Article

Discovery Medicine 2026; 38(209): 1883–1899
https://doi.org/10.24976/Discov.Med.202638209.147

Single-Cell eQTL-Guided Mendelian Randomization Reveals a Dual-Axis Immune Evasion Mechanism and Therapeutic Targets in Esophageal Cancer

Yi Li[1,†], Ziqi Wan[2,†], Qi Zhao[1], Hongwei Wang[3], Yi Ba[1,*], Lingjuan Jiang[4,*]

[1]Department of Cancer Medical Center, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100730 Beijing, China
[2]Department of Radiation Oncology, Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100021 Beijing, China
[3]Department of Gastroenterology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100730 Beijing, China
[4]Center for Biomarker Discovery and Validation, National Infrastructures for Translational Medicine, Institute of Clinical Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100730 Beijing, China
*Correspondence: bayipumch@aliyun.com (Yi Ba); jianglingjuan@pumch.cn (Lingjuan Jiang)
[†]These authors contributed equally.
Submitted: 13 April 2026 Revised: 8 May 2026 Accepted: 21 May 2026 Published: 20 June 2026

Background: The tumor microenvironment (TME) drives esophageal cancer (EC) progression and immune evasion, yet immune cell-type-specific genetic regulatory candidates associated with EC susceptibility remain incompletely characterized. This study aimed to prioritize immune cell-specific genetic regulatory candidates associated with EC susceptibility.
Methods: We integrated single-cell expression quantitative trait loci (sc-eQTL) data with East Asian EC genome-wide association studies (GWAS; n = 160,589) to perform Mendelian randomization (MR). MR-prioritized candidates were further evaluated using tumor single-cell expression profiling, *in silico* perturbation analysis, The Cancer Genome Atlas Program (TCGA) transcriptomic correlations, phenome-wide association studies (PheWAS), and drug-repositioning analysis.
Results: MR prioritized 18 gene–cell type pairs representing 16 immune cell-associated candidate genes (*F*-statistics >10), all of which remained significant within the prioritized candidate set after Benjamini–Hochberg correction. These findings supported a novel putative dual-axis immune evasion model. The *TNFSF13B*-associated inflammatory component showed monocyte-enriched tumor expression and transcriptomic associations with myeloid-derived suppressor cell (MDSC) markers *CD33* (*R* = 0.63) and *ITGAM* (*R* = 0.47). Simultaneously, *SLC25A29* showed CD8 effector memory T/NK-context MR signals but adipocyte-enriched tumor expression, suggesting a stromal metabolic–immune crosstalk hypothesis. *In silico* perturbation further linked *SLC25A29* to CD8+ T cell effector-related programs, including *GZMK*. PheWAS did not identify evident severe pleiotropic safety signals for these hubs. Drug repositioning nominated *TNFSF13B*-directed agents, including belimumab, as testable repurposing candidates requiring further validation.
Conclusions: This sc-eQTL-guided framework prioritizes immune cell-associated candidate genes and generates testable hypotheses linking germline genetic regulation, tumor single-cell expression localization, and TME remodeling in EC.

Keywords: esophageal cancer; tumor microenvironment; Mendelian randomization; single-cell eQTL; drug repositioning

Introduction

Esophageal cancer (EC) is one of the most malignant cancers of the gastrointestinal system, with over 511,000 new cases documented annually worldwide [1]. Despite combined-modality therapy, including surgical resection, chemotherapy, and radiotherapy, the overall five-year survival rate remains bleak at 15–20% [2]. The tumor microenvironment (TME) is critical for EC progression, immune evasion, and therapeutic resistance [3], a complex interaction driven by a variety of environmental factors [4]. This complex ecosystem is composed of various immune cell populations that facilitate tumor progression through immunosuppressive mechanisms, including the accumulation of pro-tumor macrophages and exhausted T cells [5].

Current EC research utilizes bulk RNA sequencing, which averages gene expression signals across heterogeneous cell populations and masks important cell-type-specific transcriptional programs. While single-cell RNA sequencing has revolutionized our ability to analyze TME at the cellular level [6], differential expression analyses based on these data are inherently prone to confounding and reverse causality. Therefore, focusing therapeutic targets solely on transcriptional changes carries a significant risk of translational failure, hindering the identification of

Copyright: © 2026 The Author(s). Published by Discovery Medicine. This is an open access article under the CC BY 4.0 license.
Note: *Discovery Medicine* stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

regulatory networks controlling immune cell function in the TME and the prioritization of causally validated therapeutic targets.

Mendelian randomization (MR) uses genetic variants as instrumental variables to infer causal relationships between exposures and outcomes, an approach that is less prone to confounding bias and comparable to randomized controlled trials [7]. In addition, the recent combination of single-cell expression quantitative trait loci (sc-eQTL) mapping and large-scale genome-wide association studies (GWAS) provides an opportunity to identify the genetic variants that modulate gene expression in a cell-type-specific manner [8,9]. This multi-omics approach allows researchers to identify causal genes acting in specific immune cell lineages, bridging genetic susceptibility and microenvironmental remodeling with single-cell resolution. However, it is still challenging to identify which cell-type-specific genetic variants causally contribute to EC development.

Given the significant knowledge gap concerning immune cell-specific causal genes in EC pathogenesis, it is therapeutically essential to uncover the genetic drivers within particular immune cell lineages. Therefore, we systematically integrated sc-eQTL data with large-scale GWAS for a comprehensive MR analysis. We further validated those findings with phenome-wide association studies (PheWAS) to evaluate the pleiotropic effects and systemic safety across various disease phenotypes [10]. Additionally, we employed multi-dimensional validation integrating The Cancer Genome Atlas Program (TCGA) clinical correlations and virtual knockout approaches to functionally characterize prioritized targets and elucidate their mechanistic roles in EC development [11]. This study aims to prioritize immune cell-associated candidate genes for EC and generate testable hypotheses linking germline genetic regulation, TME remodeling, and potential drug-repurposing opportunities, thereby providing a framework for future functional validation studies.

Materials and Methods

Study Design and Data Acquisition

We employed an integrative multi-omics study strategy that amalgamated tumor scRNA-seq data, immune cell-specific sc-eQTL data, and large-scale GWAS summary statistics. The scRNA-seq datasets were obtained from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database under accession number GSE196756, which includes six EC samples (three tumor tissues and three neighboring normal tissues). Single-cell libraries were constructed with the 10× Genomics Chromium platform and sequenced on an Illumina NovaSeq 6000 system. The summary statistics for the EC GWAS outcome were obtained from the IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/, Accession ID: ebi-a-GCST90018621), which includes 160,589 East Asian participants [12].

Single-Cell RNA Sequencing Data Processing and Clustering

We analyzed raw count matrices using the Seurat package (v4.0) in R (v4.2). Rigorous quality control (QC) standards were implemented: cells with less than 50 identified features (nFeature_RNA <50) and those demonstrating a mitochondrial gene expression ratio of 15% or greater (percent.mt ≥ 15) were omitted. Filtered gene expression data were normalized by the LogNormalize method (scale factor = 10,000), and then the top 1500 highly variable genes were identified. To correct potential confounding batch effects across samples, the *Harmony* algorithm was applied using default parameters. A K-nearest neighbor (KNN) graph was constructed by using the first 20 principal components (PCs), and cells were clustered using the FindClusters function with a resolution of 0.3. Results of high-dimensional clustering were visualized using t-distributed stochastic neighbor embedding (t-SNE).

Cell Type Identification and Differential Expression Analysis

Cell types were identified using the *SingleR* package (v2.0) with several robust reference datasets, including Blueprint Encode, Human Primary Cell Atlas, Monaco Immune Data, and Novershtern Hematopoietic Data. Automated annotations were manually curated to define final cell lineages, including but not limited to CD8+ T cells, CD4+ T cells, B cells, monocytes, fibroblasts, epithelial cells, and keratinocytes. Differentially expressed genes (DEGs) defining cell populations were identified using Wilcoxon rank-sum test (FindMarkers). Statistical significance was defined by an adjusted $p < 0.05$ and an absolute $\log_2$(Fold Change) >1. Visualization was executed using the *scRNAtoolVis* R package (v0.1.0).

Cell-Cell Communication Network Analysis

Cell-cell communication in EC TME was analyzed by the *CellChat* R package (v2.1.0) [13]. Normalized scRNA-seq expression matrices were used to calculate the probabilities of communication among defined cellular subpopulations. To ensure robust signaling inference and to minimize false-positive ligand-receptor pairs, we used *filterCommunication* to remove interactions detected in fewer than 10 cells.

Selection of Instrumental Variables for sc-eQTLs

To test the putative causality at a single-cell resolution, we extracted cell-type-specific cis-eQTLs from the OneK1K database (https://onek1k.org/) as instrumental variables. These blood-derived sc-eQTLs were used as immune-cell regulatory proxies; however, their applicability to tumor-infiltrating immune cells remains a limitation. Candidate instrumental variables were selected with a sc-eQTL significance threshold of $p < 5 \times 10^{-6}$ to keep enough cis-regulatory instruments in cell-type-

specific eQTL-based MR. To address weak-instrument bias, we checked the strength of the instrument using F-statistics. We performed a sensitivity analysis using a stricter sc-eQTL threshold of $p < 5 \times 10^{-8}$, where we had sufficient instruments. Strict linkage disequilibrium (LD) clumping ($R^2 < 0.01$, window size = 10,000 kb) was conducted using the *ieugwasr* R package (v1.1.0) to ensure the independence of the instrumental variables.

Two-Sample MR Analysis

The causal effect of immune cell-specific eGenes on EC risk was evaluated using the TwoSampleMR/MR-Base framework [14]. The MR design was based on three core assumptions: genetic instruments should be strongly associated with immune cell-specific gene expression, independent of potential confounders, and influence EC risk only through the corresponding gene expression. To address these assumptions, we selected cell-type-specific cis-eQTLs as instruments, evaluated instrument strength using F-statistics, applied stringent LD clumping, and conducted sensitivity analyses to explore potential pleiotropy and heterogeneity [15].

Following data harmonization and the exclusion of SNPs directly associated with the outcome ($p < 0.05$), causal estimates were calculated using the inverse variance weighted (IVW) method for multiple SNPs and the Wald ratio method for single SNPs. Horizontal pleiotropy was evaluated via the MR-Egger intercept test [16]; Non-significant MR-Egger intercept ($p > 0.05$) indicated no evident pleiotropic bias. Candidate gene–cell type pairs were initially prioritized at nominal $p < 0.05$. To address multiple testing among the reported nominal MR associations, Benjamini–Hochberg FDR correction was applied across the 18 reported gene–cell type pairs, with BH-FDR <0.05 considered significant within the prioritized candidate set.

