell level using the AUCell algorithm, which calculates the Area Under the Curve (AUC) score for each regulon. The regulon specificity score (RSS) developed by Suo et al.³⁷ was utilized to assess the cellular specificity of the inferred regulons.

Gene set enrichment analysis

The Gene Ontology (GO) enrichment analyses, encompassing Biological Process (BP), Molecular Function (MF), and Cellular Component (CC), as well as pathway analyses using HALLMARK gene sets from the Molecular Signatures Database (MSigDB), were conducted using the R package clusterProfiler (v4.14.6). An adjusted *p*-value cutoff of <0.05 was applied to determine significant enrichment of GO terms and HALLMARK pathways, with the top 10 of each visualized.

Construction of a gene signature from prognostic regulons

Among malignant epithelial-specific regulons, we identified transcriptional regulators that were significantly associated with patients' overall survival (OS) by univariate cox regression analysis ($p < 0.05$). Subsequently, we performed an integrative machine learning analysis using the “mime” R package (v0.13)³⁸ to construct a risk signature based on these cell-type specific transcriptional regulators, which we termed the STAD prognostic signature (SPS). The prognosis of patients was predicted in the training dataset (TCGA) and three validation datasets (ACRG, KUCM, and YUSH) by evaluating a total of 117 combinations of 10 machine learning algorithms incorporating these regulators. We further calculate Harrell's concordance index (C-index) for each model on the TCGA, ACRG, KUCM, and YUSH datasets. The model exhibiting the highest mean C-index was identified as optimal. We then stratified the STAD cohort into SPS-high and SPS-low groups using the median score as the cutoff.

Evaluation of model predictive performance and clinical utility

The improvement in model fit and predictive accuracy was assessed by comparing the SPS_stage with the conventional clinical stage using C-index and time-dependent ROC analysis. Calibration curves were generated to assess the agreement between predicted survival probabilities and actual observations. Decision Curve Analysis (DCA) was employed to evaluate the clinical net benefit across a spectrum of threshold probabilities. All of these evaluations were performed in the training dataset (TCGA) and three validation datasets (ACRG, KUCM, and YUSH).

Assessment of immune cell infiltration and tumor-related features enrichment

To comprehensively compare the immune microenvironment and tumor-related pathway activities between SPS-high and SPS-low groups, we performed the following analyses. Immune cell infiltration was deconvoluted using CIBERSORTx (<https://cibersortx.stanford.edu>), an algorithm that estimates the relative abundances of 22 immune cell types via support vector regression. For mutual validation, immune cell proportions were also quantified by single-sample gene set enrichment analysis (ssGSEA).³⁹ In parallel, ssGSEA was applied to evaluate the activity of tumor-related pathways, generating corresponding enrichment scores for each sample.³⁹ Both the immune cell fractions and pathway enrichment scores were subsequently integrated for downstream comparative statistics and visualization.

Transfections

All control and target siRNAs were purchased from GenePharma (Shanghai, China). Transfection was performed using jetPRIME (Polyplus, France) according to the manufacturer's protocol. At 48 h post-transfection, cells were collected for subsequent analysis. The siRNA sequences were provided in the Table S4.

Real-time quantitative PCR (RT-qPCR)

Total RNA was extracted using TRIzol Reagent (Invitrogen, USA) following the manufacturer's instructions. The following cDNA was generated with PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa, Japan). Subsequently, RT-qPCR was performed in duplicate using TB Green Premix Ex Taq (TaKaRa) on LightCycler 480 (Roche, Switzerland). GAPDH served as an internal control for normalization. Relative gene expression changes were calculated using the $2^{-\Delta\Delta C_t}$ method. The RT-qPCR primer sequences were listed in the Table S5.

Cell proliferation assays

Cell proliferation in our study was assessed using the Cell Counting Kit-8 (CCK-8, Dojindo, Japan) and the Cell-Light EdU Apollo In Vitro Kit (EdU, RiboBio, China) according to the standard protocol. For the CCK-8 assay, the Infinite M200pro microplate reader (Tecan, Switzerland) was employed to measure absorbance at 450 nm at 0, 1, 2, 3, and 4 days. For the EdU assay, the percentage of EdU-positive cells was quantified by high-content analysis on the ArrayScan VTI system (Thermo Fisher Scientific, USA).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses in the study were performed using R software (v4.3.3; <https://www.r-project.org/>) and GraphPad Prism (v10). Survival analyses were conducted using the R packages "survival" (v3.8.3) and "survminer" (v0.4.9). Survival curves for overall survival (OS) of STAD patients were estimated using the Kaplan-Meier (KM) method, with statistical significance determined by the log rank test. The Cox proportional hazards regression model was utilized to assess the independent predictive value of the signature and estimate the hazards ratios (HRs) and 95% confidence intervals (95% CIs), after adjustment for age, gender, and tumor stage. All bootstrap validations were performed with 1,000 iterations to ensure model robustness. Correlations between continuous variables were evaluated using Spearman's rank correlation. Group comparisons were performed using two-sided Wilcoxon rank-sum test. All *in vitro* experiments were conducted with at least three independent biological replicates. Data are represented as mean $\pm$ SD. The corresponding figure legends describe the statistical test for each experiment, the number of data points (n) analyzed, what n represents, and the number of experimental replicates performed. For single-variable comparisons across multiple groups (e.g., RT-qPCR analysis), one-way ANOVA followed by Dunnett's post hoc test was employed. For CCK-8 proliferation assays, two-way ANOVA followed by Dunnett's multiple comparisons test was used. Statistical significance was defined as $p < 0.05$. In the figures, significance is denoted by the following symbols: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.