Am J Cancer Res 2026;16(10):4387-4406
www.ajcr.us /ISSN:2156-6976/ajcr0174439

Original Article

CDX4-mediated regulation of SH3BGRL2 suppresses esophageal squamous cell carcinoma progression via apoptosis and migration

Chunlong Zheng[1*], Yuzhou Duan[1*], Nian Zhang[2], Yan Chu[3], Jianyong Sun[1]

[1]*Department of Thoracic Surgery, Tangdu Hospital, The Fourth Military Medical University, Xi'an, China;* [2]*Department of Anesthesia, Tangdu Hospital, The Fourth Military Medical University, Xi'an, China;* [3]*School of Medicine, Xi'an Siyuan University, Xi'an, China.* [*]*Equal contributors.*

Received June 2, 2026; Accepted October 11, 2026; Epub October 15, 2026; Published October 30, 2026

Abstract: Given that protein expression serves as the primary mechanism mediating gene functions, this study aimed to identify the key proteins correlated with esophageal squamous cell carcinoma (ESCC) and explore their effects on ESCC biological functions. In this study, we analyzed protein quantitative trait loci (pQTLs) based on genome-wide association studies (GWASs) and integrated data from the FinnGen, IEU openGWAS, and BioBank Japan databases to excavate genetic associations with ESCC risk, constructed a protein interaction network to clarify the correlations among pivotal genes obtained from the above analysis, adopted an eight-point drug analysis method to screen potential drug targets of these genes, and applied single-cell sequencing analysis to investigate the expression profiles of related genes in different cell pairs. Functional validation revealed that the gene SH3BGRL2 is significantly expressed in ESCC cell lines and regulates the viability, migration and apoptosis of ESCC cells. The results demonstrated that genetically predicted circulating levels of 83 plasma proteins were correlated with ESCC risk, among which 40 proteins were associated with decreased ESCC susceptibility and 43 with increased risk; four proteins including ADH1C, ADH4, GCKR and METAP1 exhibited significant genetic colocalization, and 11 genes covering ADH1C and VEGFA were screened as potential drug targets. Additionally, overexpression of SH3BGRL2 could inhibit the viability and migration of ESCC cells and promote their apoptosis, and the expression of SH3BGRL2 was regulated by the transcription factor CDX4. These findings identify candidate proteins associated with ESCC risk and support a functional relationship between CDX4 and SH3BGRL2 in the studied cell model, warranting further mechanistic and clinical validation.

Keywords: Esophageal squamous cell carcinoma, single-cell RNA sequencing, CDX4, biomarkers, therapeutic, SH3BGRL2

Introduction

Esophageal cancer is a major global health problem and ranks as the sixth leading cause of cancer death worldwide [1, 2]. The disease is particularly common in parts of East Asia and Southern Africa, where it causes a substantial burden of illness and death [3]. Although screening and treatment have improved, patients with esophageal squamous cell carcinoma (ESCC) often have poor outcomes. Many patients are diagnosed at an advanced stage, and current treatments have limited effectiveness [4-6]. These problems point to a need for new biomarkers to support early detection and help assess prognosis. New treatment targets are also needed to improve outcomes for patients with ESCC.

Protein biomarkers have received growing attention in cancer research, including studies of ESCC [7]. Proteins reflect gene activity and take part in processes that shape disease development and patient outcomes [8-10]. Studying proteins linked to the biological changes in ESCC may therefore help improve patient care and guide treatment.

One question in biomarker research is whether a protein contributes to disease or is simply

https://doi.org/10.62347/KXEX7123

associated with it. Mendelian randomization (MR) uses genetic variants as indicators of an exposure to examine possible causal links between modifiable risk factors and disease outcomes [11-13]. Researchers have increasingly used this method in ESCC studies [14-16]. By using genetic variants as instrumental variables, MR can reduce some sources of bias found in observational studies. It can also help assess whether specific proteins may affect ESCC risk and progression, providing evidence beyond an observed association [17, 18].

Single-cell technologies offer a way to examine these findings within tumors. They reveal differences between tumor cell populations, identify rare cells, and show how cell populations are organized. These features can help explain cancer development, progression, and responses to treatment [19-21]. In ESCC, single-cell RNA sequencing (scRNA-seq) measures gene expression in individual cells. These data may help identify protein biomarkers and treatment targets that could support more personalized care [22-25].

In this study, we combined a proteome-wide MR approach with circulating protein and genomic data to examine proteins that may affect ESCC risk. Because MR provides limited information about the underlying biological mechanisms, we also used colocalization analysis and summary-data-based Mendelian randomization (SMR) to help identify proteins involved in ESCC. We assessed whether these proteins could serve as targets for drug development and used single-cell expression profiles from ESCC tissues to identify the cell populations expressing their genes. We also carried out wound healing, Transwell migration, and apoptosis assays to examine the biological role of SH3BGRL2, a key protein identified in our analysis. These assays supported its role in ESCC and its potential as a treatment target. We report this study in accordance with the MDAR reporting checklist.

Material and methods

Study population and data sources

Genome-Wide Association Study (GWAS) Data: Our primary GWAS dataset for ESCC was sourced from the FinnGen Release 9 database, specifically from the dataset labeled C3_OESOPHAGUS_EXALLC. This dataset comprises data on 556 East European cases and 287,137 European controls. Validation of GWAS findings was performed using two additional datasets: one from the IEU OpenGWAS database (bbj-a-117) with 1,388 European cases and 159,201 European controls, and another from the BioBank Japan, which includes data on 1,388 East Asian cases and 159,201 East Asian controls. Genetic associations for plasma protein biomarkers were obtained from a study that synthesized data across seven major proteomic studies [26].

Single-Cell and Gene Expression Data: Single-cell RNA sequencing data was obtained from the Gene Expression Omnibus (GEO) database (GSE145370), which includes samples from 7 tumor tissues and 7 matched adjacent normal tissues. Additionally, gene expression data for 5 ESCC and 5 normal samples were extracted from GSE17351. Differential gene expression analysis was aimed at identifying key gene expression differences critical for ESCC progression.

Clinical and RNA sequencing (RNA-seq) Data from The Cancer Genome Atlas (TCGA)-ESCA: Further analyses were supplemented with RNA-seq and clinical feature data from the TCGA-ESCA dataset, accessible through the UCSC XENA platform. This dataset was integral in constructing an innovative prognostic model based on key gene expressions linked to ESCC prognosis.

Selection of genetic instrumental variables

We selected cis-protein quantitative trait loci (cis-pQTLs) as instrumental variables (IVs) for circulating proteins. We defined these variants as lead single nucleotide polymorphisms (SNPs) located within 1 megabase (Mb) of the transcription start site of the gene encoding the protein.

Criteria for SNP selection

We included SNPs associated with circulating protein levels at a genome-wide significance threshold of $P < 5 \times 10^{-8}$. Eligible SNPs were located within 1 Mb upstream or downstream of the gene coding regions. We also required an

F-statistic greater than 10, calculated as $F = \beta^2/SE^2$, to reduce the risk of using weak instruments.

We excluded SNPs in the region designated as the major histocompatibility complex (MHC) region on chromosome 6 (chr6: 5.5-34.0 Mb). This region has complex linkage disequilibrium (LD) patterns that may affect the estimates. We used PLINK to perform LD clumping at an r^2 threshold of 0.0001 and retain largely independent SNPs.

Key gene screening

We used several genetic analyses to examine whether genetically predicted circulating protein levels were linked to ESCC risk. We performed MR analysis with the TwoSampleMR R package, using genetic variants as IVs. For proteins with one IV, we estimated the effect using the Wald ratio method. For proteins with multiple IVs, we used the inverse-variance weighted (IVW) method.

We used weighted median and MR-Egger analyses to assess the sensitivity of the findings to horizontal pleiotropy. We retained results with an MR-Egger intercept test $P > 0.05$ and used an adjusted $P < 0.05$ as the threshold for statistical significance. We assessed heterogeneity with Cochran's Q test and also performed the test described in the source as a "one-time test" for proteins with multiple IVs.

We carried out confirmatory MR analyses in the FinnGen R9 and BBJ datasets, using $P < 0.05$ as the significance threshold. We combined results across cohorts with a random-effects model.

For MR associations with an adjusted $P < 0.05$, we used the coloc R package to assess whether the protein and ESCC associations shared a causal variant. This analysis used a Bayesian approach and evaluated posterior probabilities within a ± 25 kb window to help distinguish a shared signal from separate signals linked through LD.

We also performed SMR-HEIDI analysis with SMR v1.3.1 to assess these associations. We required an SMR test $P < 0.05$ and a heterogeneity in dependent instruments (HEIDI) test $P > 0.05$ to screen out associations likely to reflect separate variants in LD.

Prognostic model construction

We used multivariable Cox regression to examine the relationship between candidate targets and overall survival (OS). We performed the analysis with the survival v3.2-7 and survminer v0.4.8 R packages and calculated hazard ratios (HRs) with 95% confidence intervals (CIs). We also used Kaplan-Meier curves to describe survival and the log-rank test to assess differences between groups.

We compared differentially expressed genes from the TCGA and GEO datasets and identified five genes that were downregulated in both. We included SH3BGRL2, GSN, and PCDH1 in a three-gene prognostic model built with the TCGA-ESCA dataset. We interpreted the survival association for GSN with caution because it did not meet the proportional hazards assumption, and the survival tests gave different results. The Results section provides further details.

We calculated each patient's risk score using the following formula:

$$risk\ score = \sum_{i=1}^{n} (coefi * Xi)$$

In this formula, "coef" represents the gene coefficient estimated by multivariable Cox regression, and "X" represents the expression level of the corresponding gene.

Single-cell data analysis

We processed the single-cell transcriptomic data with Seurat and removed low-quality cells. We excluded cells in which ribosomal genes accounted for more than 20% of expression or erythroid genes accounted for more than 3%. We normalized the data with NormalizeData and identified highly variable genes with FindVariableFeatures. We scaled the expression data for these genes with ScaleData and performed principal component analysis (PCA). We used FindNeighbors and FindClusters to group the cells and manually assigned cell types based on marker genes [27].

We used FindAllMarkers to compare gene expression across cell types in ESCC tissues. This analysis assessed whether genes encoding proteins linked to ESCA were highly expressed in particular cell populations. We considered genes enriched in a cell type when

their average $\log_2$ fold change was greater than 1 and their false discovery rate (FDR)-adjusted *P*-value was below 0.05.

Functional enrichment

We used Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to examine the functions of the identified genes. We performed these analyses with the clusterProfiler package to assess the biological processes and signaling pathways associated with these genes.

Cell culture and gene overexpression and knockdown

The TE10 cell line is widely used in esophageal cancer research and has excellent transfection efficiency. TE10 exhibits the most stable overexpression levels and pronounced phenotypic changes, which helps reduce technical noise and effectively evaluate gene function. TE10 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin at 37°C in a humidified atmosphere with 5% CO_2. For transient overexpression, cells at 70% confluence were transfected with 2 μg of pcDNA3.1-SH3BGRL2 open reading frame (ORF) clone (GeneCopoeia, Product ID: T1448) or pcDNA3.1-CDX4 ORF clone (GeneCopoeia, Product ID: T1502), or with empty pcDNA3.1 vector (Thermo Fisher Scientific, Cat. No. V790-20) as a negative control, using Lipofectamine 3000 (Thermo Fisher Scientific, Cat. No. L3000001) according to the manufacturer's protocol. For SH3BGRL2 knockdown, cells were transfected with 50 nM SH3BGRL2-specific shRNA (sense: 5'-CCGGCCTGCCACCTCAGATATTTAACTCGAGTTAAATATCTGAGGTGGCAGGTTTTTG-3'; antisense: 5'-AATTCAAAAACCTGCCACCTCAGATATTTAACTCGAGTTAAATATCTGAGGTGGCAGG-3') were annealed and cloned into pLKO.1-puro lentiviral vector (Addgene, Cat. No. 8453) via AgeI and EcoRI restriction sites [28]. Non-targeting control short hairpin RNA (shRNA) (5'-ATCTCGCTTGGGCGAGAGTAAG-3') was cloned into the same vector backbone. TE10 cells at 70% confluence were transfected with 2 μg of shRNA plasmid using Lipofectamine 3000 (Thermo Fisher Scientific, Cat. No. L3000001). Sampling and assay endpoints were defined relative to the time of transfection, as specified in the individual assay sections. To assess the persistence of the intended expression changes, SH3BGRL2 overexpression and knockdown and CDX4 overexpression were evaluated by qRT-PCR and western blotting at 72 h post-transfection (Supplementary Figure 1).