Population-Sensitivity Analyses for Ancestry Mismatch

To address the ancestry mismatch between European-derived sc-eQTL instruments and the East Asian EC GWAS outcome dataset, additional population-sensitivity analyses were performed. For each harmonized SNP, we calculated the absolute effect allele frequency (EAF) difference between the exposure and outcome datasets. MR analyses were repeated after excluding SNPs with EAF differences >0.10 and >0.20. To account for ancestry-specific LD structure, instrumental variables were further re-clumped using the 1000 Genomes East Asian reference panel under the same clumping parameters as the primary analysis ($R^2 < 0.01$ within a 10,000-kb window). MR estimates were recalculated after East Asian LD re-clumping and after the combined application of East Asian LD re-clumping and EAF-difference filtering. Robustness was further evaluated using Cochran's Q test, MR-Egger intercept, leave-one-out analysis, weighted median, weighted mode, and MR-PRESSO, where sufficient SNPs were available. For single-SNP instruments, heterogeneity, MR-Egger intercept, and MR-PRESSO analyses were not applicable.

Validation of Causal Genes Using TCGA Data

To evaluate the clinical relevance of prioritized causal genes, RNA-sequencing data from the TCGA-ESCA cohort (https://portal.gdc.cancer.gov/), comprising EC and adjacent normal tissues, were analyzed. The expression matrices were strictly pre-processed by *limma* R package (v3.68.2). To ensure the consistency of the data, the gene symbols were harmonized, and the duplicate probe entries were averaged with the *avereps* function. The Wilcoxon rank-sum test was used to evaluate differences in expression levels between the tumor and normal sample groups, and statistical significance was defined as $p < 0.05$. Lastly, expression differences were visualized as boxplots with overlaid jitter points using the *ggpubr* R package (v0.6.3).

Functional Enrichment Analysis

To explore the biological functions of genes dysregulated after target perturbation, we performed pathway enrichment (Gene Ontology [GO] and Kyoto Encyclopedia of Genes and Genomes [KEGG]) for significantly dysregulated genes identified from virtual knockout experiments (adjusted $p < 0.05$). The enrichment statistical significance was defined by an adjusted $p < 0.05$ with the *clusterProfiler* R package (v4.20.0).

PheWAS

To systematically evaluate the pleiotropic effects and systemic safety profile of the prioritized causal targets, we conducted a PheWAS across a broad spectrum of continuous and binary clinical traits, utilizing the AZPheWAS web platform (https://azphewas.com/) [17]. Targets significantly correlated with major adverse complications were excluded at this step.

Virtual Single-Cell Knockout and Drug Repurposing

To elucidate the downstream mechanisms of the key causal genes, we performed virtual single-cell knockout experiments *in silico* using the *scTenifoldKnk* R package (v1.0.3) [18]. To reflect genome-wide transcriptional perturbations, gene regulatory networks (scGRNs) were simulated with 10 subnetworks (nc_nNet = 10) and 500 cells per network (nc_nCells = 500). Genes with an adjusted $p <$ 0.05 were considered significantly differentially regulated and retained for downstream functional enrichment analysis. Finally, we integrated our findings with respected pharmacogenomic databases (Open Targets [https://www.opentargets.org/], DrugBank [https://go.drugbank.com/], and DGIdb [https://www.dgidb.org/]) to identify safe antagonist or inhibitor candidates for genomics-guided drug repurposing.

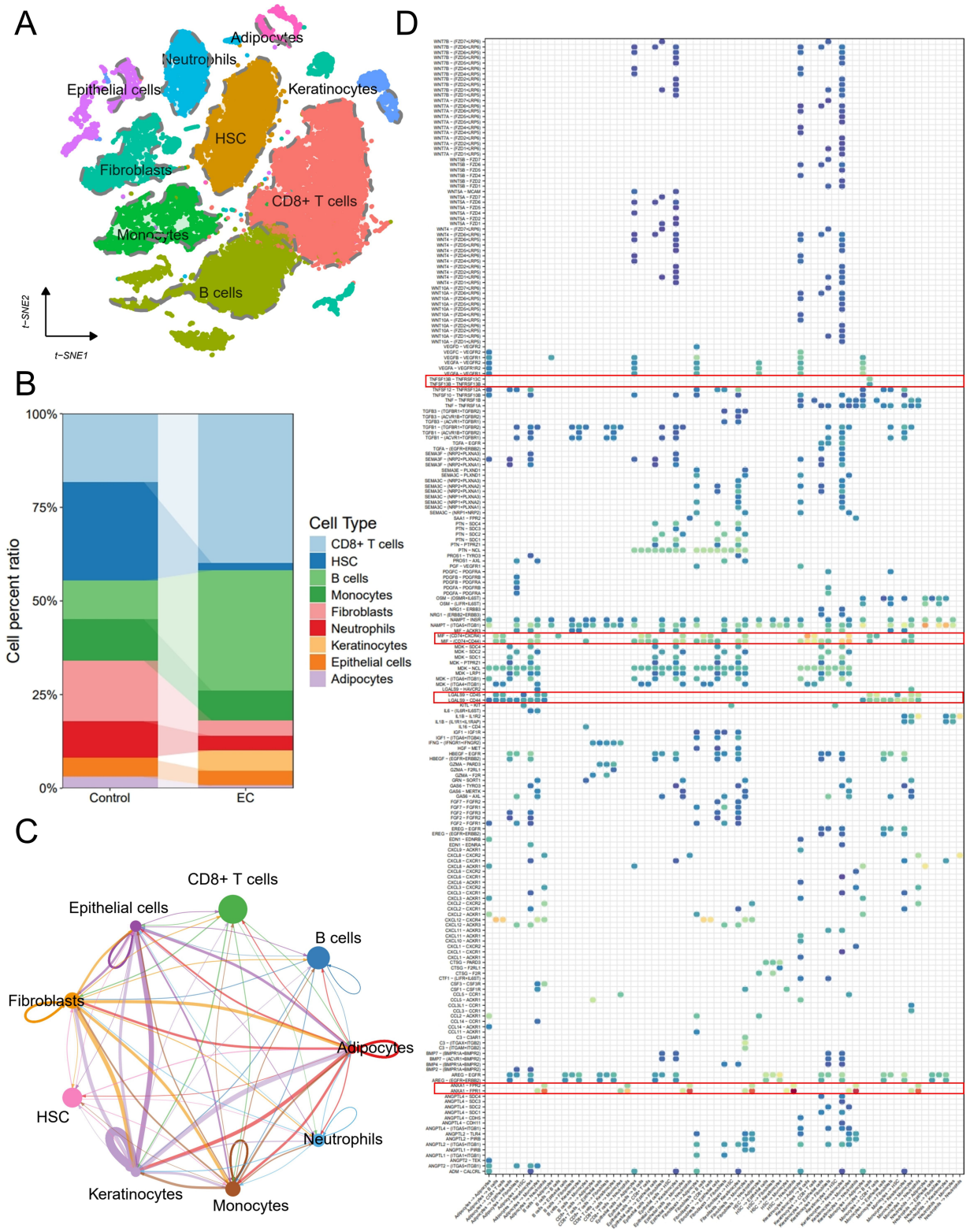


Fig. 1. Single-cell transcriptomic landscape and cell-cell communication in the EC microenvironment. (A) t-SNE visualization of 21,847 high-quality cells of six EC samples, colored by cell type annotations. (B) Stacked bar plot comparing the proportions of the cell types between adjacent normal (Control) and EC tissues. (C) CellChat-derived communication network displaying the interaction counts and strengths among different cell types. The main signaling hubs are monocytes and CD8+ T cells, and the thickness of the lines indicates the strength of interaction. (D) Bubble plot of enriched ligand-receptor signaling pathways. Dot size indicates communication probability, and color intensity indicates *p*-value significance. Boxed regions indicate representative immune-related ligand-receptor interactions described in the text. EC, esophageal cancer; t-SNE, t-distributed stochastic neighbor embedding.

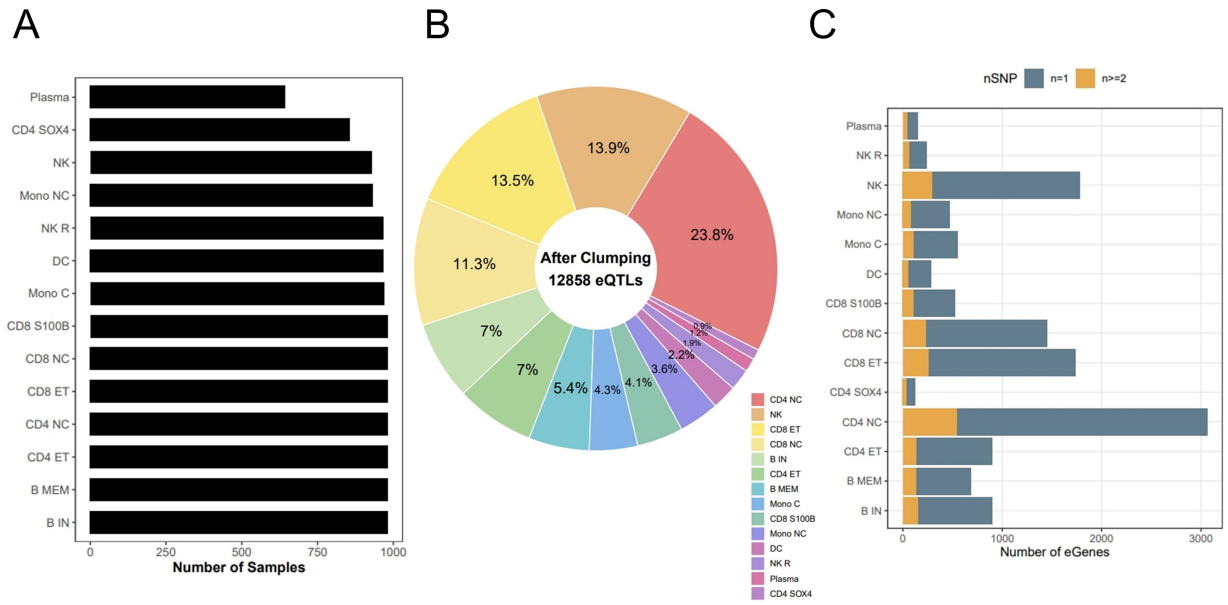


Fig. 2. Characteristics of immune cell-specific eQTL instrumental variables. (A) Bar plot showing sample sizes for each of the 14 immune cell types from the OneK1K cohort. (B) Donut chart illustrating the percentage distribution of 8733 eGenes across immune cell lineages following LD clumping. CD4 NC harbors the largest proportion (23.8%). (C) Stacked bar plot showing the distribution of SNP counts per eGene across cell types. The majority of eGenes are represented by a single SNP (n = 1). eQTL, expression quantitative trait loci; LD, Linkage disequilibrium; CD4 NC, CD4+ naive/central memory; SNP, single nucleotide polymorphism.

