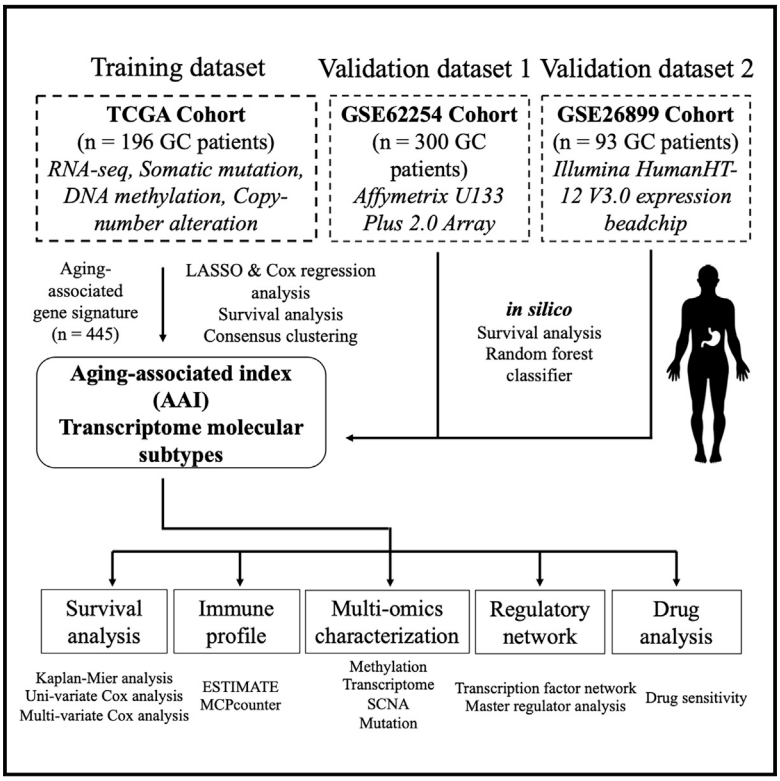
iScience

Article

Deciphering aging-associated prognosis and heterogeneity in gastric cancer through a machine learning-driven approach

Graphical abstract



Highlights

- AAI could predict prognosis, therapy response, and correlates with monocytes
- Aging-associated genes identified two molecular subtypes with unique features
- A subtype-specific regulatory network revealed *SOX7* and *ELK3* as druggable targets
- Patients with high *ELK3* expression may indicate sensitivity to FDA-approved drugs

Authors

Jiang Li, Chuanlai Yang, Yunxiao Zhang, Xiaoning Hong, Mingye Jiang, Zhongxu Zhu, Jiang Li

Correspondence

zhuzhongxu@him.cas.cn (Z.Z.), lijiang29@mail.sysu.edu.cn (J.L.)

In brief

Drugs; Chemistry; Computer science

Li et al., 2025, iScience *28*, 112316
May 16, 2025
https://doi.org/10.1016/j.isci.2025.112316

CellPress

iScience

CellPress
OPEN ACCESS

Article

Deciphering aging-associated prognosis and heterogeneity in gastric cancer through a machine learning-driven approach

Jiang Li,[1,6] Chuanlai Yang,[2,6] Yunxiao Zhang,[1,3] Xiaoning Hong,[1] Mingye Jiang,[1] Zhongxu Zhu,[4,*] and Jiang Li[1,5,7,*]
[1]Clinical Big Data Research Center, The Seventh Affiliated Hospital, Sun Yat-Sen University, Shenzhen, China
[2]Department of Science and Technology, The Second Affiliated Hospital of Soochow University, Soochow, China
[3]Department of Andrology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, China
[4]Biomics Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, China
[5]Shenzhen Key Laboratory of Chinese Medicine Active Substance Screening and Translational Research, Guangdong, Shenzhen, China
[6]These authors contribute equally
[7]Lead contact
*Correspondence: zhuzhongxu@him.cas.cn (Z.Z.), lijiang29@mail.sysu.edu.cn (J.L.)
https://doi.org/10.1016/j.isci.2025.112316

SUMMARY

Gastric cancer (GC) is a prevalent malignancy with a high mortality rate and limited treatment options. Aging significantly contributes to tumor progression, and GC was confirmed as an aging-related heterogeneous disease. This study established an aging-associated index (AAI) using a machine learning-derived gene panel to stratify GC patients. High AAI scores associated with poor prognosis and indicated potential benefits from adjuvant chemotherapy, while showing resistance to immunotherapy. Single-cell transcriptome analysis revealed that AAI was enriched in monocyte cells within the tumor microenvironment. Two distinct molecular subtypes of GC were identified through unsupervised clustering, leading to the development of a subtype-specific regulatory network highlighting *SOX7* and *ELK3* as potential therapeutic targets. Drug sensitivity analyses indicated that patients with high *ELK3* expression may respond to FDA-approved drugs (axitinib, dacarbazine, crizotinib, and vincristine). Finally, a user-friendly Shiny application was created to facilitate access to the prognostic model and molecular subtype classifier for GC.

INTRODUCTION

Gastric cancer (GC) is one of the most lethal malignancies globally, ranking as the fifth leading cause of cancer-related mortality. Each year, over a million new cases are diagnosed in East Asia alone.[1] The high mortality rate of GC is largely due to its aggressive tumor progression, diverse clinical manifestation, and limited response to treatment. Although several interventions have been developed to reduce GC-associated mortality, including endoscopic detection, gastrectomy, chemotherapy (CT), and chemoradiotherapy (CRT), the outcomes remain unfavorable.[2–4] Accurate prediction models for stratifying GC patients into distinct risk subgroups are still lacking. Given the limited treatment options and poor prognosis for GC, there is an urgent need to identify and validate novel therapeutic targets to improve patient outcomes. Developing effective models to stratify patients into distinct risk groups is crucial for enabling more targeted therapies for this highly heterogeneous cancer.

With the global population aging rapidly, understanding the fundamental mechanisms driving aging has become essential. Recently, a revised set of twelve hallmarks of aging was proposed, providing crucial insights into this complex biological process.[5] These hallmarks encompass disruptions at multiple levels, including genetic regulation across genomic, epigenetic, cellular, and organismal scales that collectively drive the functional decline associated with aging.[5,6] Aging, a major risk factor for cancer-related mortality, exhibits significant associations with several of these hallmarks, such as genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, impaired macroautophagy, deregulated nutrient-sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication, chronic inflammation, and dysbiosis. Each of these aging hallmarks represents a potential target for novel anti-aging interventions aimed at promoting healthy longevity and reducing cancer risk in older adults. Notably, aging and cancer share fundamental mechanisms, with age-related tissue changes fostering an environment that supports tumor progression.[7] Therefore, elucidating the biological mechanisms underlying aging could provide a unique perspective on improving cancer prognosis and developing novel treatments. A recent pan-cancer study revealed that GC exhibits an age-associated profile, showing that overall survival (OS) in GC patients is inversely correlated with age and that significant molecular differences exist across different age groups.[8] Despite this, the relationship between aging and GC remains

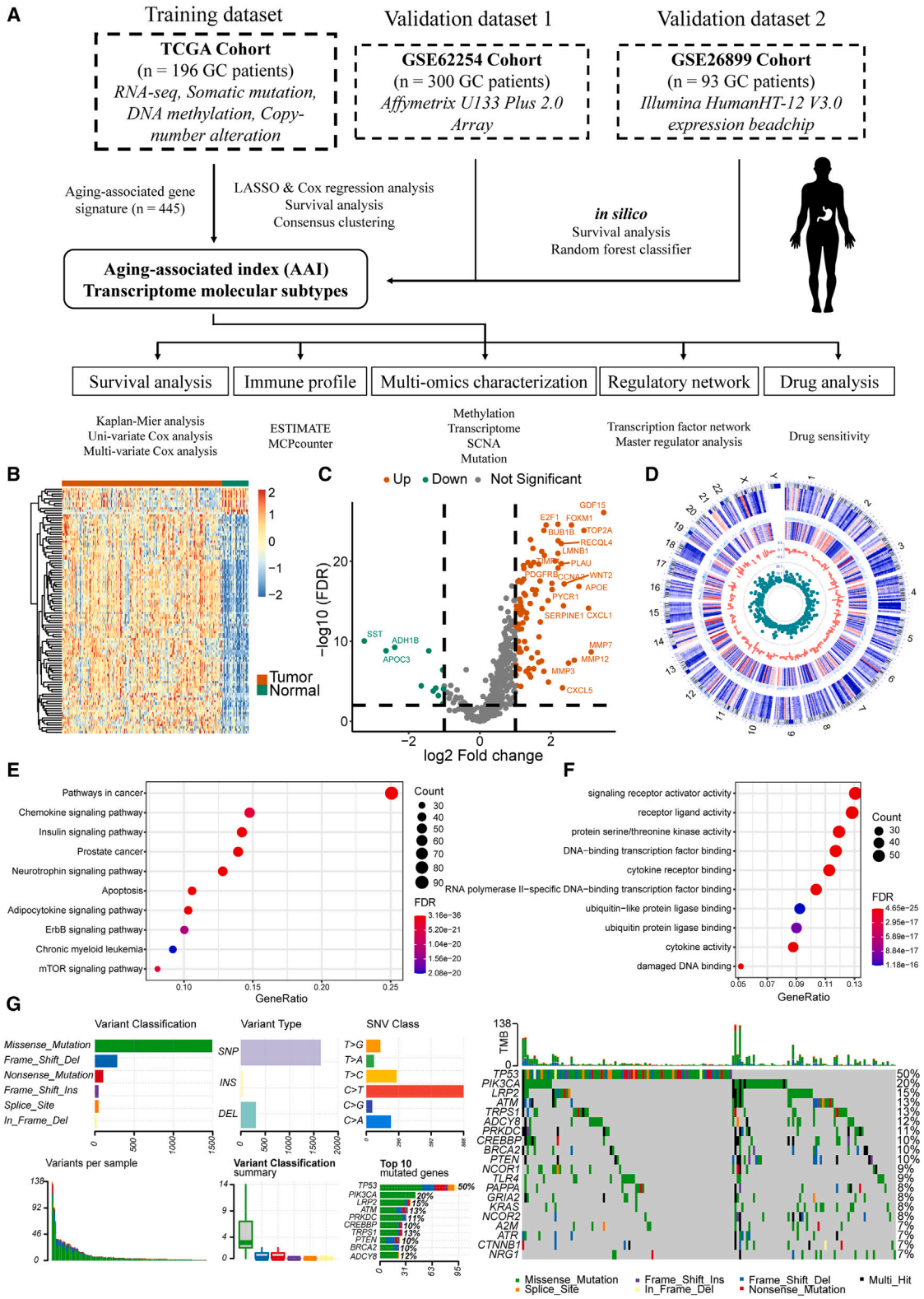


Figure 1. The schematic workflow of this study, and the variate landscape of aging-associated genes in GC patients

(A) Flowchart for comprehensive analysis of aging patterns in GC patients.

(B) Heatmap of differentially expressed aging-associated genes between tumor and adjacent normal samples in the TCGA-STAD dataset.

(legend continued on next page)

largely unexplored, and the specific roles of aging-associated patterns in GC pathogenesis have not been thoroughly investigated.