RNA isolation and quantitative reverse transcription-PCR

Total RNA was extracted with TRIzol reagent (Invitrogen) and reverse transcribed using the PrimeScript RT reagent Kit (TaKaRa). Quantitative real-time PCR was carried out on an Eppendorf realplex4 instrument (Eppendorf) using SYBR Green II (TaKaRa). The primer sequences for SH3BGRL2 (Forward 5'-CTCTTCCTCGGGCTTCGT-3' and Reverse 5'-GCATCTCCGTCATCTTCCAC-3') and ACTB (Forward 5'-CATGTACGTTGCTATCCAGGC-3' and Reverse 5'-CTCCTTAATGTCACGCACGAT-3'), synthesized by HuaGene Biotech. Each 20 μL reaction contained 10 μL SYBR Green mix, 0.4 μM of each primer, and 2 μL of cDNA template. Cycling conditions were 95°C for 30 seconds, followed by 40 cycles of 95°C for 5 seconds and 60°C for 30 seconds. Melting-curve analysis confirmed primer specificity. Three independent biological experiments were performed, and each biological replicate included three technical replicates. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method with β-actin as the internal control.

Western blot analysis

Total cellular protein was extracted from ESCC cells by lysing them in ice-cold radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease and phosphatase inhibitors (Thermo Fisher Scientific), followed by 30 minutes on ice with intermittent vortexing and centrifugation at 14,000×g for 15 minutes at 4°C. Protein concentrations were determined using the Pierce bicinchoninic acid (BCA) assay (Thermo Fisher Scientific) with BSA as standard. Equal amounts (20 μg) of denatured protein were resolved on a 5% stacking/10% resolving sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel (Bio-Rad) at 120 V for 90 minutes and transferred onto 0.45 μm polyvinylidene difluoride (PVDF) membranes (Millipore) in transfer buffer (25

mM Tris, 192 mM glycine, 20% methanol) at 100 V for 75 minutes at 4°C. Membranes were rinsed in Tris-buffered saline containing Tween-20 (TBS-T), blocked in 5% non-fat dry milk in TBS-T for 1 hour at room temperature, and incubated overnight at 4°C with primary antibodies diluted in blocking buffer: anti-SH3BGRL2 (1:1,000, Cat #150992, Abcam) and anti-β-actin (1:5,000, Cat #A5441, Sigma-Aldrich). After three 5-minute washes in TBS-T, membranes were incubated for 1 hour at room temperature with horseradish peroxidase (HRP)-conjugated goat anti-rabbit (1:5,000, Cat #7074, Cell Signaling Technology) and goat anti-mouse (1:10,000, Cat #A4416, Sigma-Aldrich) secondary antibodies in blocking buffer. Following three additional washes, bands were visualized using SuperSignal West Pico PLUS (Thermo Fisher Scientific) on a ChemiDoc MP Imaging System (Bio-Rad) and quantified by densitometry with ImageJ v1.53 (NIH).

Cell viability assay

TE10 esophageal squamous cell carcinoma (ESCC) cells were seeded in 96-well plates and transfected at 70-90% confluence. For the CCK-8 assay, the culture medium was replaced with fresh complete medium at 48 h post-transfection. At 72 h post-transfection, the medium was removed, and the cells were washed twice with phosphate-buffered saline (PBS). Fresh culture medium and Cell Counting Kit-8 (CCK-8) solution were mixed at a ratio of 10:1, and 100 µL of the mixture was added to each well. After incubation for 2 h, absorbance was measured at 450 nm to assess cell viability. Three independent biological experiments were performed, each with three technical replicate wells. Technical replicate values were averaged within each biological experiment for statistical analysis.

Apoptosis assay

We assessed apoptosis 72 hours after transfection using an Annexin V-fluorescein isothiocyanate (FITC)/propidium iodide (PI) staining kit (BD Biosciences, San Jose, CA, USA), following the manufacturer's protocol. We collected cells using trypsin without ethylenediaminetetraacetic acid (EDTA) and centrifuged them at 300×g for 5 minutes at 4°C. We washed the cell pellet twice with ice-cold phosphate-buffered saline (PBS, pH 7.4) to remove residual serum and culture medium.

We resuspended the cells in 100 µL of 1× binding buffer and adjusted the concentration to 1-5 × 10^6 cells/mL. Each sample received 5 µL of Annexin V-FITC and 10 µL of PI staining solution. We gently mixed the samples and incubated them in the dark at room temperature for 10-15 minutes. We added 400 µL of 1× binding buffer to each sample and mixed the suspension thoroughly.

We analyzed the samples immediately using a FACSCalibur flow cytometer (BD Biosciences), with a 488 nm laser for Annexin V-FITC excitation and a 633 nm laser for PI excitation. We collected at least 10,000 events per sample. We calculated the apoptotic cell percentage by adding the early apoptotic population (Annexin V^+/PI^-) and the late apoptotic/necrotic population (Annexin V^+/PI^+). We performed three independent biological experiments, each with three technical replicates, and analyzed the data using FlowJo software (TreeStar Inc., Ashland, OR, USA).

Cell Transwell assay

For the added SH3BGRL2-knockdown and CDX4-overexpression/SH3BGRL2-knockdown experiments, cells were harvested at 48 h post-transfection and seeded into uncoated Transwell inserts for migration assays or Matrigel-coated inserts for invasion assays. After incubation for 24 h, corresponding to 72 h post-transfection, the inserts were collected for staining and imaging. We discarded the medium inside the inserts and gently removed the cells remaining on the inner surface with PBS-moistened cotton swabs.

We added 1 mL of crystal violet staining solution to each well of a clean 24-well plate and placed the inserts in the wells to stain the migrated cells for 3-10 minutes. We washed the inserts two or three times with PBS to remove excess stain and photographed them to assess cell migration. We performed three independent biological experiments, each with three technical replicate inserts.

Wound-healing assay

We transfected the cells with plasmids and used a sterile pipette tip to scratch the cell monolayer 24 hours later. We gently washed the monolayer once with PBS to remove detached cells and debris. We added complete

culture medium to support cell migration and proliferation.

We captured baseline images immediately after washing. The cells were cultured for another 48 hours after scratching, and we captured the final images at 72 hours after transfection. We also took images at designated intervals to assess the rate and extent of wound closure. We performed three independent biological experiments, each with three technical replicate wells.

Luciferase experiment

TE10 cells were cultured in DMEM (Gibco, 11965092) supplemented with 10% fetal bovine serum (Gibco, 16000044) and 1% penicillin-streptomycin (Gibco, 15140122) at 37°C in 5% CO_2. For promoter-reporter assays, cells at 70% confluence were co-transfected with 500 ng of pGL3-SH3BGRL2 promoter luciferase reporter plasmid (Promega, E1751) and 50 ng of pRL-TK Renilla control plasmid (Promega, E2241), together with 1 μg of pcDNA3.1-CDX4 expression vector (GeneCopoeia, T1052) or empty pcDNA3.1 vector, using Lipofectamine 3000 (Thermo Fisher Scientific, L3000001) according to the manufacturer's instructions. At 24 h post-transfection, firefly and Renilla luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega, E1910) on a GloMax® 96 Microplate Luminometer (Promega). Relative promoter activity was calculated as the ratio of firefly to Renilla luminescence.

Transcription factor screening

The REMAP (Regulatory Element of Mammalian and Plant) database provides a comprehensive compendium of transcription factor binding profiles derived from high-throughput chromatin immunoprecipitation sequencing (ChIP-seq) experiments across multiple species. Here, we propose to leverage the REMAP database to systematically identify candidate transcription factors regulating our genes of interest. First, gene promoter and enhancer regions will be mapped to REMAP's curated peak sets to retrieve overlapping transcription factor binding sites. Next, enrichment analysis will be performed to prioritize transcription factors whose binding events are significantly overrepresented within these regulatory regions. Finally, Tumor sequencing data were used to find transcription factors that had differential expression and were correlated with the target gene as key transcription factors.

Statistical analysis

Data analysis and visualization were conducted using R software version 4.1.2. Continuous data are presented as mean ± standard deviation (SD) unless otherwise indicated. For transcriptomic analyses, differential expression was assessed using the limma package. Genes with an absolute $\log_2$ fold-change ($|\log FC|$) > 0.7 and an adjusted P value < 0.05 were considered differentially expressed. Considering the relatively limited sample size of the validation cohort, this threshold was selected to balance sensitivity and specificity and to avoid excluding biologically relevant candidates.

For comparisons between two groups, an unpaired two-tailed Student's t-test was used when the model residuals satisfied the normality assumption and the variances were homogeneous. Normality was assessed using the Shapiro-Wilk test applied to model residuals, and homogeneity of variance was evaluated using Levene's test. When the variance homogeneity assumption was violated, Welch's t-test was used. For comparisons among three or more groups, one-way analysis of variance (ANOVA) followed by Tukey's post hoc multiple-comparisons test was used when the relevant parametric assumptions were satisfied. When these assumptions were not met, appropriate non-parametric tests were applied.

Pearson's correlation coefficient was used to evaluate associations between continuous variables. For multiple correlation analyses, including immune checkpoint and immune-cell correlation analyses, P values were adjusted using the Benjamini-Hochberg false discovery rate (BH-FDR) procedure, and adjusted P values < 0.05 were considered statistically significant.

Prognostic analyses were conducted using the survival and survminer R packages. For continuous gene-expression variables, the optimal cutoff was determined using the surv_cutpoint function in the survminer package, with a minimum proportion constraint of 0.30, and patients were subsequently categorized into

high- and low-expression groups using surv_categorize. Overall survival between groups was primarily compared using the log-rank test. The proportional hazards assumption was assessed using scaled Schoenfeld residuals. When the proportional hazards assumption was violated, complementary weighted survival tests were performed as appropriate.

All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant unless otherwise specified. For analyses involving multiple testing, statistical significance was determined using the corresponding adjusted *P* value or multiple-comparison procedure described above.

Results

Proteomic Mendelian randomization analysis to identify circulating proteins associated with esophageal cancer risk

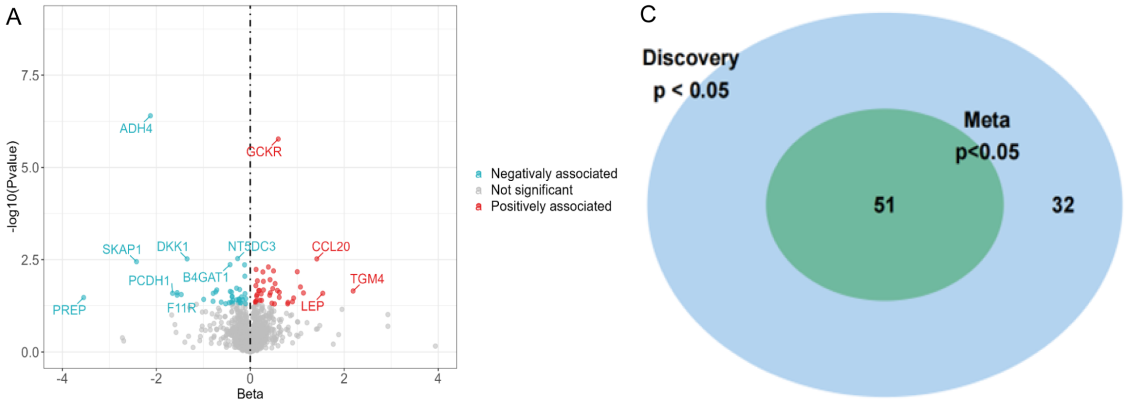
Two-sample Mendelian randomization analysis using cis-pQTL-derived circulating protein levels from the IEU OpenGWAS database revealed 83 plasma proteins significantly associated with ESCC risk (**Figure 1A**). Among these candidates, 40 proteins demonstrated protective effects while 43 exhibited risk-enhancing effects on ESCC development (**Figure 1B**). Bayesian colocalization analysis identified four proteins - ADH1C, ADH4, GCKR, and METAP1 - showing strong evidence of shared causal variants between protein expression quantitative trait loci and ESCC susceptibility (posterior probability of hypothesis 4 (PPH4) > 75%, **Table 1**). Subsequent replication analysis in validation cohorts (BBJ-Esophageal cancer and FinnGen R9) confirmed 51 proteins maintaining significant associations with ESCC risk after multiple testing correction (**Figure 1C**). Functional enrichment analysis was conducted on 51 identified proteins to elucidate their biological roles and pathways. Gene Ontology analysis revealed significant enrichment in pathways critical to carcinogenesis, including regulation of peptidase/endopeptidase activity, Toll-like receptor 4 signaling pathway, and epithelial cell proliferation. KEGG pathway analysis further demonstrated metabolic reprogramming signatures associated with fatty acid degradation, pyruvate metabolism, and glycolysis/gluconeogenesis, as well as cell adhesion molecule pathways (**Figure 1D**). To further evaluate the robustness of the association between genetically predicted circulating SH3BGRL2 levels and ESCC risk, we performed a cross-cohort analysis using the same cis-pQTL instrument, rs2812679 (F = 100). The direction of the association was consistent across the evaluated cohorts. Fixed-effect meta-analysis showed that genetically predicted higher circulating SH3BGRL2 levels were associated with a lower risk of ESCC (OR = 0.50, 95% CI: 0.28-0.89, P = 0.0188), with no substantial evidence of between-cohort heterogeneity. These findings provide additional support for the robustness of the genetic association between circulating SH3BGRL2 levels and ESCC risk (**Figure 1E**).