Results

Single-Cell Transcriptomic Atlas and Cell-Cell Communication Landscape of the EC Microenvironment

After rigorous quality control and normalization, 21,847 qualified cells were retained for further analysis to characterize the TME (**Supplementary Fig. 1**). Post batch-effect correction, unsupervised clustering and t-SNE projection revealed nine different cell populations, including CD8+ T cells, B cells, monocytes, fibroblasts, epithelial cells, adipocytes, hematopoietic stem cells (HSC), neutrophils, and keratinocytes (Fig. 1A).

We compared the cellular composition between tumor and adjacent normal tissues and found apparent shifts in composition. In particular, the proportions of key immune cell populations, including CD8+ T cells and monocytes, were higher in tumor tissues, whereas structural epithelial cells were lower (Fig. 1B). These descriptive differences suggest an active and ongoing anti-tumor immune surveillance with an expansion of the monocytic compartment that could be explained by the possible recruitment of monocytic myeloid-derived suppressor cells (M-MDSCs) contributing synergistically to immune evasion.

Ligand-receptor interactions analysis showed broad cell-cell communication with monocytes and CD8+ T cells as major signaling hubs (Fig. 1C). In particular, CellChat analysis revealed high probabilities of communication for representative immune-related ligand-receptor pairs, such as annexin A1 (*ANXA1*)-, macrophage migration inhibitory factor (*MIF*)-, galectin 9 (*LGALS9*)- and TNF superfamily member 13b (*TNFSF13B*)- related interactions. By permutation testing, these specific signaling networks were significant ($p < 0.05$) in the interactions between antigen-presenting cells (monocytes, B cells) and T cells (Fig. 1D). The abundance of these cell surface interactions suggests an important role for direct cell-cell contact and antigen presentation in the development of anti-tumor immunity in EC. Additionally, the topological centrality of monocytes and CD8+ T cells as major signaling hubs strongly suggests their potential roles as key regulators of tumor progression. The rationale for these analyses was to assess whether these highly communicative immune populations also represent major cellular targets of genetic risk variants in our subsequent MR analysis.

Characteristics of Immune Cell-Specific eQTL Instrumental Variables

To provide a solid statistical basis for causal inference, we rigorously assessed the properties of the extracted cis-eQTLs across 14 different immune cell types. In total, we discovered 12,858 independent cis-eQTLs, representing 8733 unique eGenes across all immune lineages. The expression datasets had sufficient statistical power, with sample sizes ranging from 643 cells (plasma cells) to 982 cells (B cells and CD8+ T cell subsets) (Fig. 2A).

We observed substantial biological heterogeneity in the distribution of eGenes across immune cell types. The largest fraction of eGenes was present in CD4+

naive/central memory T cells (CD4 NC: 23.8%), followed by NK cells (13.9%) and CD8+ effector T cells (13.5%) (Fig. 2B). Furthermore, analysis of the instrumental variables' composition showed that most eGenes (68.4%) were driven by one highly potent SNP, 24.7% were driven by two independent SNPs, and 6.9% were driven by three or more SNPs (Fig. 2C). This pattern of distribution demonstrates the classical genetic makeup of immune-specific gene regulation and is helpful in improving the robustness of subsequent MR analyses by reducing weak instrument bias. Importantly, the high fraction of eGenes identified in CD8+ T cells and other key immune populations is in striking concordance with their prominent roles as signaling hubs that we discovered in cell-cell communication analysis, providing a genetic basis for their phenotypic prevalence in the TME.

MR Prioritizes Immune-Specific Causal Genes for EC

In the sc-eQTL-guided MR analysis, all prioritized instruments showed adequate instrument strength (*F*-statistics ranging from 23.16 to 370.66, all significantly >10), and Cochran's Q tests showed no significant heterogeneity among analyzable multi-SNP instruments (all $p >$ 0.05).

After this stringent filtering ($p < 0.05$, pleiotropy $p \geq$ 0.05, and consistent effect direction), we successfully identified 18 unique gene-cell type pairs representing 16 distinct genes significantly associated with EC risk (Table 1). Following Benjamini-Hochberg FDR correction, all 18 nominally significant associations remained statistically significant (BH-adjusted q values of 0.04983 before rounding; **Supplementary Table 1**), supporting the robustness of the prioritized gene–cell type pairs.

To evaluate whether the sc-eQTL instrument threshold influenced the MR findings, we repeated the analysis using a more stringent sc-eQTL threshold of $p < 5 \times 10^{-8}$. Under this threshold, 17 of the 18 reported gene–cell type pairs retained at least one instrument, although many were reduced to single-SNP models. Several associations remained nominally significant, whereas others showed directionally consistent but attenuated estimates. These findings suggest that the main effect directions were broadly preserved under a stricter instrument threshold, although reduced instrument numbers limited statistical power (**Supplementary Table 2**).

For the prioritized causal eGenes, a distinct cluster of molecules involved in MHC-II was observed with *HLA-DQA2* in naive/immature B cells (OR = 1.109, 95% CI: 1.013–1.215, $p = 0.024$) and NK-recruiting cells (OR = 1.225, 95% CI: 1.021–1.470, $p = 0.029$), *HLA-DQA1* in CD8 effector memory T cells (OR = 1.162, 95% CI: 1.011–1.335, $p = 0.035$), and *HLA-DRB5* in CD4 SOX4 T cells (OR = 0.717, 95% CI: 0.539–0.954, $p = 0.022$). In addition to the canonical antigen-presenting machinery, our framework also prioritized several highly novel and therapeutically actionable targets, including *TNFSF13B* in CD4 naive/central memory T cells (OR = 1.273, 95% CI: 1.001–1.619, $p = 0.049$), *SLC25A29* in both CD8 effector memory T cells (OR = 1.169, 95% CI: 1.006–1.358, $p = 0.041$) and NK cells (OR = 1.171, 95% CI: 1.006–1.364, $p = 0.042$), *AGER* in dendritic cells (OR = 0.749, 95% CI: 0.582–0.963, $p = 0.024$), and *WARS* in naive/immature B cells (OR = 0.826, 95% CI: 0.696–0.979, $p = 0.027$) (Fig. 3A).

We used a Manhattan plot to visualize the genome-wide distribution of these eQTL associations (Fig. 3B). Furthermore, the circos plot revealed that CD8 effector memory T cells and NK cells harbored multiple EC-associated genetic regulatory candidates, highlighting these lineages as key mediators of immune-related genetic susceptibility in EC (Fig. 3C). Initial sensitivity analyses showed no evidence of significant heterogeneity or directional horizontal pleiotropy among analyzable multi-SNP instruments.

Population-Sensitivity Analyses Addressing Ancestry-Related Allele Frequency and LD Differences

Among the 47 harmonized SNP instruments supporting the 18 significant gene–cell type pairs, the median absolute EAF difference between the exposure and outcome datasets was 0.117. Specifically, 26 SNPs showed an EAF difference >0.10, 11 showed an EAF difference >0.20, and 3 showed an EAF difference >0.30, indicating non-negligible ancestry-related allele frequency differences and the need for further sensitivity evaluation (**Supplementary Table 3**).

After excluding SNPs with EAF differences >0.20, all 18 gene–cell type pairs retained at least one instrumental variable, although seven were reduced to single-SNP models (**Supplementary Table 4**). Therefore, heterogeneity tests, MR-Egger intercept analysis, and MR-PRESSO were not applicable to these single-SNP models. Under the EAF $\leq$ 0.20 threshold, several associations retained consistent effect directions compared with the original MR results. Notably, *SLC25A29* in NK cells remained significant and directionally stable, while *SLC25A29* in CD8 effector memory T cells showed a consistent but borderline association. *WARS* in naive/immature B cells and *FBXO2* in CD8 effector memory T cells also remained stable after EAF filtering (**Supplementary Table 5**).

To further account for ancestry-specific LD structure, we re-clumped the instruments using the 1000 Genomes East Asian reference panel. The major effect directions were broadly preserved after East Asian LD re-clumping and after the combined analysis applying both East Asian LD re-clumping and EAF $\leq$ 0.20 filtering (**Supplementary Fig. 2**). In particular, *SLC25A29* in NK cells, *FBXO2* in CD8 effector memory T cells, and *WARS* in naive/immature B cells showed consistent estimates across sensitivity analyses.

Table 1. Causal effects of immune cell-specific eGenes on EC risk and corresponding instrumental variable diagnostics.

Gene	Cell Type	nsnp	OR (95% CI)	IVW *p*-value	*F*-statistic	Cochran's Q *p*-value	MR-Egger Intercept *p*-value
HLA-DRB5	CD4 SOX4 T cell	2	0.717 (0.539–0.954)	0.022	36.46	0.911	NA
AGER	Dendritic Cell	2	0.749 (0.582–0.963)	0.024	47.70	0.970	NA
HLA-DQA2	Naïve/Immature B Cell	6	1.109 (1.013–1.215)	0.024	303.88	0.586	0.355
EIF2A	Natural Killer Cell	2	1.516 (1.055–2.180)	0.025	30.58	0.492	NA
WARS	Naïve/Immature B Cell	2	0.826 (0.696–0.979)	0.027	158.28	0.502	NA
HLA-DQA2	Natural Killer Recruiting Cell	3	1.225 (1.021–1.470)	0.029	71.44	0.738	0.902
TMEM258	CD4 Effector memory/TEMRA	2	0.730 (0.551–0.968)	0.029	39.21	0.834	NA
ZBTB9	CD8 S100B T cell	3	0.773 (0.610–0.979)	0.033	57.56	0.864	0.714
MAST4	CD4 Naïve/Central memory T cell	2	1.533 (1.034–2.273)	0.033	34.04	0.718	NA
FBXO2	CD8 Effector memory	2	1.796 (1.046–3.086)	0.034	30.76	0.881	NA
HLA-DQA1	CD8 Effector memory	4	1.162 (1.011–1.335)	0.035	105.21	0.453	0.844
PARP14	CD8 Naïve/Central memory T cell	2	0.653 (0.436–0.979)	0.039	23.16	0.837	NA
SLC25A29	CD8 Effector memory	2	1.169 (1.006–1.358)	0.041	370.66	0.966	NA
SLC25A29	Natural Killer Cell	2	1.171 (1.006–1.364)	0.042	365.40	0.993	NA
POLR1D	CD4 Naïve/Central memory T cell	2	0.796 (0.635–0.999)	0.049	298.87	0.759	NA
TNFSF13B	CD4 Naïve/Central memory T cell	2	1.273 (1.001–1.619)	0.049	62.07	0.819	NA
APOM	Dendritic Cell	5	1.235 (1.000–1.525)	0.050	43.21	0.709	0.999
EBPL	Naïve/Immature B Cell	2	0.723 (0.523–1.000)	0.050	43.66	0.485	NA

OR, odds ratio; CI, confidence interval; IVW, inverse variance weighted; NA, not applicable.