Here, we leveraged an aging-associated signature to establish a novel prognostic indicator, the aging-associated index (AAI), and further characterized two distinct molecular subtypes of GC with unique clinical and molecular features. These findings were independently validated across multiple public datasets. Furthermore, we used subtype-specific regulatory network inference to reveal key transcription factor (TF) master regulators driving GC invasion and metastasis, highlighting putative therapeutic targets. Moreover, drug sensitivity analysis targeting of these master regulatory TFs prioritized several FDA-approved drug candidates that could offer more precise treatments for GC subgroups. In summary, our study dissects the heterogeneity of GC and assesses its clinical prognosis, providing insights that will facilitate the selection of effective therapeutic regimens for GC patients.

RESULTS

The aging-associated signature exhibits distinct landscape patterns in GC patients

We analyzed 445 aging-associated genes from the Aging Atlas database and 196 GC tumor samples with multi-omics profiles from TCGA-STAD as a training dataset. Two independent validation cohorts with gene expression profiles and clinical information were collected from Gene Expression Omnibus (GEO) (GSE62254, n = 300, and GSE26899, n = 93). Clinical information for GC patients in all three datasets is summarized in Table S1. The schematic workflow of this study is illustrated in Figure 1A. To systematically characterize aging-associated genes, the differential analysis between tumor and adjacent normal samples was performed on TCGA-STAD training dataset, identifying 103 differentially expressed aging-associated signature genes (|log2 fold change| > 1 and BH-adjusted $p < 0.01$) (Table S2). Among these, 92 genes were upregulated and 11 were downregulated in GC tumor samples (Figures 1B and 1C).

A circus plot embedding chromosomal location, expression, DNA methylation levels, log2 fold change, and p values from the differential analysis of each gene highlighted the diverse patterns of the aging-associated signature across multiple genetic layers (Figure 1D). Moreover, the functional annotation of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) items indicated that the 445 aging-associated genes were significantly enriched in biological pathways, including NF-κB signaling pathway, cellular senescence, and DNA damage binding (Figures 1E and 1F). Genetic mutation analysis revealed that approximately 90% of GC samples in TCGA-STAD carried variant mutations, with *TP53* possessing the highest mutation rate (~50%) and other genes showing mutation frequencies ranging from 7% to 20% (Figure 1G).

A prognostic risk model based on a prioritized aging-associated signature in GC

To prioritize a gene signature for predicting the prognosis of GC patients, we applied a univariate Cox regression analysis to identify genes with significant survival associations. A total of 54 genes were selected with a $p < 0.1$ cutoff in the training dataset. We then employed the least absolute shrinkage and selection operator (LASSO) regression model with 100 permutations to refine the panel, retaining genes that appeared in more than 85 iterations, resulting in a final 20-gene panel for building the prognostic model (Figures 2A and 2B).

To assess aging-associated risk in GC patients, we built a multivariate Cox regression and introduced the aging-associated index (AAI), calculated as follows: AAI = (*AREG* * −0.03518909) + (*AXL* * −0.109384956)+ (*BDNF* * −1.14196311) + (*BIRC3* * −0.34023338) + (*CAT* * 0.49831949) + (*CETP* * 0.81387879) + (*CREB3L3* * 0.11688062) + (*EGF* * 0.53775648) + (*FGF23* * −1.42946464) + (*IL7* * −0.23905160) + (*MLH1* * 0.06625251) + (*PEX5* * −0.86896108) + (*POU1F1* * 2.45172075) + (*PTGES* * −0.16415875) + (*SERPINE1* * 0.22980775) + (*SIRT7* * −0.2548 1754) + (*SOCS2* * −0.05342394) + (*TCF3* * −0.30544728) + (*TNFRSF1A* * 1.23092773) + (*UCP1* * −4.16902399). AAI scores varied significantly in key clinical factors such as tumor stage (stage I-IV), lymph node metastasis (N0 and N1-3), tumor metastasis (M0 and M1), and survival status (alive or death) in TCGA-STAD (Figure 2C) as well as in validation cohorts GSE62254 (Figure S1A) and GSE26899 (Figure S1B). Several aging-associated genes showed significant differential expression between the high- and low-AAI groups (Figure 2D).

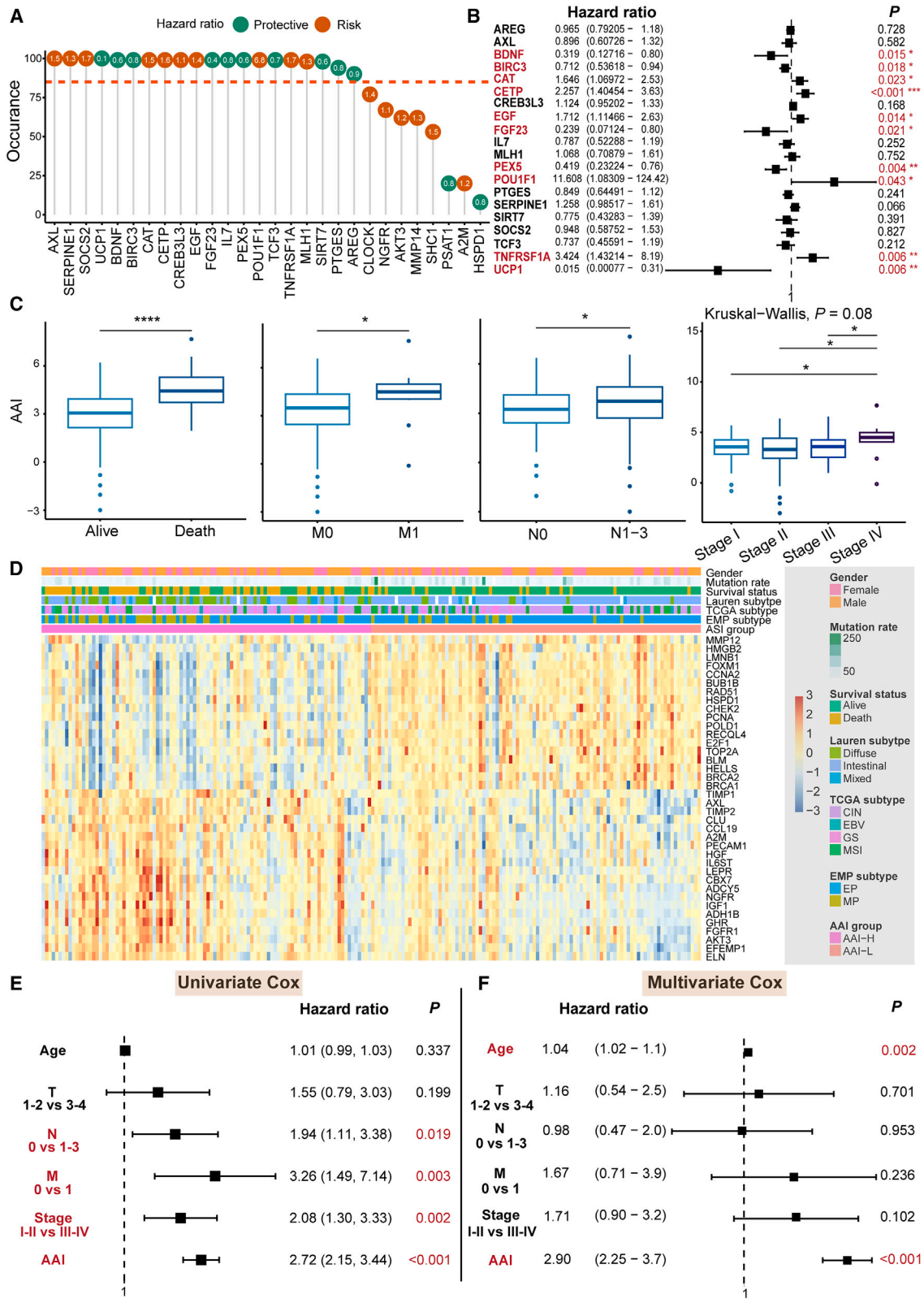
Moreover, AAI was consistently identified as an independent prognostic indicator in GC patients, even when adjusted for other clinical factors (univariate and multivariate Cox, $p < 0.001$) (Figures 2E and 2F). To further confirm the clinical and therapeutic potential of AAI, we performed a stratified survival analysis, and the results showed that high-AAI patients significantly benefited from the adjuvant chemotherapy (Figures S2A and S2B), while low-AAI patients showed no significant differences (Figures S2C and S2D). And similar results were observed from an anti-PD1/PD-L1 therapy cohort, where the high-AAI group displayed a significantly poorer OS (Figure S2E). Additionally, patients with higher AAI scores tended to be more resistant to the anti-PD-1/PD-L1 therapy in the IMvigor210CoreBiologies dataset, with significant enrichment in group with progressive or stable disease status ($p = 4.44 \times 10^{-3}$, Chi-square test) (Figure S2F). These findings highlighted the great potential of AAI as an effective indicator for guiding chemotherapy and immunotherapy decisions. Overall, AAI could serve as a quantitative risk metric and a promising prognostic indicator for assessing clinical outcomes in GC patients.

(C) Volcano plot showed the DEGs between tumor and adjacent normal samples. Points with gene symbols are obvious DEGs (BH-adjusted $p < 0.001$ and |log2 fold change| > 3).

(D) The genomic location, gene expression profiles, and DNA methylation levels of aging-associated DEGs in the TCGA-STAD dataset.

(E and F) Functional annotations on DEGs of (E) GO terms and (F) KEGG analysis.

(G) The somatic mutation landscape of aging-associated genes. The top20 mutated genes were presented in the oncoplot.



(legend on next page)

Validation of the aging-associated prognostic prediction model in bulk and single-cell datasets

To assess the prognostic value of AAI scores in GC patients, we stratified patients based on the median AAI and created Kaplan-Meier (KM) plots to compare survival difference using the log rank test. Our results indicated that GC patients with higher AAI scores had significantly lower survival probabilities, with poorer OS and disease-free survival (DFS) ($p < 0.05$, log rank tests) in the TCGA-STAD training dataset (Figure 3A).

We further validated AAI's prognostic capability in two independent GC datasets, GSE62254 and GSE26899, using median AAI values specific to each dataset. Consistent with the training dataset, patients in the high- AAI subgroup had a shorter OS and relapse-free survival (RFS) (all $p < 0.05$, log rank tests) in both GSE62254 (Figure 3B) and GSE26899 (Figure 3C).

Additionally, we comprehensively explored AAI in a single-cell dataset (GSE163558) from seven GC patients. Major cell types were annotated based on known marker genes (Figures 3D and S3A). Dot plot showed that AAI was highly composed in macrophages and monocytes (Figures 3E and S3B), highlighting a strong association between AAI and tumor immunity, particularly given the role of these myeloid-derived cells. Notably, AAI scores varied significantly in GC cells based on metastatic status (Figure 3F), underscoring AAI's potential as a predictive tool for tumor progression and metastasis, with higher AAI values observed in advanced-stage GC patients.