The influence of crucial circulating proteins on ESCC development

Using transcriptomic data from TCGA-ESCA and GSE17351, we next evaluated the expression patterns of the 51 proteins that remained significant after replication and meta-analysis. The TCGA-ESCA cohort contained 10 upregulated and 13 downregulated candidate genes, whereas GSE17351 contained 1 upregulated and 9 downregulated candidate genes (**Figure 2A**, **2B**). Intersection of the differentially expressed candidates from the two datasets identified five genes - SH3BGRL2, GSN, ALAD, PCDH1, and ACAT1 - that showed concordant downregulation in ESCC compared with normal tissues (**Figure 2C**).

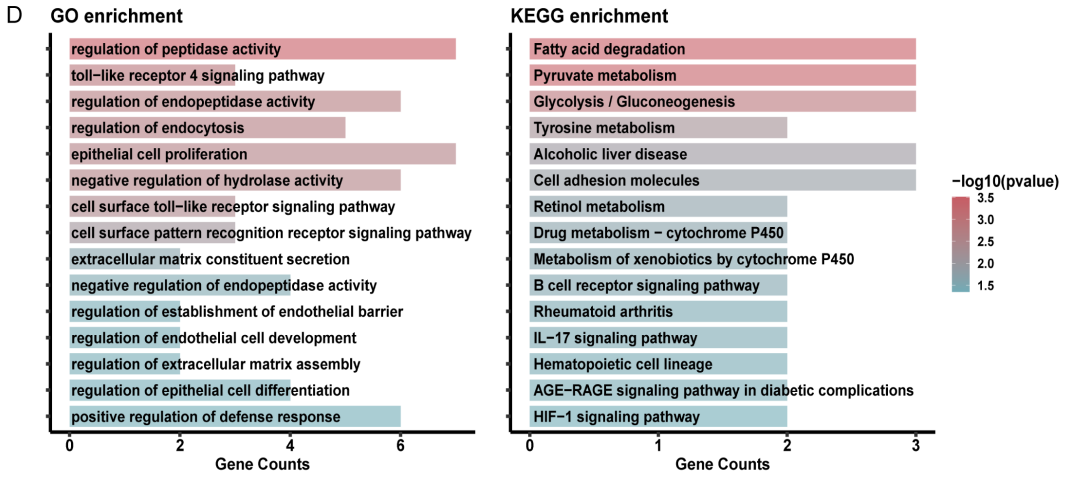
To further prioritize these five candidates, we compared the magnitude of their expression changes in the two independent datasets. SH3BGRL2 showed the largest absolute $\log_2$ fold-change among the five candidates in both datasets, with a logFC of -2.293 in TCGA-ESCA (FDR = 0.000471) and -4.765 in GSE17351 (FDR = 0.0135) (**Figure 2D**). Together with the consistent protective direction observed in the cross-cohort MR analysis, these findings supported the prioritization of SH3BGRL2 for subsequent functional investigation.

Further prognostic analysis showed that higher expression of SH3BGRL2 and PCDH1 was associated with more favorable overall survival, whereas the association for GSN was weaker (**Figure 2E**). To formally assess the proportional hazards assumption, scaled Schoenfeld residual tests were performed for the three prognostic genes. No significant violation of the propor-



B

Protein	Method	Nsnp	OR (95% CI)	P-value
TGM4	Wald ratio	1	8.92(1.37-58.34)	0.022 *
LEP	Wald ratio	1	4.68(1.21-18.18)	0.026 *
CCL20	Wald ratio	1	4.12(1.62-10.5)	0.003 *
FAM172A	Wald ratio	1	3.1(1.15-8.36)	0.025 *
DAPP1	Wald ratio	1	2.91(1.21-7)	0.017 *
CCDC126	Wald ratio	1	2.72(1.32-5.6)	0.007 *
C1orf198	Wald ratio	1	2.52(1.07-5.92)	0.035 *
S100A14	Wald ratio	1	2.46(1.03-5.87)	0.044 *
TRIM3	Wald ratio	1	2.23(1.02-4.9)	0.046 *
STX16	Wald ratio	1	0.23(0.06-0.85)	0.028 *
CIAPIN1	Wald ratio	1	0.21(0.05-0.82)	0.025 *
F11R	Wald ratio	1	0.21(0.05-0.85)	0.029 *
PCDH1	Wald ratio	1	0.19(0.04-0.82)	0.026 *
ADH4	Wald ratio	1	0.12(0.05-0.27)	<0.001 ***
SKAP1	Wald ratio	1	0.09(0.02-0.45)	0.004 *
PREP	Wald ratio	1	0.03(0-0.76)	0.034 *
ADH1C	Wald ratio	1	0.02(0-0.06)	<0.001 ***
METAP1	Wald ratio	1	0(0-0.03)	<0.001 ***



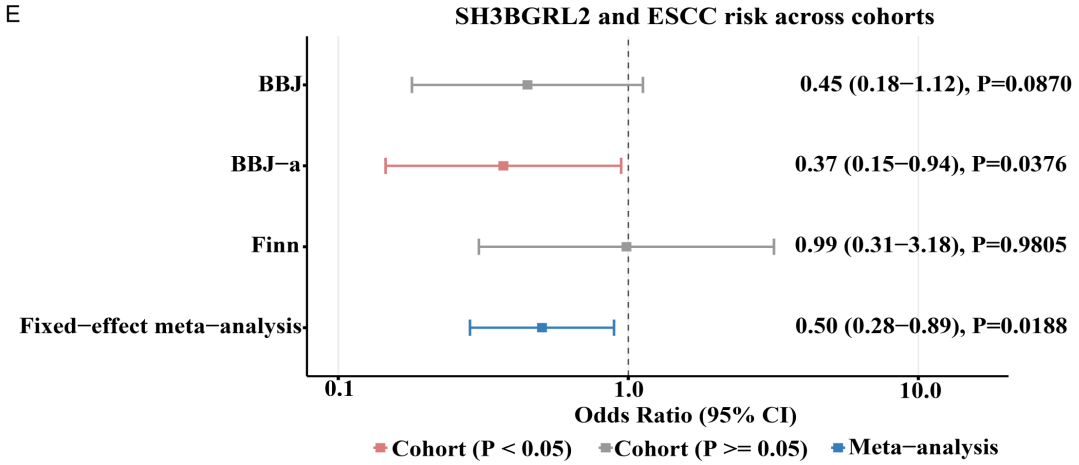


Figure 1. Marker gene screening for esophageal squamous cell carcinoma based on proteome-wide Mendelian randomization. A. Selection of esophageal squamous cell carcinoma (ESCC) Marker Genes based on IEU OpenGWAS Database. B. Forest plot of Top 18 genes influencing ESCC. C. Key Marker Genes stably influencing ESCC after validation. D. Bar chart of functional enrichment analysis of 51 genes. E. Cross-cohort analysis of the association between genetically predicted circulating SH3BGRL2 levels and ESCC risk using the cis-pQTL instrument rs2812679. $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

Table 1. Colocalization results in bbj-a-117 dataset

nsnps	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf	Protein
1	2.39E-37	4.82E-06	4.96E-35	0	0.999995182	ADH1C
1	1.20E-35	0.000242243	4.96E-35	0	0.999757757	ADH4
1	2.65E-220	0.000841776	3.14E-220	0	0.999158224	GCKR
1	1.02E-08	3.19E-06	3.19E-06	0	0.999993613	METAP1

tional hazards assumption was detected for SH3BGRL2 (P = 0.4498) or PCDH1 (P = 0.7234), and the log-rank test was therefore retained as the primary method for these two genes (Supplementary Figure 2A, 2B). In contrast, the proportional hazards assumption was significantly violated for GSN (P = 0.0251; Supplementary Figure 2C). The conventional log-rank test for GSN showed no statistically significant difference in overall survival (P = 0.1444). Because of the non-proportional hazards pattern, we additionally performed a Prentice-weighted log-rank test and a Rényi supremum test, which yielded P = 0.0636 and P = 0.0373, respectively. Although the Rényi supremum test suggested a potential survival difference between the GSN expression groups, neither the conventional nor the Prentice-weighted log-rank test reached statistical significance. Given the violation of the proportional hazards assumption and the differing results across tests, the prognostic association of GSN should be interpreted cautiously.

Prognostic model of key expression quantitative trait locus (eQTL) proteins can predict the prognosis of ESCC

Utilizing three potential prognostic genes, we developed an ESCC prognosis model using least absolute shrinkage and selection operator (LASSO) regression. Our analysis revealed that this model effectively predicts the prognosis of ESCC (**Figure 3A**, P < 0.001, HR = 0.32). To evaluate the predictive ability of the model at different time points, we utilized time series-based receiver operating characteristic (ROC) analysis (**Figure 3B**). The results showed a weaker prediction ability in the first year (AUC = 0.68), but as time increases, the model's effectiveness improves and reaches its peak in the third year (AUC = 0.89). Additionally, we constructed a comprehensive predictive model that incorporates other clinical parameters and visualized it using a nomogram. The higher the model score, the poorer the prognostic effect (**Figure 3C**). To further understand the mechanism of the prognostic model, we performed differential expression analysis by dividing the model into high and low expression groups. Our analysis revealed 450 highly expressed genes and 493 low expression genes (Supplementary Figure 3). Functional enrichment analysis of these genes showed a significant impact on xenobiotic metabolism (cytochrome P450), estrogen signaling, and IL-17 signaling (**Figure 3D**).