1. *F*-statistic: The instrumental variable strength was quantified using the *F*-statistic, calculated as the mean of individual SNP *F*-statistics ($F = \beta^2/SE^2$). All prioritized targets exhibited an *F*-statistic substantially greater than 10, strictly ruling out the possibility of weak instrument bias.

2. Cochran's Q *p*-value: This statistic assesses the heterogeneity of instrumental variables applied in the IVW approach. A *p*-value > 0.05 indicates no significant heterogeneity.

3. MR-Egger Intercept *p*-value: This test is used to detect directional horizontal pleiotropy. A *p*-value > 0.05 suggests no significant pleiotropic effects. Values reported as NA refer to causal models comprising exactly two single nucleotide polymorphisms (SNPs, nsnp = 2). The MR-Egger intercept test is mathematically impossible for these particular targets due to insufficient degrees of freedom. However, the robust *F*-statistics and absence of heterogeneity (Cochran's Q $p > 0.05$) strongly support the validity and robustness of their causal IVW estimates.

4. Gene symbols are defined as follows: *HLA-DRB5*, major histocompatibility complex, class II, DR beta 5; *AGER*, advanced glycosylation end-product specific receptor; *HLA-DQA2*, major histocompatibility complex, class II, DQ alpha 2; *EIF2A*, eukaryotic translation initiation factor 2A; *WARS*, tryptophanyl-tRNA synthetase 1; *TMEM258*, transmembrane protein 258; *ZBTB9*, zinc finger and BTB domain containing 9; *MAST4*, microtubule-associated serine/threonine kinase family member 4; *FBXO2*, F-box protein 2; *HLA-DQA1*, major histocompatibility complex, class II, DQ alpha 1; *PARP14,* poly(ADP-ribose) polymerase family member 14; *SLC25A29*, solute carrier family 25 member 29; *POLR1D*, RNA polymerase I and III subunit D; *TNFSF13B*, TNF superfamily member 13b; *APOM*, apolipoprotein M; *EBPL*, emopamil binding protein like.

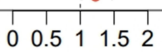
Among analyzable multi-SNP instruments, Cochran's Q, MR-Egger intercept, MR-PRESSO, and leave-one-out analyses did not reveal significant heterogeneity, directional horizontal pleiotropy, outlier-driven effects, or clear single-SNP-driven patterns (**Supplementary Table 6**). Overall, these analyses suggest that the nominal MR-prioritized associations were not solely driven by ancestry-related EAF or LD differences, although several associations should be interpreted as exploratory due to attenuation after filtering, limited instrument numbers, and lack of global FDR significance.

Single-Cell Expression Profiling and Cellular Localization of Core Causal Targets

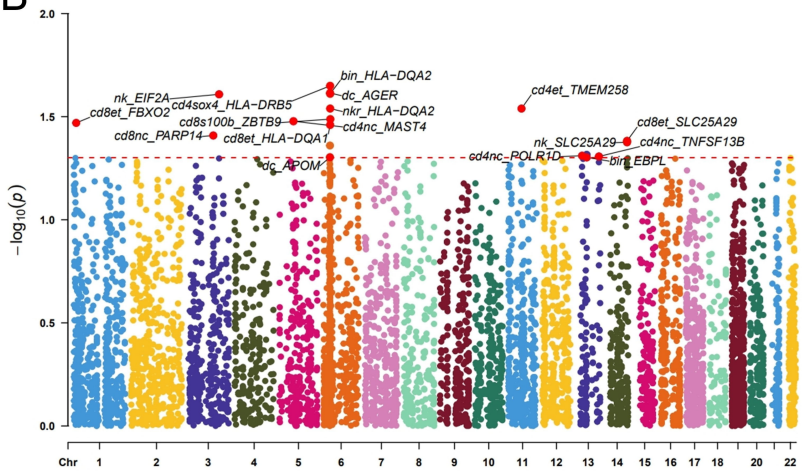
While MR analysis prioritized the genetic causality of these targets, their actual transcriptional activity within the complex TME required further single-cell validation. Among the 16 prioritized causal genes, members of the Human Leukocyte Antigen (HLA) family, specifically *HLA-DRB5* (OR = 0.717, $p = 0.022$) and *HLA-DQA2* (OR = 1.109, $p = 0.024$), exhibit particular robust nominal associations. Although HLA family genes showed relatively strong nominal MR associations, subsequent downstream characterization focused on non-HLA candidates with clearer interpretability and translational tractability.

A

id	Cell	Gene	nsnp	method	pval	OR(95% CI)
bin	Naïve/Immature B Cell	HLA−DQA2	6	IVW	**0.024**	1.109 (1.013 to 1.215)
bin	Naïve/Immature B Cell	WARS	2	IVW	**0.027**	0.826 (0.696 to 0.979)
bin	Naïve/Immature B Cell	EBPL	2	IVW	0.050	0.723 (0.523 to 1.000)
cd4et	CD4 Effector memory/TEMRA	TMEM258	2	IVW	**0.029**	0.730 (0.551 to 0.968)
cd4nc	CD4 Naive/Central memory T cell	MAST4	2	IVW	**0.033**	1.533 (1.034 to 2.273)
cd4nc	CD4 Naive/Central memory T cell	POLR1D	2	IVW	**0.049**	0.796 (0.635 to 0.999)
cd4nc	CD4 Naive/Central memory T cell	TNFSF13B	2	IVW	**0.049**	1.273 (1.001 to 1.619)
cd4sox4	CD4 SOX4 T cell	HLA−DRB5	2	IVW	**0.022**	0.717 (0.539 to 0.954)
cd8et	CD8 Effector memory	FBXO2	2	IVW	**0.034**	1.796 (1.046 to 3.086)
cd8et	CD8 Effector memory	HLA−DQA1	4	IVW	**0.035**	1.162 (1.011 to 1.335)
cd8et	CD8 Effector memory	SLC25A29	2	IVW	**0.041**	1.169 (1.006 to 1.358)
cd8nc	CD8 Naive/Central memory T cell	PARP14	2	IVW	**0.039**	0.653 (0.436 to 0.979)
cd8s100b	CD8 S100B T cell	ZBTB9	3	IVW	**0.033**	0.773 (0.610 to 0.979)
dc	Dendritic Cell	AGER	2	IVW	**0.024**	0.749 (0.582 to 0.963)
dc	Dendritic Cell	APOM	5	IVW	0.050	1.235 (1.000 to 1.525)
nk	Natural Killer Cell	EIF2A	2	IVW	**0.025**	1.516 (1.055 to 2.180)
nk	Natural Killer Cell	SLC25A29	2	IVW	**0.042**	1.171 (1.006 to 1.364)
nkr	Natural Killer Recruiting Cell	HLA−DQA2	3	IVW	**0.029**	1.225 (1.021 to 1.470)



B



C

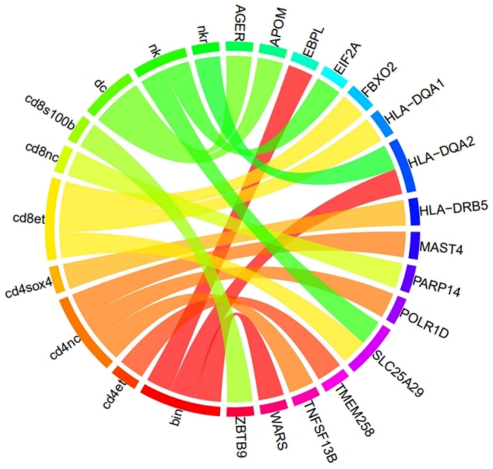


Fig. 3. MR identifies immune-specific causal genes for EC. (A) Summary table and corresponding forest plot summarizing the MR results for 18 significant gene-cell type pairs, including ORs, 95% CIs, and *p*-values. Red dots indicate risk-associated genes (OR >1), and blue dots indicate protective genes (OR <1). (B) Manhattan plot visualizing the genome-wide distribution of eQTL associations with EC risk. Significant loci are explicitly annotated. The dashed red line indicates the genome-wide significance threshold. (C) Circos plot illustrating the relationships between the 18 causal genes (right ring) and their corresponding immune cell origins (left ring). Ribbon width represents the strength of association. MR, Mendelian randomization; OR, odds ratio; CI, confidence interval.

Consequently, we targeted *SLC25A29* and *TNFSF13B* for further analysis as our top candidates. We selected these because they are druggable at the molecular level: *TNFSF13B* has an FDA-approved inhibitor (belimumab [19]), and *SLC25A29* encodes a mitochondrial arginine/lysine transporter involved in cancer-related metabolic remodeling [20,21]. While the established roles of HLA family genes in antigen presentation are well-documented, *SLC25A29* and *TNFSF13B* serve as relatively uncharacterized upstream regulatory hubs in EC.

We first analyzed their spatial expression distribution in the annotated EC scRNA-seq atlas. As demonstrated in Fig. 4A,B, *SLC25A29* was mainly expressed in adipocytes, with relatively lower expression in other cell types. Interestingly, bulk-level analysis of the TCGA-ESCA cohort showed similar overall expression of *SLC25A29* in normal and tumor tissues (**Supplementary Fig. 3**), but its MR-identified causal signal and adipocyte-restricted expression suggest a specialized role within the stromal niche, which may be masked in bulk transcriptomic profiles. On the other hand, the clinical validation using the TCGA-ESCA cohort showed a significant difference in the expression of *TNFSF13B* (p = 0.0091, Fig. 4C), supporting its dysregulation during esophageal carcinogenesis. Consistent with its known role in B cell activation and innate immune regulation, *TNFSF13B* displayed predomi-

nant expression in the monocytic compartment at the single-cell resolution (Fig. 4D–F). Importantly, we considered the immune regulatory contexts inferred from MR and the localization of tumor single-cell expression as complementary but separate layers of evidence. Based on this cross-compartment mapping, we proposed a putative "dual-axis" immune evasion model involving the *TNFSF13B*-associated inflammatory axis and *SLC25A29*-related stromal metabolic–immune crosstalk. Notably, the MR signal of *SLC25A29* was extracted from CD8+ effector memory T cells and NK cells, and its major expression is in adipocytes in the EC single-cell atlas, suggesting that *SLC25A29* may represent a stromal niche interacting with genetically susceptible immune effector cells, motivating further *in silico* perturbation of this hub.

Functional Enrichment and Phenome-Wide Safety Assessment of Target Genes

PheWAS analysis of 17,361 binary traits and 1419 quantitative traits indicated that most of the causal genes, including *HLA-DQA2*, *AGER*, and *TNFSF13B*, showed no significant associations with severe adverse outcomes such as cardiovascular diseases, neurodegenerative disorders, or malignancies, thereby providing preliminary reassurance regarding systemic safety in the context of therapeutic targeting (Fig. 5A,B).