Identification of two molecular subtypes with distinct biological and clinical characteristics

To explore the heterogeneity within GC tumors, we utilized unsupervised clustering on the 54 aging-related prognostic genes identified in the training dataset. Through consensus clustering and optimal cluster analysis, we identified two aging-related molecular subtypes, Cluster1 and Cluster2 (Figures 4A and S4A–S4C; Table S3), with the optimal number confirmed by the NbClust algorithm (Figure S4D). Silhouette and principal component analysis (PCA) analyses further demonstrated distinct expression profiles between the two subtypes (Figures 4B and 4C).

Notably, the survival analysis revealed that Cluster2 patients had significantly poorer OS and DFS compared to Cluster1 ($p < 0.05$, log rank test) (Figures 4D and 4E). We then compared the aging-related subtypes with other subtyping systems, and found that most of Cluster2 tumors were enriched in the mesenchymal phenotype (MP), genomic stability (GS), and diffuse clusters ($p < 0.05$, Chi-square tests) (Figure 4F).

To validate the robustness and reproducibility of the unsupervised clustering, we built a random forest classifier based on the top 1000 candidate genes with high median absolute deviation (MAD) across TCGA-STAD samples, then applied it to the independent GSE62254 and GSE28699 datasets. Consistently, patients in Cluster1 subgroup exhibited significantly favorable survival outcomes across both validation datasets (all $p < 0.05$, log rank tests), while Cluster2 displayed a strong correlation with MP and diffuse subtypes ($p < 0.05$, Chi-square test) (Figures 4G–4L).

Distinct molecular landscapes of GC molecular subtypes identified through multi-omics characterization

To systematically dissect the clinical and biological characteristics of the two molecular subtypes identified, we conducted a multi-omics analysis on the TCGA-STAD cohort. Cluster2, the subtype associated with poorer prognosis, showed significantly higher AAI scores in both TCGA-STAD and GSE62254 cohorts (Figures S5A–S5C). Strikingly, patients in Cluster2 subtype tended to be younger than those in Cluster1 (Figure 5A), and this observation was further validated in the GEO dataset (GSE62254 and GSE26899) (Figures S5D and S5E), indicating that early-onset patients may have a predisposition to develop into this aggressive GC subtype.

We further found that Cluster2 was characterized by distinct molecular features, including hypermethylation (Figure 5B) and a low number of somatic copy number alterations (SCNAs) (Figure 5C). Somatic mutation analysis in TCGA-STAD revealed high mutation frequencies in *TP53* and *TTN* genes across both subtypes (Figure 5D). Except *CDH11*, most of the genes were significantly mutated in the Cluster1 subtype, indicating an unstable genomic status (Figure 5E). Gene set enrichment analysis (GSEA) on the TCGA-STAD expression data highlighted elevated epithelial to mesenchymal transition (EMT), VEGF, immune-related pathways, and stromal content in Cluster2. In contrast, Cluster1 showed significant upregulation in cell cycle related signature genes (Figure 5F).

Furthermore, given the immune-related distinctions between the two subtypes, we predicted tumor infiltration profiles using ESTIMATE and MCPcounter algorithms. Compared to Cluster1, Cluster2 exhibited a higher presence of immune cell proportions, particularly T cells, $CD8^+$ T cells, and endothelia cells (Figure 5G), along with elevated immune and stromal scores, while Cluster1 showed higher tumor purity (Figure 5H). Such immune landscapes were further validated in the GSE62254 (Figure S5F) and GSE26899 (Figure S5G) datasets, highlighting the reliability of the two molecular subtype classification. More specifically, the higher proportion of monocyte and myeloid cells in Cluster2 corresponded with our previous single-cell analysis, where

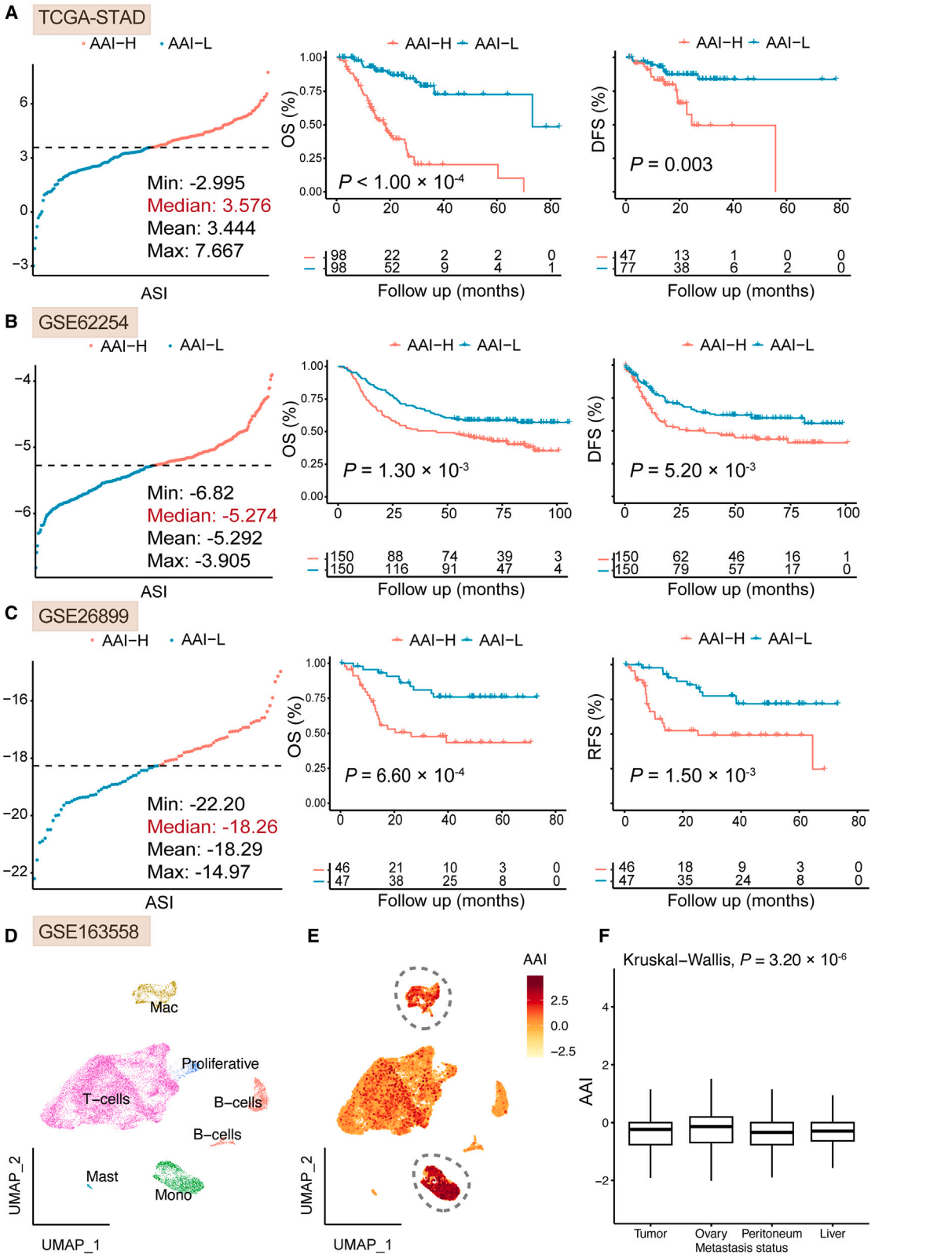
Figure 2. The aging-associated index was established for GC patients

(A) A 20-gene panel was prioritized using machine learning method.

(B) The prognostic significances of selected genes using multi-variate Cox regression model.

(C) The AAI significantly associated with several clinicopathological factors, such as survival status, tumor metastasis, lymph node metastasis, and tumor stage. Data are represented as mean ± SEM.

(D) A heatmap showed the significant aging-associated DEGs between high and low AAI group (stratified by the median value of AAI). The AAI served as a significant independent prognostic factor when combined with other clinical information in (E) univariate Cox and (F) multi-variate Cox regression analysis. The Wilcoxon test was used for comparisons between two groups and statistical significance was indicated with asterisks: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; ns denotes not significant. A p value of <0.05 was considered statistically significant.



(legend on next page)

AAI was highly associated with myeloid-related clusters (monocyte and macrophage), and Cluster2 had higher AAI scores compared to Cluster1.

In addition, immune checkpoint gene expression analysis between subtypes in TCGA-STAD (Figure S6A) and GSE62254 datasets (Figure S6B) showed consistent upregulation of immunosuppressive signature genes in the Cluster2, indicating an elevated level of immunosuppression in this subtype.

Establishment of a subtype-specific regulatory network uncovers two master regulator TFs in aggressive GC subtype

To reveal the potential regulatory mechanisms driving the aggressive GC subtype, Cluster2, we constructed a TF regulatory network inference by integrating TF and potential target gene expression profiles. Differentially expressed TFs ($|\log_2$ fold change$| > 0.5$ and BH-adjusted $p < 0.001$) and mRNAs (($|\log_2$ fold change$| > 1$ and BH-adjusted $p < 0.05$) between the two subtypes were prioritized as master regulators and target genes. These elements were Z-normalized and merged into a regulatory network inference using the R package "RTN", followed by 1,000 permutations and weak interaction filtering to ensure robustness (Figure 6A).

Moreover, to pinpoint specific master regulatory TFs linked to Cluster2's regulatory mechanism, a master regulatory analysis was performed based on a hypergeometric test against curated EMT signature genes. Nine candidate TFs potentially driving Cluster2 progression were identified ($p < 0.05$) (Table S4), among which two TFs, *ELK3* and *SOX7*, were further confirmed to have prognostic relevance (univariate Cox $p < 0.05$) (Figure 6B). Patients with high expression of the two master regulators exhibited the poorest survival outcomes (Figures 6C and 6D), and both TFs were significantly upregulated in Cluster2 (Figures 6E and 6F).

Additionally, receiver-operating characteristic (ROC) curves based on *ELK3* and *SOX7* expression levels demonstrated high predictive accuracy for GC molecular subtype classification (Figure 6G). Both TFs were upregulated in Cluster2, which were consistent with the expression levels (Figures S7A and S7B) and survival outcomes (Figures S7C–S7F) in the GSE62254 validation dataset. Additionally, we performed Pearson correlation analysis in the TCGA-STAD dataset, revealing significant negative correlations between gene expression and DNA methylation levels of *ELK3* and *SOX7* (all $p < 0.05$, Figures S7G and S7H). These findings suggest that epigenetic changes in Cluster2 subtype patients modulate TF expression levels, thereby altering downstream target gene expression and promoting tumor invasiveness. Moreover, the immunohistochemistry staining results from The Human Protein Atlas (HPA) database[9] validated the protein expression levels of *ELK3* and *SOX7* (Figure 6H).

In summary, our regulatory network analysis highlights key TFs involved in the aggressive Cluster2 subtype, underscoring their potential as diagnostic biomarkers. Further validation of these findings in larger independent cohorts is needed to establish their clinical utility.

Prioritization of drug candidates targeting key master regulator TFs through drug sensitivity analysis

Following the identification of master regulatory TFs in the subtype-specific regulatory network inference, we next prioritized potential drug candidates targeting these TFs, aiming to advance personalized therapies for GC. We collected drug sensitivity scores and gene expression profiles for 32 GC cell lines from the CellMinerCDB database. After filtering out drugs with over half of missing scores across cell lines, we calculated the area under the curve (AUC) of each drug response, using expression levels of *SOX7* and *ELF3* as predictors and stratifying drug activity scores by their median values. Our findings revealed several drugs that may be repurposed to target GC patient subgroups (Figures 7A–7C; Table S5).

