

Supplementary data

A clinically relevant SLC2A1-associated malignant epithelial cell state predicts prognosis and immunotherapy response in lung adenocarcinoma

Linqian Song^{1,2}, Shiyun Xing^{1,2}, Peijie Li^{1,2}, Yunliang Cao^{1,2}, Hu Ma^{1,2}, *

¹Department of Oncology, The Second Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou 563000, P. R. China;

²Key Laboratory for Cancer Prevention and Treatment of Guizhou Province, Zunyi, Guizhou 563000, P. R. China.

* Corresponding author

Correspondence to: Hu Ma, Email: mahuab@163.com; ORCID: 0000-0003-1654-1576.

Supplemental Tables and Figures

List of Supplementary Figures

Supplementary Figure S1. Single-cell transcriptomic features of T cell subpopulations.

Supplementary Figure S2. Landscape features of myeloid cells.

Supplementary Figure S3. Distribution of epithelial subclusters across paired normal and LUAD samples.

Supplementary Figure S4. Additional characterization of Scissor⁺ epithelial cells relative to background epithelial cells.

Supplementary Figure S5. Stability and feature contribution analyses of prognostic genes in the LASSO–Cox model.

Supplementary Figure S6. Schematic model summarizing the proposed biological framework associated with the SLC2A1-related Scissor⁺ epithelial state.

List of Supplementary Tables

Supplementary Table S1. Canonical marker genes used for cell-type annotation in single-cell RNA sequencing analysis.

Supplementary Table S2. Sequences of siRNAs and primers used in this study.

Supplementary Table S3. Summary of datasets used in this study.

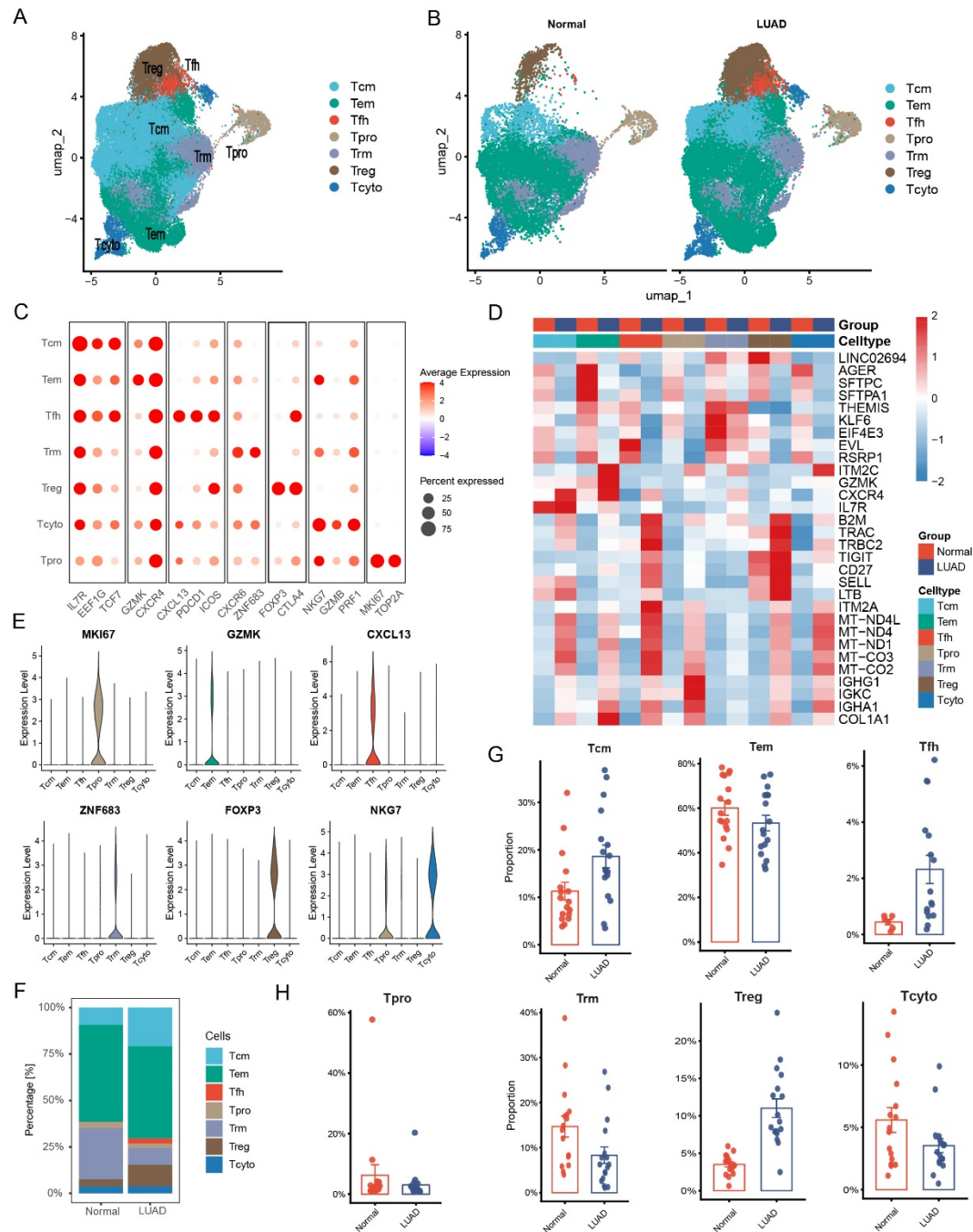
Supplementary Table S4. Numbers of major cell types identified in each LUAD and normal lung sample.

