Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

EIF5A marks an immune-inflamed epithelial state and restrains proliferation and migration in gastric cancer cells

Zhilin Liu
Ningbo No. 2 Hospital
Yunjie Chen
Ningbo No. 2 Hospital
Feng Ren
Ningbo No. 2 Hospital
Kepu Zheng
Ningbo No. 2 Hospital
Yanghui Wen
15257842040@163.com

Ningbo No. 2 Hospital

Research Article

Keywords: EIF5A, gastric cancer, tumor microenvironment, tumor mutation burden, epithelial-mesenchymal transition, single-cell RNA sequencing

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11638300/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

EIF5A marks an immune-inflamed epithelial state and restrains proliferation and migration in gastric cancer cells

Zhilin Liu[1†], Yunjie Chen[1†], Feng Ren[1], Kepu Zheng[1*], Yanghui Wen[1*]

1 Department of General Surgery, Ningbo No. 2 Hospital, Ningbo, Zhejiang 315010, China

†These authors contributed equally to this work.

Corresponding authors: Kepu Zheng, PhD, E-mail: hxjachie@163.com, Department of General Surgery, Ningbo No. 2 Hospital, Ningbo, Zhejiang 315010, China. Tel.: +86 18787070851.

Yanghui Wen, MS, E-mail: 15257842040@163.com, Department of General Surgery, Ningbo No. 2 Hospital, Ningbo, Zhejiang 315010, China. Tel.: +86 15257842040.

Abstract

Background: Gastric cancer (GC), most frequently represented as stomach adenocarcinoma (STAD), is clinically and molecularly heterogeneous, and biomarkers linking tumor-cell state, immune context and therapeutic response are needed. EIF5A encodes a hypusine-modified translation factor with context-dependent functions in cancer, but its significance in GC remains insufficiently defined.

Methods: We integrated The Cancer Genome Atlas-STAD, Gene Expression Omnibus validation cohorts (GSE15459, GSE62254 and GSE84437), public immunotherapy resources, GSE183904 single-cell RNA-sequencing data, and matched somatic mutation, copy-number variation, DNA methylation,

proteomic and intratumoral microbiome datasets. Samples were stratified by EIF5A expression. Survival, clinicopathological associations, tumor mutation burden (TMB), functional programs, immune features, upstream regulation and microbiome features were evaluated. EIF5A overexpression was further tested in EIF5A-low GC cell lines using qRT-PCR, CCK-8, colony formation, transwell migration and wound-healing assays.

Results: EIF5A expression was increased in immunotherapy responders and TMB-high tumors, with favorable survival associations in selected subgroups, particularly male patients, advanced-stage disease and diffuse-type tumors. EIF5A-high tumors showed epithelial differentiation, translation/proteostasis, cell-cycle and DNA damage surveillance programs, with reduced Epithelial-Mesenchymal Transition-, invasion- and focal adhesion-related signals at transcriptomic and proteomic levels. They also exhibited enhanced cytotoxic, checkpoint, T-cell and NK-cell signatures alongside Treg and exhaustion signals, indicating an immune-inflamed but regulated microenvironment. Copy-number and methylation alterations only partly explained EIF5A expression, while perturbation evidence and cross-cohort validation identified HMGA1 as a candidate upstream regulator. The HMGA1-EIF5A axis was associated with chemokine networks, immune infiltration and selective intratumoral microbial changes. Single-cell analysis localized EIF5A mainly to cycling and malignant epithelial cells, where it correlated with proliferation, glycolysis and hypoxia rather than epithelial-mesenchymal transition. EIF5A overexpression in HGC-27 and

SNU-719 cells suppressed cell viability, colony formation, transwell migration and wound closure.

Conclusions: EIF5A marks an epithelial, immune-inflamed and relatively less invasive state in GC and may help refine biomarkers for prognosis and immunotherapy response. The HMGA1-EIF5A regulatory relationship and the suppressive effects of EIF5A overexpression support further mechanistic validation and therapeutic exploration, suggesting a potential anti-tumor role for EIF5A in GC.

Keywords: EIF5A; gastric cancer; tumor microenvironment; tumor mutation burden; epithelial-mesenchymal transition; single-cell RNA sequencing

Background

Gastric cancer (GC), usually represented as stomach adenocarcinoma (STAD) in genomic cohorts, remains a major cause of cancer death worldwide. Surgery, chemotherapy, targeted therapy and immunotherapy have improved treatment, but advanced disease still carries poor long-term outcomes [1,2]. Molecular classification has shown why a single clinical label is insufficient. TCGA divides gastric adenocarcinoma into Epstein-Barr virus-positive, microsatellite instability, genomically stable and chromosomal instability subtypes, whereas the ACRG system identifies microsatellite instability, mesenchymal-like, TP53-active and TP53-inactive groups [3,4]. These systems are biologically useful, but markers that

connect tumor-cell state, immune context and clinical outcome are still needed.

Checkpoint blockade has changed the treatment of advanced GC, but only some patients obtain durable benefit. PD-1 blockade is more active in tumors with microsatellite instability-high status, Epstein-Barr virus positivity, PD-L1 positivity or other immune-active features [5-9]. TMB can reflect immunogenicity and has been associated with checkpoint inhibitor benefit in several cancers, although its prognostic meaning depends on tumor type and treatment setting [10,11]. In GC, response is rarely explained by one marker alone. Chemokines, cytotoxic lymphocytes, checkpoint molecules, myeloid cells, epithelial differentiation and microbial features may act together. A useful biomarker should therefore capture part of this combined biology, not just one isolated measurement.

EIF5A encodes eukaryotic translation initiation factor 5A, a translation factor modified by hypusination through the polyamine-hypusine circuit. Although EIF5A was originally described as an initiation factor, later work has implicated it in translation elongation, termination and the synthesis of difficult peptide motifs, including polyproline-rich sequences [12-17]. These activities can influence proteostasis, collagen production, cell-cycle progression, stress responses, apoptosis, immune function and cancer-related phenotypes [12-18]. The direction of its effect is not fixed. In GC, many reports focus on EIF5A2 rather than EIF5A/EIF5A1, and EIF5A2 has

been linked to tumor aggressiveness and EMT-related programs [19-22]. Whether EIF5A itself marks proliferation, epithelial differentiation, immune activity or invasion in GC remains unresolved.

Here, we examined EIF5A in GC using bulk transcriptomic, genomic, methylation, proteomic, microbiome, immunotherapy-response and single-cell datasets. The analysis focused on prognosis, immune activity, TMB, epithelial-mesenchymal state, upstream regulation and tumor-cell behavior. We also overexpressed EIF5A in GC cell lines with low endogenous expression. Across these analyses, EIF5A tracked an epithelial, immune-inflamed tumor state and showed suppressive effects on proliferation and migration in vitro.

Methods

Acquisition of TCGA-STAD Somatic Mutation Data

Somatic mutation data for gastric adenocarcinoma were retrieved from the Genomic Data Commons (GDC) using the TCGAbiolinks R package (Project: TCGA-STAD; Data Type: Masked Somatic Mutation; Workflow: Aliquot Ensemble Somatic Variant Merging and Masking). A chunked download strategy was implemented to mitigate network instability. Mutation data were organized following the GDC hierarchical structure for downstream MAF parsing and mutational signature analysis.

Preprocessing of TCGA-STAD Transcriptomic Data

Raw RNA-seq count matrices underwent gene symbol mapping and deduplication: when multiple transcript IDs mapped to a single gene, the entry with the highest cohort-wide expression was retained. Expression matrices were $\log_2$(count+1)-transformed and quality-assessed through sample distribution plots and inter-sample correlation heatmaps to identify outliers. Clinical data integration involved sample ID standardization, overall survival (OS) extraction, and exclusion of cases with extremely short survival (≤30 days) to eliminate perioperative mortality confounders. Expression and clinical data were rigorously matched, retaining core variables with high completeness (age, sex, TNM stage) to construct a standardized analytical dataset.

Preprocessing of External Validation Cohorts (GSE62254, GSE15459, GSE84437)

Three GEO GC expression datasets underwent standardized preprocessing for cross-platform validation. Probe-to-gene mapping utilized GPL570 (Affymetrix) and GPL6947 (Illumina) annotations, removing blank entries and retaining primary symbols for multi-gene probes. Expression values were assessed for scale distribution and uniformly $\log_2(x+1)$-transformed to ensure comparability. Multi-probe aggregation strategies were platform-specific: median for Affymetrix and mean for Illumina, yielding gene-level expression matrices. Clinical data harmonization involved sample ID validation, unified event definitions for survival outcomes (OS/DFS), and

variable type standardization for key covariates (molecular subtype, TNM stage). Samples with complete expression-clinical data were retained and order-aligned to construct a cross-cohort comparable dataset.

Data Source and Preprocessing

TCGA-STAD served as the discovery cohort, with three GEO GC datasets (GSE15459, GSE62254, GSE84437) as validation cohorts. All cohorts comprised gene-level expression matrices with clinical data including overall survival time (OS.time) and event status (OS: 1=death/0=alive/censored). The preprocessing workflow involved: sample ID alignment followed by exclusion of cases with missing survival data or extremely short survival (≤30 days), and uniform $\log_2(x+1)$ transformation after distribution assessment. GEO cohorts were merged on common genes and batch-corrected using ComBat (sva package), with removal of genes containing missing values or near-zero variance (var<0.01) prior to correction. The batch-corrected expression matrix underwent gene-wise Z-score standardization ($\mu=0$, $\sigma=1$) to eliminate inter-cohort scale variations.

EIF5A Expression Stratification and Survival Analysis

EIF5A expression was extracted to construct survival data frames (sample ID, OS.time, OS, EIF5A_expr). Optimal cutpoints were independently determined per cohort (TCGA and GEO) using surv_cutpoint (survminer package, minprop=0.3 for adequate minimum group size), dichotomizing samples into high-expression (EIF5A>cutpoint) and low-expression

(EIF5A≤cutpoint) groups. Survival differences were assessed via Kaplan-Meier curves with log-rank tests, and hazard ratios (HR) with 95% confidence intervals were estimated using Cox proportional hazards models. To validate the clinical robustness of EIF5A's prognostic value, stratified survival analyses were performed within subgroups defined by age (<65/≥65 years), TNM stage (I-II/III-IV), Lauren classification, and sex.

Immune Microenvironment Characterization

Sample-level immune scores were computed from predefined marker gene sets (checkpoint, T cell, NK cell, macrophage, Treg, exhaustion, cytotoxic) by averaging expression (colMeans) of available genes within each set. Between-group differences (EIF5A-high vs. -low) were assessed using Wilcoxon rank-sum tests. Effector/suppressor ratios were calculated to quantify immune functional balance: Cytotoxic/Exhaustion and Cytotoxic/(Macrophage+Treg), with $\varepsilon=1\times10^{-6}$ to prevent division by zero, followed by statistical comparisons and boxplot visualization.

Differential Expression Analysis with Adaptive Threshold Strategy

Genome-wide differential expression between EIF5A-high and -low groups was performed using limma: intercept-free design matrices (~0+group) were contrasted (High-Low) and fitted via lmFit and eBayes. To ensure statistical power for enrichment analyses, an adaptive threshold selection strategy was employed: adjusted p-values (adj.P.Val, FDR<0.05) were prioritized with dynamic |logFC| thresholds (0.5→0.4→0.3→0.25). DEGs

were defined as upregulated (logFC>threshold with p-value criteria, enriched in EIF5A-high group) and downregulated (logFC<-threshold with p-value criteria, enriched in EIF5A-low group). Volcano plots highlighted top 20 genes with maximal fold changes (10 per direction), with subtitles indicating the final threshold type and DEG counts.

Functional Enrichment Analysis Strategy

Differential genes were converted from SYMBOL to ENTREZID (org.Hs.eg.db, bitr function) for separate enrichment of upregulated and downregulated genes in GO Biological Processes (enrichGO, ont="BP") and KEGG pathways (enrichKEGG, organism="hsa"), with BH correction for multiple testing ($p<0.05$, $qvalue<0.2$). As the intersection of DEGs between batch-corrected TCGA and GEO cohorts typically failed to meet preset criteria for robust enrichment (total≥80 with ≥20 per direction), enrichment analyses were performed independently in each cohort, with top 10 terms visualized via dotplots. Common biological pathways were identified by comparing enrichment results across cohorts to extract core regulatory signals.

Tumor Mutation Burden Analysis

Tumor mutation burden (TMB) was calculated from TCGA-STAD GDC somatic mutation data (Masked Somatic Mutation, MAF format). MAF files were recursively scanned and read, then merged for TMB calculation using maftools (mutations per megabase, captureSize=38Mb for whole-exome

sequencing coverage). Sample matching employed exact matching with full TCGA barcodes (Tumor_Sample_Barcode). TMB differences between EIF5A-high and -low groups were assessed using Wilcoxon rank-sum tests, with boxplots overlaid with jittered points displaying distributions annotated with mean±SD and p-values.

Copy Number Variation Analysis of EIF5A

The copy number variation (CNV) driving effect of EIF5A in GC was assessed using the TCGA-STAD cohort. Expression and CNV data were derived from RNA-seq transcriptomes and DNAcopy-segmented CNV data, respectively. Based on EIF5A genomic annotation (Chr17: 7,306,999–7,312,463), CNV segments covering this region were extracted, with segment-averaged copy numbers serving as sample-level CNV quantitative metrics. Expression and CNV data were matched via patient identifiers, retaining only primary tumor samples. Samples were stratified by CNV amplitude using GISTIC2.0 thresholds into copy number loss (Loss), neutral (Neutral), and gain (Gain) groups. Spearman rank correlation assessed the relationship between CNV levels and mRNA expression, while non-parametric tests (Kruskal-Wallis/Wilcoxon) compared expression differences across CNV status groups to evaluate dose-dependent effects of copy number alterations. Results were visualized through correlation scatter plots (with regression lines and 95% confidence intervals), CNV status-grouped boxplots, and statistical summary tables, systematically

characterizing EIF5A CNV distribution patterns and their regulatory impact on transcriptional expression.