Notably, *ELK3* expression predicted sensitivity to the FDA-approved drug axitinib (AUC = 0.81 and ANOVA $p = 0.038$), a VEGFR inhibitor used for epithelial ovarian cancer therapy.[10] Another FDA-approved drug, dacarbazine, used as a chemotherapeutic agent for metastatic melanomas,[11] also showed sensitivity in cell lines with high expression of *ELK3* (AUC = 0.76 and ANOVA $p = 0.047$). Additionally, crizotinib, approved for *ALK*-positive or *ROS1*-positive NSCLC,[12] was more sensitive to the elevated *ELK3* expression (AUC = 0.75 and ANOVA $p = 0.007$), while vincristine, an FDA-approved anticancer drug, also showed higher efficacy in high-*ELK3*-expressing GC cell lines (AUC = 0.84 and ANOVA $p = 0.035$). Structural diagrams of the molecules of the identified four drug candidate compounds were obtained from the chEMBL database (https://www.ebi.ac.uk/) (Figures S8A–S8D). These findings reveal that the activity of these candidate drugs strongly correlates with the expression levels of *ELK3* and *SOX7* (Figure 7D).

In conclusion, our drug sensitivity analysis based on the TF regulatory network provides promising, subtype-specific therapeutic targets that could serve as both prognostic and predictive biomarkers for GC.

DISCUSSION

Cancers are often considered diseases of aging due to the close association between aging and dysregulation at molecular and cellular levels. While the extended hallmarks of aging provide valuable insights into cancer biology, few studies have comprehensively examined aging-related processes in GC. In this study, we derived an aging-related signature from curated public database and introduced a prognostic marker, the AAI, to predict

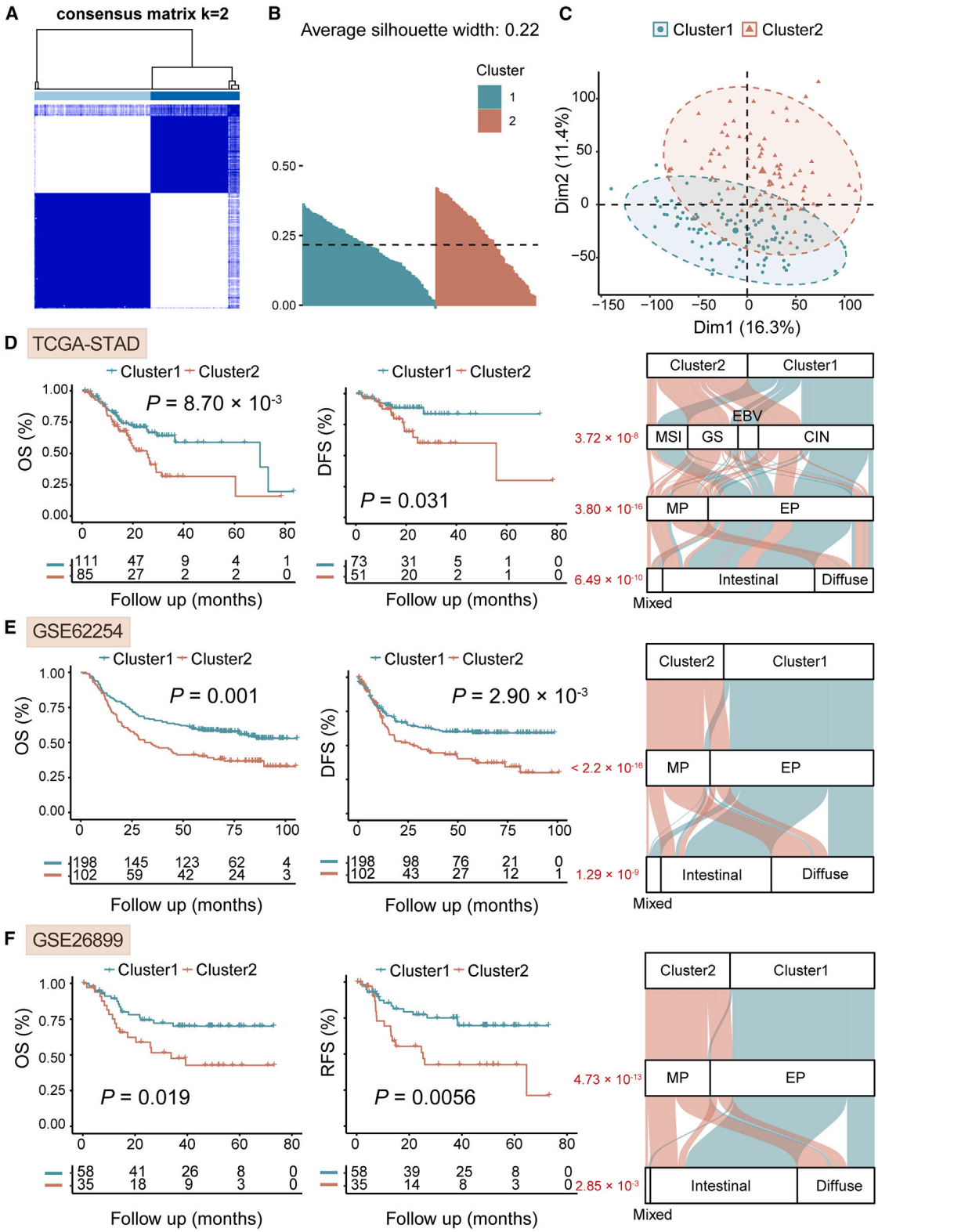
Figure 3. The AAI could be utilized as a prognosis and metastasis predictor of GC patients in multiple datasets

(A–C) The GC patients with higher AAI scores showed poorer OS and RFS in (A) the TCGA-STAD discovery dataset, and two other independent validation datasets, (B) GSE62254 and (C) GSE26899 (patients were stratified by the median value of AAI, separately).

(D) UMAP plot visualization of cell subtypes from seven GC patients. Cell subtypes were annotated by marker expression levels.

(E) UMAP plot visualization of the distribution of the AAI in GC cells.

(F) Violin plot of the AAI value in different tumor status of GC patients. Data are represented as mean ± SEM. Survival curves were compared via log rank test.



(legend on next page)

high-risk GC patients. The predictive power of AAI was successfully validated and testified in both bulk and single-cell datasets. Moreover, we identified two distinct molecular subtypes of GC, each with unique prognostic and clinical characteristics. To further uncover the regulatory mechanisms underlying these aging-related GC subtypes, we conducted an integrative network inference analysis, followed by master regulator analysis to prioritize key TFs, *ELK3* and *SOX7*, that drive the aggressive Cluster2 subtype. Finally, drug sensitivity analysis highlighted several FDA-approved drug candidates that could be repurposed to target these master regulatory TFs, suggesting their potential as therapeutic options for GC.

In this study, we prioritized a panel of 20 aging-related genes (*AREG*, *AXL*, *BDNF*, *BIRC3*, *CAT*, *CETP*, *CREB3L3*, *EGF*, *FGF23*, *IL7*, *MLH1*, *PEX5*, *POU1F1*, *PTGES*, *SERPINE1*, *SIRT7*, *SOCS2*, *TCF3*, *TNFRSF1A*, and *UCP1*) using a machine learning approach. The AAI generated by their expression levels significantly predicted GC prognosis across multiple datasets. For example, higher expression of the *EGFR* ligand *AREG* was associated with favorable survival outcomes in stage II/III GC patients enrolled in the adjuvant chemotherapy trial,[13] which was consistent with findings from our discovery dataset. A recent study on *AXL* revealed that its activation by the cancer-associated fibroblast (CAF) induced gene *GAS6* could promote GC aggressiveness,[14] while *BDNF*, derived from CAFs, activated TrkB-Nrf2 signaling pathway in GC cells, contributing to resistance to anlotinib in GC.[15] Further, *BIRC3* was identified as a promising biomarker for esophageal adenocarcinoma patients.[16] The presence of *EGF* correlates with gastric wall invasion and lymph node metastasis in GC, as well as EMT promotion in pancreatic cancer.[17] The fibroblast growth factor *FGF2* played a key role in tumorigenesis, with higher expression levels linked to poorer OS and serving as a biomarker for bone metastasis prognosis.[18] *IL7*, essential for immune regulation, particularly influenced B cell development and the proliferation of memory and naive T cells.[19,20] *MLH*, a gene responsible for DNA mismatch repair, was found hypermethylated and related to oxaliplatin resistance in GC patients.[21] High *PEX5* expression, interacting with miR-31-5p, was found to correlate with poor prognosis in liver cancer via Wnt/β-catenin signaling and HR pathways.[22] Recent study showed that *POU1F1* promoted GC progression via regulating macrophage proliferation, migration, and polarization, highlighting its vital relationship with tumor-associated macrophages.[23] *SERPINE1* promotes tumor progression and angiogenesis by activating the VEGFR-2 signaling pathway by interacting with miR-145-5p, a miRNA sponged by the ceRNA *NKX2-1-AS1* in GC.[24] Studies have shown that downregulation of *SIRT7* in breast cancer led to lung metastasis and could serve as a prognostic marker for metastasis-free survival.[25,26] Our findings showed the same survival outcomes, indicating that *SIRT7* could serving as a protective factor in GC. Additionally, a previous study showed that exosome circDIDO1 suppressed GC progression by regulating the miR-1307-3p/*SOSC2* axis, highlighting that the high expression levels of *SOSC2* promoted GC invasiveness.[27] *TCF3* functions as a tumor suppressor and has been found to be epigenetically silenced by EZH2 and DNMT3B in endometrial cancer.[28] Our results showed that higher expression levels of *TNFRSF1A* associated with worse OS, consistent with reports that elevated *TNFRSF1A* m^6A and protein levels correlate with poor survival in esophageal squamous cell carcinoma.[29] Previous research has shown that overexpression of *UCP1* could inhibit proliferation and metastasis in triple-negative breast cancer,[30] and our results similarly indicate that *UCP1* was upregulated in GC samples. Moreover, five out of 20 AAI genes (*AREG*, *AXL*, *BIRC3*, *PTGES*, and *SERPINE1*) overlapped with differentially expressed genes in mesenchymal GC cell lines,[31] demonstrating that our identified AAI gene panel is also associated with epigenomic changes in GC.