Supplementary Table S5. Distribution of epithelial subclusters in each LUAD and normal sample.

Supplementary Table S6. Clinical and transcriptomic datasets used in this study.

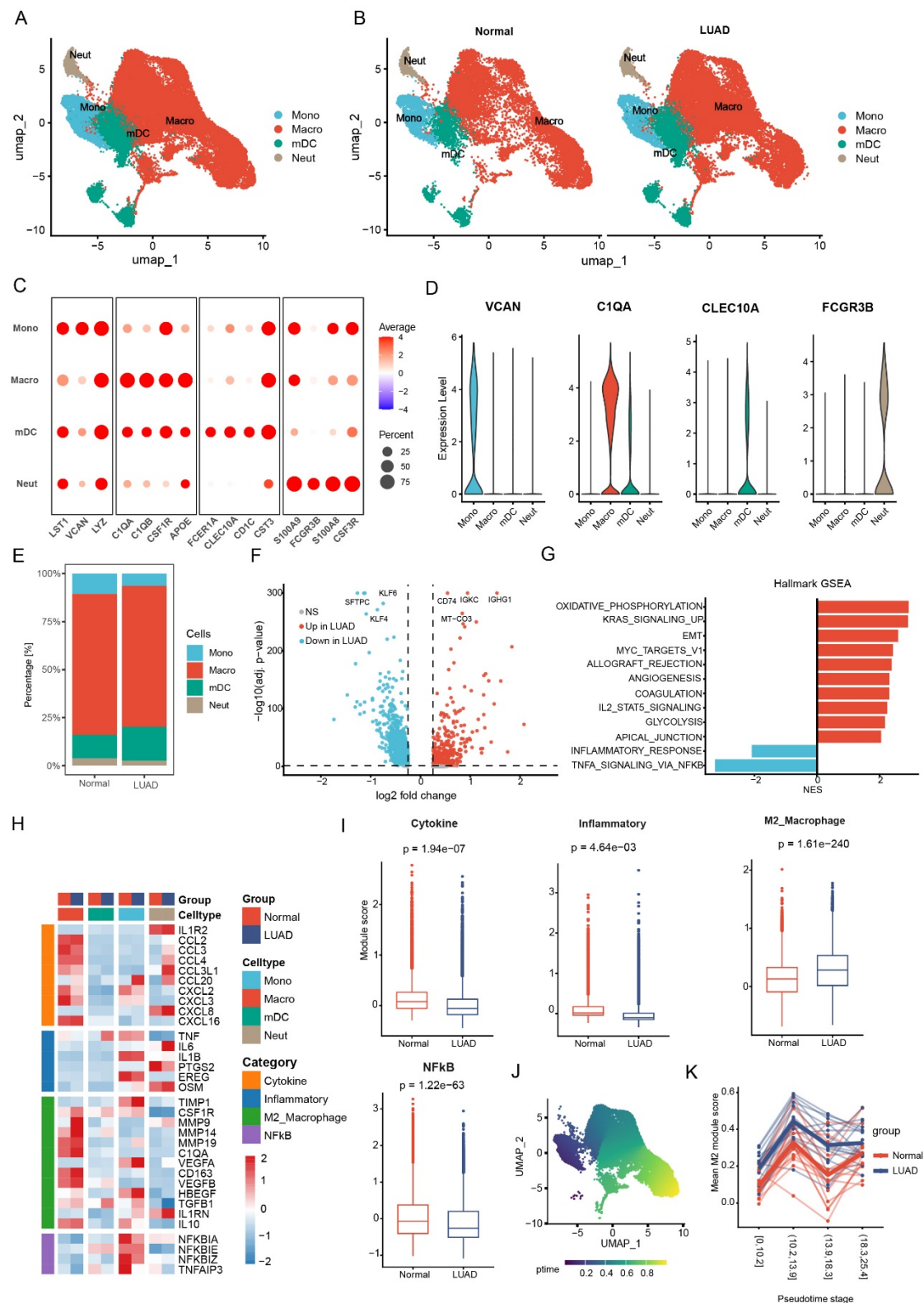
Supplementary Table S7. Main MR results.