GO enrichment analysis of the differentially expressed genes following *in silico* virtual knockout of *SLC25A29* in adipocytes showed significant enrichment for terms related to antigen processing and presentation, MHC class II protein complex assembly, MHC protein complex-related terms, and peptide antigen binding (Fig. 5C). KEGG pathway analysis also showed enrichment of immune-related pathways, such as antigen processing and presentation, graft-versus-host disease, inflammatory bowel disease, and regulation of lipolysis in adipocytes (Fig. 5D). The compensatory upregulation of antigen processing pathways following metabolic disruption suggests that stromal *SLC25A29* might indirectly impact immune surveillance via lipid-metabolism-related stromal remodeling rather than direct MHC-II modulation.

Virtual Knockout Simulation and Drug Repositioning Potential

To computationally investigate the potential stromal-immune crosstalk driven by our spatial findings, we next simulated disruption of the *SLC25A29*-driven adipocyte niche. A targeted virtual deletion of *SLC25A29* specifically in adipocytes was performed using the *scTenifoldKnk* machine learning algorithm, which identified 187 differentially regulated genes. Such *in silico* perturbation resulted in the notable collapse of critical downstream regulatory nodes like *LEPR*, *LRRK2*, and *APOA1*, which are fundamentally involved in lipid metabolism as well as the structural remodeling of the TME (Fig. 6A,B).

In parallel, virtual knockout of *TNFSF13B* in the monocyte subpopulation revealed 142 differentially regulated genes. Interestingly, this targeted deletion led to the conspicuous perturbation of potent pro-inflammatory and myeloid-derived suppressor markers, including *AC245128.3*, *S100A9*, and *SLC25A37* (Fig. 6C,D). This specific transcriptomic change suggests that *TNFSF13B* acts to limit or dampen these potent local inflammatory responses in the TME, consistent with its biological function in regulating B cell survival and immune homeostasis.

Based on these findings, we systematically explored the therapeutic potential of targeting these causal genes by drug repositioning. Using authoritative pharmacogenomic databases, we identified several FDA-approved and investigational compounds targeting the prioritized genes (Table 2). Notably, Belimumab and other *TNFSF13B*-directed agents, as well as *AGER*-related compounds such as azeliragon, emerged as testable repurposing candidates for EC. HLA-targeting agents were not included in the current drug repurposing table, given the extreme polymorphism, complex LD structure, functional redundancy, and potential risk of autoimmunity associated with direct HLA modulation.

Collectively, this multi-omics framework not only identifies immune-specific causal genes in EC but also provides actionable drug repurposing opportunities, accelerating the translation of genomic discoveries into clinical applications.

Clinical Correlation and Cell-Intrinsic Functional Validation of the Proposed Dual-Axis Model

To validate the mechanisms proposed by our *in silico* models, we first applied the real-world bulk transcriptomic data of the TCGA-ESCA cohort. Expression of *TNFSF13B* was highly significantly positively correlated with classical MDSC markers in the monocytic inflammatory axis, such as *CD33* ($R = 0.63$, $p < 2.2 \times 10^{-16}$) and *ITGAM* ($R = 0.47$, $p = 1.2 \times 10^{-11}$) (Fig. 7A,B). The robust clinical correlation supports the notion that monocytic *TNFSF13B* overexpression is responsible for MDSC accumulation and innate immunosuppression.

Interestingly, *SLC25A29* showed a statistically significant but modest negative correlation with the immune checkpoint *CD274* (encoding *PD-L1*) of the TCGA-ESCA cohort ($R = -0.17$, $p = 0.02$; Fig. 7C) for the stromal metabolic-immune axis. Although this macroscopic clinical association is captured by TCGA data, it does not reflect cell-intrinsic functional defects. To address this limitation, a high-resolution virtual perturbation was further performed. *In silico* deletion of *SLC25A29* in CD8+ effector memory T cells, corresponding to the immune-cell regulatory context underlying its MR signal, selectively perturbed two effector-related nodes, *GZMK* and *NEU1*, without causing global transcriptional collapse (adjusted $p < 0.05$, Fig. 7D). The profound perturbation of *GZMK*, a

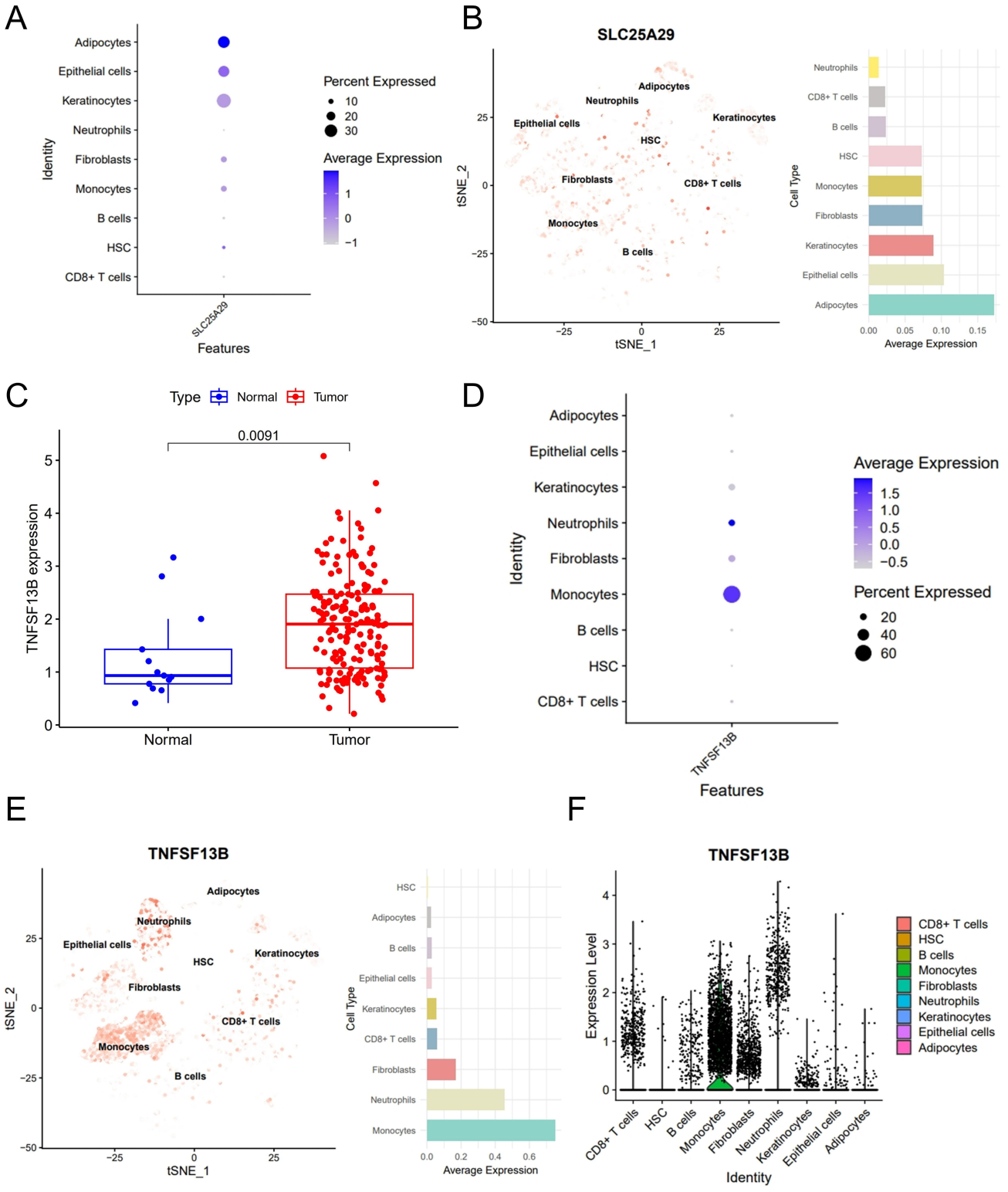


Fig. 4. Single-cell spatial expression profiling, cellular localization, and clinical validation of core causal targets. (A,B) *SLC25A29* single-cell transcriptomic expression. (A) Bubble plot of percent expressed and average expression levels. (B) The t-SNE feature plot and average expression bar plot illustrating the specific enrichment of *SLC25A29* in adipocytes. (C) Clinical validation of *TNFSF13B* in TCGA-ESCA bulk RNA-seq cohort with significant differential expression between normal esophageal tissues and tumor samples. (D–F) *TNFSF13B* single-cell transcriptomic expression. (D) Bubble plot, and (E) t-SNE feature plot along with an average expression bar plot. (F) The violin plot showing the specific enrichment of *TNFSF13B* mainly in monocytes.