DNA Methylation Analysis of EIF5A

DNA methylation-transcription coupling and prognostic analyses were conducted using the TCGA-STAD cohort. Methylation data were obtained from Illumina HumanMethylation450 BeadChip platform (β values), with matched mRNA expression matrices, paired via TCGA sample barcodes and restricted to primary tumors. EIF5A-associated CpG sites were precisely mapped using 450K probe annotations, excluding probes potentially confounded by homologous genes or pseudogenes to establish an EIF5A-specific methylation probe set. CpG sites were mapped to functional regions (promoter/5′UTR/first exon/gene body/3′UTR) using Illumina annotations to support region-specific mechanistic interpretation.

Spearman rank correlation assessed the association between methylation levels at each CpG site and EIF5A mRNA expression, defining significantly correlated sites as candidate functional loci. Correlation results were stratified by probe functional regions to characterize the EIF5A "methylation regulatory landscape," with emphasis on differential regulatory patterns between gene body and promoter regions. Results were visualized panoramically to reveal coupling structures and directional characteristics between CpG sites across functional regions and transcriptional output.

Survival analyses were performed for significantly correlated CpG sites. Patients were stratified by methylation levels, with Kaplan-Meier curves estimated and group differences assessed via log-rank tests. Cox proportional hazards models calculated hazard ratios (HR) with 95% confidence intervals to quantify associations between methylation stratification and mortality risk. Multi-site prognostic effects were summarized in forest plots, integrated with functional region annotations for mechanistic interpretation. This workflow established an evidence chain from "functional annotation—expression coupling—prognostic assessment" to identify key CpG sites with regulatory significance and clinical relevance.

Identification and Validation of EIF5A Transcriptional Regulators

A "perturbation evidence integration—cross-cohort expression coupling—independent cohort survival verification" framework was employed to systematically identify upstream regulators of EIF5A. Perturbation data from GPSAdb2.0 were aggregated across studies to summarize EIF5A expression responses following genetic perturbations (CRISPR knockout/RNAi), with cross-study aggregation by regulatory direction and effect size. Effects were weighted by experimental strategy reliability, with candidate sets constrained by study count and directional consistency to prioritize regulators with causal implications.

Spearman correlations between candidate regulators and EIF5A were calculated in TCGA-STAD and independent GEO cohorts, requiring

consistent correlation directions across cohorts with significance thresholds to ensure reproducibility, and validating concordance between perturbation evidence and tumor sample correlation directions. Key regulators were subjected to overall survival (OS) analyses in both cohorts: Kaplan-Meier curves with log-rank tests assessed group differences, and Cox proportional hazards models estimated HR with 95% confidence intervals; stratification employed data-driven optimal cutpoints with median-split sensitivity validation.

To evaluate the combined prognostic effect of the "regulator—EIF5A axis," four-quadrant joint stratification based on dual gene expression (high-high/high-low/low-high/low-low) was constructed to compare survival differences across phenotypes, validating the clinical relevance and stability of this regulatory relationship. Results were presented via correlation scatter plots, stratified survival curves, and hazard ratio forest plots.

Immune Microenvironment Association Analysis of EIF5A and Regulatory Factors

Immune microenvironment associations of EIF5A and upstream regulators were analyzed in TCGA and GEO cohorts, integrating immune infiltration quantification, immune gene expression coupling, group-based differential testing, and tumor mutation burden (TMB) assessment. Preprocessed expression matrices, clinical data, and candidate regulator gene sets (merged with EIF5A) were loaded. Single-sample enrichment strategy

(ssGSEA) calculated immune cell infiltration scores: gene-wise standardized expression matrices were averaged within predefined immune cell marker gene sets to derive cell-type infiltration scores, constructing immune infiltration profiles across cohorts.

Expression matrices for chemokines, chemokine receptors, and immune regulatory genes (immune checkpoints/HLA/inflammatory and immunosuppressive cytokines) were extracted. Spearman rank correlations assessed at the sample level: (i) associations between EIF5A/regulators and immune infiltration scores; (ii) associations between EIF5A/regulators and chemokine/immune regulatory gene expression. Significant correlations were filtered by preset thresholds and visualized via heatmaps, lollipop plots, and bubble plots.

Samples were stratified into high/low groups by target gene expression (data-driven optimal cutpoints, with median as fallback when unstable), with Wilcoxon rank-sum tests comparing immune infiltration score differences and significance annotation. TMB analysis calculated mutation burden from TCGA-STAD MAF data (maftools, captureSize=38Mb), comparing TMB differences by gene expression groups (Wilcoxon test) and assessing Spearman correlations between TMB and gene expression, presented via boxplots and scatter plots. Statistical analyses were performed in R (data.table/dplyr for processing, ggplot2/ggpubr/pheatmap

for visualization, survival/survminer for cutpoint and survival analyses, maftools for TMB calculation).

Microbiome Association Analysis

We assessed associations between HMGA1 and the tumor-associated microbiome in TCGA-STAD using multi-omic data. Genus-level relative abundance, matched transcriptome expression, and clinical data were imported; TCGA barcodes were harmonized by truncation to the first 14 characters and intersected across data types, retaining primary tumors. HMGA1 expression was dichotomized by the cohort median (HMGA1-High vs HMGA1-Low). Genera with mean relative abundance>1×10^-4 were tested between groups using Wilcoxon rank-sum tests, with log2 fold change computed using a small pseudocount to avoid division by zero; taxa with $P<0.05$ were reported as differentially abundant. Alpha diversity (Shannon, Simpson, richness) was computed from the full community matrix and compared by Wilcoxon tests. Beta diversity was quantified using Bray–Curtis distances; within-group dispersion (mean pairwise distance or PERMDISP) was assessed, community structure was visualized by PCoA, and overall group differences were tested by PERMANOVA with 999 permutations, with P-values annotated on the ordination. Visualization included stacked bar charts of the top 10 genera, LDA-like composite scores for differential taxa ($|\log2FC| \times -\log10\ P$) displayed as direction-colored bar plots, and grouped bar plots of differential taxa with significance labels. At the correlation level, Spearman tests evaluated associations between

HMGA1 expression and abundances of significant genera, displayed as scatter plots. All analyses were performed in R (data handling/visualization via data.table/dplyr/ggplot2; ecological statistics via vegan/phyloseq; permutation testing via adonis2 and related functions).

EIF5A-Associated Proteomics and Prognostic Model Development

Proteomics, transcriptomics, and clinical survival data from the TCGA-STAD cohort were integrated to systematically characterize biological differences and prognostic implications between EIF5A-high and -low expressing tumors. Following sample harmonization and standardization, patients were stratified into high- and low-expression groups by median EIF5A expression. Mechanistic dissection focused on four predefined biological axes: (i) translation and proteostasis; (ii) DNA damage response and apoptosis; (iii) stemness/EMT/Wnt/FAK/PKC oncogenic signaling; (iv) immune and inflammatory stress. For key protein markers within each axis, Spearman correlations with EIF5A were assessed and between-group protein level differences were compared using Wilcoxon rank-sum tests, with dual "correlation + differential" evidence supporting mechanistic inference.

FOXO3A activity was quantified by calculating the ratio of total FOXO3A to inhibitory phospho-FOXO3A_pS318S321 (total protein/phosphoprotein ratio), comparing FOXO3A activity changes between EIF5A groups to infer tumor suppressor signal intensity. For prognosis, survival analyses were performed separately for proteins positively/negatively correlated with

EIF5A to identify representative favorable or unfavorable prognostic markers, with multidimensional visualization summarizing "EIF5A correlation direction—survival risk" concordance.

A protein-based prognostic risk model was constructed via stratified train/test split (7:3). Prognostically relevant proteins ($P<0.05$ by univariate Cox) underwent LASSO-Cox regularization, with λ.min determining optimal feature subsets. Risk scores (risk score=$\Sigma\beta_i \times$protein expression$_i$) were calculated, with patients dichotomized into high/low-risk groups by median and discrimination validated in the test set (Kaplan-Meier curves/log-rank tests/time-dependent ROC-AUC). This workflow used mechanistic axes as scaffolds to connect molecular correlations, group differences, survival outcomes, and predictive modeling, constructing a comprehensive evidence chain for EIF5A-associated tumor biology and clinical significance.

Tumor Microenvironment Heterogeneity Analysis via Single-Cell Transcriptomics

A systematic single-cell RNA sequencing analysis workflow was established to dissect cellular heterogeneity in tumor versus normal tissue microenvironments. GSE183904 dataset was retrieved from GEO, with sample attributes extracted from series_matrix files to establish sample-phenotype mappings. Quality control employed standardized thresholds: 500-6000 detected genes, ≤20% mitochondrial gene proportion, and ≥3 minimum cell count. Data normalization used log-normalization

(log$_2$[(counts×10^4/total_counts)+1]), with the top 2000 highly variable genes identified via variance stabilization transformation for dimensionality reduction.

Batch effect correction employed the Harmony algorithm, achieving soft clustering and linear correction through iterative optimization while balancing cluster purity and batch mixing. UMAP nonlinear dimensionality reduction was performed on the first 30 Harmony-corrected embedding dimensions, with Louvain algorithm (resolution 0.5) constructing shared nearest neighbor graphs for community detection. COSG algorithm identified top 100 cluster-specific marker genes by optimizing cosine similarity and expression sparsity, with clusters annotated into five major lineages (epithelial, stromal, endothelial, immune, proliferative) based on canonical markers.

Cellular composition differential analysis calculated cell-type proportions at the sample level, with Wilcoxon rank-sum tests comparing normal versus tumor group distributions and Benjamini-Hochberg correction for multiple testing. Target gene EIF5A expression analysis extracted log-normalized layer data, calculating sample-level means within all cells and individual subpopulations for statistical inference to avoid cell-level pseudoreplication.

Association Analysis Between EIF5A Expression and Malignant Features via Single-Cell Transcriptomics

A systematic single-cell RNA sequencing analysis workflow was established to dissect EIF5A expression patterns in tumor epithelial cells and associations with malignant features. Normalized expression matrices were extracted from pre-analyzed tumor epithelial cell data (with CNV annotations), with GetAssayData function retrieving log-normalized layer expression values for EIF5A and regulatory factor HMGA1 at single-cell resolution. Gene expression stratification employed median dichotomization to partition cells into high/low expression groups, offering superior statistical power compared to quartile-based methods, with stratification performed at both global and within-sample levels to validate robustness.

Functional feature scoring integrated seven biological pathway categories: proliferation signaling (20 genes including MKI67/TOP2A/PCNA), epithelial-mesenchymal transition EMT (23 genes including VIM/CDH2/SNAI1/ZEB1), hypoxia response (22 genes including VEGFA/SLC2A1/HK2/LDHA), endoplasmic reticulum stress UPR (20 genes including HSPA5/XBP1/ATF4/DDIT3), inflammatory response (20 genes including IL1B/IL6/TNF/NFKB1), interferon response (22 genes including ISG15/MX1/OAS1/IFIT1), and glycolysis (20 genes including HK1/PFKL/ALDOA/ENO1/PKM), with AddModuleScore function calculating per-cell module scores. Cell cycle scoring utilized Seurat's built-in

CellCycleScoring function based on S-phase and G2M-phase marker gene sets to compute S.Score/G2M.Score and determine cell cycle phase.

Statistical inference proceeded at cellular and sample levels: cell-level Wilcoxon rank-sum tests compared functional feature differences between EIF5A-high and -low groups; sample-level Spearman correlation analyses assessed association strength between aneuploidy proportions, gene expression, and functional scores after calculating per-sample means to avoid cell-level pseudoreplication. Effect sizes were quantified via Cohen's d to measure practical group difference magnitudes ($d=(\mu_1-\mu_2)/\sigma_pooled$, where σ_pooled represents pooled standard deviation).

Cell culture, transfection and qRT-PCR

Human normal gastric epithelial cells (GES-1) and GC cell lines (MKN45, AGS, HGC-27 and SNU-719) were cultured in recommended complete medium supplemented with 10% fetal bovine serum and antibiotics at 37 degrees C in a humidified 5% CO2 incubator. HGC-27 and SNU-719 were chosen for functional assays because their endogenous EIF5A expression was relatively low. Cells were transfected with EIF5A overexpression plasmid (oe-EIF5A) or the corresponding negative control plasmid (oe-NC) using a lipid-based protocol according to the manufacturer's instructions.

Total RNA was extracted and reverse-transcribed into cDNA. qRT-PCR used SYBR Green-based detection, and relative EIF5A expression was normalized

to beta-actin with the 2^-DeltaDeltaCt method. Three independent biological replicates were included unless otherwise stated.

Cell proliferation, colony formation, migration and wound-healing assays

Cell viability was measured by CCK-8 assay at the indicated time points after transfection. For colony formation, transfected cells were seeded at low density, cultured until visible colonies formed, fixed, stained and counted. Transwell migration assays used chambers without Matrigel; migrated cells on the lower membrane surface were fixed, stained, imaged and quantified. For wound-healing assays, confluent monolayers were scratched with a sterile pipette tip, washed to remove detached cells and cultured in low-serum medium. Wound closure was photographed at 0, 24 and 48 h and quantified as wound-healing rate.

Statistical analysis

Statistical analyses were performed primarily in R. Two-group comparisons used Wilcoxon rank-sum tests or Student's t-tests as appropriate. Multiple-group comparisons used Kruskal-Wallis tests with post hoc pairwise testing when applicable. Spearman correlation was used for non-parametric association analyses. Survival differences were assessed with log-rank tests and Cox proportional hazards models. $P<0.05$ was considered statistically significant unless otherwise specified, and multiple testing was adjusted with the Benjamini-Hochberg method.