Supplementary Table S8. Forest plot data.



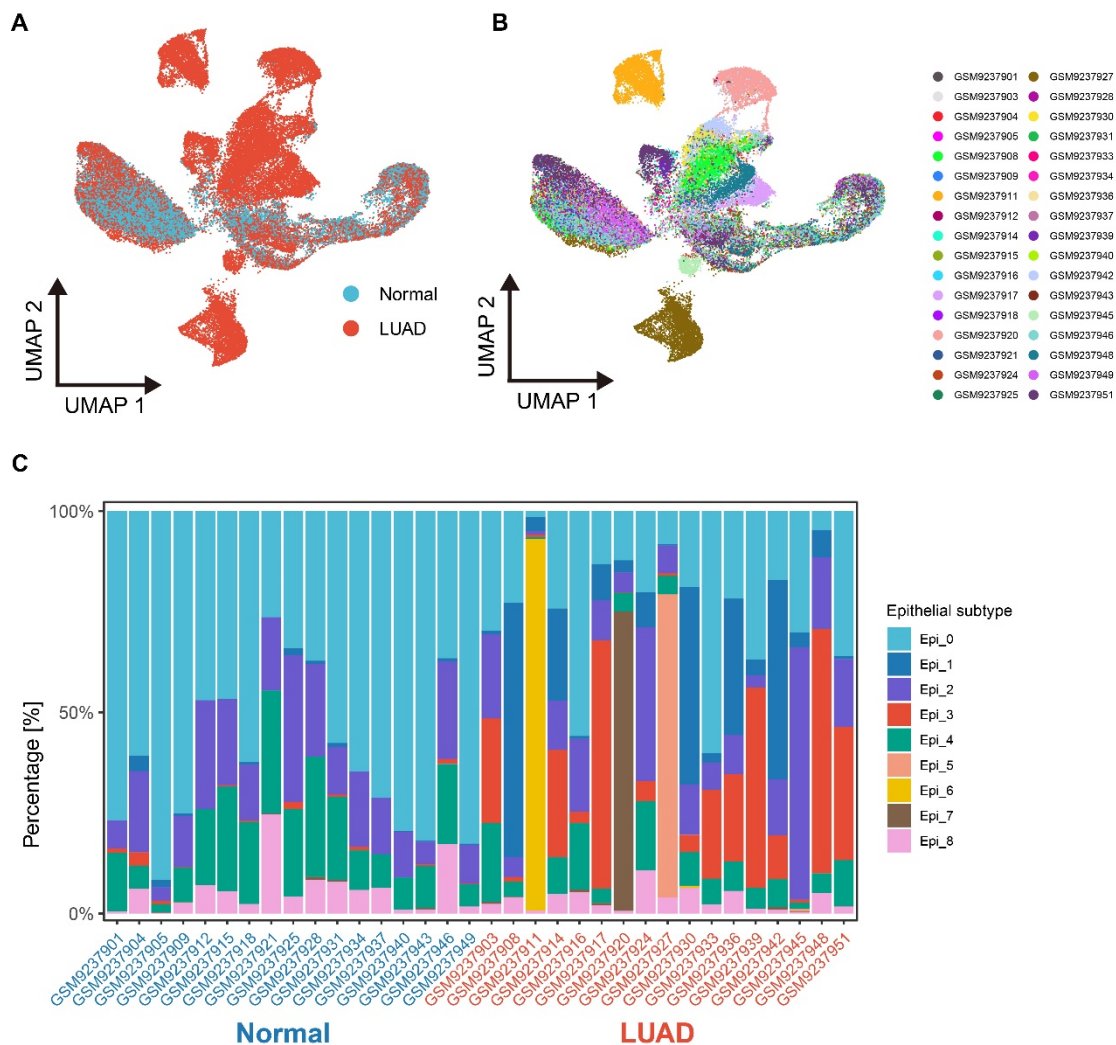
Supplementary Figure S1. Single-cell transcriptomic features of T cell subpopulations. (A) UMAP visualization of T cell subpopulations. (B) UMAP distribution of T cells in normal lung tissues (Normal) and lung adenocarcinoma tissues (LUAD). (C) Dot plot showing the expression of representative marker genes for each T cell subpopulation. (D) Heatmap showing gene expression features of different T cell subpopulations in the Normal and LUAD groups. (E) Violin plots showing the expression levels of six representative genes across T cell subpopulations. (F) Distribution of the proportions of T cell subpopulations in the Normal and LUAD groups. (G) The proportions of Tcm, Tem, and Tfh cells between the Normal and LUAD groups. (H) Comparison of the proportions of Tpro, Trm, Treg, and Tcyto cells between the