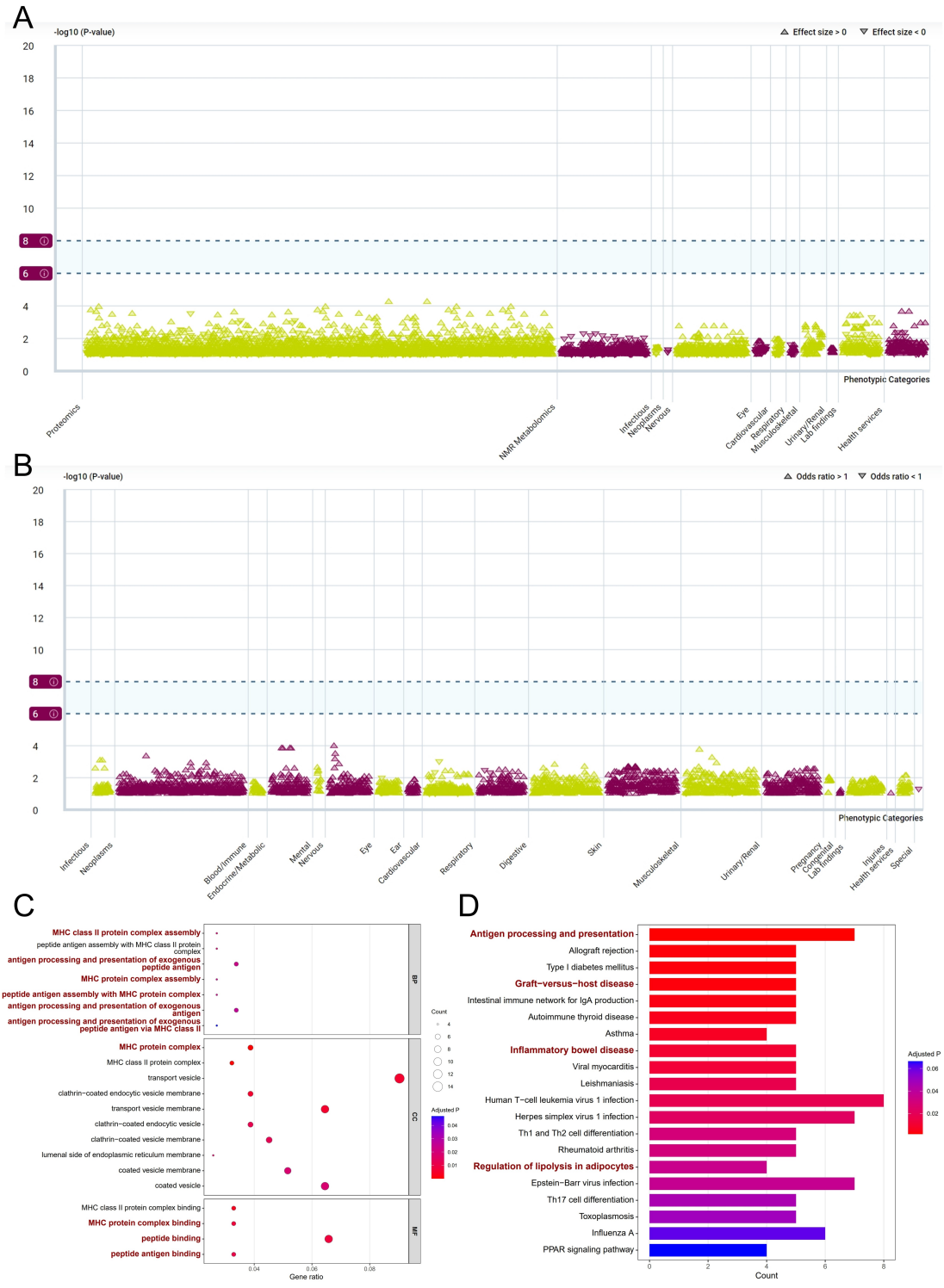


Fig. 5. Safety evaluation of PheWAS and functional enrichment of target genes. (A,B) PheWAS scatter plots to assess the pleiotropic effects and systemic safety of the prioritized causal genes across thousands of continuous (A) and binary (B) clinical phenotypes. Significance thresholds after Bonferroni correction are indicated by dashed lines. No significant associations with severe adverse outcomes were found. (C) Bar plot of significantly enriched GO terms for the DEGs identified after virtual knockout of *SLC25A29* in adipocytes. (D) The bar plot presents the KEGG pathways that were significantly enriched in the differentially expressed genes observed following the virtual knockout. Pathways discussed in the main text are highlighted in dark red bold font for clarity. PheWAS, phenome-wide association studies; GO, Gene Ontology; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Table 2. Drug repositioning candidates for non-HLA MR-prioritized causal genes.

Target Gene	Drug Name	Mechanism of Action	Primary Indication	Clinical Status	Source
TNFSF13B	Belimumab	Inhibitor (Antibody)	Systemic lupus erythematosus, Neuromyelitis optica	Approved	Open Targets, DrugBank
TNFSF13B	Tabalumab	Inhibitor (Antibody)	Multiple myeloma, Autoimmune diseases	Phase III	Open Targets
TNFSF13B	Atacicept	Inhibitor (Protein)	Lupus nephritis, Autoimmune diseases	Investigational	Open Targets
TNFSF13B	Blisibimod	Inhibitor	IgA nephropathy, SLE	Investigational	Open Targets
AGER	Azeliragon	RAGE Antagonist (Small molecule)	Alzheimer's disease, Glioblastoma multiforme	Failed Phase III in Alzheimer's disease due to lack of efficacy on cognitive and functional endpoints; exploratory candidate requiring EC-specific preclinical validation	Open Targets, DrugBank, DGIdb
AGER	Pyridoxamine	Inhibitor	Diabetic complications	Investigational	DrugBank, DGIdb
WARS	Tryptophan analogues	Target Inhibitor	Metabolic & nutritional interventions	Approved/Investigational	DrugBank

key cytotoxic effector, provides computational support for a possible link between *SLC25A29*-related regulation and CD8+ T cell effector programs.

Discussion

In this study, we integrated single-cell transcriptomic profiling, immune cell-specific sc-eQTL-guided MR [8], and systematic drug repurposing [22] to elucidate the mechanisms underlying EC immune evasion. Using large-scale East Asian EC GWAS data (160,589 samples) [23] and the OneK1K sc-eQTL resource [8], we identified 18 immune cell-specific causal gene–cell type pairs that correspond to 16 distinct genes. Our approach extends correlation-based single-cell studies by integrating immune cell-specific genetic regulation with EC GWAS data, providing a framework for single-cell-informed causal prioritization. The findings support a putative dual-axis immune evasion model in EC, comprising a *TNFSF13B*-associated inflammatory axis with monocyte-enriched tumor expression and an *SLC25A29*-mediated stromal metabolic-immune axis with adipocyte-enriched tumor expression.

Our work addresses three major limitations of traditional cancer immunogenetics. First, although many EC risk loci have been found by traditional GWAS, most of them are located in non-coding regions and cannot be attributed to particular cell types [23]. In bulk tissue eQTL-MR, averaging expression signals across heterogeneous populations obscures the cell-type-specific regulatory networks [9]. We prioritized genetic regulatory signals in circulating immune-cell contexts (including CD8+ effector memory T cells, NK cells, and monocytes) using immune cell-specific sc-eQTLs [8]. Because these eQTLs were derived from peripheral blood, not tumor-infiltrating lymphocytes, our results should be considered a projection of genetically controlled immune programs onto the EC microenvironment, rather than direct evidence of eQTL activity within tumor-infiltrating cells. Accordingly, for *TNFSF13B* and *SLC25A29*, the MR-inferred immune-cell contexts were interpreted together with tumor scRNA-seq expression localization as a cross-compartment mapping strategy, rather than as evidence that the genetic regulatory cell type and the dominant tumor-expressing cell type must be identical. Second, previous single-cell studies in EC have provided valuable atlases [24,25] but were inherently correlational; they cannot distinguish cause from consequence. Our MR analysis prioritizes *TNFSF13B* and *SLC25A29* as upstream regulatory candidates rather than downstream markers. Third, while the MHC region harbors the strongest EC associations, we strategically prioritized non-HLA targets (e.g., *TNFSF13B*, *SLC25A29*, *AGER*) for downstream biological characterization and drug-repurposing analysis. Although several HLA-region genes showed nominal MR associations, they were not advanced as therapeutic repurposing targets because HLA genes are highly polymorphic, central to antigen presentation, and located within a complex LD-rich region with substantial functional redundancy [26]. Thus, HLA-related findings were retained as genetic observations rather than actionable drug-repurposing candidates.

The dual-axis mechanism was first proposed in an independent CellChat-based analysis of cell-cell communication [27], where monocytes were identified as central signaling hubs in the EC microenvironment. *TNFSF13B* was convergently prioritized with the MR signal from CD4 naive/central memory T cell contexts in our follow-up sc-

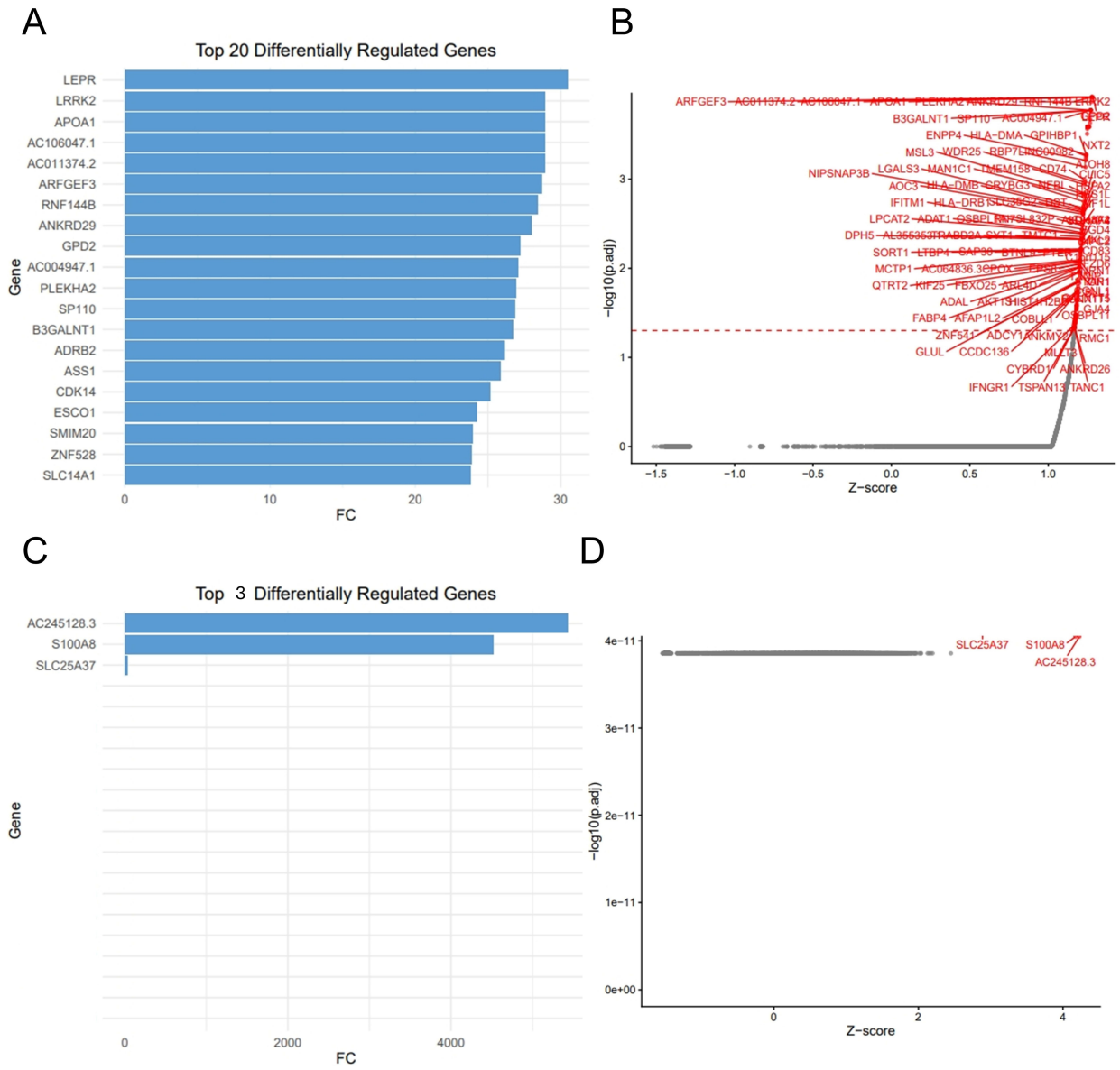


Fig. 6. ***In silico*** **virtual single-cell knockout simulation reveals downstream mechanisms.** (A,B) Differentially regulated genes following *SLC25A29* virtual knockout in adipocytes: (A) Bar plot showing the top 20 differentially expressed genes (including *LEPR*, *LRRK2*, *APOA1*), and (B) Volcano plot displaying global transcriptomic alterations. (C,D) Differentially regulated genes following *TNFSF13B* virtual knockout in monocytes: (C) Bar plot showing the top differentially expressed genes (including *AC245128.3*, *S100A9*, and *SLC25A37*), and (D) Volcano plot. Genes with adjusted $p < 0.05$ are considered significantly differentially regulated.

eQTL-guided MR analysis, whereas its tumor expression was enriched in monocytes. This convergence of phenotypic network analysis and MR-based prioritization raises the possibility that genetic regulatory signals may converge on highly communicative immune nodes in the EC microenvironment. *TNFSF13B* is conventionally a B-cell survival factor [28], but single-cell spatial profiling revealed that *TNFSF13B* is mainly expressed in monocytes and not B cells in the EC TME. *In silico* virtual knockout of monocyte-enriched *TNFSF13B* resulted in marked regulation of *S100A9*, a hallmark of MDSC activation [29]. These computational predictions were further supported by TCGA-ESCA correlation analysis, which demonstrated robust positive associations between *TNFSF13B* and classical MDSC markers *CD33* and *ITGAM*. Taken together, these results are consistent with a *TNFSF13B*-associated paracrine inflammatory model where monocyte-enriched *TNFSF13B* might contribute to MDSC-related immunosuppressive programs.