Results

EIF5A Links Immunotherapy Response to Tumor Immunogenicity and Subgroup-Specific Survival in GC

To systematically assess the relationship between EIF5A and immunotherapy benefit alongside GC clinical-molecular characteristics, we first examined prior immunotherapy cohort evidence showing elevated EIF5A expression in responders [5], suggesting potential associations with antitumor immune activation or sensitivity **(Figure 1A)**. In the TCGA cohort, EIF5A-high patients exhibited significantly elevated tumor mutation burden (TMB) (Wilcoxon $P=8.311\times10^{-7}$; low 3.9±6.4 vs. high 10.2±18.5 mutations/Mb), providing genomic support for associations between EIF5A and immunotherapy benefit **(Figure 1B)**. Clinical stratification revealed higher EIF5A expression in Stage I-II versus Stage III-IV within the merged GEO cohort (Kruskal-Wallis $P=2.871\times10^{-5}$), while TCGA showed a trend without significance (P=0.077) **(Figure 1C)**, and Lauren subtype analysis indicated significant differences across subtypes in GEO (Kruskal-Wallis $P=1.707\times10^{-4}$) with intestinal type generally higher than diffuse, whereas TCGA exhibited similar directionality without overall significance (P=0.17) **(Figure 1D)**. Notably, the GEO validation set comprised three Asian cohorts (GSE62254/GSE15459/GSE84437) while TCGA predominantly represented Western populations, with differences in molecular lineage composition, clinical stratification proportions, treatment strategies, and platform characteristics potentially influencing reproducibility of EIF5A-

clinical signals, thus these inconsistencies are considered reasonable sources of cohort compositional heterogeneity rather than racial causal conclusions. Sex analysis showed higher EIF5A in females versus males in TCGA (Kruskal-Wallis P=0.011) but not in GEO (P=0.73) **(Figure 1E)**, while age stratification (<65 vs. ≥65 years) showed no significant differences in either cohort (GEO P=0.09; TCGA P=0.15) **(Figure 1F)**. Prognostic analyses revealed marked subgroup-specific EIF5A effects: sex stratification showed associations between high EIF5A and improved overall survival (OS) primarily in males (GEO P<0.0001; TCGA P=0.0045) with no significance in females (GEO P=0.15; TCGA P=0.55) **(Figure 1G, Supplementary Figure 1A)**, stage stratification demonstrated protective associations enriched in advanced Stage III-IV patients (GEO P<0.0001; TCGA P=0.032) without significance in early Stage I-II (GEO P=0.18; TCGA P=0.099) **(Figure 1H, Supplementary Figure 1A)**, and Lauren subtype stratification suggested inter-cohort heterogeneity with high EIF5A indicating better OS in Mixed (P=0.047) and Intestinal (P=0.00054subgroups in GEO **(Figure 1I)** but no association in TCGA Intestinal subtype (P=0.75) **(Supplementary Figure 1A**), whereas Diffuse subtype showed more consistent results with strong correlations in GEO (P<0.0001) **(Figure 1J left)** and Kaplan-Meier Plotter surgery-only populations **(Figure 1J right)**. Age stratification revealed EIF5A protective signals in both elderly/non-elderly populations with slight intensity variations (GEO: ≥65 years P<0.0001, <65 years P=0.076; TCGA: <65

years P=0.015, ≥65 years P=0.065) **(Figure 1K, Supplementary Figure 1A)**. In summary, EIF5A consistently associated with immunotherapy response and elevated TMB, demonstrating more stable prognostic value in male, advanced-stage, and diffuse-type subgroups, while compositional differences between Asian-predominant GEO and Western-predominant TCGA cohorts may partially explain cohort-specific stratification signals.

High EIF5A Expression Reshapes the Transcriptomic Landscape and Maintains Epithelial Phenotypic Features

In two independent cohorts (GEO and TCGA), we stratified samples based on EIF5A expression levels (high/low) and performed differential expression analysis. The results revealed 469 differentially expressed genes (DEGs) in the GEO cohort, comprising 110 upregulated and 359 downregulated genes in the EIF5A-high group **(Figure 2A)**. In the TCGA cohort, a more substantial number of DEGs was identified, including 276 upregulated and 20,400 downregulated genes **(Figure 2A)**. These findings indicate that elevated EIF5A expression is accompanied by extensive transcriptomic remodeling, and this phenomenon demonstrates remarkable consistency and reproducibility across datasets from different sources. To elucidate the biological implications of these DEGs, we further conducted GO and KEGG enrichment analyses **(Figure 2B)**. In the GEO cohort, downregulated genes associated with high EIF5A expression were primarily enriched in processes related to extracellular matrix remodeling and cell adhesion/migration, including extracellular matrix/structure organization, cell–substrate

adhesion, and focal adhesion assembly, along with muscle contraction-related processes (muscle system process/muscle contraction). Correspondingly, KEGG pathway analysis highlighted enrichment in Vascular smooth muscle contraction, cGMP–PKG signaling, Wnt signaling, Complement and coagulation cascades, and Renin–angiotensin system, pointing to alterations in cytoskeleton/contractile functions, matrix interactions, and microenvironment-related signaling. Conversely, upregulated genes associated with high EIF5A expression showed significant GO enrichment in mitotic and chromosome segregation processes, such as chromosome segregation, nuclear division, sister chromatid segregation, and microtubule cytoskeleton organization involved in mitosis. At the KEGG level, these genes were prominently enriched in Cell cycle and p53 signaling pathway, suggesting that elevated EIF5A expression is closely linked to cell cycle progression and activation of proliferative programs.

In the TCGA cohort, upregulated genes associated with high EIF5A expression were enriched in GO terms related to epithelial differentiation and keratinization processes, including epidermal cell differentiation, keratinocyte differentiation, keratinization, and intermediate filament cytoskeleton organization, accompanied by enrichment in chromatin/nucleosome organization processes (protein–DNA complex organization, nucleosome organization/assembly). KEGG analysis revealed significant enrichment in Cornified envelope formation, ATP-dependent

chromatin remodeling, Cell cycle, and inflammation-related pathways such as IL-17 signaling pathway. In contrast, downregulated genes associated with high EIF5A expression were predominantly enriched in GO/KEGG terms related to sensory receptors and neurohumoral signaling, including sensory perception of smell, Olfactory transduction, Neuroactive ligand–receptor interaction, cAMP signaling pathway, and Taste transduction. This suggests that elevated EIF5A expression is accompanied by relative silencing of "non-core tumor/tissue-specific" signaling modules, with transcriptional programs shifting toward networks associated with epithelial structure maintenance and proliferation.

Furthermore, correlation analysis in the TCGA dataset revealed that EIF5A expression exhibited significant positive correlations with epithelial marker genes KRT8, KRT18, and CLDN7, while showing significant negative correlations with epithelial-mesenchymal transition (EMT)-related genes VIM, CDH2, FOXC2, ACTA2, ZEB1, ZEB2, LAMA4, and TWIST2 **(Figure 2C)**. Consistently, EIF5A expression demonstrated a strong positive correlation with epithelial signature scores ($R=0.58$, $P=1.5\times 10^{-37}$) and a negative correlation with mesenchymal/EMT signature scores ($R=-0.27$, $P=3\times 10^{-8}$) **(Figure 2C–D)**. Taken together with our previous observation that high EIF5A expression is associated with improved prognosis, these findings collectively demonstrate that EIF5A overexpression synergizes with "epithelial phenotype maintenance/differentiation" and "cell cycle

activation" while antagonizing the EMT program. This suggests that EIF5A may maintain the epithelial phenotype of tumor cells by reshaping epithelial structure-related transcriptional networks and restricting mesenchymal transdifferentiation, thereby influencing their biological behavior and ultimately contributing to improved clinical outcomes.

Proteomic Signature of EIF5A Overexpression: Coordinated Translational Enhancement and EMT Suppression

To validate EIF5A-related molecular networks at the protein level, we systematically analyzed the association patterns between EIF5A and key signaling pathway proteins using TCGA proteomic data. In the "Translation and Growth" module, EIF5A exhibited significant positive correlations with classical protein synthesis and translational initiation axes: p70S6K1, p90RSK, EIF4E, MNK1, and 4EBP1 all increased with elevated EIF5A expression, suggesting that EIF5A overexpression parallels a protein phenotype of "enhanced translational capacity/protein synthesis output" **(Figure 3A).** Notably, the m6A reader protein YTHDF2 also positively correlated with EIF5A, suggesting this association may extend to mRNA metabolism and translational fate regulation. However, EIF5A negatively correlated with several "growth signal convergence" nodes, including c-KIT, RICTOR, PAK4, Cyclin D1, SGK3, and PRAS40, indicating that EIF5A overexpression does not represent a "global upregulation of oncogenic growth signals," but rather reflects selective coupling with "translation/protein synthesis programs" accompanied by suppressive

regulation of pro-invasive or specific pathway branches.In the "Damage/Apoptosis" module, EIF5A positively correlated with DNA damage response and genome stability-related proteins: CHK2, VHL, MSH2, and cell cycle checkpoint-associated CDK1_pY15 were upregulated with increased EIF5A expression, suggesting that EIF5A-high samples tend to activate and maintain DNA damage surveillance, repair, and checkpoint phenotypes **(Figure 3B)**. This finding echoes the association between EIF5A overexpression and favorable prognosis observed in Figure 1, implying that genome stability maintenance may be one mechanism underlying its prognostic protective effect.In the "Oncogenic proteins" module, EIF5A demonstrated widespread negative correlations with multiple oncogenic, stemness, and migration-invasion-related proteins **(Figure 3C)**. Particularly noteworthy is the overall negative correlation pattern with multiple key nodes of classical EMT/invasion and cell adhesion-cytoskeletal remodeling. This systematic negative correlation pattern further corroborates the antagonistic relationship between EIF5A and EMT programs observed at the transcriptomic level, providing direct evidence at the protein functional level.

In the tumor immune regulation dimension, TCGA proteomic data revealed significant positive correlations between EIF5A and PD-L1 ($r=0.202$, $P=4.5\times 10^{-4}$), as well as with IRF-3 ($r=0.151$, $P=0.0088$) **(Figure 3D)**. This suggests that EIF5A-associated protein states not only link intracellular translation/oncogenic-EMT networks but also couple with tumor immune

regulatory endpoints: on one hand, PD-L1 elevation points to potential immune checkpoint activation; on the other hand, IRF-3 elevation suggests that innate immunity/interferon-related transcriptional axes may participate in shaping or feedback regulation of EIF5A-related phenotypes.Regarding the negative correlation between EIF5A and the tumor suppressor FOXO3A, we performed in-depth phosphorylation state analysis **(Figure 3E)**. Although total FOXO3A protein levels negatively correlated with EIF5A, its inactivated form FOXO3A_pS318S321 also negatively correlated with EIF5A. To assess the actual transcriptional activity of FOXO3A, we constructed a FOXO3A activity ratio (Activity Ratio= (FOXO3A) - Z(FOXO3A_pS318S321), where higher values represent stronger FOXO3A tumor suppressive activity). Group comparisons showed no significant difference in FOXO3A activity ratio between EIF5A-high and -low groups (P=0.36), suggesting that EIF5A-related prognostic and pathway associations are not primarily driven by FOXO3A activity changes, but more likely explained by the aforementioned combinatorial network of translational enhancement, oncogenic/EMT suppression, and immune checkpoint activation.

To systematically evaluate the association between proteomic features and prognosis, we performed quadrant mapping of each protein's correlation coefficient with EIF5A against its univariate survival risk **(Figure 3F)**. Results showed: proteins in the upper-left quadrant (negatively correlated with EIF5A and associated with worse prognosis, EIF5A^{-}/Bad) showed

obvious clustering, including invasion-related nodes such as FAK_pY397, PAK4, and c-KIT, suggesting that when EIF5A is low, these multi-pathway oncogenic/EMT/invasion nodes are more likely to be at high levels and drive risk elevation; proteins in the lower-right quadrant (positively correlated with EIF5A and associated with better prognosis, EIF5A^{+}/Good) enriched DNA repair and translation-related proteins such as CHK2 and MSH2, forming a correspondence of "EIF5A elevation—protective protein program elevation."

Finally, we constructed a prognostic prediction model based on EIF5A-related protein features **(Figure 3G)**. Using LASSO-Cox regression for feature selection and penalized regression, we ultimately incorporated core proteins to establish a risk scoring system. Time-dependent ROC curve analysis showed that the model performed best in 5-year survival prediction: the test set AUC reached 0.733 (5-year), superior to 1-year (0.505) and 3-year (0.635); corresponding training set AUCs were 0.932 (5-year), 0.788 (1-year), and 0.784 (3-year). This result indicates that EIF5A-related protein features are more sensitive to "long-term outcomes," suggesting they more likely capture stable biological states driving long-term tumor evolution rather than merely reflecting short-term fluctuations.

In summary, systematic analysis based on TCGA proteomic data reveals that EIF5A overexpression coordinately activates translational enhancement and genome stability maintenance programs at the protein functional level,

while systematically suppressing EMT/invasion-related focal adhesion, cytoskeletal remodeling, and oncogenic signaling networks, and exhibiting complex coupling relationships with tumor immune checkpoints. This multidimensional proteomic signature not only validates transcriptomic findings at the mechanistic level but also provides direct functional protein evidence for the prognostic protective role of EIF5A.

High EIF5A expression associates with a complex immune microenvironment characterized by immune activation and checkpoint upregulation

To systematically evaluate the relationship between EIF5A expression and tumor immune microenvironment status, we quantified immune-related gene signatures in tumor samples and performed stratified comparisons between EIF5A-high and EIF5A-low groups (median split) across two independent cohorts (GEO and TCGA). We constructed a multi-dimensional immune signature panel encompassing: immune checkpoint (Checkpoint) signature comprising PDCD1, CD274, PDCD1LG2, CTLA4, LAG3, TIGIT, and HAVCR2; cytotoxic (Cytotoxic) signature composed of GZMA, GZMB, PRF1, and IFNG; and signatures representing major immune cell populations, including T cells (CD3D, CD3E, CD4, CD8A, CD8B), B cells (CD19, CD79A, MS4A1), NK cells (NCAM1, NKG7, KLRD1, GNLY), macrophages (CD68, CD163, CD14, CSF1R), dendritic cells (ITGAX, CD1C, CLEC4C), regulatory T cells/Tregs (FOXP3, IL2RA), and T cell exhaustion (Exhaustion; PDCD1, CTLA4, LAG3, HAVCR2, TIGIT, TOX).

In the GEO cohort, the EIF5A-high group exhibited a stronger immune effector-associated phenotype **(Figure 4A)**: both Checkpoint and Cytotoxic signatures were significantly elevated ($P<0.01$), accompanied by enhanced NK cell signals ($P<0.05$). Concurrently, Treg-related signals were also upregulated ($P<0.05$), while myeloid-related components, particularly the dendritic cell (Dendritic) signature, showed a significant decline ($P<0.01$). This pattern suggests that elevated EIF5A parallels enhanced immune effector functions, yet antigen presentation and immune regulatory axes may undergo remodeling, forming a "high-effector-high-regulation" equilibrium state. In the TCGA cohort, these associations were independently validated with even greater consistency **(Figure 4B)**: the EIF5A-high group similarly demonstrated significantly enhanced Checkpoint and Cytotoxic signals (both $P<0.001$), concurrent with upregulation of T cell ($P<0.01$), NK cell ($P<0.05$), and immunosuppressive/exhaustion-related signatures, including activation of Treg ($P<0.01$) and T cell exhaustion (Exhaustion, $P<0.01$) programs. Notably, the dendritic cell (Dendritic) signature was also significantly reduced in the TCGA cohort ($P<0.001$), highly consistent with the GEO findings and further supporting the association between high EIF5A expression and diminished antigen-presenting cell populations.