Normal and LUAD groups.



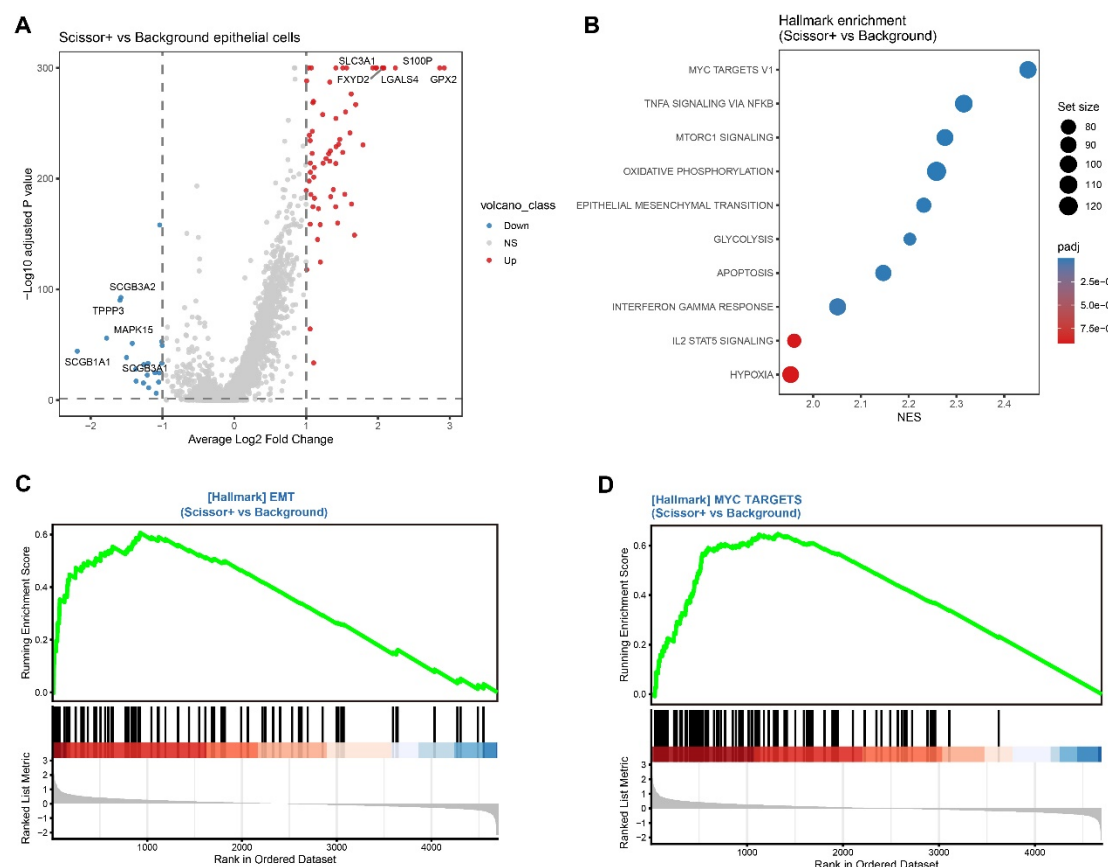
Supplementary Figure S2. Landscape features of myeloid cells. (A) UMAP visualization of myeloid cell subpopulations, including monocytes (Mono), macrophages (Macro), myeloid dendritic cells (mDC), and neutrophils (Neut). (B) UMAP distribution of myeloid cells in