The two identified molecular subtypes had diverse biological properties and clinical outcomes. A simplified overview of these subtypes is shown in Figure 8. Cluster1, characterized by a low AAI and favorable survival outcome, was significantly enriched in epithelial and cell cycle pathways. Mutational analysis confirmed that Cluster1 possessed more genomic variations, indicating active genome alterations. In contrast, Cluster2, characterized by a higher AAI, higher stromal and immune contents, and poor survival prognosis, was enriched in EMT and wound response pathways. Immune infiltration and checkpoint gene analysis revealed that Cluster2 had an immunosuppressive microenvironment, underscoring its distinct immune landscape. Of note, Cluster2 also had a younger patient demographic compared to Cluster1, aligning with the rising incidence of early-onset GC in individuals under 40 in the U.S.[32] This subtype, marked by high metastasis rates,[33] likely encompasses early-onset GC patients who experience poorer survival outcomes and heightened mesenchymal prosperity. To elucidate the regulatory mechanism driving Cluster2's aggressiveness, an integrative network inference was established and prioritized two clinically relevant TFs, *ELK3* and *SOX7*, as potential contributors to its progression and metastasis. Specifically, *ELK3* has also been shown to be upregulated in the GC cell lines of mesenchymal subtype.[31] This finding aligns with our results showing that *ELK3* is upregulated in the Cluster2 subtype, which is strongly associated with the MP. Recently, studies showed that enhancers could interact with TFs to regulate aging-related pathways such as senescence, inflammation, and stress response,[34,35] and also restoring enhancer activity through epigenetic modulation or targeted gene editing could offer novel strategies to counteract aging-associated transcriptional dysregulation and improve cellular resilience.[36] These previous studies, combined with our TF regulatory network findings, suggest that the aging-associated gene signature may be

Figure 4. A two-class molecular subtypes were identified using unsupervised clustering

(A) GC patients were classified into two molecular clusters using gene expression profiles in the TCGA-STAD discovery dataset.

(B) The silhouette analysis demonstrated that the two identified subtypes were identical internally.

(C) The PCA analysis showed that the two subtypes could be clearly separated. The Cluster2 subtype exhibited poorer OS and RFS compared to the Cluster1 subtype, and significantly associated with other GC molecular subtypes in the (D) TCGA-STAD, (E) GSE62254 and (F) GSE26899 datasets. Survival curves were compared via log rank test.

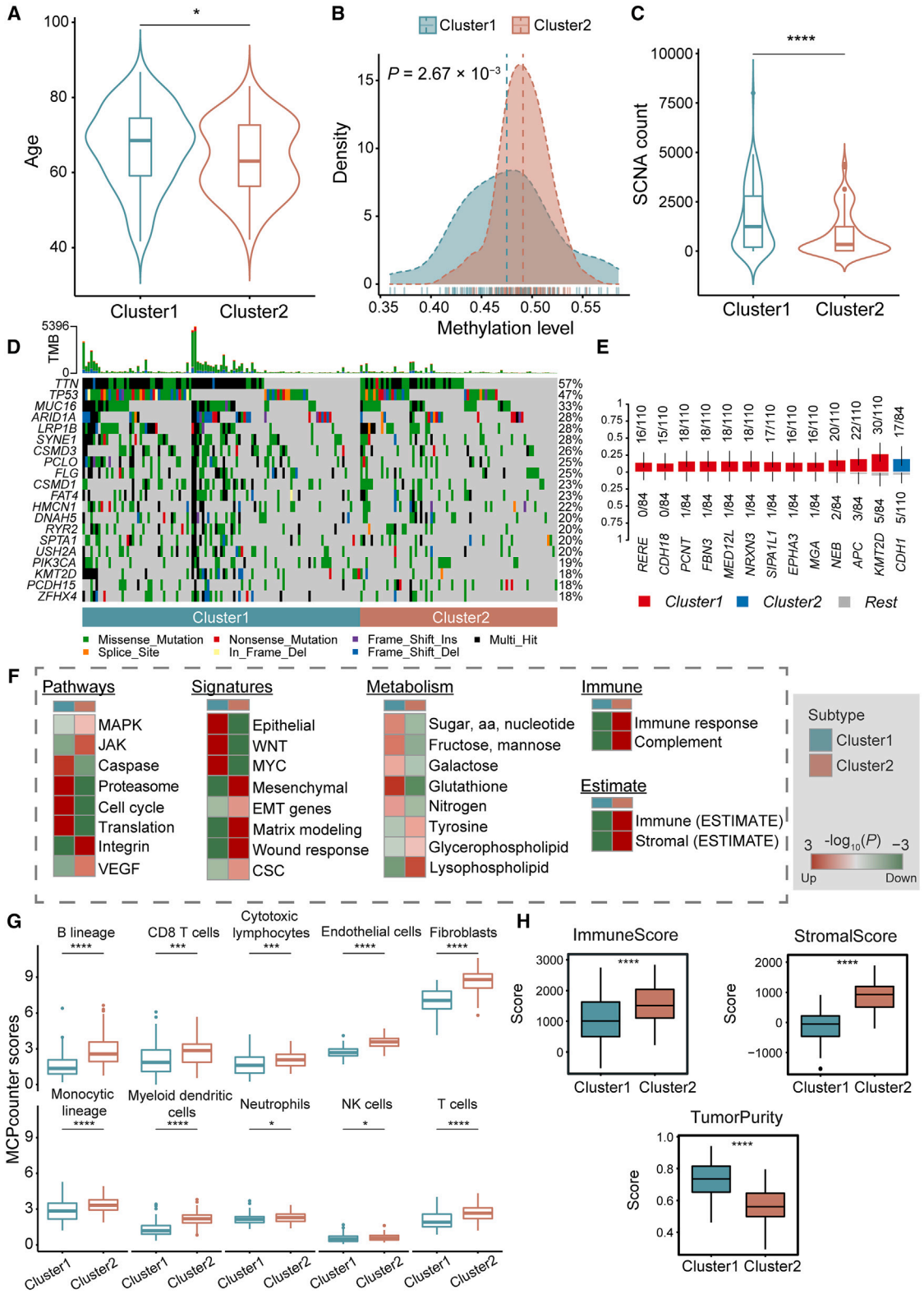


Figure 5. The multi-omics characteristics of the identified GC molecular subtypes

(A–C) The Cluster2 subtype patients had significant (A) younger population, (B) higher DNA methylation levels, and (C) less SCNA numbers. Data are represented as mean ± SEM.

(legend continued on next page)

modulated by multi-level regulatory activities, offering novel insights into the molecular mechanisms underlying aging and identifying potential therapeutic targets for age-related diseases. Also, the drug sensitivity analysis prioritized several FDA-approved drugs, including axitinib, crizotinib, dacarbazine, and vincristine, as promising treatments targeting the two TFs for Cluster2 patients.

In conclusion, we established a novel prognostic model and identified two distinct molecular subtypes with well-defined multi-omics features in GC through a comprehensive analysis of aging-associated patterns. Our regulatory network analysis pinpointed two potential druggable targets that may support individualized therapies for GC patients. Additionally, the shared prognostic model and molecular subtype classifier (https://jli-bioinfo.shinyapps.io/AAI_Aging/) provide a resource for exploring aging-related features in other datasets. We believe our comprehensive analysis lays the groundwork for the development of more personalized, targeted treatments for GC, informed by an enhanced understanding of the intricate connections between aging-associated processes and GC tumor heterogeneity. Further prospective investigations based on these insights have significant potential to advance precision oncology strategies for this disease.

Limitations of the study

While our study provides valuable insights, it still has several limitations. Although we independently validated the AAI prognostic model and identified GC molecular subtypes, this remains a retrospective analysis. Validation in larger, multi-center prospective cohorts will be essential to confirm their clinical utility. Additionally, the drug sensitivity predictions based on the identified master regulatory TFs require experimental validation in GC cell line models to substantiate their therapeutic potential.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jiang Li (lijiang29@mail.sysu.edu.cn).

Materials availability

This study did not generate new unique reagents or materials.

Data and code availability

- This paper analyses existing, publicly available data. These accession numbers and corresponding links for the datasets are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This research was funded by the Shenzhen Science and Technology Program, grant number JCYJ20230807110302006; by the Shenzhen Key Laboratory of Chinese Medicine Active substance screening and Translational Research, grant number ZDSYS20220606100801003; by the Seventh Affiliated Hospital of Sun Yat-Sen University, grant number ZSQYBRJH0024.

AUTHOR CONTRIBUTIONS

J.L. (0000-0001-9226-0536): conceptualization, software, methodology, investigation, formal analysis, visualization, data curation, writing—original draft. C.Y.: software, investigation, visualization, data curation, writing—review and editing. Y.Z., X.H., and M.J.: investigation, writing—review and editing. Z.Z. and J.L. (0000-0002-9791-2553): conceptualization, writing—review and editing, funding acquisition, project administration, supervision. All authors have reviewed and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Data collection and pre-processing
  - Construction and validation of aging-associated signature
  - Discovery and validation of molecular subtypes
  - Regulatory network and master regulator analysis
  - Estimation of tumor infiltration patterns
  - Functional enrichment analysis
  - Drug sensitivity analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

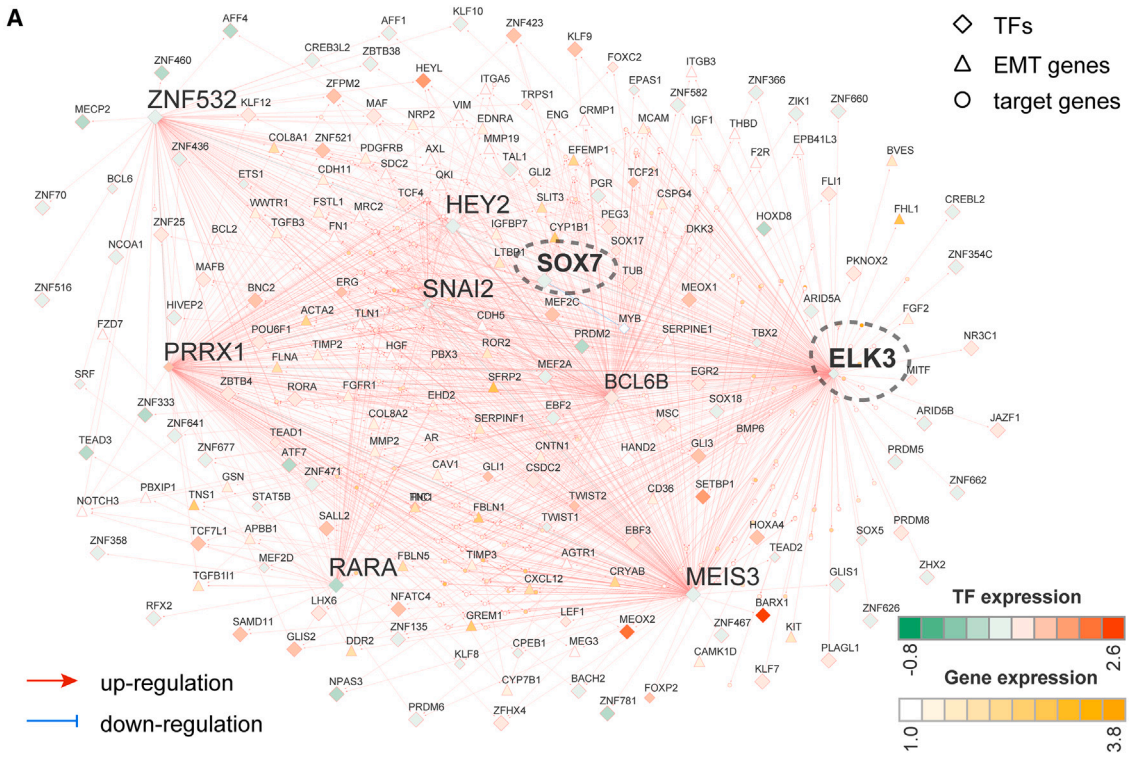
Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.112316.