Collectively, cross-cohort analysis reveals that EIF5A expression stratification robustly marks a complex immune ecological niche at the transcriptomic level: on one hand, high EIF5A expression accompanies

immune checkpoint activation and enhanced cytotoxic programs, indicating the presence of active anti-tumor immune responses within the tumor tissue; on the other hand, immunosuppressive components (Tregs, T cell exhaustion) and reduced antigen-presenting cells occur simultaneously, suggesting that immune effector and immune evasion mechanisms may coexist. This dual "immune activation-immune suppression" phenotype likely reflects a dynamic immune equilibrium state in EIF5A-high tumors, providing important insights for understanding its prognostic significance and potential responsiveness to immunotherapy.

Multi-omics analysis identifies HMGA1 as a key upstream regulator of EIF5A overexpression

To elucidate the potential upstream drivers of aberrant EIF5A overexpression, we conducted systematic analyses across three molecular layers in the TCGA cohort: copy number variation (CNV), DNA methylation, and functional perturbation evidence integration.

First, we evaluated the correlation between EIF5A copy number (segment mean) and mRNA expression levels in TCGA samples. The results revealed a weak positive correlation (Spearman $r=0.17$, $P=5.0\times10^{-4}$) **(Figure 5A)**, suggesting that CNV may contribute to EIF5A expression but with limited effect size. Stratified comparison by CNV status (Loss, Neutral, Gain) demonstrated that both Neutral and Gain groups exhibited significantly elevated EIF5A expression compared to the Loss group (Loss vs. Neutral:

P=3.1×10⁻⁴; Loss vs. Gain: P=0.017); however, no significant difference was observed between Gain and Neutral groups (P=0.31) **(Figure 5B)**, indicating that EIF5A expression elevation does not further increase linearly with copy number gain. Consistently, density analysis of CNV distribution across the entire cohort showed that EIF5A copy numbers in most samples were concentrated in the Neutral range **(Figure 5C)**, supporting the conclusion that "EIF5A overexpression is not primarily driven by copy number amplification."

Given the limited explanatory power of CNV, we further focused on TCGA methylation data to systematically evaluate the associations between CpG site methylation across different genomic regions of EIF5A and its mRNA expression. We employed lollipop plots to display the Spearman correlation coefficients and significance levels between CpG probe methylation and EIF5A expression across promoter distal regions, gene body, and 3' untranslated region (3'UTR) **(Figure 5D)**. The results revealed several methylation sites significantly associated with EIF5A expression across different genomic regions; however, the overall correlation strength remained weak to moderate, with no strongly correlated sites observed. Subsequently, we performed univariate survival analysis (using optimal cutpoint stratification) for statistically significant CpG sites and found that most sites did not exhibit stable prognostic associations. Only cg01418441 showed a prognosis-related signal **(Figure 5E)**, which may be located in the gene body region, with its methylation level positively coupled with

EIF5A expression and associated with better survival outcomes. However, given that its correlation with EIF5A expression remained weak, this suggests that methylation alterations may participate in EIF5A expression regulation but are unlikely to serve as the dominant driver explaining sustained EIF5A overexpression.

To obtain upstream regulatory clues with stronger causal implications, we systematically queried the GPSAdb2.0 database for "which genes' functional perturbations induce EIF5A expression changes" and employed a three-step screening framework: "perturbation evidence integration—cross-cohort expression correlation validation—independent cohort survival verification" for candidate gene identification **(Figure 5F)**. We performed cross-study aggregation analysis of EIF5A expression responses under different perturbation strategies (including CRISPR knockout, RNAi silencing, etc.) and obtained a set of candidate regulators with directional consistency and multi-study support. Building upon this foundation, we conducted expression correlation validation in TCGA and GEO tumor cohorts, combined with prognostic analysis for priority ranking, ultimately identifying HMGA1 (High Mobility Group AT-hook 1) as the key upstream candidate gene: HMGA1 exhibited a strong positive correlation with EIF5A in the TCGA cohort (Spearman $r=0.89$, $P<2.2\times10^{-16}$, $n=404$), which was independently validated in the GEO cohort ($r=0.44$, $P<2.2\times10^{-16}$, $n=913$) **(Figure 5G)**. Furthermore, high HMGA1 expression was significantly associated with better overall survival in the TCGA cohort (log-rank

P=5.5×10⁻⁴; optimal cutpoint=6.036) (Figure 5H), and similar prognostic protective effects were observed in the GEO dataset **(Supplementary Figure 2A**), with prognostic directions highly consistent with EIF5A-associated results.

In summary, multi-omics integrative analysis indicates that both CNV and DNA methylation may exert limited regulatory effects on EIF5A expression but fail to explain the primary source of EIF5A overexpression. In contrast, results based on functional perturbation evidence and cross-cohort validation strongly support that the "HMGA1–EIF5A regulatory axis" more likely constitutes the key upstream pathway for aberrant EIF5A expression, and this regulatory relationship demonstrates clear clinical relevance and prognostic significance. This finding provides an important molecular basis for understanding EIF5A regulatory mechanisms in GC and offers a theoretical foundation for therapeutic strategies targeting the HMGA1-EIF5A axis.

Systematic associations between the HMGA1-EIF5A regulatory axis and tumor mutational burden, immune infiltration, and chemokine networks

To further validate and extend the associations between the "HMGA1-EIF5A regulatory axis" and immune-related phenotypes, we conducted systematic evaluations across four dimensions in the TCGA and independent GEO cohorts: tumor mutational burden (TMB), immune infiltration

characteristics, chemokine networks, and immune checkpoint signaling. In the TCGA cohort, stratification by HMGA1 expression levels revealed that the high-expression group exhibited significantly elevated overall TMB levels. Continuous variable correlation analysis also demonstrated a significant positive correlation between HMGA1 and TMB **(Supplementary Figure 2B)**. This trend was highly consistent with the immune-related patterns observed for EIF5A, suggesting that this regulatory axis may accompany a tumor biological background characterized by "high mutational burden—enhanced immunogenicity." Further immune infiltration analysis showed that the HMGA1-high group exhibited significantly elevated infiltration scores across multiple immune cell types, including $CD8^+$ T cells, NK cells, macrophages, and dendritic cells **(Supplementary Figure 2C)**, indicating that elevated HMGA1 associates with a more "immune-inflamed/immune-activated" tumor microenvironment state. Considering that immune cell recruitment largely depends on the precise regulation of chemokine networks, we further conducted systematic screening and correlation assessment of chemokine expression profiles in the TCGA cohort. Results revealed that HMGA1 and EIF5A exhibited highly consistent directionality in chemokine expression regulation, with positive correlations observed for most chemokines **(Supplementary Figure 2D)**. In the independent GEO cohort, the above "regulator-chemokine" coupling relationships were robustly validated **(Supplementary Figure 2E)**. Particularly noteworthy, key chemokines closely associated with T cell

recruitment, CXCL9 and CXCL10, showed consistent significant positive associations with both EIF5A and HMGA1 across two independent cohorts, strongly supporting that this regulatory axis may participate in shaping enhanced immune infiltration microenvironment phenotypes through strengthening T cell chemotactic signaling networks. Finally, we further examined immune checkpoint molecular correlation networks in both TCGA and GEO cohorts. Results showed that PDCD1 (PD-1) and CD274 (PD-L1) exhibited stable positive correlations in both independent cohorts **(Supplementary Figure 2F-G)**, both significantly correlated with HMGA1-EIF5A axis expression levels. This finding echoes the "immune checkpoint-related phenotype" in Figure 4 of the main text, suggesting that under the immune-inflamed background associated with the HMGA1-EIF5A axis, enhanced immune effector functions may coexist with synchronous checkpoint axis activation, thereby presenting an "inflamed but suppressed" immune ecological niche characteristic commonly seen in immunotherapy-sensitive populations. This finding not only provides multi-layered evidence for understanding the immune regulatory mechanisms of the HMGA1-EIF5A axis, but also offers potential molecular marker combinations for predicting immunotherapy response.

Intratumoral microbiota features associated with the HMGA1-EIF5A axis: preliminary evidence for immune-microbiome co-remodeling

Building upon the established association between the HMGA1-EIF5A axis and favorable prognosis, we further constructed a more comprehensive

landscape across three interconnected levels—"host transcriptional regulation—immune microenvironment—intratumoral microecology"—to evaluate whether microecological features may serve as accompanying characteristics of this axis-related immune phenotype. Based on intratumoral microbial profiling from the TCMA/TCGA-STAD cohort, GC samples exhibited marked individual heterogeneity at the genus level, with dominant genera including Prevotella, Helicobacter, Streptococcus, Haemophilus, Lactobacillus, Fusobacterium, Veillonella, and Leptotrichia **(Figure 6A)**. Although stratification by HMGA1 expression levels (median split) revealed no significant overall differences in α-diversity indices (Shannon, Simpson, Richness), and β-diversity analysis (Bray-Curtis distance) did not show strong separation (Adonis test showed a marginal trend) **(Figure 6B)**, principal coordinate analysis (PCoA) visualization suggested some degree of structural shift between the two groups along PCoA1/PCoA2 dimensions, albeit with partial overlap **(Figure 6C)**. This pattern suggests that HMGA1-associated microecological changes are more likely manifested as "selective remodeling of specific genera" rather than global diversity reorganization.

To identify key discriminant microbial taxa, we employed Linear Discriminant Analysis Effect Size (LEfSe) to screen for genera associated with HMGA1 stratification. Results revealed that Leptotrichia, Lachnoanaerobaculum, Streptococcus, and Peptoanaerobacter all exhibited significantly higher relative abundance in the HMGA1-low group **(Figure**

6D). Further continuous-variable correlation validation between differential genera abundance and HMGA1 expression levels showed that all these genera displayed negative correlation trends with HMGA1, with Streptococcus showing the most significant and consistent negative correlation **(Figure 6E)**. It is important to emphasize that reports on Streptococcus in GC intratumoral microbiome studies are relatively complex, exhibiting context-dependency. Previous GC subtyping studies integrating microbiome with multi-omics features suggest that microbial subtypes characterized by high Streptococcus abundance are associated with worse clinical outcomes. Our results showing significant enrichment of Streptococcus in the HMGA1-low group (relatively poorer prognosis) and its significant negative correlation with HMGA1 expression are highly consistent with previous literature reports, further supporting that high Streptococcus abundance may serve as a microecological feature associated with unfavorable prognosis. Similarly, the enrichment patterns of Leptotrichia, Lachnoanaerobaculum, and Peptoanaerobacter in the HMGA1-low group also suggest that these genera may collectively constitute a microecological signature combination associated with relatively poorer prognosis.

Integrating microecological results with the main narrative of this study: previous analyses have demonstrated that the HMGA1-EIF5A axis is highly consistent with immune-inflamed features (including chemokine network activation, enhanced immune cell infiltration, and checkpoint axis

upregulation) and favorable prognostic direction. Building upon this, Figure 6 further reveals that the high-expression state of this axis is not only coupled with immune phenotypes but also accompanied by systematic shifts in genus-level microbial profiles—particularly the relative reduction of multiple genera such as Streptococcus potentially associated with poor prognosis. Although this association does not manifest as significant overall diversity changes, the consistent trend among "differential genera—HMGA1 expression—clinical outcomes" provides an important complementary perspective: the HMGA1-EIF5A-associated tumor ecological niche may simultaneously encompass coordinated remodeling of host transcriptional regulation, immune recruitment/activation, and microecological composition, collectively pointing toward more favorable clinical outcomes. This multi-layered integrative pattern offers new insights for understanding the complexity of GC prognosis and developing comprehensive biomarker strategies.

Single-cell transcriptomics reveals preferential EIF5A expression in malignant epithelial cells associated with proliferation and metabolic adaptation

To dissect the cell type-specific expression pattern of EIF5A and its relationship with tumor cell states at single-cell resolution, we performed in-depth analysis of the GC single-cell dataset GSE183904. UMAP dimensionality reduction and clustering partitioned cells into major populations including epithelial cells, stromal cells, endothelial cells,

immune cells, and proliferating cells **(Figure 7A)**. Expression heatmap of signature genes for each subset showed: epithelial cells specifically expressed EPCAM, KRT8, and MUC5AC; stromal cells were enriched for ACTA2, DCN, and COL1A1; endothelial cells expressed PECAM1, VWF, and CDH5; immune cells were marked by CD68, PTPRC, and CD3D; proliferating cells highly expressed cell cycle markers MKI67, TOP2A, and CDK1 **(Figure 7B)**, validating the accuracy of cell type annotation. Systematic evaluation of EIF5A expression distribution across different cell populations using dotplot revealed that EIF5A was most prominently expressed in proliferating cells, followed by epithelial cells **(Figure 7C)**, suggesting a closer association between EIF5A and cell proliferation programs as well as tumor epithelial cell states. Comparison of EIF5A expression levels across major cell types by tissue origin (normal vs. tumor) showed that epithelial cells, stromal cells, and immune cells from tumor tissues all exhibited elevated EIF5A expression **(Figure 7D)**. Given the overall expression level in epithelial cells and their direct correspondence to tumor parenchymal cells, subsequent analyses focused primarily on the role of EIF5A in epithelial cells. To distinguish malignant from non-malignant epithelial cell states, we inferred copy number variations in tumor tissues using the inferCNV algorithm and classified cells into aneuploid and diploid categories. Results showed that aneuploid cells exhibited significantly higher EIF5A expression than diploid cells **(Figure 7E)**, and the proportion of EIF5A-high cells was markedly higher in the aneuploid population

(Figure 7F), suggesting that EIF5A upregulation accompanies aneuploidy-associated genomic instability features. At the functional level, we employed gene set scoring strategies to evaluate associations between EIF5A expression and multiple key biological programs **(Figure 7G)**. EIF5A showed relatively consistent positive correlations with glycolysis, proliferation, and hypoxia-related scores, while displaying negative correlations with EMT and endoplasmic reticulum stress (UPR/ER stress) scores, indicating that EIF5A tends to mark tumor cell states characterized by metabolic adaptation and active proliferation rather than mesenchymal transdifferentiation programs.Consistent with these findings, cell cycle phase analysis revealed that EIF5A expression was relatively low in G1 phase cells but enriched in S and G2M phases **(Figure 7H)**. Notably, even when comparing only G1 phase cells, aneuploid cells still exhibited significantly higher EIF5A expression than diploid cells **(Figure 7I)**, indicating that the association between EIF5A and malignant copy number variation status is not simply attributable to cell cycle progression differences.