normal lung tissues (Normal) and lung adenocarcinoma tissues (LUAD). (C) Dot plot showing the expression of representative marker genes for each myeloid cell subpopulation. (D) Violin plots showing the expression levels of representative marker genes across different myeloid cell subpopulations. (E) Proportions of myeloid cell subpopulations in the Normal and LUAD groups. (F) Volcano plot showing differentially expressed genes in myeloid cells between the Normal and LUAD groups. (G) Gene set enrichment analysis (GSEA) based on the Hallmark gene sets. (H) Heatmap showing the expression of module genes related to cytokine signaling, inflammatory response, M2 macrophages, and NF- κ B signaling across different myeloid cell subpopulations in the Normal and LUAD groups. (I) Comparison of the scores of the four functional modules between the Normal and LUAD groups. (J) UMAP visualization of pseudotime analysis of myeloid cells. (K) Changes in the M2 macrophage module score across different pseudotime stages in the Normal and LUAD groups.



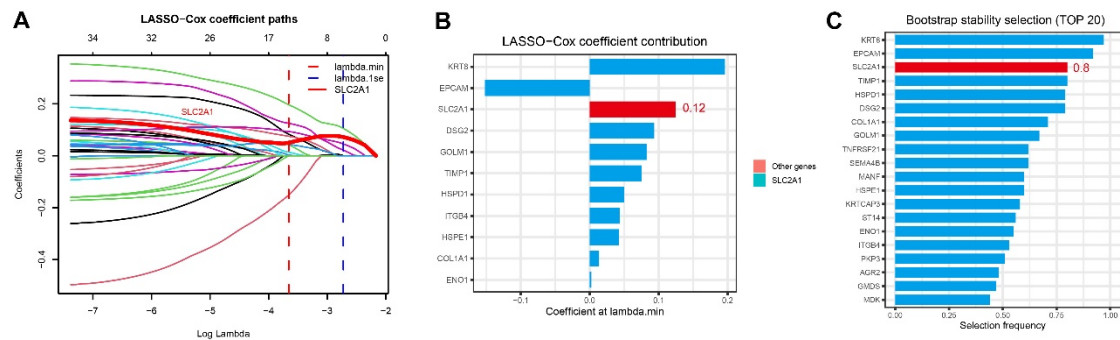
Supplementary Figure S3. Distribution of epithelial subclusters across paired normal and LUAD samples. (A) UMAP visualization of epithelial cells derived from paired tumor-adjacent normal lung tissues and LUAD tumor samples in the integrated single-cell RNA-seq dataset.

Cells are colored according to tissue origin (Normal vs LUAD). (B) UMAP visualization of epithelial cells colored by individual sample identity, showing donor distribution across epithelial populations. (C) Relative proportions of epithelial subclusters (Epi_0–Epi_8) across individual samples. Samples are grouped according to tissue type (Normal and LUAD). The analysis was performed to evaluate donor contributions within epithelial subclusters and to assess whether identified epithelial states were dominated by a limited number of samples. Most epithelial subclusters were detected across multiple donors, supporting the robustness of the identified epithelial programs despite inter-sample heterogeneity.