SH3BGRL2 and PCDH1 are related to tumor immune microenvironment

Genetic factors affecting the immune microenvironment and associated immune checkpoints can influence patient outcomes. To assess the effects of three specific genes on the tumor immune microenvironment, we examined their correlation with immune checkpoints. Our analysis indicated a negative correlation between SH3BGRL2 and PCDH1 with genes such as TNFRSF18, CD44, and PDCD1LG2, and a strong positive correlation with BTNL3. In contrast, GSN showed minimal association with immune checkpoints (**Figure 4A**). Additionally, we assessed patient immune cell scores to explore the relationship between the three genes and immune cells. Our findings demonstrated a positive correlation between SH3BGRL2 and PCDH1 with regulatory T cells

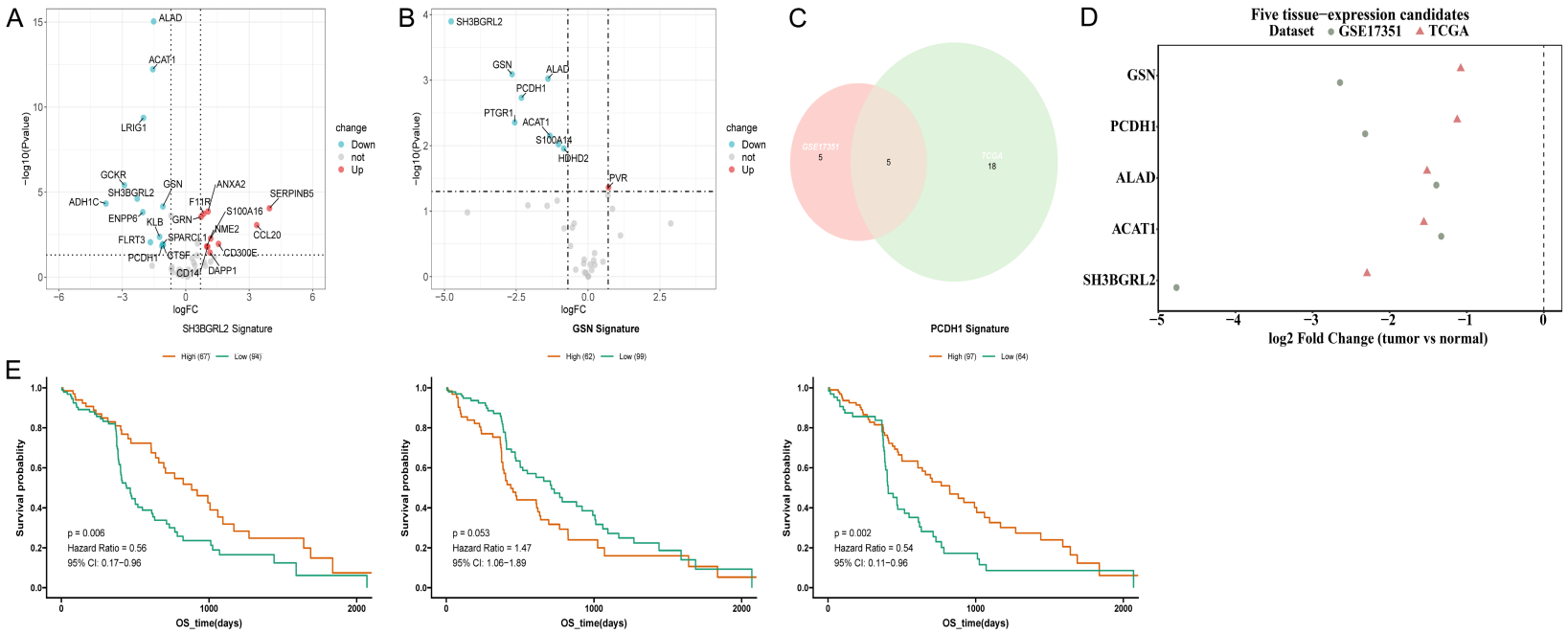


Figure 2. Integrated transcriptomic analysis prioritizes SH3BGRL2 among five candidate genes in ESCC. A. Volcano plot showing differentially expressed genes among the replicated candidates in the TCGA-ESCA dataset. B. Volcano plot showing differentially expressed genes among the replicated candidates in the GSE17351 dataset. C. Venn diagram showing the five genes with concordant differential expression in both datasets: SH3BGRL2, GSN, ALAD, PCDH1, and ACAT1. D. Comparison of the $\log_2$ fold changes of the five candidate genes between ESCC tumor and normal tissues in TCGA-ESCA and GSE17351. SH3BGRL2 showed the largest absolute expression difference across the two datasets. E. Kaplan-Meier overall survival analyses of SH3BGRL2, PCDH1, and GSN in ESCC patients.

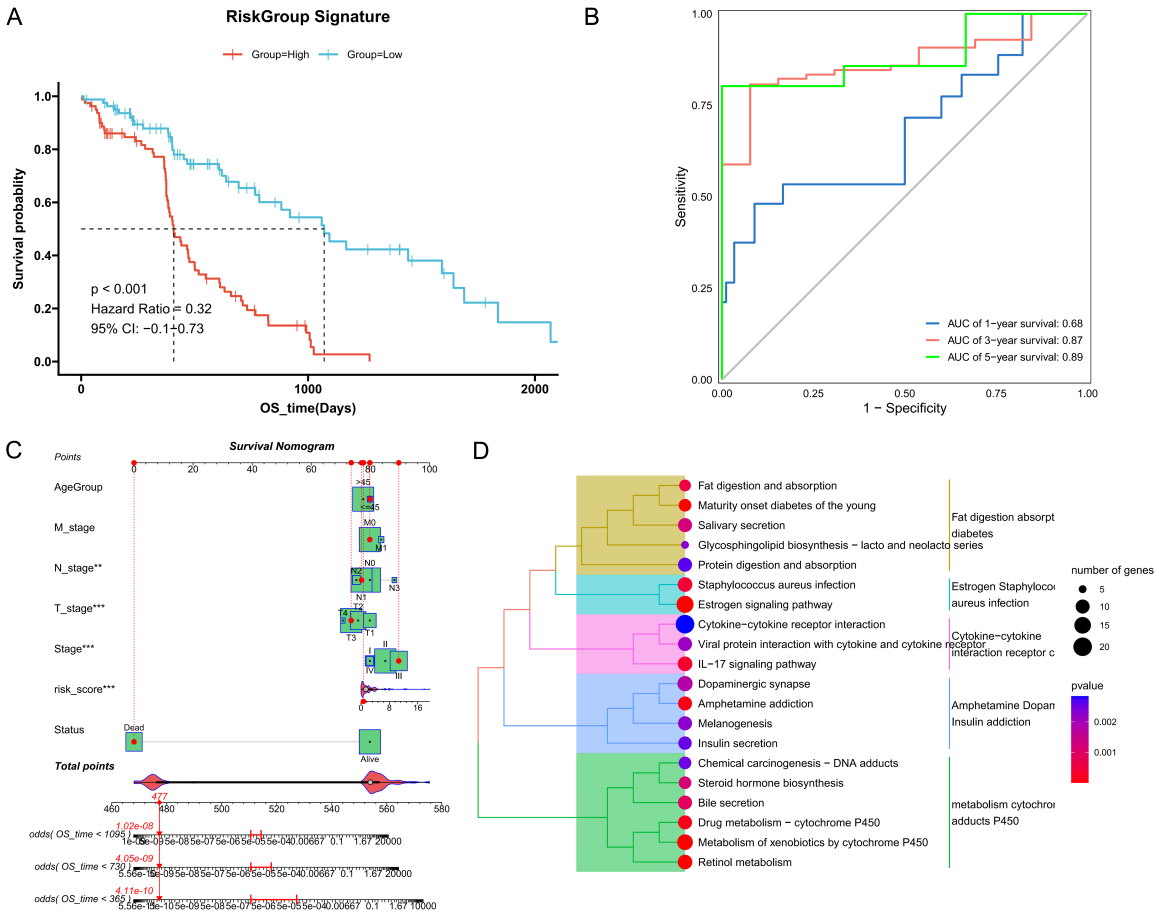


Figure 3. Key three gene prognosis models affect esophageal squamous cell carcinoma (ESCC) prognosis. A. Three gene-based prognostic model affects the prognosis of ESCC; B. Observe the diagnostic efficacy of prognostic model on ESCC based on time series receiver operating characteristic (ROC); C. Nomogram of prognostic diagnosis combined with other clinical parameters; D. KEGG enrichment analysis graph of potential mechanism of prognostic model.

(Tregs) (**Figure 4B**). To determine whether these genes are expressed in immune cells or exclusively in tumor cells, we utilized the single-cell sequencing dataset GSE145370 to examine gene expression characteristics. Following annotation, we identified expression in six immune cell types: B cells, dendritic cells, macrophages, monocytes, natural killer (NK) cells, and T cells (**Figure 4C**). Upon examining gene expression in immune cells, we discovered that SH3BGRL2 and PCDH1 were not expressed in immune cells, indicating their primary function in tumor cells. Conversely, GSN was predominantly expressed in monocytes (**Figure 4D**).

Effects of SH3BGRL2 overexpression and knockdown on ESCC cell phenotypes

SH3BGRL2 was prioritized for functional investigation because its transcript showed the largest absolute expression difference among the five candidates in both datasets, and its role in ESCC remains insufficiently characterized. To explore its functional role, we overexpressed SH3BGRL2 in ESCC cell lines (**Figure 5A-C**) and performed several functional assays to evaluate its effects. At 72 h after transfection, qRT-PCR and western blot analyses showed increased SH3BGRL2 mRNA and protein levels in the SH3BGRL2-overexpression group and reduced levels in the SH3BGRL2-knockdown group (Supplementary Figure 1A and 1B). CDX4 mRNA and protein levels were also increased following CDX4 overexpression (Supplementary Figure 1C and 1D). These results support the persistence of the intended expression changes at 72 h after transfection.

In ESCC TE10 cells, SH3BGRL2 overexpression resulted in significant inhibitory effects on cel-

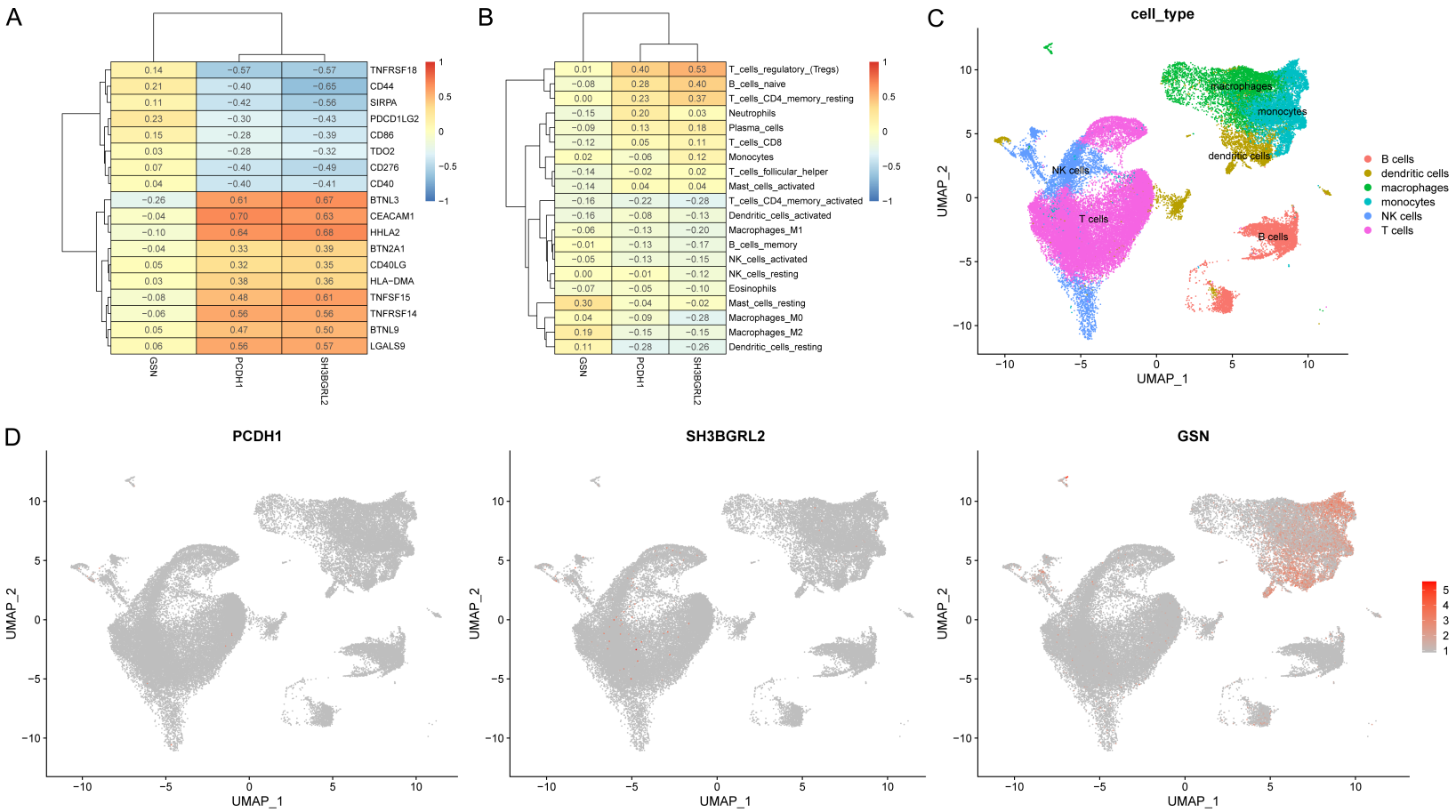


Figure 4. SH3BGRL2 and PCDH1 affect the immune microenvironment of ESCC tumors. A. Heatmap for associations between three key genes and immune checkpoints; B. Analysis of associations between three key genes and immune cell scores; C. Dimensional reduction map for cell annotation for single-cell sequencing; D. Expression characteristics of three key genes in different immune cells. *P* values were adjusted using the Benjamini-Hochberg false discovery rate procedure.