Finally, we validated the expression coupling between EIF5A and HMGA1 at the sample level, showing a significant positive correlation **(Figure 7J)**, highly consistent with the "HMGA1-EIF5A regulatory axis" established in bulk cohorts, further supporting at single-cell resolution the conclusion that this axis accompanies malignant states and proliferation-related programs in tumor epithelial cells.

EIF5A inhibits GC cell proliferation and migration in vitro

To further investigate the biological function of EIF5A in gastric cancer (GC), we first evaluated EIF5A mRNA expression in normal gastric epithelial cells (GES-1) and GC cell lines (MKN45, AGS, HGC-27, and SNU-719) using quantitative real-time PCR (qRT-PCR). The results demonstrated that EIF5A expression was significantly lower in GC cell lines compared with GES-1 cells, with particularly low expression observed in HGC-27 and SNU-719 cells **(Figure 8A)**. To assess the functional role of EIF5A, HGC-27 and SNU-719 cells were transfected with EIF5A overexpression plasmids (oe-EIF5A). qRT-PCR analysis confirmed successful overexpression of EIF5A in both cell lines compared with the negative control group (oe-NC) **(Figure 8B)**.

Subsequently, CCK8 assay and colony formation assays were performed to inhibit the effect of EIF5A on GC cell proliferation. The results revealed that EIF5A overexpression significantly decreased the number of colonies formed by HGC-27 and SNU-719 cells compared with the oe-NC group **(Figures 8C, D)**, indicating that EIF5A suppresses the proliferative capacity of GC cells.

The migratory ability of GC cells was further examined using transwell migration assays. Compared with control cells, EIF5A overexpression significantly decreased the number of migrated cells in both HGC-27 and SNU-719 cell lines **(Figures 8E)**. These findings collectively indicate that EIF5A suppresses the migratory potential of GC cells.To further validate the

effect of EIF5A on cell migration, wound-healing assays were conducted. Consistent with the transwell migration results, EIF5A-overexpressing HGC-27 and SNU-719 cells exhibited markedly decelerated wound closure at both 24 h and 48 h after scratching **(Figures 8F)**. Quantitative analysis demonstrated significantly lower wound-healing rates in the oe-EIF5A group than in the oe-NC group.

Taken together, our results demonstrate that EIF5A functions as a tumor suppressor in GC by reducing tumor cell proliferation and migration, suggesting that EIF5A may inhibit GC progression and metastasis.

Discussion

In this study, EIF5A emerged as a marker of a distinct GC context rather than a simple binary indicator of high or low risk. EIF5A-high tumors exhibited higher TMB, stronger cytotoxic and checkpoint-associated immune signatures, and attenuated EMT/invasion-associated programs across transcriptomic, genomic, proteomic, microbiome, and single-cell analyses. Although EIF5A was associated with favorable survival in selected subgroups, this association was not uniform across all patients. This finding has clinical relevance, as EIF5A should not be considered a universal stand-alone prognostic biomarker for GC.

The favorable association of EIF5A may appear counterintuitive given the involvement of eIF5A in translation and cell-cycle biology. However, our data do not support a simple oncogene-like model. EIF5A was positively

correlated with translational machinery and checkpoint/surveillance proteins, including CHK2 and MSH2, but negatively associated with FAK, PAK4, c-KIT, and several EMT/invasion-related nodes. One plausible interpretation is that EIF5A-high tumors remain proliferative while retaining epithelial features and immunologic visibility, rather than transitioning toward a mesenchymal, matrix-remodeling phenotype. This interpretation is consistent with the poor prognosis of mesenchymal-like GC [4] and with previous evidence linking EMT and stromal remodeling to metastasis, immune exclusion, and therapeutic resistance [23,24].

Notably, these immune-related findings should not be interpreted as evidence of a uniformly immune-hot phenotype. Although cytotoxic, T-cell, and NK-cell signatures were increased, checkpoint, Treg, and exhaustion-related scores were also elevated. This pattern suggests immune recognition accompanied by concurrent inhibitory pressure. It may explain why EIF5A is associated with immunotherapy response and higher TMB without conferring a consistent prognostic effect across all clinical subgroups. In practice, EIF5A may be more informative when integrated with TMB, checkpoint expression, chemokine activity, and molecular subtype rather than used as a stand-alone biomarker.

The HMGA1-related findings are intriguing but should be interpreted with caution. HMGA1 is generally recognized as a chromatin architectural factor with oncogenic roles in several cancers, including GC, where high

expression has been associated with proliferation, metastasis, and poor prognosis in some studies [25-28]. In the present cohorts, however, HMGA1 was strongly coupled to EIF5A and shared favorable prognostic and immune-inflamed associations. This discrepancy may reflect cohort composition, immune context, tumor-cell state, or a more specific HMGA1-EIF5A program not captured by broad descriptions of HMGA1 biology. Direct HMGA1 perturbation, EIF5A quantification, and chromatin-binding assays will be required before a regulatory mechanism can be established.

The microbiome analysis should also be regarded as hypothesis-generating. HMGA1-high and HMGA1-low tumors did not show strong separation in global diversity; however, several genera, including Streptococcus, were enriched in HMGA1-low tumors and negatively correlated with HMGA1 expression. This observation is compatible with recent studies of tumor-specific microbiomes and their associations with intratumoral bacteria, immunity, prognosis, and therapeutic response [29-32]. In GC, Streptococcus-enriched states have been linked to epithelial disruption, EMT, immune modulation, and disease progression [31,32]. Nevertheless, sequencing-derived microbial signals are susceptible to low biomass, contamination, tissue processing, and filtering strategies. Validation using contamination-controlled microbiome assays is therefore warranted.

Single-cell analysis localized EIF5A predominantly to cycling and malignant epithelial cells. In aneuploid epithelial cells, EIF5A was associated with

proliferation, glycolysis, and hypoxia, but not EMT. Cell-line experiments further showed that restoration of EIF5A in EIF5A-low HGC-27 and SNU-719 cells reduced proliferation, colony formation, and migration. These findings should not necessarily be interpreted within a single linear model. In tumors, EIF5A may mark cycling epithelial cells that retain epithelial/proteostatic features; in EIF5A-low cell lines, EIF5A restoration may reactivate programs that limit colony formation and motility. Knockdown, rescue, and hypusination-focused assays will be needed to define the underlying mechanism.

Conclusions

EIF5A identifies a GC subgroup characterized by epithelial differentiation, translational/proteostatic activity, immune inflammation, higher TMB, and reduced EMT/invasion programs. HMGA1 represents a plausible upstream candidate and may link EIF5A to immune recruitment and selected microbiome differences. EIF5A overexpression suppressed proliferation and migration in GC cells, although the underlying mechanism remains to be defined. At this stage, EIF5A is best regarded as a candidate component of multiparameter prognostic and immunotherapy biomarker models.

List of abbreviations

ACRG, Asian Cancer Research Group; AUC, area under the curve; BH, Benjamini–Hochberg; CCK-8, Cell Counting Kit-8; CNV, copy number variation; DEG, differentially expressed gene; DFS, disease-free survival;

EIF5A, eukaryotic translation initiation factor 5A; EMT, epithelial–mesenchymal transition; FDR, false discovery rate; GC, gastric cancer; GDC, Genomic Data Commons; GEO, Gene Expression Omnibus; GO, Gene Ontology; HMGA1, high mobility group A1; HR, hazard ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; MAF, mutation annotation format; NK, natural killer; OS, overall survival; PCoA, principal coordinates analysis; qRT-PCR, quantitative reverse transcription polymerase chain reaction; ROC, receiver operating characteristic; SD, standard deviation; ssGSEA, single-sample gene set enrichment analysis; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas; TMB, tumor mutation burden; TNM, tumor–node–metastasis; Treg, regulatory T cell; UMAP, uniform manifold approximation and projection; UPR, unfolded protein response.

Declarations

Ethics approval and consent to participate

This study analyzed publicly available, de-identified human datasets and performed in vitro experiments using established cell lines; no new human participants, human tissue specimens or animal experiments were involved. Ethical approval and informed consent for the public datasets were obtained by the original studies, as applicable.

Consent for publication

Not applicable.

Availability of data and materials

The public datasets analyzed in this study are available from the Genomic Data Commons/TCGA-STAD resource and the NCBI Gene Expression Omnibus, including GSE15459, GSE62254, GSE84437 and GSE183904. Additional public resources analyzed include GEPIA2, Kaplan-Meier Plotter, GPSAdb 2.0 and public proteomic, methylation, copy-number, immunotherapy-response and intratumoral microbiome datasets described in the Materials and Methods. Processed data and experimental data generated during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Funding

This study was supported by the Key Medical Discipline of Ningbo (2026-A27) and the Zhu Xiushan Talent Award Fund (2023HMJQ04F).

Authors' contributions

Zhilin Liu: Formal analysis, data curation, visualization, validation, conceptualization, methodology, and writing - original draft. Yunjie Chen: Methodology, Investigation, Validation, Data curation, Visualization, Software. Feng Ren: Methodology, Investigation, Validation, Data curation, Visualization, Software. Kepu Zheng: Methodology, Investigation,

Validation, Data curation, Visualization, Software. Yanghui Wen: Funding acquisition, project administration, conceptualization, methodology, supervision, writing - original draft, writing - review & editing.

Acknowledgements

Not applicable.

Clinical trial number

Not applicable.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229-263. https://doi.org/10.3322/caac.21834.

2. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. Lancet. 2020;396:635-648. https://doi.org/10.1016/S0140-6736(20)31288-5.

3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202-209. https://doi.org/10.1038/nature13480.

4. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21:449-456. https://doi.org/10.1038/nm.3850.

5. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat Med. 2018;24:1449-1458. https://doi.org/10.1038/s41591-018-0101-z.

6. Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomized, open-label, phase 3 trial. Lancet. 2021;398:27-40. https://doi.org/10.1016/S0140-6736(21)00797-2.

7. Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, et al. Safety and efficacy of pembrolizumab monotherapy in patients with previously treated advanced gastric and gastroesophageal junction cancer: phase 2 clinical KEYNOTE-059 trial. JAMA Oncol. 2018;4:e180013. https://doi.org/10.1001/jamaoncol.2018.0013.

8. Bang YJ, Kang YK, Catenacci DV, Muro K, Fuchs CS, Geva R, et al. Pembrolizumab alone or in combination with chemotherapy as first-line therapy for patients with advanced gastric or gastroesophageal junction adenocarcinoma: results from the phase II nonrandomized KEYNOTE-059 study. Gastric Cancer. 2019;22:828-837. https://doi.org/10.1007/s10120-018-00909-5.

9. Janjigian YY, Kawazoe A, Yañez P, Li N, Lonardi S, Kolesnik O, et al. The KEYNOTE-811 trial of dual PD-1 and HER2 blockade in HER2-positive

gastric cancer. Nature. 2021;600:727-730. https://doi.org/10.1038/s41586-021-04161-3.

10. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019;51:202-206. https://doi.org/10.1038/s41588-018-0312-8.

11. Valero C, Lee M, Hoen D, Wang J, Nadeem Z, Patel N, et al. The association between tumor mutational burden and prognosis is dependent on treatment context. Nat Genet. 2021;53:11-15. https://doi.org/10.1038/s41588-020-00752-4.

12. Saini P, Eyler DE, Green R, Dever TE. Hypusine-containing protein eIF5A promotes translation elongation. Nature. 2009;459:118-121. https://doi.org/10.1038/nature08034.

13. Gutierrez E, Shin BS, Woolstenhulme CJ, Kim JR, Saini P, Buskirk AR, et al. eIF5A promotes translation of polyproline motifs. Mol Cell. 2013;51:35-45. https://doi.org/10.1016/j.molcel.2013.04.021.

14. Schuller AP, Wu CC, Dever TE, Buskirk AR, Green R. eIF5A functions globally in translation elongation and termination. Mol Cell. 2017;66:194-205. https://doi.org/10.1016/j.molcel.2017.03.003.

15. Barba-Aliaga M, Mena A, Espinoza V, Apostolova N, Costell M, Alepuz P. Hypusinated eIF5A is required for the translation of collagen. J Cell Sci. 2021;134:jcs258643. https://doi.org/10.1242/jcs.258643.

16. Nakanishi S, Cleveland JL. Targeting the polyamine-hypusine circuit for the prevention and treatment of cancer. Amino Acids. 2016;48:2353-2362. https://doi.org/10.1007/s00726-016-2275-3.

17. Mathews MB, Hershey JWB. The translation factor eIF5A and human cancer. Biochim Biophys Acta. 2015;1849:836-844. https://doi.org/10.1016/j.bbagrm.2015.05.002.

18. Nakanishi S, Cleveland JL. The many faces of hypusinated eIF5A: cell context-specific effects of the hypusine circuit and implications for human health. Int J Mol Sci. 2024;25:8171. https://doi.org/10.3390/ijms25158171.

19. Meng QB, Kang WM, Yu JC, Liu YQ, Ma ZQ, Zhou L, et al. Overexpression of eukaryotic translation initiation factor 5A2 (EIF5A2) correlates with cell aggressiveness and poor survival in gastric cancer. PLoS One. 2015;10:e0119229. https://doi.org/10.1371/journal.pone.0119229.

20. Yang Q, Ye Z, Zhang Q, Zhao Z, Yuan H. Expression of eukaryotic translation initiation factor 5A-2 (eIF5A-2) associated with poor survival in gastric cancer. Tumour Biol. 2016;37:1189-1195. https://doi.org/10.1007/s13277-015-3894-0.

21. Sun J, Xu Z, Lv H, Wang Y, Wang L, Ni Y, et al. eIF5A2 regulates the resistance of gastric cancer cells to cisplatin via induction of EMT. Am J Transl Res. 2018;10:4269-4279.