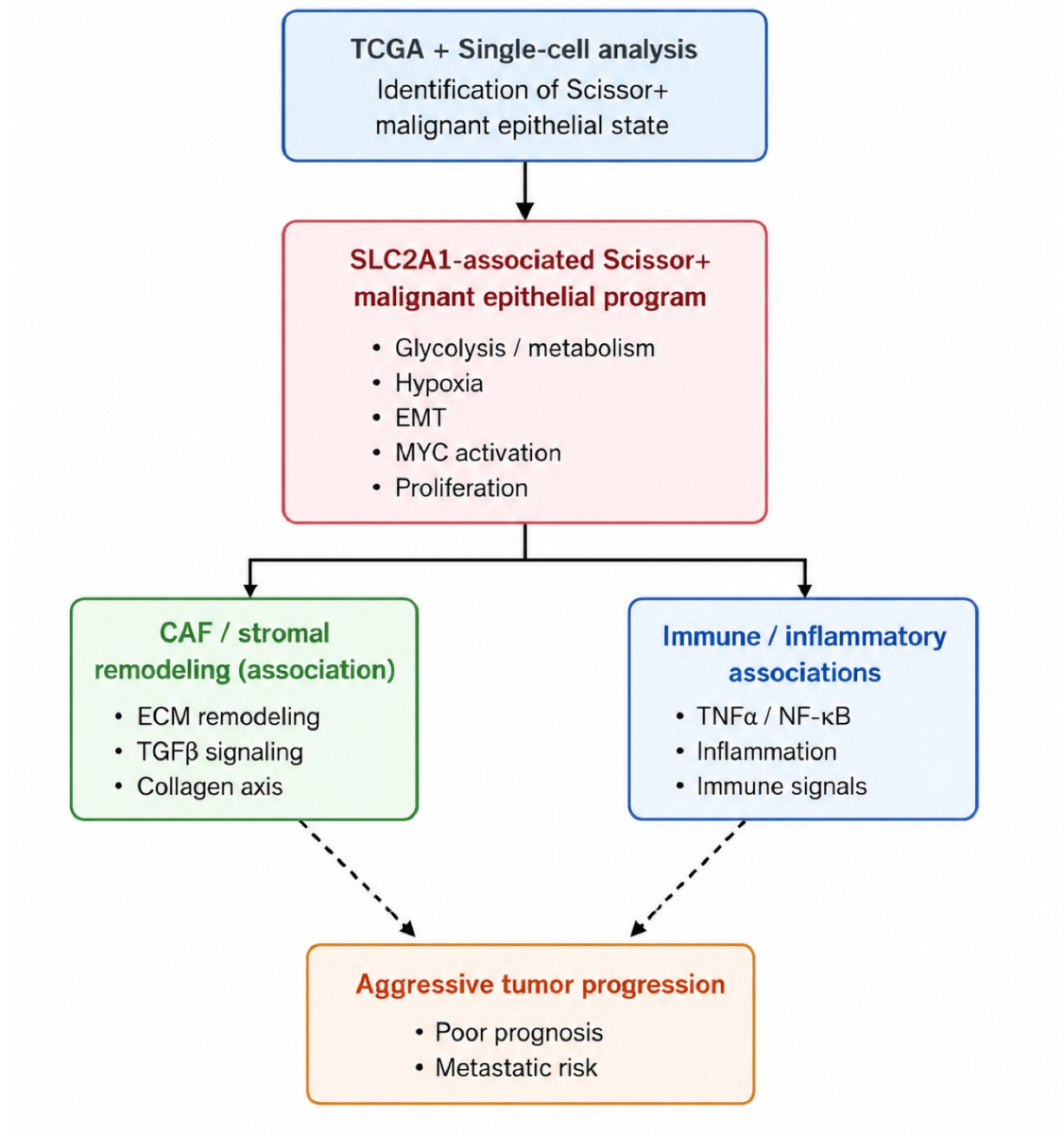


Supplementary Figure S4. Additional characterization of Scissor⁺ epithelial cells relative to background epithelial cells. (A) Differential expression analysis comparing Scissor⁺ epithelial cells with background epithelial cells. Red dots indicate significantly upregulated genes, whereas blue dots indicate significantly downregulated genes. (B) Hallmark pathway enrichment analysis of genes upregulated in Scissor⁺ epithelial cells relative to background epithelial cells. Enriched pathways included MYC targets, TNF α /NF- κ B signaling, mTORC1 signaling, oxidative phosphorylation, epithelial–mesenchymal transition (EMT), glycolysis, hypoxia, and inflammatory signaling. (C–D) Gene set enrichment analysis (GSEA) demonstrating significant enrichment of EMT-related signatures (C) and MYC target programs (D) in Scissor⁺ epithelial cells compared with background epithelial cells. These analyses were performed to further determine whether the prognostic associations identified by Scissor were

driven by intrinsic epithelial malignant programs rather than solely by differences between Scissor⁺ and Scissor⁻ populations.