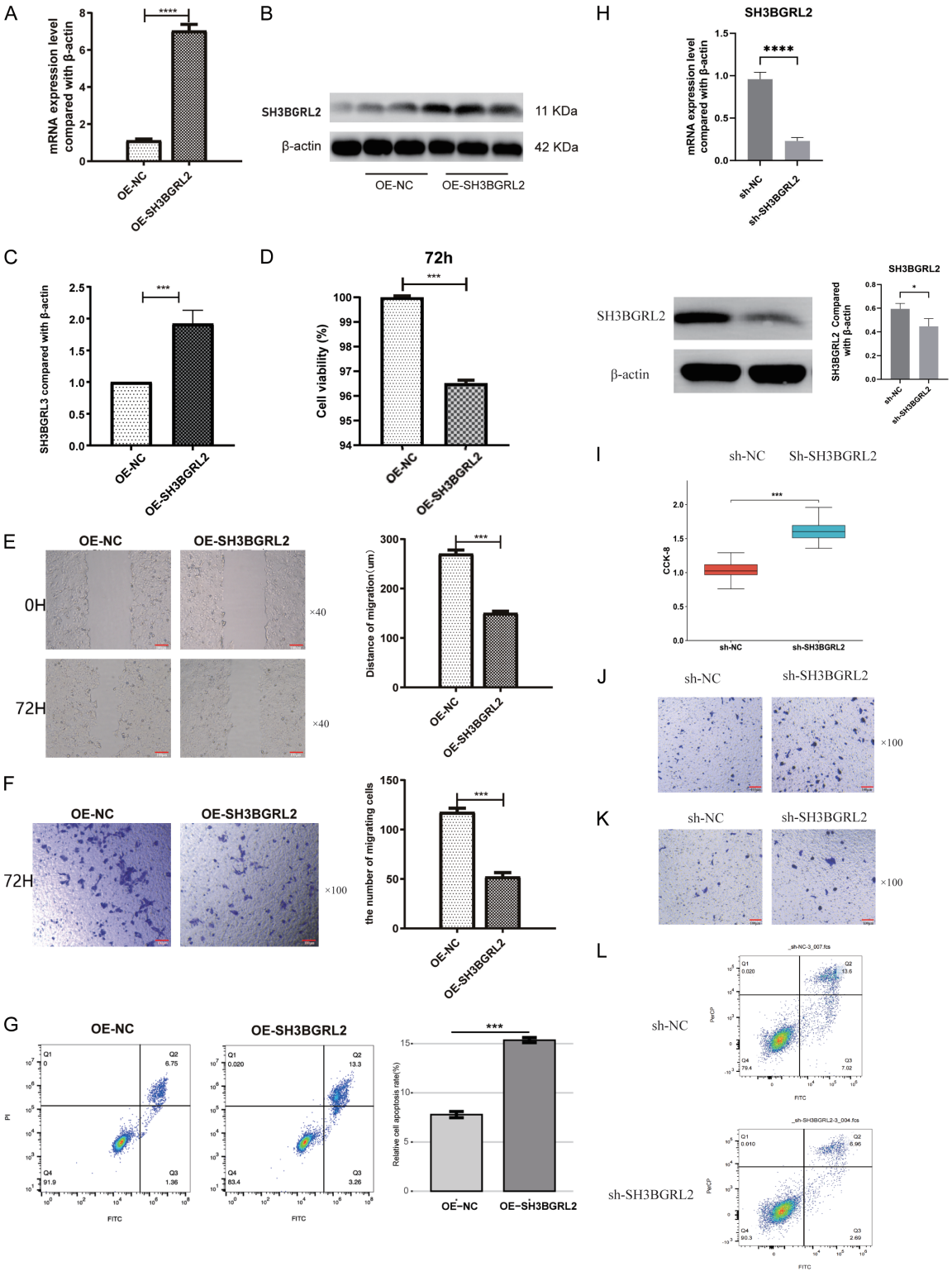


Figure 5. Effects of SH3BGRL2 overexpression and knockdown on esophageal squamous cell carcinoma (ESCC) cell phenotypes. (A) Quantitative reverse transcription PCR (qRT-PCR) analysis of SH3BGRL2 mRNA expression following overexpression. (B, C) Representative western blots and quantification of SH3BGRL2 protein expression following overexpression. (D) Cell viability measured using the Cell Counting Kit-8 (CCK-8) assay. (E) Representative wound-healing images and quantification of wound closure. (F) Representative Transwell migration images and quantification of migrated cells. (G) Representative flow cytometry plots and quantification of Annexin V-fluorescein isothio-

cyanate (FITC)/propidium iodide (PI)-stained cells following SH3BGRL2 overexpression. (H) Validation of SH3BGRL2 knockdown by qRT-PCR and western blotting, with protein quantification. (I-L) Effects of SH3BGRL2 knockdown on cell viability (I), migration (J), invasion (K), and apoptosis (L). Images in (E) were acquired at ×40 magnification, and images in (F), (J), and (K) were acquired at ×100 magnification. Scale bars are shown in the corresponding panels. ** for $P < 0.01$, *** for $P < 0.001$, and **** for $P < 0.0001$.

lular functions. The CCK-8 assay showed a marked decrease in cell viability, indicating that SH3BGRL2 may be crucial in regulating cell proliferation in ESCC (**Figure 5D**). SH3BGRL2 overexpression reduced wound closure in the wound-healing assay (**Figure 5E**). The Transwell assay also showed fewer migrated cells after SH3BGRL2 overexpression (**Figure 5F**). Flow cytometry showed a higher proportion of apoptotic cells (**Figure 5G**), suggesting that SH3BGRL2 may promote cell death and help suppress tumor growth.

We also examined SH3BGRL2 knockdown to assess the effects of reducing its expression. Quantitative reverse transcription PCR and western blotting confirmed lower SH3BGRL2 mRNA and protein levels in the sh-SH3BGRL2 group than in the sh-NC group (**Figure 5H**). SH3BGRL2 knockdown increased cell viability (**Figure 5I**), migration and invasion (**Figure 5J** and **5K**), while reducing the proportion of apoptotic cells (**Figure 5L**). These findings support the results of the overexpression experiments.

SH3BGRL2 expression is regulated by CDX4

To further investigate the regulatory mechanism underlying the reduced expression of SH3BGRL2 in ESCC, we first examined CDX4 and SH3BGRL2 protein expression in normal esophageal epithelial cells and two ESCC cell lines. CDX4 expression was higher in normal esophageal epithelial cells and lower in both ESCC cell lines. SH3BGRL2 showed a similar expression pattern, with higher expression in normal esophageal epithelial cells and lower expression in the two ESCC cell lines (**Figure 6A**). These findings provided additional evidence that the reduced expression of SH3BGRL2 in ESCC was accompanied by reduced CDX4 expression.

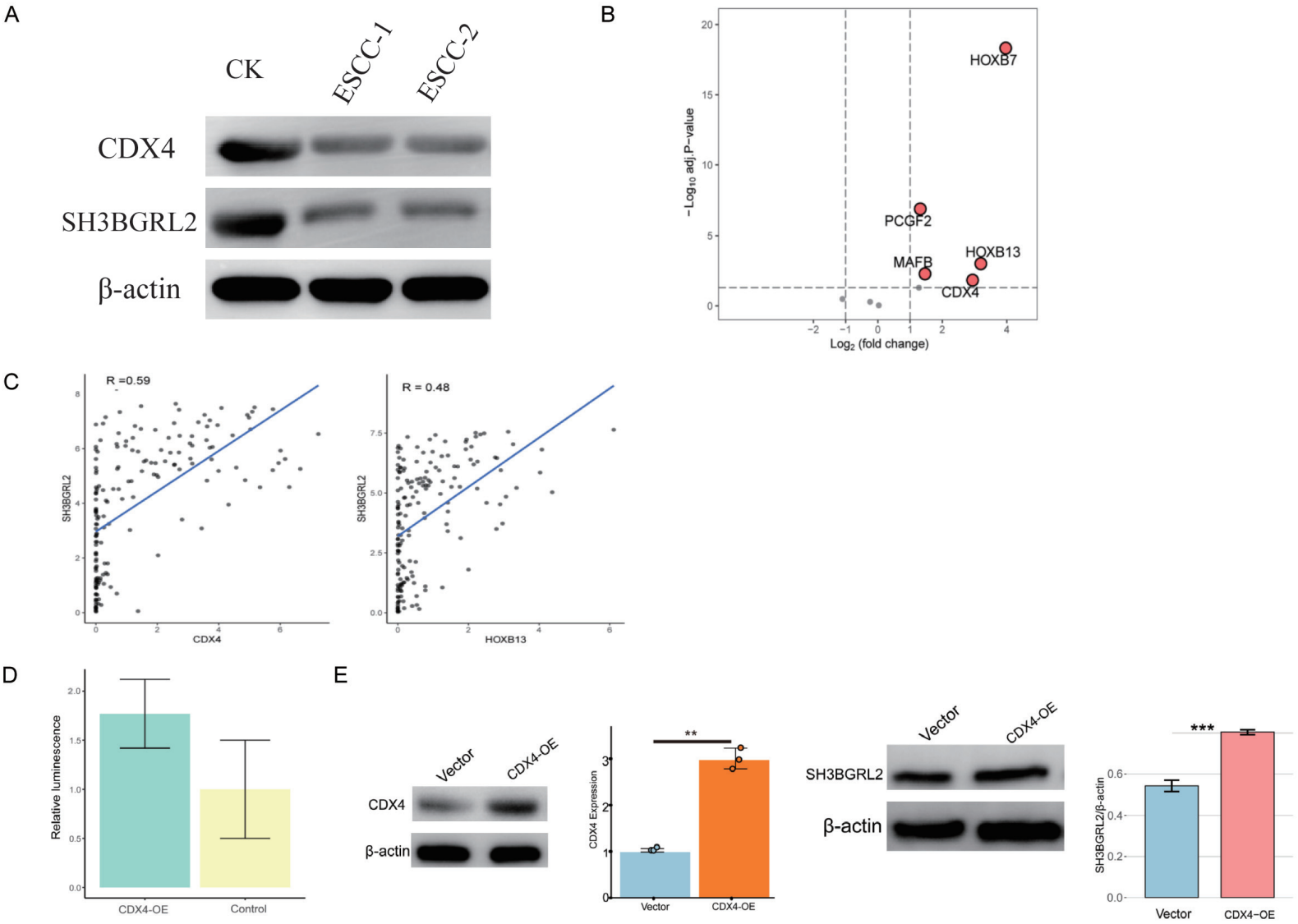
We examined whether CDX4 could regulate SH3BGRL2 expression. A search of the REMAP database identified nine candidate transcription factors that might regulate SH3BGRL2. Differential expression analysis narrowed this group to five factors with high expression in the tumor dataset used in our study (**Figure 6B**). Among these candidates, CDX4 expression showed a significant positive correlation with SH3BGRL2 expression (**Figure 6C**). We therefore selected CDX4 for further study as a possible upstream regulator of SH3BGRL2.