22. Tian SB, Yu JC, Liu YQ, Kang WM, Ma ZQ, Ye X, et al. MiR-30b suppresses tumor migration and invasion by targeting EIF5A2 in gastric

cancer. World J Gastroenterol. 2015;21:9337-9347.
https://doi.org/10.3748/wjg.v21.i31.9337.
23. Nieto MA, Huang RYJ, Jackson RA, Thiery JP. EMT: 2016. Cell.
2016;166:21-45. https://doi.org/10.1016/j.cell.2016.06.028.
24. Pastushenko I, Blanpain C. EMT transition states during tumor
progression and metastasis. Trends Cell Biol. 2019;29:212-226.
https://doi.org/10.1016/j.tcb.2018.12.001.
25. Shah SN, Resar LMS. The high mobility group A1 (HMGA1)
transcriptome in cancer and development. Curr Mol Med. 2016;16:353-393.
https://doi.org/10.2174/1566524016666160316152147.
26. Wang L, Zhang J, Xia M, Liu C, Zu X, Zhong J. High mobility group A1
(HMGA1): structure, biological function, and therapeutic potential. Int J Biol
Sci. 2022;18:4414-4431. https://doi.org/10.7150/ijbs.72952.
27. Yang Q, Wang Y, Li M, Wang Z, Zhang J, Dai W, et al. HMGA1 promotes
gastric cancer growth and metastasis by transactivating SUZ12 and
CCDC43 expression. Aging (Albany NY). 2021;13:16043-16061.
https://doi.org/10.18632/aging.203130.
28. Akaboshi S, Watanabe S, Hino Y, Sekita Y, Xi Y, Araki K, et al. HMGA1 is
induced by Wnt/beta-catenin pathway and maintains cell proliferation in
gastric cancer. Am J Pathol. 2009;175:1675-1685.
https://doi.org/10.2353/ajpath.2009.090069.
29. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The
human tumor microbiome is composed of tumor type-specific intracellular

bacteria. Science. 2020;368:973-980.
https://doi.org/10.1126/science.aay9189.
30. Yang X, An H, He Y, Fu G, Jiang Z. Comprehensive analysis of microbiota signature across 32 cancer types. Front Oncol. 2023;13:1127225. https://doi.org/10.3389/fonc.2023.1127225.
31. Wang G, Wang H, Ji X, Wang T, Zhang Y, Jiang W, et al. Intratumoral microbiome is associated with gastric cancer prognosis and therapy efficacy. Gut Microbes. 2024;16:2369336. https://doi.org/10.1080/19490976.2024.2369336.
32. Yuan L, Pan L, Wang Y, Zhao J, Fang L, Zhou Y, et al. Characterization of the landscape of the intratumoral microbiota reveals that Streptococcus anginosus increases the risk of gastric cancer initiation and progression. Cell Discov. 2024;10:117. https://doi.org/10.1038/s41421-024-00746-0.

Figures and legends

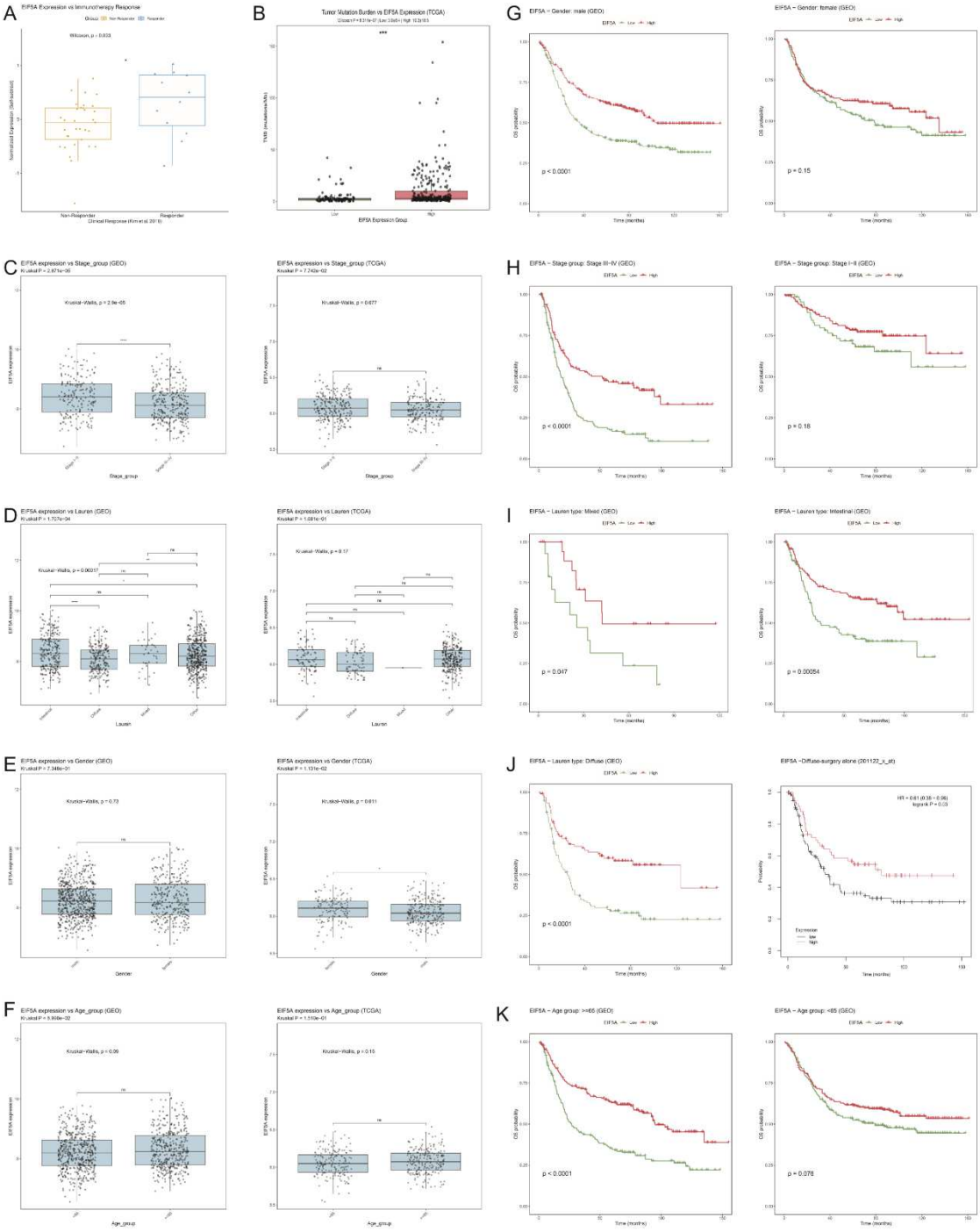
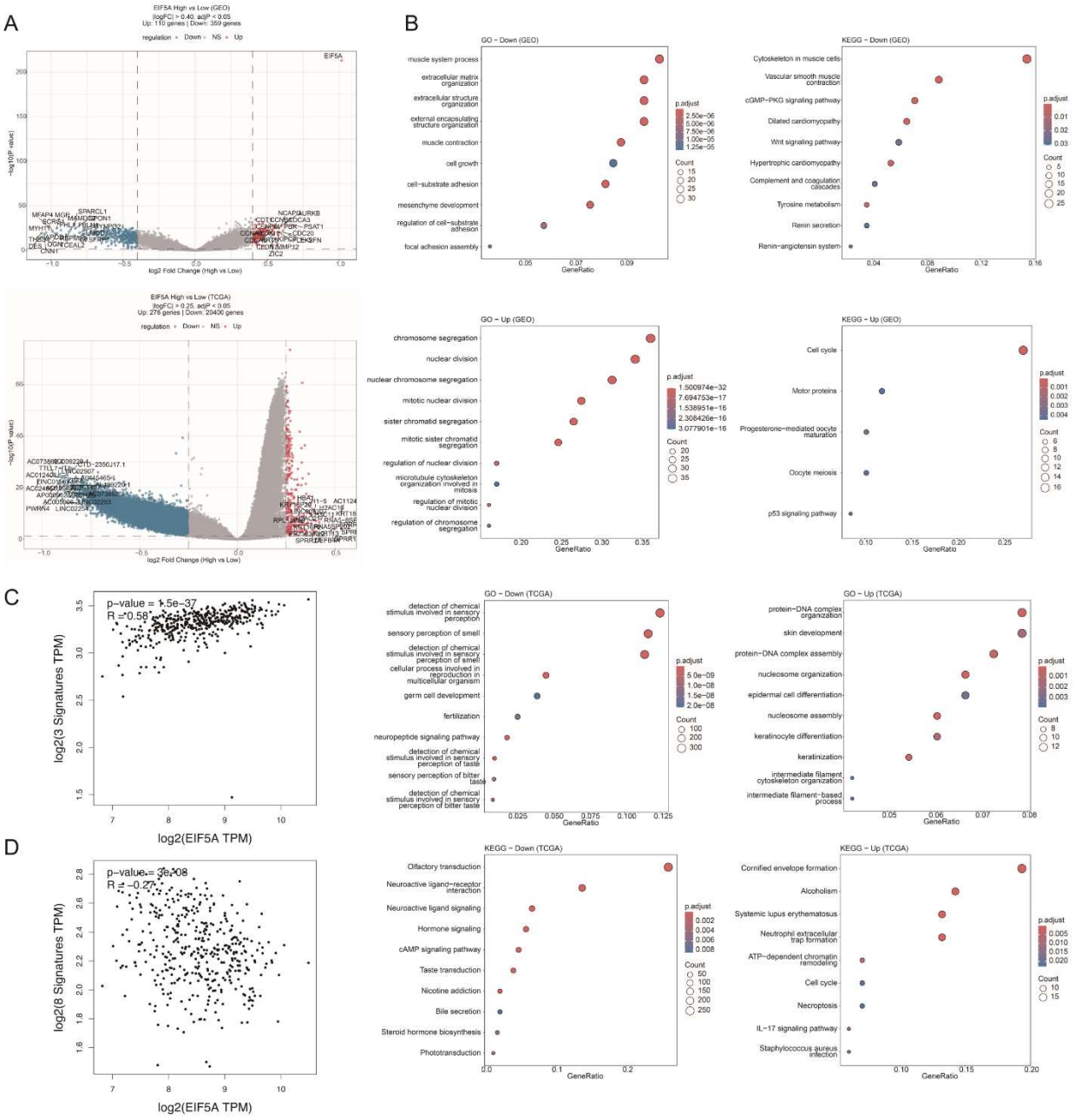


Figure 1. EIF5A expression, immunotherapy response, tumor mutation burden, clinicopathological features, and subgroup prognosis. A, EIF5A expression in immunotherapy responders and non-responders using data

from Kim et al. [5]. B, Tumor mutation burden (TMB) in TCGA groups stratified by median EIF5A expression. C–F, EIF5A expression by tumor stage (I–II versus III–IV; C), Lauren classification (intestinal, diffuse, and mixed; D), sex (E), and age (<65 versus ≥65 years; F). GEO and TCGA cohorts are shown on the left and right, respectively. G–I, Kaplan–Meier overall survival curves comparing EIF5A-high and EIF5A-low groups in GEO subgroups defined by sex (male, left; female, right; G), stage (III–IV, left; I–II, right; H), and Lauren classification (mixed, left; intestinal, right; I). J, Survival associations in diffuse-type tumors in GEO (left) and surgery-only patients in the Kaplan–Meier Plotter database (right). K, GEO survival curves stratified by age (<65 years, left; ≥65 years, right). Two-group comparisons used Wilcoxon rank-sum tests; multiple-group comparisons used Kruskal–Wallis tests; survival comparisons used log-rank tests. ns, not significant.



1262 Figure 2. Transcriptomic programs and epithelial–mesenchymal
1263 associations linked to EIF5A expression in GC. A, Volcano plots showing
1264 differentially expressed genes between EIF5A-high and EIF5A-low groups in
1265 the GEO cohort and the TCGA-STAD cohort. B, Gene Ontology (GO) and
1266 Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of
1267 genes upregulated or downregulated between EIF5A-high and EIF5A-low

groups in the GEO cohort (top) and TCGA-STAD cohort (bottom). Representative enriched terms are displayed as bubble plots. C, Correlation analyses (Spearman) between EIF5A and epithelial-associated markers (KRT8, KRT18 and CLDN7) in TCGA-STAD using GEPIA2. D, Correlation analyses (Spearman) between EIF5A and mesenchymal/EMT-associated markers (VIM, CDH2, ACTA2, ZEB1, ZEB2, TWIST2, FOXC2 and LAMA4) in TCGA-STAD using GEPIA2.