In parallel, we proposed a stromal metabolic-immune axis centered on SLC25A29, a mitochondrial carrier for arginine and lysine transport [30] whose role in tumor immunology was previously unknown. MR indicated that the *SLC25A29* genetic regulatory signal was derived from

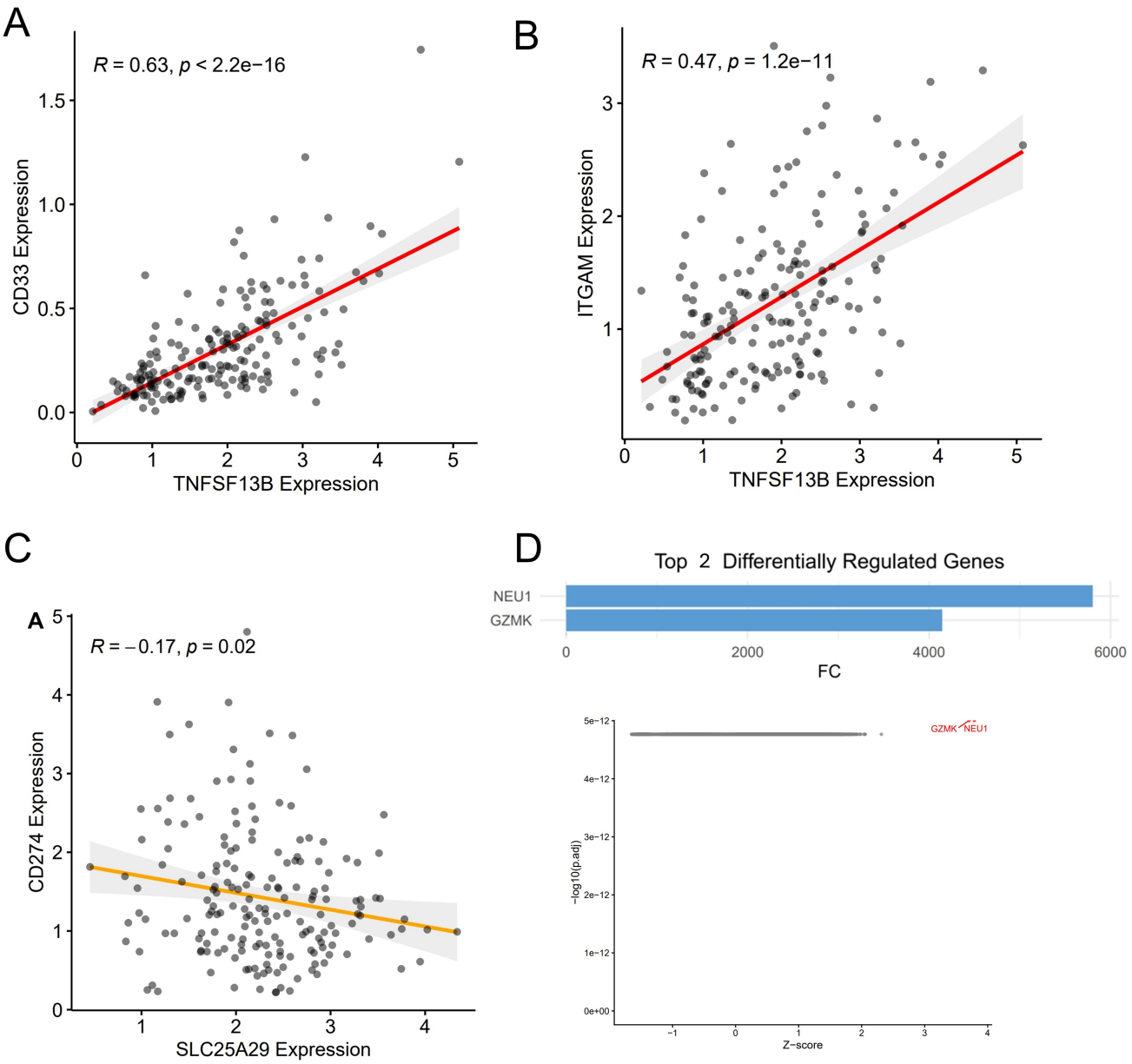


Fig. 7. Clinical correlation and cell-intrinsic functional validation of the proposed dual-axis immune evasion model. (A,B) Validation of the monocytic inflammatory axis using the TCGA-ESCA cohort. (C) Clinical correlation of the stromal metabolic-immune axis. (D) Cell-intrinsic functional validation via high-resolution virtual perturbation.

circulating CD8+ T and NK cells. Paradoxically, single-cell expression localized *SLC25A29* almost exclusively to adipocytes within the EC microenvironment. Rather than a contradiction, we propose that this reflects trans-cellular metabolic crosstalk: adipocyte-enriched *SLC25A29* may shape a nutrient-deprived stromal niche that suppresses the cytotoxic capacity of genetically susceptible infiltrating T cells. Virtual knockout of *SLC25A29* in CD8+ effector memory T cells specifically affected *GZMK*, a core cytotoxic effector, without global transcriptional collapse. Simultaneously, adipocyte-specific *SLC25A29* deletion perturbed lipid metabolic regulators (*LEPR*, *LRRK2*, *APOA1*) and induced compensatory enrichment of antigen processing pathways. Interestingly, *SLC25A29* and *CD274* (PD-L1) were slightly negatively correlated in TCGA analysis. Given the small effect size, this finding should be regarded as an exploratory transcriptomic observation. Further functional studies are needed to determine whether this relationship is a PD-L1-dependent or PD-L1-independent immune regulatory feature [31]. Based on the adipocyte-enriched expression of *SLC25A29* and its computational link to CD8+ T cell effector programs, we tentatively proposed a "metabolic exhaustion" hypothesis, where adipocyte-associated lipid remodeling and nutrient competition may contribute to impairing both CD8+ T cell cytotoxicity and antigen presentation [32]. These findings identify *TN-*

FSF13B-related inflammatory programs and *SLC25A29*-related stromal metabolic features as two candidate components of the EC immune microenvironment.

Although the PD-1/PD-L1 pathway is the current focus of EC immunotherapy, durable clinical benefit remains limited in many patients. Objective response rates are approximately 47% to 58% for first-line chemoimmunotherapy regimens and only 11% to 25% for monotherapies in later-line settings [33,34]. Such clinical heterogeneity highlights the need to explore immune-regulatory programs in addition to canonical checkpoint signaling. To assess the translational potential of our identified targets, we performed PheWAS on thousands of binary and continuous clinical phenotypes [17,35]. Analysis showed that the prioritized causal genes (e.g., *TNFSF13B*, *AGER*, and *HLA-DQA2*) have no significant pleiotropic associations with severe cardiovascular, neurological, or autoimmune disorders, suggesting a favorable safety profile. Belimumab, an FDA-approved anti-*TNFSF13B* monoclonal antibody for systemic lupus erythematosus, has a decade of safety data [36] and may provide a pharmacological tool to test the *TNFSF13B*-associated inflammatory axis. Azeliragon, an oral AGER antagonist [37], previously failed in Phase III Alzheimer's disease trials after failing to meet cognitive and functional efficacy endpoints. Enthusiasm for direct clinical repurposing is limited, but the AD failure may not directly translate to EC due to distinct disease biology and clinical endpoints. Thus, azeliragon is presented here as an exploratory, testable RAGE-targeting candidate needing EC-specific preclinical validation. These repurposing candidates may provide rapid, lower-risk pathways to the clinic, as human genetic support doubles the rate of successful clinical translation [38]. We propose a hypothesis-generating framework for future testing, in which monocyte-enriched EC subtypes could be probed for *TNFSF13B*/*AGER*-related inflammatory features and potential interactions with PD-1/PD-L1 blockade in preclinical models.

Several limitations should be acknowledged. First, the sc-eQTL data derive primarily from European-ancestry individuals and peripheral blood immune cells, whereas the EC GWAS is East Asian and the scRNA-seq dataset represents tumor tissues; therefore, both ancestry mismatch and tissue-context differences between circulating and tumor-infiltrating immune cells may affect generalizability. Second, all functional predictions were computational and require experimental validation. For *SLC25A29*, the mismatch between CD8+ T/NK cell-contexts MR signal and adipocyte-enriched tumor expression indicates that the proposed stromal metabolic–immune crosstalk model remains hypothesis-generating. Third, our scRNA-seq dataset included only six EC patients, which may not capture the full inter-tumor heterogeneity. Fourth, although the reported nominal MR associations remained significant after BH-FDR correction within the prioritized candidate set, the screening-level nature of the MR framework and limited SNP numbers for several associations warrant cautious interpretation. Future studies should also focus on the development of ancestry-matched sc-eQTL resources, larger tumor single-cell cohorts and experimental validation using adipocyte-immune co-culture systems, tumor organoids, genetic perturbation models and preclinical evaluation of candidate therapeutic combinations.

Conclusions

Collectively, this study identified *TNFSF13B*-associated inflammatory programs and *SLC25A29*-related stromal metabolic features as two candidate components of the EC immune microenvironment. These findings link germline genetic regulation to the organization of the TME and provide a foundation for improving prioritization of immune–stromal targets in EC.

Abbreviations

CD4 NC, CD4+ naive/central memory; DEGs, Differentially expressed genes; EC, Esophageal cancer; GEO, Gene Expression Omnibus; GO, Gene Ontology; GWAS, Genome-wide association studies; HSC, Hematopoietic stem cells; IRB, Institutional review board; IVW, Inverse variance weighted; KEGG, Kyoto Encyclopedia of Genes and Genomes; KNN, K-nearest neighbor; LD, Linkage disequilibrium; MDSCs, Myeloid-derived suppressor cells; MR, Mendelian randomization; PCs, Principal components; PheWAS, Phenome-wide association studies; QC, Quality control; sc-eQTL, Single-cell expression quantitative trait loci; scGRNs, Gene regulatory networks; TME, Tumor microenvironment; t-SNE, t-distributed stochastic neighbor embedding.