To further evaluate this regulatory relationship, we performed a dual-luciferase reporter assay using the SH3BGRL2 promoter. CDX4 overexpression significantly increased SH3BGRL2 promoter-driven luciferase activity compared with the vector control (**Figure 6D**). Consistent with the reporter assay, CDX4 overexpression increased SH3BGRL2 protein expression (**Figure 6E**). Together, these findings support a transcriptional regulatory relationship between CDX4 and SH3BGRL2.

We next examined whether SH3BGRL2 contributes to the cellular effects of CDX4. Western blot analysis showed that SH3BGRL2 knockdown attenuated the increase in SH3BGRL2 protein associated with CDX4 overexpression (**Figure 6F**). CDX4 overexpression reduced cell viability, whereas simultaneous SH3BGRL2 knockdown partially reversed this effect (**Figure 6G**). CDX4 overexpression also reduced migration and invasion and increased the proportion of apoptotic cells. These effects were partially reversed by simultaneous SH3BGRL2 knockdown (**Figure 6H-J**). Together, these findings support a contribution of SH3BGRL2 to the effects of CDX4 on ESCC cell viability, migration, invasion, and apoptosis.

Discussion

The proteome-wide MR analysis evaluated genetically predicted circulating plasma protein levels, whereas the transcriptomic and functional analyses examined SH3BGRL2 expression and cellular phenotypes in ESCC tissues and cultured cells. These findings provide evidence at distinct biological levels and should be interpreted within their respective contexts. Further investigation of the relationship between circulating and tumor-tissue SH3BGRL2 protein levels will help clarify how these findings relate to one another.


A
CK
ESCC-1
ESCC-2
CDX4
SH3BGRL2
β-actin
B
HOXB7
PCGF2
MAFB
HOXB13
CDX4
−Log10 adj.P−value
Log2 (fold change)
C
R =0.59
SH3BGRL2
CDX4
R = 0.48
SH3BGRL2
HOXB13
D
Relative luminescence
CDX4-OE
Control
E
Vector
CDX4-OE
CDX4
β-actin
CDX4 Expression
**
Vector
CDX4-OE
Vector
CDX4-OE
SH3BGRL2
β-actin
***
SH3BGRL2/β-actin
Vector
CDX4-OE

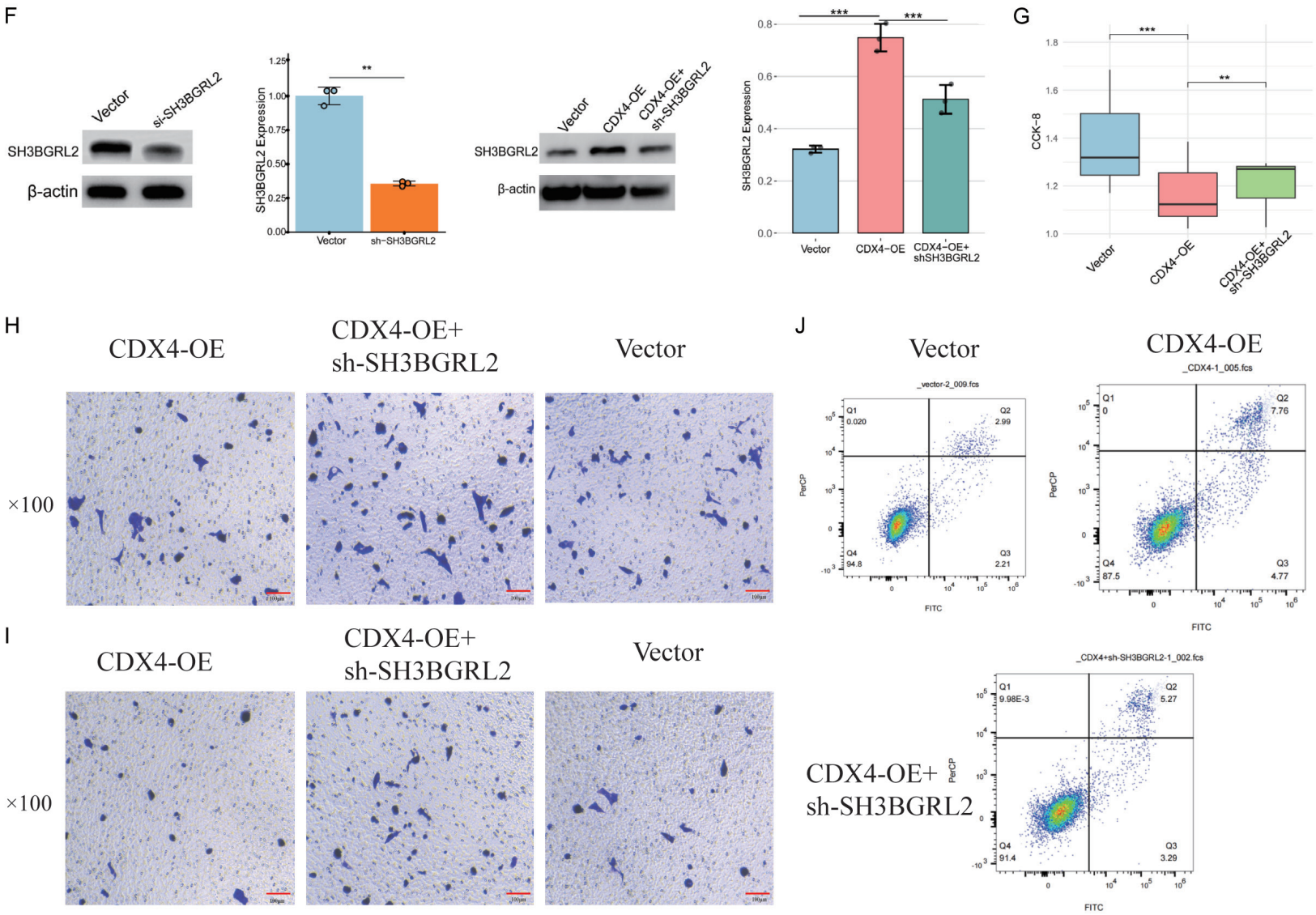


Figure 6. CDX4 regulates SH3BGRL2 expression, and SH3BGRL2 contributes to CDX4-associated cellular phenotypes. (A) Representative western blots and quantification of CDX4 and SH3BGRL2 protein expression in Het-1A and the esophageal squamous cell carcinoma (ESCC) cell lines TE-10 and KYSE-150. (B) Volcano plot showing differential expression of candidate transcription factors. The horizontal and vertical axes represent $\log_2$ fold change and $-\log_{10}$ adjusted *P* value, respec-

tively. (C) Correlations of SH3BGRL2 expression with CDX4 and HOXB13 expression, with correlation coefficients of 0.59 and 0.48, respectively. (D) Relative SH3BGRL2 promoter-driven luciferase activity following CDX4 overexpression, normalized to Renilla luciferase activity. (E) Representative western blots and quantification of CDX4 and SH3BGRL2 following CDX4 overexpression. (F) Western blot validation of SH3BGRL2 expression in the control, CDX4-overexpression, and CDX4-overexpression plus SH3BGRL2-knockdown groups. (G-J) Cell viability measured using the Cell Counting Kit-8 (CCK-8) assay (G), Transwell migration (H), Matrigel-coated Transwell invasion (I), and apoptosis assessed by Annexin V-fluorescein isothiocyanate (FITC)/propidium iodide (PI) flow cytometry (J) in the three groups. Images in (H) and (I) were acquired at ×100 magnification. Scale bars are shown in the corresponding panels. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

The specific role of SH3BGRL2 in inhibiting migration and invasion, coupled with its ability to promote apoptosis, underscores its potential as a therapeutic target in ESCC. The ability of SH3BGRL2 to reduce cell migration and invasion directly impacts ESCC's propensity to metastasize, which is often the deadliest aspect of the disease [28, 29]. Moreover, its role in promoting apoptosis addresses a fundamental cancer hallmark by triggering the programmed death of cancerous cells, potentially limiting tumor growth and spread [30]. The dual functionality of SH3BGRL2 suggests that it could play a significant role in both slowing ESCC progression and reducing the likelihood of metastasis, making it an attractive candidate for targeted therapy. Given the aggressive nature of ESCC and the current challenges in treatment, especially in advanced stages, targeting SH3BGRL2 could lead to the development of a novel therapeutic approach that specifically mitigates these key aspects of ESCC pathology [31]. Further research should, therefore, aim to understand the molecular mechanisms through which SH3BGRL2 exerts these effects. It would be beneficial to investigate how SH3BGRL2 interacts with other molecular pathways involved in cell cycle regulation and apoptosis. Additionally, while CDX4 has been largely studied in leukemia, its function in solid tumors remains underexplored. Previous work demonstrated that CDX4 is normally present in early hematopoietic progenitors but is aberrantly overexpressed in about 25% of acute myeloid leukemia (AML) patients, with enforced CDX4 expression boosting murine hematopoietic stem cell (HSC) self-renewal in vitro and causing AML-like disease in roughly half of transplanted mice [32, 33]. This study is the first to validate CDX4 as a regulator of SH3BGRL2 in ESCC, broadening the CDX family's relevance in solid tumors and providing preliminary functional evidence for the CDX4-SH3BGRL2 regulatory relationship in ESCC cells.

The insights gleaned from the drug-target interaction analysis using the Drug-Gene Interaction Database (DGIdb) database have offered promising avenues for therapeutic interventions in ESCC [34]. Our analysis identified several gene targets linked to existing drugs, including VEGFA, CCN2, and CD22. These links suggest possible treatment options for further study in ESCC. VEGFA is targeted by anti-angiogenic agents such as bevacizumab, which points to the role of blood vessel formation in ESCC and the possible value of blocking this process [35-38]. The link between CCN2 and pamrevlumab supports interest in targeting fibrosis-related pathways in the tumor microenvironment [39-42]. This approach is relatively new in ESCC research [43, 44]. CD22 is a target of inotuzumab ozogamicin, which raises the possibility of studying B-cell-related treatment approaches in ESCC based on their use in blood cancers [45-48], These findings provide a basis for exploring new uses of existing drugs in ESCC. Repurposing approved drugs may help shorten the path from laboratory research to clinical use [49, 50].

Combining MR with single-cell profiling allowed us to examine these findings from different perspectives. MR helps reduce the influence of confounding factors, while single-cell profiling reveals differences between cells in the tumor microenvironment. Together, these methods strengthen the evidence for the identified biomarkers and help explain their roles in specific cell populations. Such differences can be difficult to detect in bulk tissue data [51-54].

Our study also has limitations. First, MR assumes that the genetic variants used as instruments affect the outcome only through the exposure, which in this study was circulating protein levels. Pleiotropy may violate this assumption if the variants influence the outcome through other pathways. In addition, the lack of paired plasma and tumor-tissue

SH3BGRL2 protein measurements limits our ability to assess concordance between these compartments. Future studies incorporating matched plasma and tumor samples are needed to clarify this relationship and strengthen the interpretation of the genetic and tissue-level findings. Second, the differential gene expression analyses and prognostic model development were based on sequencing data and require further validation in independent patient cohorts. Single-cell methods also have limited sensitivity and do not capture the full transcriptome of every cell, which may affect the interpretation and reproducibility of the findings. Third, the functional experiments were conducted in a single ESCC cell line, and further validation in additional cell lines and animal models is needed. Although the reporter and expression experiments support a regulatory relationship between CDX4 and SH3BGRL2, direct binding of CDX4 to the endogenous SH3BGRL2 regulatory region was not established. Finally, the clinical relevance of CDX4, including its prognostic and clinicopathological significance in ESCC, has not been comprehensively evaluated and requires investigation in independent patient cohorts. Therefore, the present findings support the biological and functional relevance of the CDX4-SH3BGRL2 relationship, while its clinical significance remains to be established.

Conclusion

Our study shows how genomic and proteomic methods can help us understand ESCC and identify possible treatment targets. We identified protein biomarkers and highlighted the role of SH3BGRL2 in limiting tumor progression. These findings add to our understanding of ESCC biology and suggest new directions for diagnosis and treatment. Further work using these methods may support more precise care and help improve patient outcomes.

Acknowledgements

This work was supported by Provincial level project (Grant Nos. 2024SF-YBXM-112); Tangdu Hospital Science Foundation (Grant Nos. 2023BTDQN014 and 2021LCYJ024).