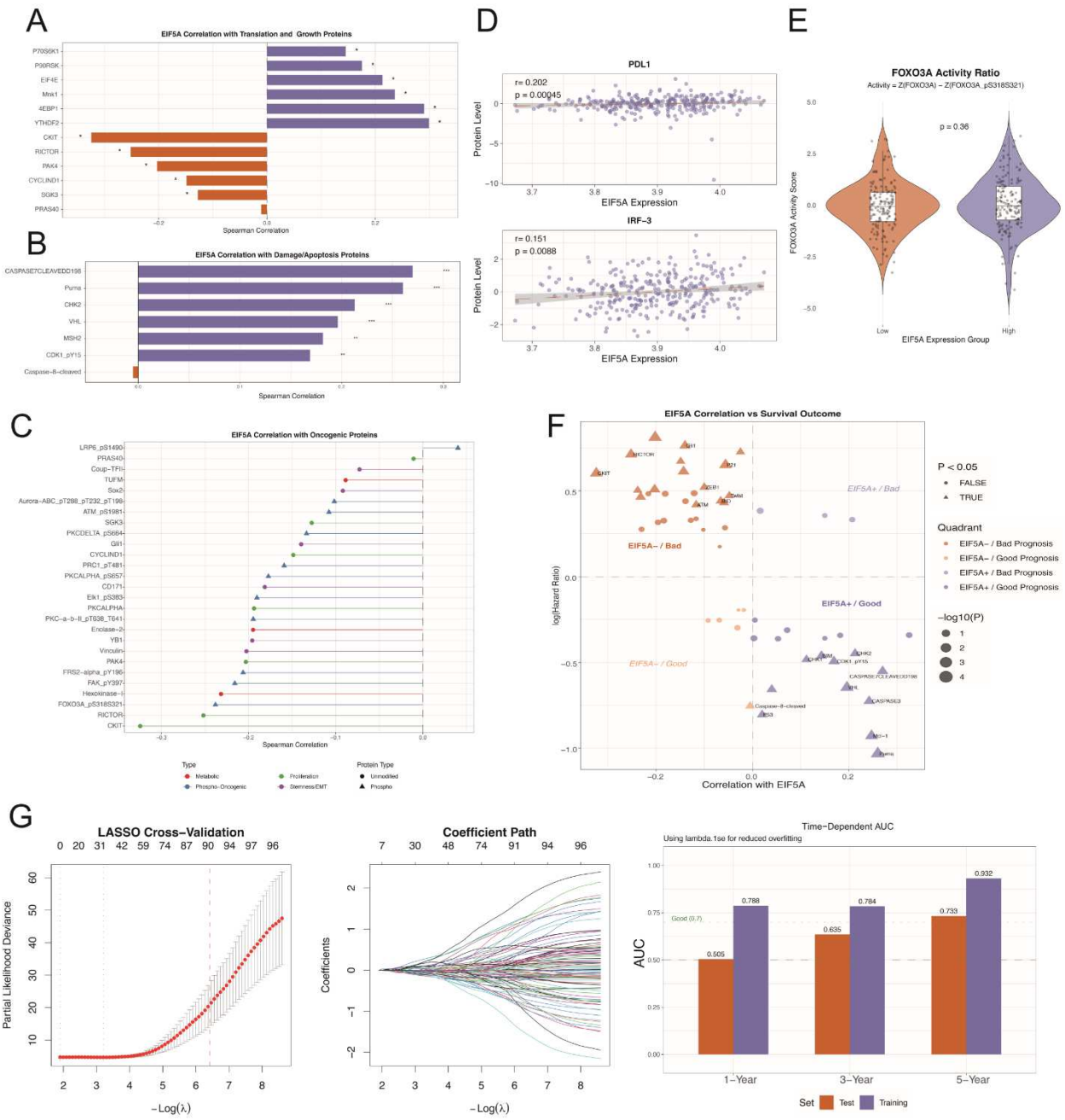
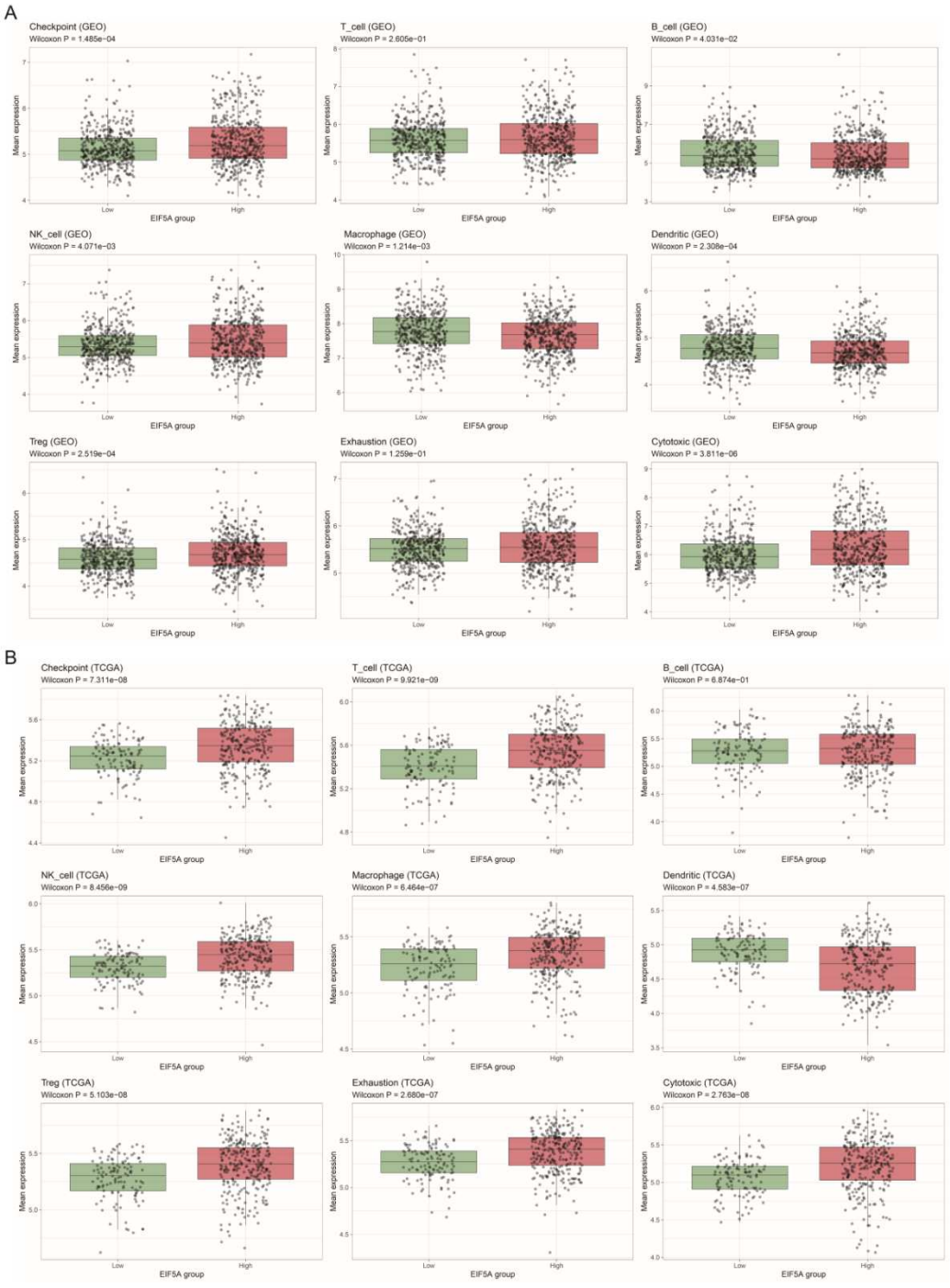


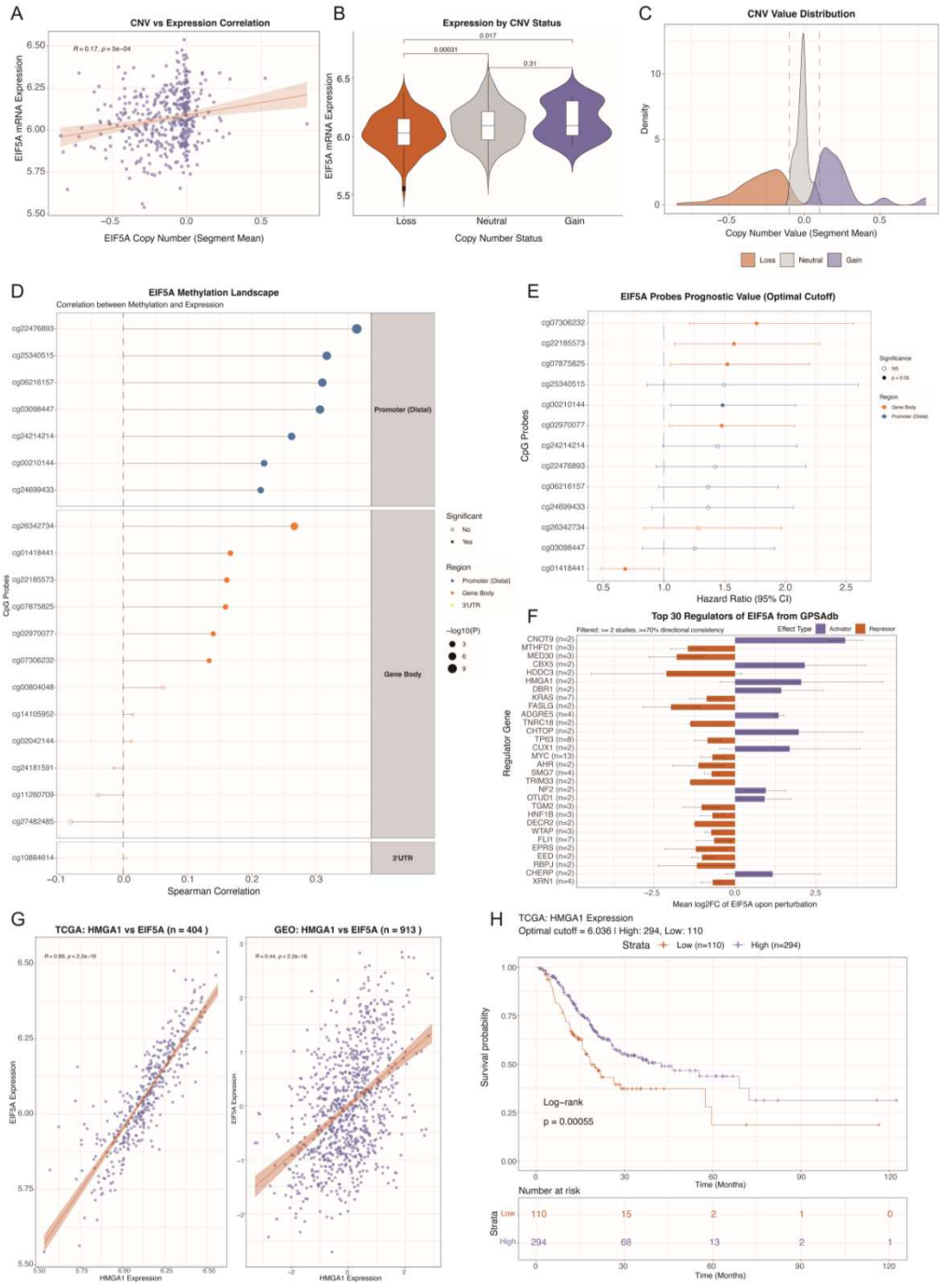
Figure 3. Proteomic landscape and prognostic protein networks associated with EIF5A expression in GC. A, Bar plots depicting correlations between EIF5A and proteins involved in translation and growth signaling pathways in the TCGA-STAD proteomic dataset. B, Bar plots displaying correlations between EIF5A and proteins related to DNA damage response, apoptosis,

and cell cycle checkpoint regulation. C, Lollipop plot illustrating correlations between EIF5A and oncogenic proteins, including EMT-related factors, focal adhesion components, stemness markers, and metabolic enzymes. D, Scatter plots showing correlations between EIF5A and immune checkpoint/innate immunity-related proteins (PD-L1 and IRF-3) at the protein level in TCGA-STAD. E, Comparison of FOXO3A activity ratio (calculated as Z(FOXO3A) − Z(FOXO3A_pS318S321)) between EIF5A-high and EIF5A-low groups, with statistical significance assessed by Wilcoxon rank-sum test. F, Quadrant plot mapping the Pearson correlation coefficient between each protein and EIF5A (x-axis) against the univariate survival hazard ratio (log(HR), y-axis). Proteins are categorized into four quadrants based on their associations with EIF5A expression and prognostic impact. G, LASSO-Cox regression model construction (left) and time-dependent receiver operating characteristic (ROC) curves (right) for 1-year, 3-year, and 5-year survival prediction in both training and test cohorts, with corresponding area under the curve (AUC) values displayed.



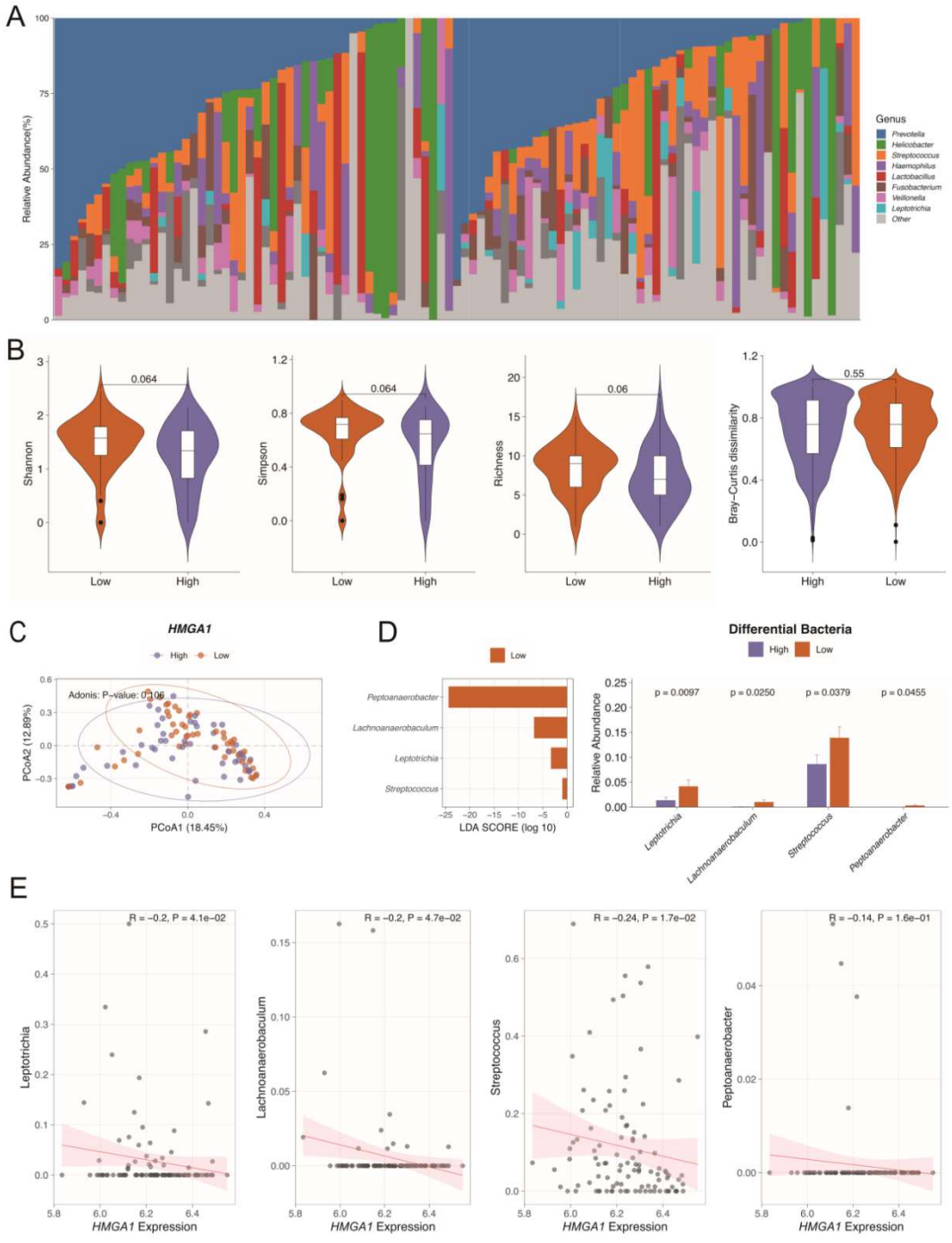
1296 Figure 4. Association between EIF5A expression and tumor immune
1297 microenvironment features in GC. A, Box plots comparing immune-related
1298 gene signature scores between EIF5A-high and EIF5A-low groups in the
1299 GEO cohort, including checkpoint, cytotoxic, T cell, B cell, NK cell,

macrophage, dendritic cell, regulatory T cell (Treg), and T cell exhaustion signatures. Statistical significance was assessed by Wilcoxon rank-sum test. B, Box plots comparing immune-related gene signature scores between EIF5A-high and EIF5A-low groups in the TCGA-STAD cohort, validating the immune microenvironment association patterns observed in the GEO cohort. Statistical significance was assessed by Wilcoxon rank-sum test.