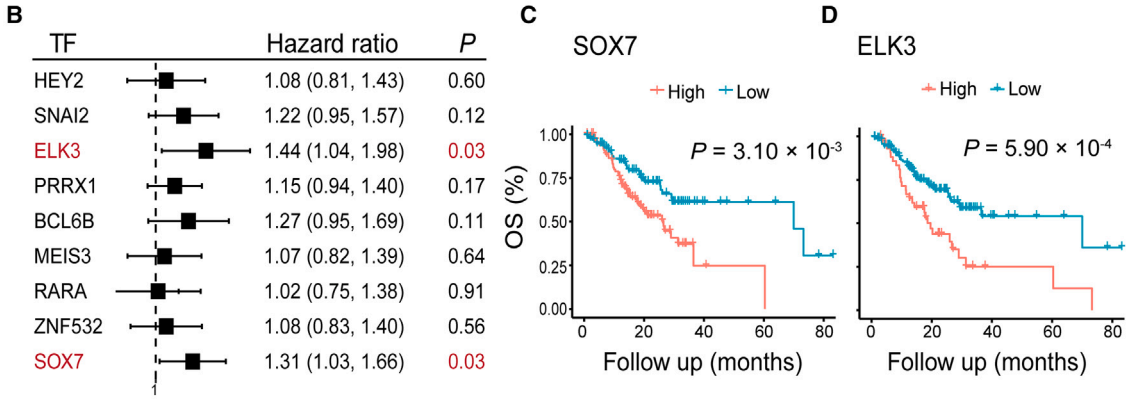
Received: December 9, 2024
Revised: February 24, 2025
Accepted: March 25, 2025
Published: March 28, 2025

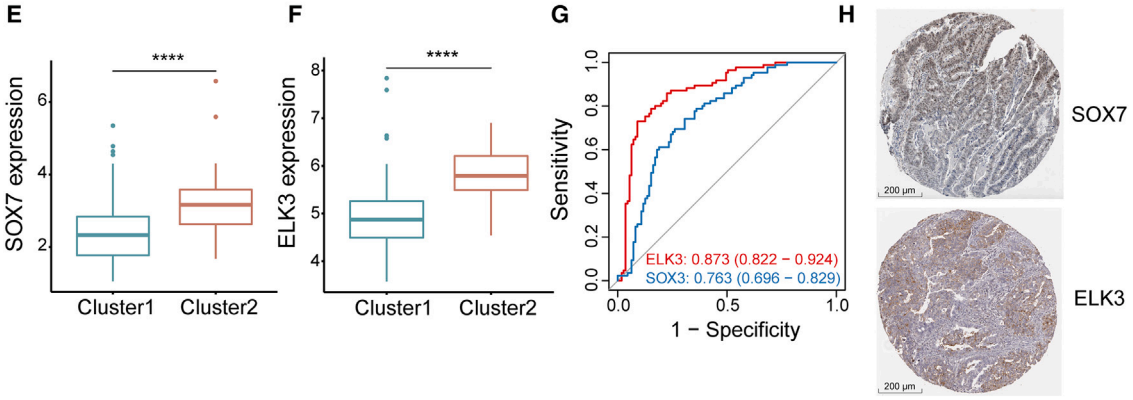
REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *71*, 209–249.
2. Joshi, S.S., and Badgwell, B.D. (2021). Current treatment and recent progress in gastric cancer. CA Cancer J. Clin. *71*, 264–279.

(D and E) The Cluster1 subtype had higher somatic mutated genes, compared to the Cluster2 subtype.
(F) The gene set mRNA enrichment analysis showing signatures of special interest of GC patients in different molecular subtypes, including characteristic gene signatures, canonical pathways, immune signatures, metabolic pathways as well as immune and stromal cell admixture in tumor samples (inferred by the ESTIMATE algorithm).
(G) The Cluster2 subtype associated with higher immune infiltration contents, compared to the Cluster1 subtype. Data are represented as mean ± SEM.
(H) The Cluster1 subtype had higher tumor purity content, whereas the Cluster2 subtype had higher immune and stromal cell admixtures. Data are represented as mean ± SEM. The Wilcoxon test was used for comparisons between two groups and statistical significance was indicated with asterisks: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; ns denotes Not Significant. A p value of <0.05 was considered statistically significant.







(legend on next page)

3. Smyth, E.C., Nilsson, M., Grabsch, H.I., van Grieken, N.C., and Lordick, F. (2020). Gastric cancer. Lancet *396*, 635–648.
4. Ho, S.W.T., and Tan, P. (2019). Dissection of gastric cancer heterogeneity for precision oncology. Cancer Sci. *110*, 3405–3414.
5. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2023). Hallmarks of aging: An expanding universe. Cell *186*, 243–278.
6. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217.
7. Montégut, L., López-Otín, C., and Kroemer, G. (2024). Aging and cancer. Mol. Cancer *23*, 106.
8. Shah, Y., Verma, A., Marderstein, A.R., White, J., Bhinder, B., Garcia Medina, J.S., and Elemento, O. (2021). Pan-cancer analysis reveals molecular patterns associated with age. Cell Rep. *37*, 110100.
9. Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., et al. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. Science *367*, eaay5947. https://doi.org/10.1126/science.aay5947.
10. Paik, E.S., Kim, T.-H., Cho, Y.J., Ryu, J., Choi, J.-J., Lee, Y.-Y., Kim, T.-J., Choi, C.-H., Kim, W.Y., Sa, J.K., et al. (2020). Preclinical assessment of the VEGFR inhibitor axitinib as a therapeutic agent for epithelial ovarian cancer. Sci. Rep. *10*, 4904.
11. Brusco, I., Li Puma, S., Chiepe, K.B., da Silva Brum, E., de David Antoniazzi, C.T., de Almeida, A.S., Camponogara, C., Silva, C.R., De Logu, F., de Andrade, V.M., et al. (2020). Dacarbazine alone or associated with melanoma-bearing cancer pain model induces painful hypersensitivity by TRPA1 activation in mice. Int. J. Cancer *146*, 2797–2809.
12. Mansfield, A.S., Wei, Z., Mehra, R., Shaw, A.T., Lieu, C.H., Forde, P.M., Drilon, A.E., Mitchell, E.P., Wright, J.J., Takebe, N., et al. (2022). Crizotinib in patients with tumors harboring ALK or ROS1 rearrangements in the NCI-MATCH trial. npj Precis. Oncol. *6*, 13.
13. Ichikawa, W., Terashima, M., Ochiai, A., Kitada, K., Kurahashi, I., Sakuramoto, S., Katai, H., Sano, T., Imamura, H., and Sasako, M. (2017). Impact of insulin-like growth factor-1 receptor and amphiregulin expression on survival in patients with stage II/III gastric cancer enrolled in the Adjuvant Chemotherapy Trial of S-1 for Gastric Cancer. Gastric Cancer. *20*, 263–273.
14. Bae, C.A., Ham, I.-H., Oh, H.J., Lee, D., Woo, J., Son, S.-Y., Yoon, J.H., Lorens, J.B., Brekken, R.A., Kim, T.-M., et al. (2020). Inhibiting the GAS6/AXL axis suppresses tumor progression by blocking the interaction between cancer-associated fibroblasts and cancer cells in gastric carcinoma. Gastric Cancer. *23*, 824–836.
15. Jin, Z., Lu, Y., Wu, X., Pan, T., Yu, Z., Hou, J., Wu, A., Li, J., Yang, Z., Li, C., et al. (2021). The cross-talk between tumor cells and activated fibroblasts mediated by lactate/BDNF/TrkB signaling promotes acquired resistance to anlotinib in human gastric cancer. Redox Biol. *46*, 102076.
16. Piro, G., Giacopuzzi, S., Bencivenga, M., Carbone, C., Verlato, G., Frizziero, M., Zanotto, M., Mina, M.M., Merz, V., Santoro, R., et al. (2015). TAK1-regulated expression of BIRC3 predicts resistance to preoperative chemoradiotherapy in oesophageal adenocarcinoma patients. Br. J. Cancer *113*, 878–885.
17. Sheng, W., Chen, C., Dong, M., Wang, G., Zhou, J., Song, H., Li, Y., Zhang, J., and Ding, S. (2017). Calreticulin promotes EGF-induced EMT in pancreatic cancer cells via Integrin/EGFR-ERK/MAPK signaling pathway. Cell Death Dis. *8*, e3147.
18. Mansinho, A., Ferreira, A.R., Casimiro, S., Alho, I., Vendrell, I., Costa, A.L., Sousa, R., Abreu, C., Pulido, C., Macedo, D., et al. (2019). Levels of circulating fibroblast growth factor 23 (FGF23) and prognosis in cancer patients with bone metastases. Int. J. Mol. Sci. *20*, 695.
19. Parrish, Y.K., Baez, I., Milford, T.-A., Benitez, A., Galloway, N., Rogerio, J.W., Sahakian, E., Kagoda, M., Huang, G., Hao, Q.-L., et al. (2009). IL-7 Dependence in human B lymphopoiesis increases during progression of ontogeny from cord blood to bone marrow. J. Immunol. *182*, 4255–4266.
20. Fry, T.J., and Mackall, C.L. (2005). The many faces of IL-7: from lymphopoiesis to peripheral T cell maintenance. J. Immunol. *174*, 6571–6576.
21. Li, Y., Yang, Y., Lu, Y., Herman, J.G., Brock, M.V., Zhao, P., and Guo, M. (2015). Predictive value of CHFR and MLH1 methylation in human gastric cancer. Gastric Cancer. *18*, 280–287.
22. Wen, J., Xiong, K., Aili, A., Wang, H., Zhu, Y., Yu, Z., Yao, X., Jiang, P., Xue, L., and Wang, J. (2020). PEX5, a novel target of microRNA-31-5p, increases radioresistance in hepatocellular carcinoma by activating Wnt/β-catenin signaling and homologous recombination. Theranostics *10*, 5322–5340.
23. Tang, C., Lei, X., Xiong, L., Hu, Z., and Tang, B. (2021). HMGA1B/2 transcriptionally activated-POU1F1 facilitates gastric carcinoma metastasis via CXCL12/CXCR4 axis-mediated macrophage polarization. Cell Death Dis. *12*, 422. https://doi.org/10.1038/s41419-021-03703-x.
24. Teng, F., Zhang, J.-X., Chen, Y., Shen, X.-D., Su, C., Guo, Y.-J., Wang, P.-H., Shi, C.-C., Lei, M., Cao, Y.-O., and Liu, S.Q. (2021). LncRNA NKX2-1-AS1 promotes tumor progression and angiogenesis via upregulation of SERPINE1 expression and activation of the VEGFR-2 signaling pathway in gastric cancer. Mol. Oncol. *15*, 1234–1255.
25. Tang, X., Shi, L., Xie, N., Liu, Z., Qian, M., Meng, F., Xu, Q., Zhou, M., Cao, X., Zhu, W.-G., and Liu, B. (2017). SIRT7 antagonizes TGF-β signaling and inhibits breast cancer metastasis. Nat. Commun. *8*, 318. https://doi.org/10.1038/s41467-017-00396-9.
26. Huo, Q., Chen, S., Zhuang, J., Quan, C., Wang, Y., and Xie, N. (2023). SIRT7 downregulation promotes breast cancer metastasis via LAP2α-induced chromosomal instability. Int. J. Biol. Sci. *19*, 1528–1542.
27. Guo, Z., Zhang, Y., Xu, W., Zhang, X., and Jiang, J. (2022). Engineered exosome-mediated delivery of circDIDO1 inhibits gastric cancer progression via regulation of MiR-1307-3p/SOCS2 Axis. J. Transl. Med. *20*, 326. https://doi.org/10.1186/s12967-022-03527-z.
28. Gui, T., Liu, M., Yao, B., Jiang, H., Yang, D., Li, Q., Zeng, X., Wang, Y., Cao, J., Deng, Y., et al. (2021). TCF3 is epigenetically silenced by EZH2 and DNMT3B and functions as a tumor suppressor in endometrial cancer. Cell Death Differ. *28*, 3316–3328.
29. Li, R., Zeng, L., Zhao, H., Deng, J., Pan, L., Zhang, S., Wu, G., Ye, Y., Zhang, J., Su, J., et al. (2022). ATXN2-mediated translation of TNFR1 promotes esophageal squamous cell carcinoma via m6A-dependent manner. Mol. Ther. *30*, 1089–1103.
30. Xia, J., Chu, C., Li, W., Chen, H., Xie, W., Cheng, R., Hu, K., and Li, X. (2022). Mitochondrial protein UCP1 inhibits the malignant behaviors of triple-negative breast cancer through activation of mitophagy and pyroptosis. Int. J. Biol. Sci. *18*, 2949–2961.