Disclosure of conflict of interest

None.

Address correspondence to: Jianyong Sun, Department of Thoracic Surgery, Tangdu Hospital, The Fourth Military Medical University, Xinsi Road 569, Xi'an 710038, Shaanxi, China. E-mail: sunjianyong2@163.com; Yan Chu, School of Medicine, Xi'an Siyuan University, Shui'an Road 28, Xi'an 710038, Shaanxi, China. E-mail: sunshineyan269@163.com

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Siegel RL, Giaquinto AN and Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024; 74: 12-49.

[3] Pickens A and Orringer MB. Geographical distribution and racial disparity in esophageal cancer. Ann Thorac Surg 2003; 76: S1367-1369.

[4] Huang FL and Yu SJ. Esophageal cancer: risk factors, genetic association, and treatment. Asian J Surg 2018; 41: 210-215.

[5] Domper Arnal MJ, Ferrández Arenas Á and Lanas Arbeloa Á. Esophageal cancer: risk factors, screening and endoscopic treatment in Western and Eastern countries. World J Gastroenterol 2015; 21: 7933-7943.

[6] Waters JK and Reznik SI. Update on management of squamous cell esophageal cancer. Curr Oncol Rep 2022; 24: 375-385.

[7] Rao D, Lu H, Wang X, Lai Z, Zhang J and Tang Z. Tissue-derived exosome proteomics identifies promising diagnostic biomarkers for esophageal cancer. Elife 2023; 12: e86209.

[8] Jing Z, Chen K and Gong L. The significance of exosomes in pathogenesis, diagnosis, and treatment of esophageal cancer. Int J Nanomedicine 2021; 16: 6115-6127.

[9] Li Y, Liu J, Cai XW, Li HX, Cheng Y, Dong XH, Yu W and Fu XL. Biomarkers for the prediction of esophageal cancer neoadjuvant chemoradiotherapy response: a systemic review. Crit Rev Oncol Hematol 2021; 167: 103466.

[10] Liu W, Xie L, He YH, Wu ZY, Liu LX, Bai XF, Deng DX, Xu XE, Liao LD, Lin W, Heng JH, Xu X, Peng L, Huang QF, Li CY, Zhang ZD, Wang W, Zhang GR, Gao X, Wang SH, Li CQ, Xu LY, Liu W and Li EM. Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting. Nat Commun 2021; 12: 4961.

[11] Birney E. Mendelian randomization. Cold Spring Harb Perspect Med 2022; 12: a041302.

[12] Bowden J and Holmes MV. Meta-analysis and Mendelian randomization: a review. Res Synth Methods 2019; 10: 486-496.

[13] Sekula P, Del Greco M F, Pattaro C and Köttgen A. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol 2016; 27: 3253-3265.
[14] Gao X, Wang Z, Liu B and Cheng Y. Causal association of gut microbiota and esophageal cancer: a Mendelian randomization study. Front Microbiol 2023; 14: 1286598.
[15] Zhang X, Yang X, Zhang T, Yin X, Man J and Lu M. Association of educational attainment with esophageal cancer, Barrett's esophagus, and gastroesophageal reflux disease, and the mediating role of modifiable risk factors: a Mendelian randomization study. Front Public Health 2023; 11: 1022367.
[16] Zhou J, Fang P, Liang Z, Li X, Luan S, Xiao X, Gu Y, Shang Q, Zhang H, Yang Y, Chen L, Zeng X and Yuan Y. Causal relationship between lung diseases and risk of esophageal cancer: insights from Mendelian randomization. J Cancer Res Clin Oncol 2023; 149: 15679-15686.
[17] Wu H, Ma T, Li D, He M, Wang H and Cui Y. Circulating vascular endothelial growth factor and cancer risk: a bidirectional mendelian randomization. Front Genet 2022; 13: 981032.
[18] Yang Z, Guo L, Sun Y, Huang Y, Li J, Lin Y, Zhang X, Wu D and Luo Y. Investigation of the causal relationship between Interleukin-6 signaling and gastrointestinal tract cancers: a Mendelian randomization study. Dig Liver Dis 2024; 56: 679-686.
[19] Gohil SH, Iorgulescu JB, Braun DA, Keskin DB and Livak KJ. Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. Nat Rev Clin Oncol 2021; 18: 244-256.
[20] Ma RY, Black A and Qian BZ. Macrophage diversity in cancer revisited in the era of single-cell omics. Trends Immunol 2022; 43: 546-563.
[21] Massimino M, Martorana F, Stella S, Vitale SR, Tomarchio C, Manzella L and Vigneri P. Single-cell analysis in the omics era: technologies and applications in cancer. Genes (Basel) 2023; 14: 1330.
[22] Guo D, Sheng K, Zhang Q, Li P, Sun H, Wang Y, Lyu X, Jia Y, Wang C, Wu J, Zhang X, Wang D, Sun Y, Huang S, Yu J and Zhang J. Single-cell transcriptomic analysis reveals the landscape of epithelial-mesenchymal transition molecular heterogeneity in esophageal squamous cell carcinoma. Cancer Lett 2024; 587: 216723.
[23] Ji G, Yang Q, Wang S, Yan X, Ou Q, Gong L, Zhao J, Zhou Y, Tian F, Lei J, Mu X, Wang J, Wang T, Wang X, Sun J, Zhang J, Jia C, Jiang T, Zhao MG and Lu Q. Single-cell profiling of response to neoadjuvant chemo-immunotherapy in surgically resectable esophageal squamous cell carcinoma. Genome Med 2024; 16: 49.
[24] Jia Y, Zhang B, Zhang C, Kwong DL, Chang Z, Li S, Wang Z, Han H, Li J, Zhong Y, Sui X, Fu L, Guan X and Qin Y. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in esophageal squamous cell carcinoma. Adv Sci (Weinh) 2023; 10: e2204565.
[25] Sui X, Chen C, Zhou X, Wen X, Shi C, Chen G, Liu J, He Z, Yao Y, Li Y and Gao Y. Integrative analysis of bulk and single-cell gene expression profiles to identify tumor-associated macrophage-derived CCL18 as a therapeutic target of esophageal squamous cell carcinoma. J Exp Clin Cancer Res 2023; 42: 51.
[26] Sun J, Zhao J, Jiang F, Wang L, Xiao Q, Han F, Chen J, Yuan S, Wei J, Larsson SC, Zhang H, Dunlop MG, Farrington SM, Ding K, Theodoratou E and Li X. Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. Genome Med 2023; 15: 75.
[27] Zhao WJ, Liu Y, Guo J and Liu H. Prognostic biomarkers identification in esophageal cancer based on WGCNA and single-cell analysis. Dis Markers 2022; 2022: 6595778.
[28] Yin L, Li W, Xu A, Shi H, Wang K, Yang H, Wang R and Peng B. SH3BGRL2 inhibits growth and metastasis in clear cell renal cell carcinoma via activating hippo/TEAD1-Twist1 pathway. EBioMedicine 2020; 51: 102596.
[29] Nie Z, Cai S, Wei Z, Li Y, Bian L, Wang C and Wang C. SH3BGRL2 functions as a crucial tumor suppressor in glioblastoma tumorigenesis. Biochem Biophys Res Commun 2021; 547: 148-154.
[30] Hecker M, Thamilarasan M, Koczan D, Schröder I, Flechtner K, Freiesleben S, Füllen G, Thiesen HJ and Zettl UK. MicroRNA expression changes during interferon-beta treatment in the peripheral blood of multiple sclerosis patients. Int J Mol Sci 2013; 14: 16087-16110.
[31] Li DD, Deng L, Hu SY, Zhang FL and Li DQ. SH3BGRL2 exerts a dual function in breast cancer growth and metastasis and is regulated by TGF-β1. Am J Cancer Res 2020; 10: 1238-1254.
[32] Bansal D, Scholl C, Fröhling S, McDowell E, Lee BH, Döhner K, Ernst P, Davidson AJ, Daley GQ, Zon LI, Gilliland DG and Huntly BJ. Cdx4 dysregulates Hox gene expression and generates acute myeloid leukemia alone and in cooperation with Meis1a in a murine model. Proc Natl Acad Sci U S A 2006; 103: 16924-16929.
[33] Fernandes P, Waldron N, Chatzilygeroudi T, Naji NS and Karantanos T. Acute erythroid leukemia: from molecular biology to clinical outcomes. Int J Mol Sci 2024; 25: 6256.
[34] Yang W, Zhao X, Han Y, Duan L, Lu X, Wang X, Zhang Y, Zhou W, Liu J, Zhang H, Zhao Q, Hong L and Fan D. Identification of hub genes and therapeutic drugs in esophageal squamous cell carcinoma based on integrated bioinformatics strategy. Cancer Cell Int 2019; 19: 142.

[35] Ma S, Lu CC, Yang LY, Wang JJ, Wang BS, Cai HQ, Hao JJ, Xu X, Cai Y, Zhang Y and Wang MR. ANXA2 promotes esophageal cancer progression by activating MYC-HIF1A-VEGF axis. J Exp Clin Cancer Res 2018; 37: 183.

[36] Pan Y, Liu J, Gao Y, Guo Y, Wang C, Liang Z, Wu M, Qian Y, Li Y, Shen J, Lu C and Ma S. FBXW7 loss of function promotes esophageal squamous cell carcinoma progression via elevating MAP4 and ERK phosphorylation. J Exp Clin Cancer Res 2023; 42: 75.

[37] Coltrini D, Ronca R, Belleri M, Zardi L, Indraccolo S, Scarlato V, Giavazzi R and Presta M. Impact of VEGF-dependent tumour micro-environment on EDB fibronectin expression by subcutaneous human tumour xenografts in nude mice. J Pathol 2009; 219: 455-462.

[38] Ku GY, Bains MS, Park DJ, Janjigian YY, Rusch VW, Rizk NP, Yoon SS, Millang B, Capanu M, Goodman KA and Ilson DH. Phase II study of bevacizumab and preoperative chemoradiation for esophageal adenocarcinoma. J Gastrointest Oncol 2016; 7: 828-837.

[39] Black SA Jr, Palamakumbura AH, Stan M and Trackman PC. Tissue-specific mechanisms for CCN2/CTGF persistence in fibrotic gingiva: interactions between cAMP and MAPK signaling pathways, and prostaglandin E2-EP3 receptor mediated activation of the c-JUN N-terminal kinase. J Biol Chem 2007; 282: 15416-15429.

[40] Gifford CC, Tang J, Costello A, Khakoo NS, Nguyen TQ, Goldschmeding R, Higgins PJ and Samarakoon R. Negative regulators of TGF-β1 signaling in renal fibrosis; pathological mechanisms and novel therapeutic opportunities. Clin Sci (Lond) 2021; 135: 275-303.

[41] Isshiki T, Naiel S, Vierhout M, Otsubo K, Ali P, Tsubouchi K, Yazdanshenas P, Kumaran V, Dvorkin-Gheva A, Kolb MRJ and Ask K. Therapeutic strategies to target connective tissue growth factor in fibrotic lung diseases. Pharmacol Ther 2024; 253: 108578.

[42] Rayego-Mateos S, Marquez-Exposito L, Basantes P, Tejedor-Santamaria L, Sanz AB, Nguyen TQ, Goldschmeding R, Ortiz A and Ruiz-Ortega M. CCN2 activates RIPK3, NLRP3 inflammasome, and NRF2/oxidative pathways linked to kidney inflammation. Antioxidants (Basel) 2023; 12: 1541.

[43] Lambi AG, DeSante RJ, Patel PR, Hilliard BA, Popoff SN and Barbe MF. Blocking CCN2 reduces established palmar neuromuscular fibrosis and improves function following repetitive overuse injury. Int J Mol Sci 2023; 24: 13866.

[44] Leask A. Breathe, breathe in the air: the anti-CCN2 antibody pamrevlumab (FG-3019) completes a successful phase II clinical trial for idiopathic pulmonary fibrosis. J Cell Commun Signal 2019; 13: 441-442.