1306 Figure 5. Copy number variation, DNA methylation, and candidate upstream
1307 regulation of EIF5A. A, Scatter plot showing correlation analysis between
1308 EIF5A copy number (segment mean) and EIF5A mRNA expression levels in
1309 the TCGA-STAD cohort (Spearman correlation). B, Box plots comparing

EIF5A expression levels stratified by EIF5A copy number variation (CNV) status (Loss, Neutral, Gain) in the TCGA-STAD cohort. Statistical significance was assessed by Wilcoxon rank-sum test. C, Density distribution plot displaying the overall distribution pattern of EIF5A copy number variation (segment mean) in the TCGA-STAD cohort. D, Lollipop plot showing correlation analyses between CpG site methylation levels and EIF5A expression across different genomic regions (promoter distal, gene body, 3' untranslated region) based on TCGA methylation data. Dot size represents statistical significance ($-\log_{10}P$) and color indicates genomic region. E, Forest plot presenting overall survival risk analysis results for CpG sites significantly correlated with EIF5A expression, displayed as hazard ratios (HR) with 95% confidence intervals (CI), evaluated by Cox proportional hazards regression model. F, Integration of upstream gene functional perturbation experiments (including CRISPR knockout, RNAi silencing, etc.) from the GPSAdb2.0 database, displaying the top 30 candidate upstream regulators and their perturbation effect directions on EIF5A expression. G, Scatter plots validating the correlation and cross-cohort consistency between the key candidate regulator HMGA1 and EIF5A expression in the TCGA-STAD cohort (left) and independent GEO cohort (right) (Spearman correlation). H, Kaplan-Meier survival curves showing overall survival differences between HMGA1-high and HMGA1-low expression groups in the TCGA-STAD cohort. Statistical significance was assessed by log-rank test.



1333 Figure 6. Intratumoral microbiome features associated with HMGA1

1334 expression stratification in GC. A, Stacked bar plot showing the relative

1335 abundance distribution of intratumoral microbiota at the genus level in GC

1336 samples from the TCMA/TCGA-STAD cohort, displaying major dominant

genera and inter-sample heterogeneity. B, Violin plots comparing intratumoral microbial diversity between HMGA1-high and HMGA1-low groups (median split), including α-diversity indices (Shannon, Simpson, Richness) and β-diversity (Bray-Curtis dissimilarity). C, Principal coordinate analysis (PCoA) ordination plot based on Bray-Curtis distance showing microbial community structure distribution patterns of HMGA1-high and HMGA1-low groups along PCoA1 and PCoA2 dimensions. D, Linear discriminant analysis effect size (LEfSe) analysis identifying differential genera associated with HMGA1 expression stratification. Left: Linear discriminant analysis (LDA) scores ($\log_{10}$) of differential genera, representing their contribution to inter-group discrimination. Right: Box plots comparing relative abundance differences of differential genera between HMGA1-high and HMGA1-low groups. E, Scatter plots showing correlation analyses between relative abundance of differential genera and HMGA1 expression levels, with association trends assessed by Spearman correlation.

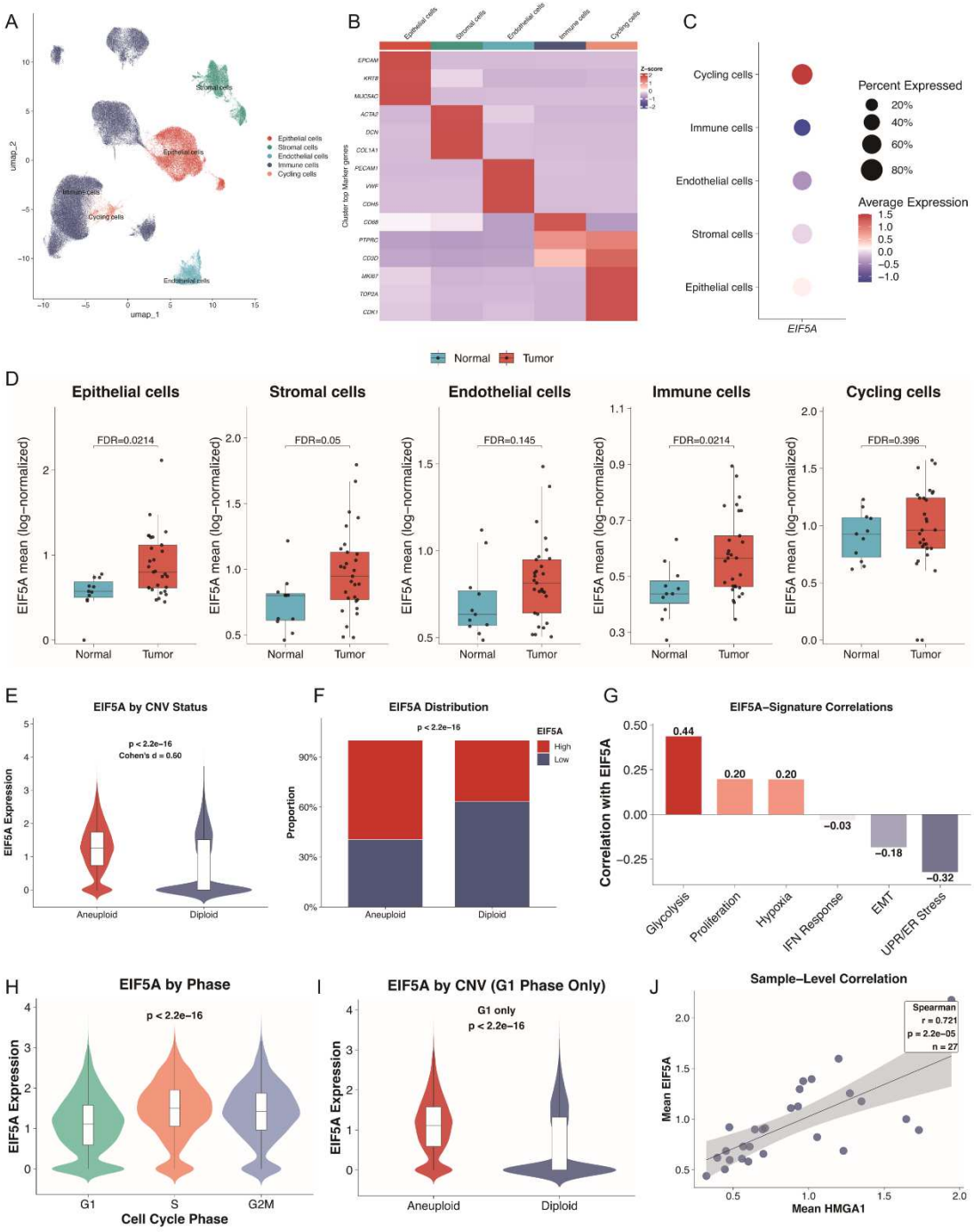
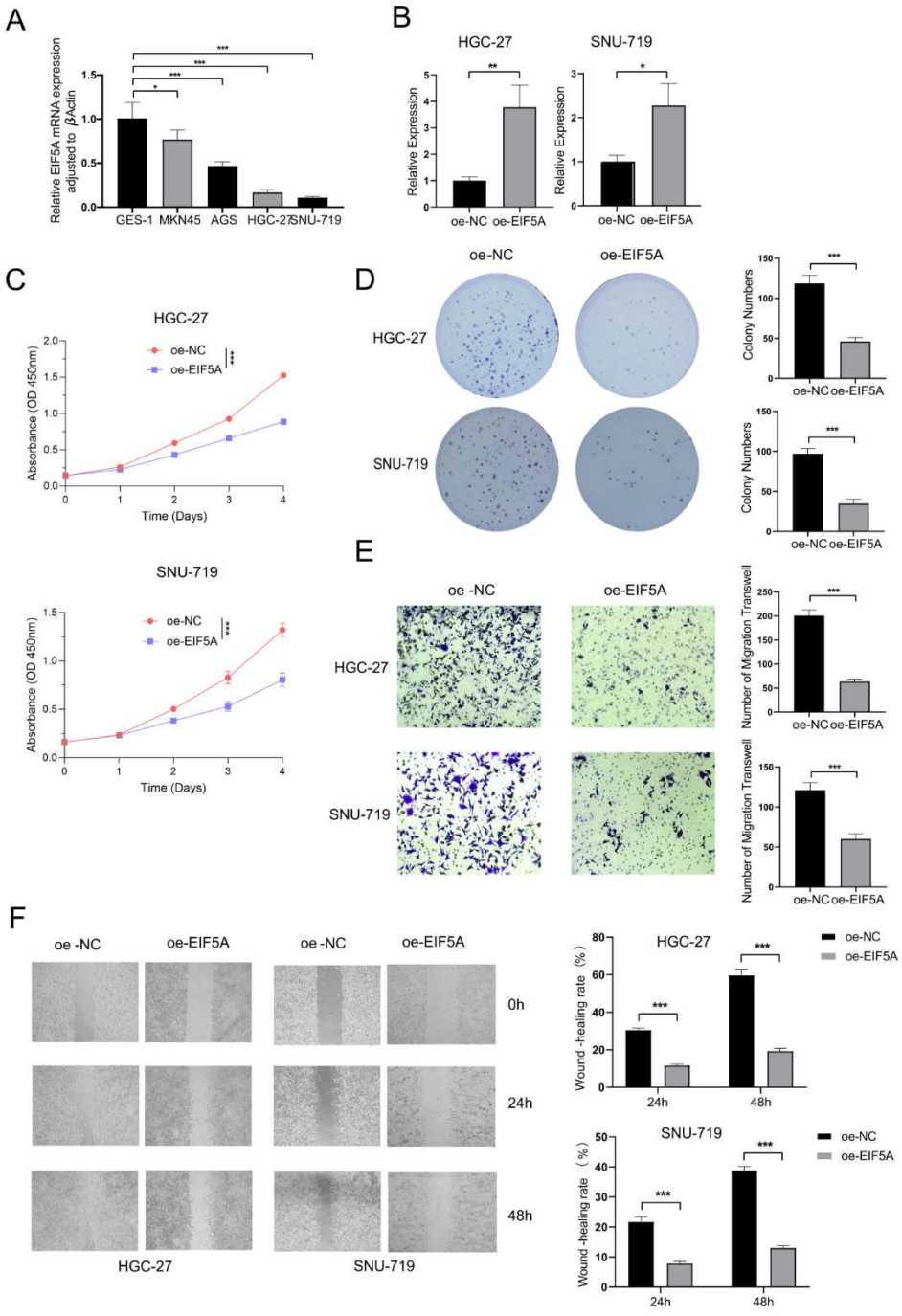


Figure 7. Single-cell EIF5A expression and associations with malignant state and proliferation programs. A, UMAP dimensionality reduction plot showing clustering distribution of major cell populations (epithelial cells, stromal cells, endothelial cells, immune cells, and proliferating cells) in the

GC single-cell dataset (GSE183904). B, Heatmap displaying expression patterns of signature genes for each cell population, used for cell type annotation validation. C, Dot plot showing the expression abundance and percentage of EIF5A across different cell populations. D, Box plots comparing average EIF5A expression levels across major cell types between normal and tumor tissue origins. E, Box plots comparing EIF5A expression between aneuploid and diploid cells, classified based on copy number variation (CNV) status inferred by inferCNV algorithm in tumor tissue cells. F, Stacked bar plot showing the proportion distribution of EIF5A-high and EIF5A-low cells in aneuploid and diploid populations. G, Scatter plots showing correlations between EIF5A expression and multiple biological processes (EMT, hypoxia, ER stress, inflammation, interferon response, glycolysis, oxidative phosphorylation, etc.) evaluated by gene set scoring. H, Violin plots displaying EIF5A expression distribution patterns across different cell cycle phases (G1, S, G2M). I, Box plot comparing EIF5A expression differences between aneuploid and diploid cells restricted to G1 phase only. J, Scatter plot showing correlation analysis between EIF5A and HMGA1 expression at the sample level.



1375 Figure 8. EIF5A inhibits GC cell proliferation and migration in vitro. A, The
1376 relative mRNA expression levels of EIF5A in normal gastric epithelial cells
1377 (GES-1) and GC cell lines (MKN45, AGS, HGC-27, and SNU-719) detected
1378 by qRT-PCR. B, qRT-PCR validation of EIF5A overexpression efficiency in

HGC-27 and SNU-719 cells. C, CCK8 assay results for HGC-27 and SNU-719 cells after EIF5A overexpression. D, Representative images and quantitative analysis of clone formation assay in HGC-27 and SNU-719 cells after EIF5A overexpression. E, Representative images and quantitative analysis of transwell migration assay in HGC-27 and SNU-719 cells after EIF5A overexpression. F, Representative images and quantitative analysis of wound healing rate in HGC-27 and SNU-719 cells after EIF5A overexpression. Data are shown as mean ± standard deviation from three independent experiments. Statistical significance was determined using Student's t-test. *P < 0.05, **P < 0.01, ***P < 0.001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- figures1.pdf
- figures2.pdf