Figure 6. The transcription factor regulatory network was established to identify subtype-specific potential target TF

(A) A regulatory network inference was established using the expression profiles of TF and mRNA.

(B) Univariate Cox analysis revealed two prognostically significant potential TFs.

(C) Patients with higher expression levels of *SOX7* showed poorer OS (patients stratified by optional value).

(D) Patients with higher expression levels of *ELK3* showed poorer OS (patients stratified by optional value). Survival curves were compared via log rank test.

(E and F) The identified TFs, (E) *ELK3* and (F) *SOX7* were upregulated in the Cluster2 subtype (Data are represented as mean ± SEM), (G) could serve as diagnostic biomarkers, and (H) further detected at protein expression levels in HPA database (scale bar: 200 μm). The Wilcoxon test was used for comparisons between two groups and statistical significance was indicated with asterisks: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; ns denotes not significant. A p value of <0.05 was considered statistically significant.

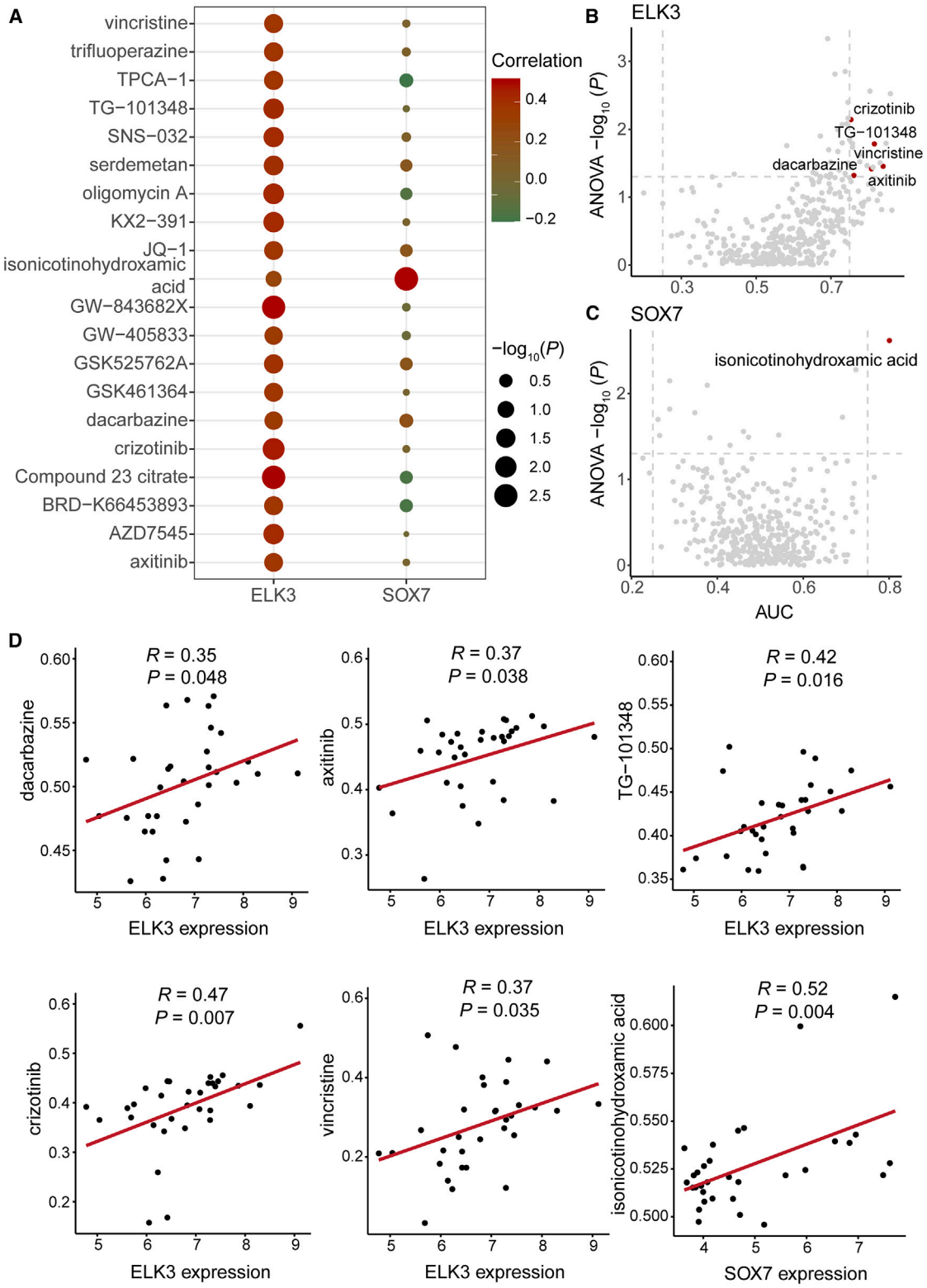


Figure 7. Drug sensitivity analysis evaluating the predictive values of the two identified TFs

(A) The heatmap showed the correlations between the drug sensitivity scores and the TF expression levels.

(B and C) The drug sensitivity analysis of (B) *ELK3* and (C) *SOX7*.

(D) The correlations between the two TFs and drug sensitivity scores.

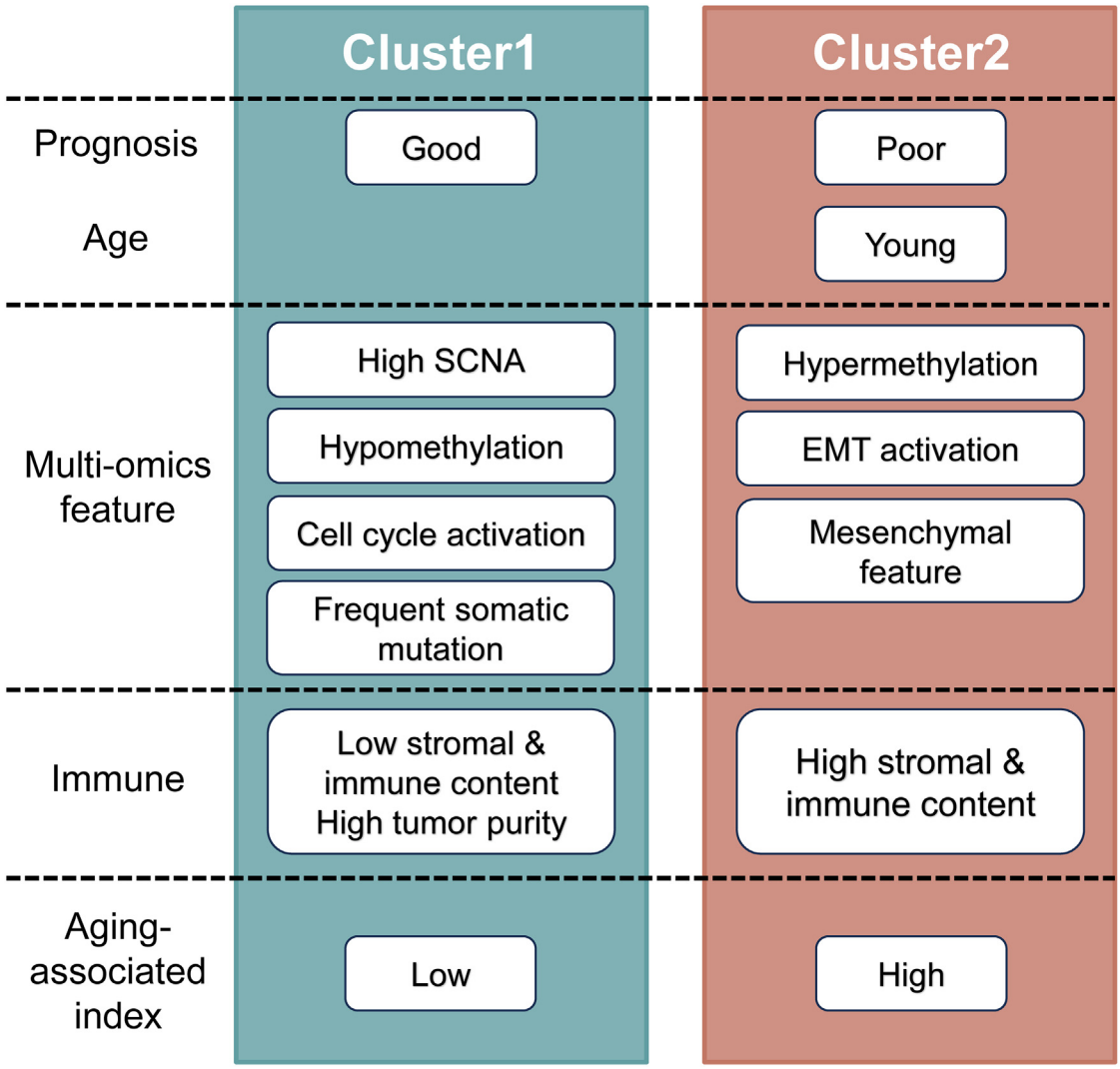


Figure 8. Overview of aging-associated molecular subtypes of GC

We identified two molecular subtypes utilizing aging-associated patterns in GC, and the clinical differences, multi-omics characteristics, and immune infiltration of the two subtypes were comprehensively explored.

31. Razavi-Mohseni, M., Huang, W., Guo, Y.A., Shigaki, D., Ho, S.W.T., Tan, P., Skanderup, A.J., and Beer, M.A. (2024). Machine learning identifies activation of RUNX/AP-1 as drivers of mesenchymal and fibrotic regulatory programs in gastric cancer. Genome Res. *34*, 680–695.