[45] O'Brien MM, Ji L, Shah NN, Rheingold SR, Bhojwani D, Yuan CM, Xu X, Yi JS, Harris AC, Brown PA, Borowitz MJ, Militano O, Kairalla J, Devidas M, Raetz EA, Gore L and Loh ML. Phase II trial of inotuzumab ozogamicin in children and adolescents with relapsed or refractory B-cell acute lymphoblastic leukemia: children's oncology group protocol AALL1621. J Clin Oncol 2022; 40: 956-967.

[46] Rubinstein JD and O'Brien MM. Inotuzumab ozogamicin in B-cell precursor acute lymphoblastic leukemia: efficacy, toxicity, and practical considerations. Front Immunol 2023; 14: 1237738.

[47] Thota S and Advani A. Inotuzumab ozogamicin in relapsed B-cell acute lymphoblastic leukemia. Eur J Haematol 2017; 98: 425-434.

[48] Yurkiewicz IR, Muffly L and Liedtke M. Inotuzumab ozogamicin: a CD22 mAb-drug conjugate for adult relapsed or refractory B-cell precursor acute lymphoblastic leukemia. Drug Des Devel Ther 2018; 12: 2293-2300.

[49] Rogers JE and Ajani JA. Perspectives on the pharmacological management of esophageal cancer: where are we now and where do we need to go? Expert Opin Pharmacother 2022; 23: 1893-1902.

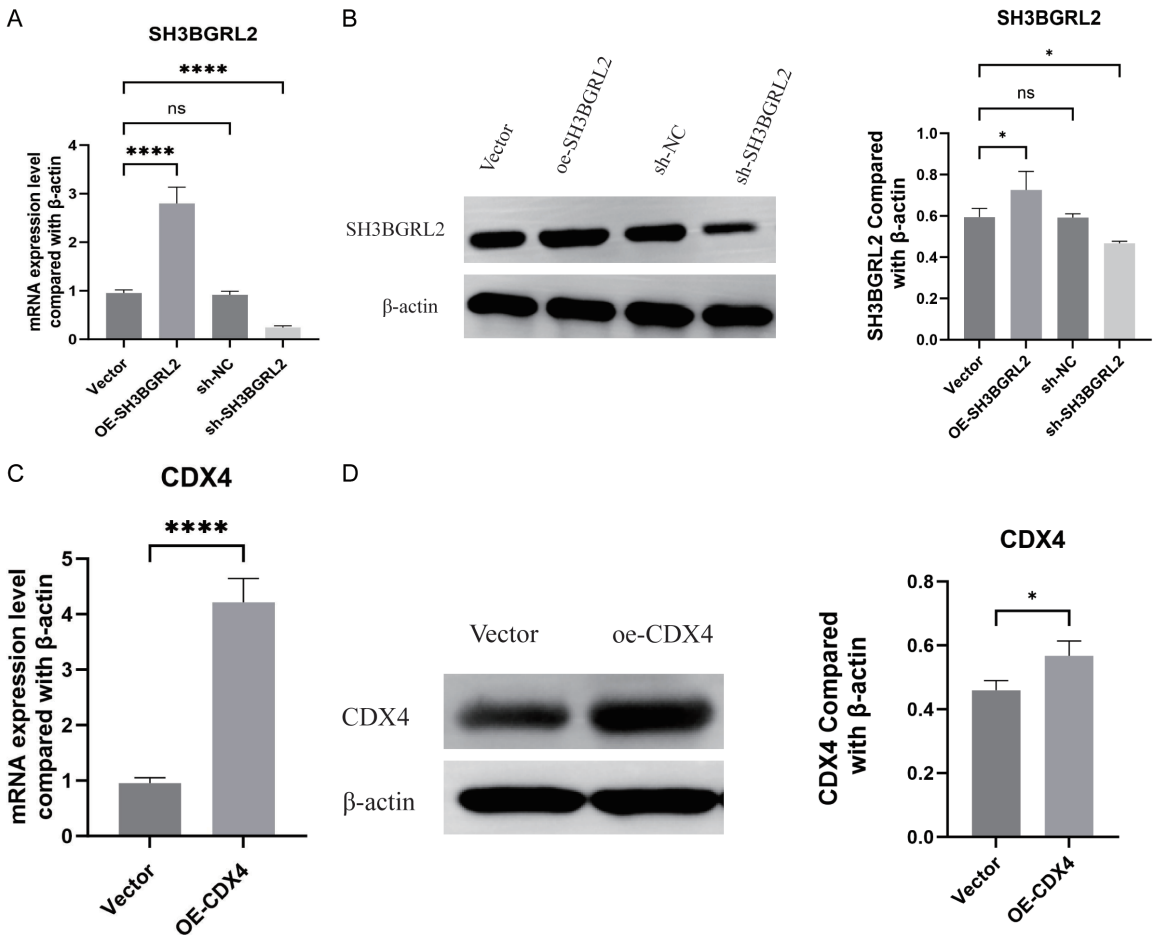
[50] Zhou D, Wu X, Liu X, He S, Ni J, Chen B and Mu D. The pharmacological mechanism of β-elemene in the treatment of esophageal cancer revealed by network pharmacology and experimental verification. Sci Rep 2023; 13: 12160.

[51] Chen H, Luo H, Tian T, Li S and Jiang Y. Integrated analyses of single-cell transcriptome and Mendelian randomization reveal the protective role of resistin in sepsis survival in intensive care unit. Int J Mol Sci 2023; 24: 14982.

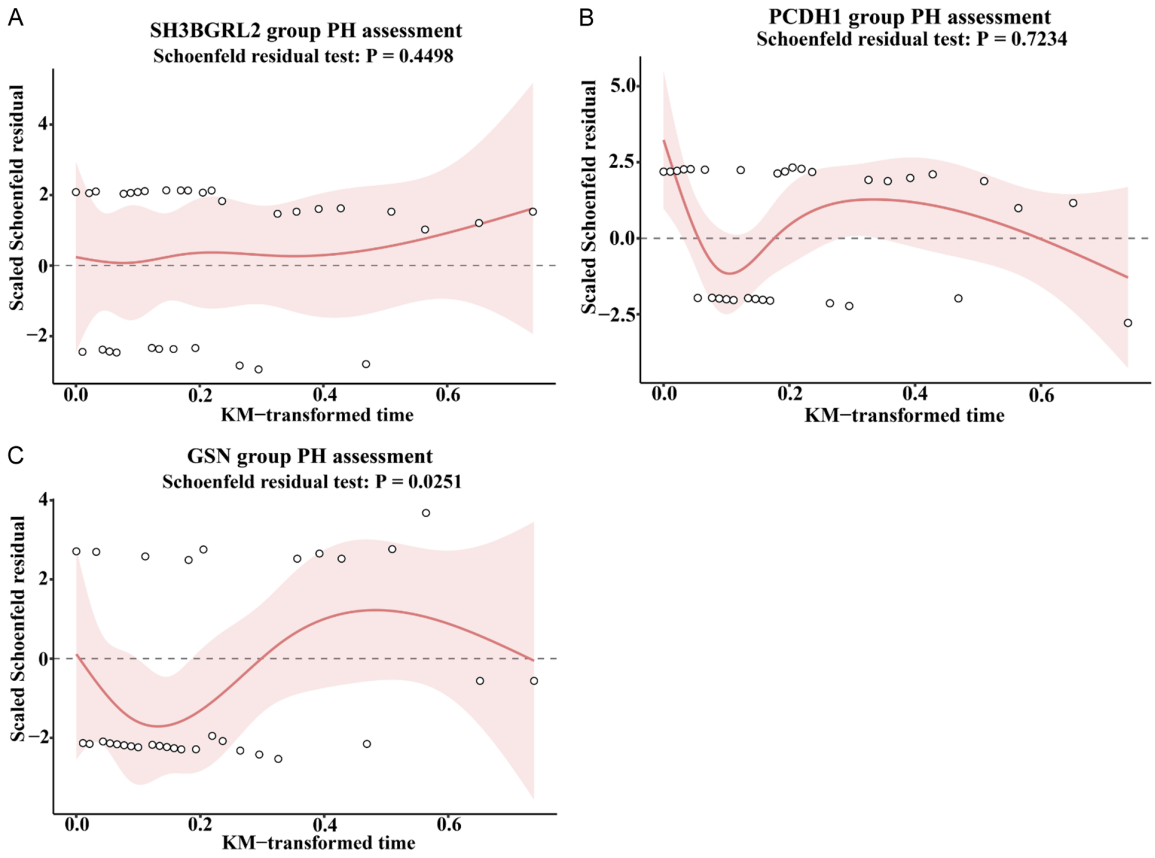
[52] Cui HK, Tang CJ, Gao Y, Li ZA, Zhang J and Li YD. An integrative analysis of single-cell and bulk transcriptome and bidirectional mendelian randomization analysis identified C1Q as a novel stimulated risk gene for Atherosclerosis. Front Immunol 2023; 14: 1289223.

[53] Yang S, Guo J, Kong Z, Deng M, Da J, Lin X, Peng S, Fu J, Luo T, Ma J, Yin H, Liu L, Liu J, Zha Y, Tan Y and Zhang J. Causal effects of gut microbiota on sepsis and sepsis-related death: insights from genome-wide Mendelian randomization, single-cell RNA, bulk RNA sequencing, and network pharmacology. J Transl Med 2024; 22: 10.

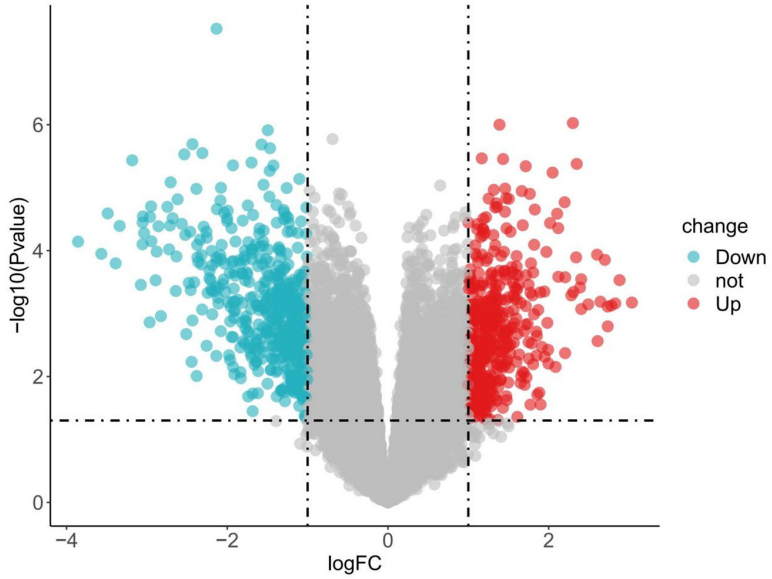
[54] Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, Lu Q, Rowson A, Taylor TRP, Clarke L, Maccora K, Chen C, Cook AL, Ye CJ, Fairfax KA, Hewitt AW and Powell JE. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science 2022; 376: eabf3041.



Supplementary Figure 1. Validation of SH3BGRL2 and CDX4 expression changes at 72 h after transfection. A. Quantitative reverse transcription PCR (qRT-PCR) analysis of SH3BGRL2 mRNA expression in the vector, SH3BGRL2-overexpression, non-targeting shRNA control (sh-NC), and SH3BGRL2-knockdown groups. B. Representative western blots and corresponding quantification of SH3BGRL2 protein expression in the same four groups. C. qRT-PCR analysis of CDX4 mRNA expression in the vector and CDX4-overexpression groups. D. Representative western blots and corresponding quantification of CDX4 protein expression in these two groups. β-actin served as the reference for mRNA normalization and as the loading control for western blot analysis. Brackets indicate the groups compared. ns, not significant; *P < 0.05; ****P < 0.0001. OE, overexpression; shRNA, short hairpin RNA.



Supplementary Figure 2. Assessment of the proportional hazards assumption for gene-expression groups. Plots show scaled Schoenfeld residuals for SH3BGRL2 (A), PCDH1 (B), and GSN (C). The corresponding proportional hazards test *P* values were 0.4498, 0.7234, and 0.0251, respectively. The horizontal axis represents Kaplan-Meier-transformed time.



Supplementary Figure 3. Differential expression volcanic map of prognostic model.