Availability of Data and Materials

The datasets used or analyzed during the current study are available from the corresponding authors upon reasonable request.

Author Contributions

YL and ZW conducted the investigation, developed the methodology, curated and analyzed the data, and wrote the original draft. QZ contributed to data validation, and supported data curation and formal analysis. HW developed the software, prepared the visualizations, and assisted with the investigation. YB and LJ conceptualized the study, acquired funding, administered and supervised the project. All authors contributed to the important editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work to take public responsibility for appropriate portions of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This work was funded by the National High Level Hospital Clinical Research Funding (2025-PUMCH-A-198, 2025-PUMCH-A-101), and the Peking Union Medical College Hospital Talent Cultivation Program (Category D, UHB11899).

Conflict of Interest

The authors declare no conflict of interest.

Declaration of Generative AI and AI-Assisted Technologies in Manuscript Preparation

During the preparation and revision of the manuscript, the authors used AI-assisted language tools only to improve grammar, readability, and clarity. All AI-assisted outputs were checked, edited, and verified by the authors. The authors are responsible for the final content of the manuscript. No AI tool was used to generate original data, perform statistical analyses, create figures, or draw scientific conclusions.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.24976/Discov.Med.202638209.147.

References

[1] Qi L, Sun M, Liu W, Zhang X, Yu Y, Tian Z, *et al*. Global esophageal cancer epidemiology in 2022 and predictions for 2050: A comprehensive analysis and projections based on GLOBOCAN data. Chinese Medical Journal. 2024; 137: 3108–3116. https://doi.org/10.1097/CM9.0000000000003420.

[2] Zhang C, Chen L, Xiu Y, Zhang H, Zhang Y, Ying W. Burden of esophageal cancer in global, regional and national regions from 1990 to 2021 and its projection until 2050: results from the GBD study 2021. Frontiers in Oncology. 2025; 14: 1518567. https://doi.org/10.3389/fonc.2024.1518567.

[3] Chen X, Zhao Y, Wang Y, Wang X, Liu Y, Liu Z, *et al*. Single-cell atlas of the esophageal squamous cell carcinoma immune ecosystem to predict immunotherapy response. Signal Transduction and Targeted Therapy. 2025; 10: 348. https://doi.org/10.1038/s41392-025-02446-x.

[4] Abbas M, Tangney M. The oncobiome; what, so what, now what? Microbiome Research Reports. 2025; 4: 16. https://doi.org/10.20517/mrr.2024.89.

[5] Huang R, Kang T, Chen S. The role of tumor-associated macrophages in tumor immune evasion. Journal of Cancer Research and Clinical Oncology. 2024; 150: 238. https://doi.org/10.1007/s00432-024-05777-4.

[6] Yang Z, Tian H, Chen X, Li B, Bai G, Cai Q, *et al*. Single-cell sequencing reveals immune features of treatment response to neoadjuvant immunochemotherapy in esophageal squamous cell carcinoma. Nature Communications. 2024; 15: 9097. https://doi.org/10.1038/s41467-024-52977-0.

[7] Moodie EEM, le Cessie S. Instrumental Variables Analysis and Mendelian Randomization for Causal Inference. The Journal of Infectious Diseases. 2025; 231: 556–558. https://doi.org/10.1093/infdis/jiae357.

[8] Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, *et al*. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science (New York, N.Y.). 2022; 376: eabf3041. https://doi.org/10.1126/science.abf3041.

[9] Kang JB, Raveane A, Nathan A, Soranzo N, Raychaudhuri S. Methods and Insights from Single-Cell Expression Quantitative Trait Loci. Annual Review of Genomics and Human Genetics. 2023; 24: 277–303. https://doi.org/10.1146/annurev-genom-101422-100437.

[10] Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, *et al*. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010; 26: 1205–1210. https://doi.org/10.1093/bioinformatics/btq126.

[11] Wang B, Xu Y, Wan AH, Wan G, Wang QP. Integrating genome-wide CRISPR screens and in silico drug profiling for targeted antidote development. Nature Protocols. 2024; 19: 2739–2770. https://doi.org/10.1038/s41596-024-00995-z.

[12] Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, *et al*. A cross-population atlas of genetic associations for 220 human phenotypes. Nature Genetics. 2021; 53: 1415–1424. https://doi.org/10.1038/s41588-021-00931-x.

[13] Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. Nature Protocols. 2025; 20: 180–219. https://doi.org/10.1038/s41596-024-01045-4.

[14] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, *et al*. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018; 7: e34408. https://doi.org/10.7554/eLife.34408.

[15] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology (Cambridge, Mass.). 2017; 28: 30–42. https://doi.org/10.1097/EDE.0000000000000559.

[16] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015; 44: 512–525. https://doi.org/10.1093/ije/dyv080.

[17] Wang L, Zhang X, Meng X, Koskeridis F, Georgiou A, Yu L, *et al*. Methodology in phenome-wide association studies: a systematic review. Journal of Medical Genetics. 2021; 58: 720–728. https://doi.org/10.1136/jmedgenet-2021-107696.

[18] Osorio D, Zhong Y, Li G, Xu Q, Yang Y, Tian Y, *et al*. scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. Patterns (New York, N.Y.). 2022; 3: 100434. https://doi.org/10.1016/j.patter.2022.100434.

[19] Singh JA, Shah NP, Mudano AS. Belimumab for systemic lupus erythematosus. The Cochrane Database of Systematic Reviews. 2021; 2: CD010668. https://doi.org/10.1002/14651858.CD010668.pub2.

[20] Porcelli V, Fiermonte G, Longo A, Palmieri F. The human gene SLC25A29, of solute carrier family 25, encodes a mitochondrial transporter of basic amino acids. The Journal of Biological

Chemistry. 2014; 289: 13374–13384. https://doi.org/10.1074/jbc.M114.547448.

[21] Zhang H, Wang Q, Gu J, Yin L, Liang S, Wu L, *et al*. Elevated mitochondrial SLC25A29 in cancer modulates metabolic status by increasing mitochondria-derived nitric oxide. Oncogene. 2018; 37: 2545–2558. https://doi.org/10.1038/s41388-018-0139-x.

[22] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, *et al*. Drug repurposing: progress, challenges and recommendations. Nature Reviews. Drug Discovery. 2019; 18: 41–58. https://doi.org/10.1038/nrd.2018.168.

[23] Wu C, Hu Z, He Z, Jia W, Wang F, Zhou Y, *et al*. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nature Genetics. 2011; 43: 679–684. https://doi.org/10.1038/ng.849.

[24] Hu C, Li X, Liang W, Li S, Huang X, Chen J, *et al*. Deciphering the transcription factor-dominated ecosystem during esophageal squamous cell carcinoma progression at the single-cell level. International Journal of Molecular Sciences. 2026; 27: 4433. https://doi.org/10.3390/ijms27104433.

[25] Li X, Yuan J, Gao M, Liu J, Wang Q, Zhang Y, *et al*. Spatio-temporal proteomic landscape reveals early warning signals of esophageal squamous cell carcinoma progression. Advanced Science. 2026; 13: e14343. https://doi.org/10.1002/advs.202514343.

[26] Garrido-Mesa J, Brown MA. Antigen-driven T cell responses in rheumatic diseases: insights from T cell receptor repertoire studies. Nature Reviews. Rheumatology. 2025; 21: 157–173. https://doi.org/10.1038/s41584-025-01218-9.

[27] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, *et al*. Inference and analysis of cell-cell communication using CellChat. Nature Communications. 2021; 12: 1088. https://doi.org/10.1038/s41467-021-21246-9.

[28] Smulski CR, Eibel H. BAFF and BAFF-Receptor in B Cell Selection and Survival. Frontiers in Immunology. 2018; 9: 2285. https://doi.org/10.3389/fimmu.2018.02285.

[29] Veglia F, Sanseviero E, Gabrilovich DI. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. Nature Reviews. Immunology. 2021; 21: 485–498. https://doi.org/10.1038/s41577-020-00490-y.

[30] Palmieri F. The mitochondrial transporter family SLC25: identification, properties and physiopathology. Molecular Aspects of Medicine. 2013; 34: 465–484. https://doi.org/10.1016/j.mam.2012.05.005.

[31] McKendry RT, Spalluto CM, Burke H, Nicholas B, Cellura D, Al-Shamkhani A, *et al*. Dysregulation of Antiviral Function of CD8(+) T Cells in the Chronic Obstructive Pulmonary Disease Lung. Role of the PD-1-PD-L1 Axis. American Journal of Respiratory and Critical Care Medicine. 2016; 193: 642–651. https://doi.org/10.1164/rccm.201504-0782OC.

[32] Kiran S, Kumar V, Kumar S, Price RL, Singh UP. Adipocyte, Immune Cells, and miRNA Crosstalk: A Novel Regulator of Metabolic Dysfunction and Obesity. Cells. 2021; 10: 1004. https://doi.org/10.3390/cells10051004.

[33] Pu W, Li S, Zhang J, Huang J, Li J, Jiang Y, *et al*. The efficacy and safety of PD-1/PD-L1 inhibitors in combination with chemotherapy as a first-line treatment for unresectable, locally advanced, HER2-negative gastric or gastroesophageal junction cancer: a meta-analysis of randomized controlled trials. Frontiers in Immunology. 2025; 16: 1566939. https://doi.org/10.3389/fimmu.2025.1566939.

[34] Hong Y, Ding ZY. PD-1 Inhibitors in the Advanced Esophageal Cancer. Frontiers in Pharmacology. 2019; 10: 1418. https://doi.org/10.3389/fphar.2019.01418.

[35] Zhang X, Li X, He Y, Law PJ, Farrington SM, Campbell H, *et al*. Phenome-wide association study (PheWAS) of colorectal cancer risk SNP effects on health outcomes in UK Biobank. British Journal of Cancer. 2022; 126: 822–830. https://doi.org/10.1038/s41416-021-01655-9.

[36] Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzová D, *et al*. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. Arthritis and Rheumatism. 2011; 63: 3918–3930. https://doi.org/10.1002/art.30613.

[37] Burstein AH, Sabbagh M, Andrews R, Valcarce C, Dunn I, Altstiel L. Development of Azeliragon, an Oral Small Molecule Antagonist of the Receptor for Advanced Glycation Endproducts, for the Potential Slowing of Loss of Cognition in Mild Alzheimer's Disease. The Journal of Prevention of Alzheimer's Disease. 2018; 5: 149–154. https://doi.org/10.14283/jpad.2018.18.

[38] Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, *et al*. The support of human genetic evidence for approved drug indications. Nature Genetics. 2015; 47: 856–860. https://doi.org/10.1038/ng.3314.