32. Anderson, W.F., Camargo, M.C., Fraumeni, J.F., Jr., Correa, P., Rosenberg, P.S., and Rabkin, C.S. (2010). Age-specific trends in incidence of noncardia gastric cancer in US adults. JAMA *303*, 1723–1728.

33. Mun, D.-G., Bhin, J., Kim, S., Kim, H., Jung, J.H., Jung, Y., Jang, Y.E., Park, J.M., Kim, H., Jung, Y., et al. (2019). Proteogenomic characterization of human early-onset gastric cancer. Cancer Cell. *35*, 111–124.e10.

34. Sen, P., Shah, P.P., Nativio, R., and Berger, S.L. (2016). Epigenetic mechanisms of longevity and aging. Cell *166*, 822–839.

35. Luo, R., Yan, J., Oh, J.W., Xi, W., Shigaki, D., Wong, W., Cho, H.S., Murphy, D., Cutler, R., Rosen, B.P., et al. (2023). Dynamic network-guided CRISPRi screen identifies CTCF-loop-constrained nonlinear enhancer gene regulatory activity during cell state transitions. Nat. Genet. *55*, 1336–1346.

36. Ponnappan, S., and Ponnappan, U. (2011). Aging and immune function: molecular mechanisms to interventions. Antioxidants Redox Signal. *14*, 1551–1585.

37. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678.

38. Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, 271–281.e7.

39. Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., III, Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature *554*, 544–548.

40. 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.

41. Aging Atlas Consortium (2021). Aging Atlas: a multi-omics database for aging biology. Nucleic Acids Res. *49*, D825–D830.

42. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. *33*, 1–22.

43. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

44. Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: An*R*package for determining the relevant number of clusters in a data set. J. Stat. Softw. *61*, 1–36. https://doi.org/10.18637/jss.v061.i06.

45. Breiman, L. (2001). Mach. Learn. *45*, 5–32.

46. Shen, W.-K., Chen, S.-Y., Gan, Z.-Q., Zhang, Y.-Z., Yue, T., Chen, M.-M., Xue, Y., Hu, H., and Guo, A.-Y. (2023). AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res. *51*, D39–D45.

47. Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E., et al. (2011). MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. *12*, R126.

48. Zhao, M., Kong, L., Liu, Y., and Qu, H. (2015). dbEMT: an epithelial-mesenchymal transition associated gene resource. Sci. Rep. *5*, 11459.

49. Castro, M.A.A., Wang, X., Fletcher, M.N.C., Meyer, K.B., and Markowetz, F. (2012). RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. Genome Biol. *13*, R29.

50. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

51. Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. *17*, 218.

52. Wang, X., Terfve, C., Rose, J.C., and Markowetz, F. (2011). HTSanalyzeR: an R/Bioconductor package for integrated network analysis of high-throughput screens. Bioinformatics *27*, 879–880.

53. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

54. Reinhold, W.C., Wilson, K., Elloumi, F., Bradwell, K.R., Ceribelli, M., Varma, S., Wang, Y., Duveau, D., Menon, N., Trepel, J., et al. (2023). CellMinerCDB: NCATS is a web-based portal integrating public cancer cell line databases for pharmacogenomic explorations. Cancer Res. *83*, 1941–1952.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA-STAD	TCGA	https://xenabrowser.net/datapages/
GSE62254	Gene Expression Omnibus (GEO)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254
GSE26899	Gene Expression Omnibus (GEO)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26899
GSE163558	Gene Expression Omnibus (GEO)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163558
IMvigor210CoreBiologies	GitHub	https://github.com/SiYangming/IMvigor210CoreBiologies
Cancer Therapeutics Response Portal	CellMinerCDB	https://discover.nci.nih.gov/rsconnect/cellminercdb/
Structural diagrams of the molecules of drugs	chEMBL	https://www.ebi.ac.uk/
Aging-related genes	Aging Altas	https://ngdc.cncb.ac.cn/aging/index
EMT signature genes	dbEMT	https://bioinfo-minzhao.org/dbemt/dbemt1/index.html
Protein levels of *SOX7* and *ELK3*	The Human Protein Atlas	https://www.proteinatlas.org/
Transcription factors	AnimalTFDB	http://bioinfo.life.hust.edu.cn/AnimalTFDB4/
Software and algorithms		
TCGAmutations	GitHub	https://github.com/PoisonAlien/TCGAmutations
AnnoProbe	CRAN	https://cran.r-project.org/web/packages/AnnoProbe/
Seurat 4.4.0	GitHub	https://github.com/satijalab/seurat/releases/tag/v4.4.0
survival	CRAN	https://cran.r-project.org/web/packages/survival/
glmnet	CRAN	https://cran.r-project.org/web/packages/glmnet/
consensusClusterplus	Bioconductor	https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
cluster	CRAN	https://cran.r-project.org/web/packages/cluster/
NbClust	CRAN	https://cran.r-project.org/web/packages/NbClust/
randomForest	CRAN	https://cran.r-project.org/web/packages/randomForest/
limma	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/limma.html
RTN	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/RTN.html
RedeR	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/RedeR.html
estimate	MD Anderson	https://r-forge.r-project.org/projects/estimate/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MCPcounter	GitHub	https://github.com/ebecht/MCPcounter
HTSanalyzeR2	GitHub	https://github.com/CityUHK-CompBio/HTSanalyzeR2
clusterProfiler	Bioconductor	https://bioconductor.org/packages/clusterProfiler
impute	Bioconductor	https://bioconductor.org/packages/release/bioc/html/impute.html
pROC	CRAN	https://cran.r-project.org/web/packages/pROC/
survminer	CRAN	https://cran.r-project.org/web/packages/survminer/
R version 4.1.3	R software	https://www.r-project.org/

METHOD DETAILS

Data collection and pre-processing

For the discovery dataset, we acquired TCGA-STAD gene expression profiles, DNA methylation array, copy number alteration and clinical information from the University of California Santa Cruz (UCSC) database.[37] Gene expression profiles were then converted to transcripts per million (TPM) and log2-transformed, yielding data from 195 GC patients. Somatic mutation data for the TCGA-STAD dataset was acquired using the R package 'TCGAmutations'.[38] Processed microarray gene expression profiles and clinical information for two validation datasets (GSE62254 and GSE26899) were downloaded from the Gene Expression Omnibus (GEO) database. Probe-level data was mapped to gene levels using the 'AnnoProbe' R package, with the highest value retained for genes with multiple probes. For the anti-PDL1 therapy cohort from Mariathasan et al., data from 348 patients with metastatic urothelial tumours treated with the anti-PD-L1 agent were obtained via the R package 'IMvigor210CoreBiologies',[39] followed by log2-transformation. Patients with treatment response data and overall survival times longer than one month were selected for further analysis. Single-cell RNA transcriptome data of seven GC patients was collected from GSE163558 and analysed using the R package "Seurat" (version 4.4.0).[40]

Construction and validation of aging-associated signature

We compiled a published set of 445 aging-associated genes from the Aging Atlas database for subsequent model analysis.[41] Genes with significant prognostic values ($P < 0.1$) were identified via univariate Cox regression analysis using the R package 'survival'. Then, the least absolute shrinkage and selection operator (LASSO) Cox regression method was adopted to reduce the number of candidates with 100 times iterations using the R package 'glmnet'.[42] Genes that appeared in over 85 iterations were selected and prioritized for the final signature panel. Eventually, a multivariate Cox regression risk model was developed using these prioritized signature genes, and the aging-associated index (AAI) was calculated for each sample as follows:

$$AAI = \sum_{i=1}^{N} \beta_i * \epsilon_i$$

Where β_i represents the coefficient retrieved from the risk model, ϵ_i is the expression of each candidate gene, and N is the number of signature genes. Samples were classified into the high- and low-AAI groups based on the median AAI value, with survival differences evaluated by Kaplan-Meier (KM) method. The clinical significance of AAI in each dataset was evaluated via the univariate Cox regression.

Discovery and validation of molecular subtypes

The abovementioned prognostic genes were adopted for downstream clustering analysis using the R package 'consensusClusterplus',[43] setting "maxK" to 5 and "clusterAlg" to "km", while keeping other parameters as default. Silhouette analysis using the R package 'cluster' was conducted to confirm the cluster consistency, and the R package 'NbClust'[44] determined the optimal cluster count. For the validation of GC molecular subtypes in other datasets, a random forest classifier was trained using the top 1,000 genes based on median absolute deviation (MAD), and the R package 'randomForest'[45] was used to classify subtypes.

Regulatory network and master regulator analysis

We integrated the gene expression profiles of transcript factors (TFs)[46] and mRNAs to construct a regulatory network. Differentially expressed TFs between identified molecular subtypes (Benjamini-Hochberg [BH]-adjusted $P < 0.001$ & $|\log2 \text{ fold change}| > 0.5$) were selected as potential regulators, while differentially expressed genes (DEGs) were identified as potential targets (BH-adjusted

$P < 0.05$ & |log2 fold change| > 1) in the training dataset, using the R package 'limma'. Z-normalized expression profiles of these potential TFs and target genes were then processed with the R package 'RTN',[47] with 1,000 permutations, bootstrap resampling, and weak interaction filtering. Master regulator analysis was subsequently performed based on the hypergeometric test with epithelial-to-mesenchymal transition (EMT) signature genes.[48] The regulatory network was visualized with the R package 'RedeR'.[49]

Estimation of tumor infiltration patterns

The ESTIMATE algorithm was employed to evaluate the immune cell content, stromal content and tumor purity scores for each GC sample.[50] The abundances of infiltrating immune and stromal cell subpopulations were estimated using the R package 'MCPcounter',[51] with the gene expression profiles as input.

Functional enrichment analysis

The gene set enrichment analysis (GSEA) was performed using the R package 'HTSanalyzeR2' with 1,000 permutations.[52] Analyses of biological pathways, including the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) gene sets, were implemented using the R package 'clusterProfiler'.[53]

Drug sensitivity analysis

Drug activity data, including investigational drugs and corresponding gene expression profiles, were downloaded from the Cancer Therapeutics Response Portal (CTRP) through CellMinerCDB database.[54] Drugs lacking activity scores in over half of the GC cell lines were excluded, and missing data was imputed by the R package 'impute'. To investigate the predictive power of master regulator TFs, cell lines were stratified by median drug activity scores, and the area under the curve (AUC) was calculated from TF expressions. Moreover, a univariate linear regression model was constructed to predict drug activity scores based on TF gene expression levels, with ANOVA tests evaluating the overall significance. A gene was significantly associated with drug sensitivity if ANOVA $P < 0.05$ and AUC ≥ 0.70 or with resistance if ANOVA $P < 0.05$ and AUC ≤ 0.30.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were conducted using R software (v.4.1.3). The Wilcoxon test was used for comparisons between two groups, while the Kruskal-Wallis test was applied for multiple groups. Receiver-operating characteristic (ROC) curve analysis was performed using the R package 'pROC'. Survival curves were plotted using KM analysis and compared via log-rank test, with visualizations created by the R package 'survminer'. Statistical significance was indicated with asterisks: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; ns denotes Not Significant. A P-value of < 0.05 was considered statistically significant.