Supplementary Figure S5. Stability and feature contribution analyses of prognostic genes in the LASSO-Cox model. (A) LASSO coefficient trajectories of candidate prognostic genes across different penalization parameters. Vertical dashed lines indicate the optimal lambda values (lambda.min and lambda.1se). SLC2A1 is highlighted in red. (B) Relative coefficient contributions of genes retained in the LASSO-Cox model at lambda.min. SLC2A1 exhibited a relatively strong positive coefficient compared with other selected prognostic genes. (C) Bootstrap-based stability selection analysis showing the selection frequencies of the top prognostic genes across repeated resampling iterations. SLC2A1 demonstrated a high selection frequency, supporting the robustness and reproducibility of its prioritization within the prognostic modeling framework.



Supplementary Figure S6. Schematic model summarizing the proposed biological framework associated with the SLC2A1-related Scissor⁺ epithelial state. Integrated analyses of TCGA-LUAD bulk transcriptomic data and single-cell RNA-seq data identified an SLC2A1-associated Scissor⁺ malignant epithelial state characterized by glycolytic/metabolic reprogramming, hypoxia signaling, epithelial–mesenchymal transition (EMT), MYC activation, and proliferative features. This epithelial program was further associated with stromal remodeling signatures, including extracellular matrix (ECM) remodeling, TGFβ signaling, and collagen-related pathways, as well as inflammatory and immune-related signaling pathways such as TNFα/NF-κB activation. Collectively, these features were associated with aggressive tumor progression, unfavorable prognosis, and increased metastatic potential.

Supplementary Table S1. Canonical marker genes used for cell-type annotation in single-cell RNA sequencing analysis.

cellMarker	Marker Genes
T_cells	CD3E, TRAC, CD3D
B_cells	MS4A1, CD79A, CD19
Plasma	MZB1, JCHAIN, IGHA1
Endothelial	PECAM1, VWF
Fibroblast	COL1A1, COL1A2, DCN
Epithelial	EPCAM, KRT7, KRT8
Myeloid	LYZ, CD14, SPI1
Tcm(T_cells)	IL7R, EEF1G, TCF7
Tem(T_cells)	GZMK, CXCR4
Tfh(T_cells)	CXCL13, PDCD1, ICOS
Trm(T_cells)	CXCR6, ZNF683
Treg(T_cells)	FOXP3, CTLA4
Tcyto(T_cells)	NKG7, GZMB, PRF1
Tpro(T_cells)	MKI67, TOP2A
Mono(Myeloid)	LST1, VCAN, LYZ
Macro(Myeloid)	C1QA, C1QB, CSF1R, APOE
mDC(Myeloid)	FCER1A, CLEC10A, CD1C, CST3
Neut(Myeloid)	S100A9, FCGR3B, S100A8, CSF3R

Supplementary Table S1. Canonical marker genes used for cell-type annotation in single-cell RNA sequencing analysis.

This table lists the representative marker genes used to define major cell types, including T cells, B cells, plasma cells, myeloid cells, fibroblasts, endothelial cells, and epithelial cells, as well as subtype-specific markers for T-cell and myeloid cell subpopulations.

Supplementary Table S2. Sequences of siRNAs and primers used in this study.

	Forward (5'-3')	Reverse (5'-3')
si-RNA		
NC-SLC2A1	UUCUCCGAACGUGUCACGUTT	ACGUGACACGUUCGGAGAATT
Si1-SLC2A1	CCTGCAGTTTGGCTACAACA	GTGGACCCATGTCTGGTTG
Si2-SLC2A1	GCUCUGUUGAUGAAGAUUATT	UAAUCUUCAUCAACAGAGCTT
Si3-SLC2A1	CCUUGUGCUUCAACAUCAATT	UUGAUGUUGAAGCACAAGGTT

RT-PCR

SLC2A1	TGGCATCAACGCTGTCTTCT	AACAGCGACACGACAGTGAA
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC

Supplementary Table S2. Sequences of siRNAs and primers used in this study.

This table provides the sequences of small interfering RNAs (siRNAs) targeting SLC2A1 and the corresponding negative control, as well as primer sequences used for RT-qPCR analysis of SLC2A1 and GAPDH.

Supplementary Table S3. Summary of datasets used in this study.

Category	Dataset	Platform/Database	Data Source	Data Availability
Single-cell RNA sequencing	GSE308103	GEO	Gene Expression Omnibus (NCBI)	Publicly available
External validation cohort	GSE31210	GEO	GEO	Publicly available
External validation cohort	GSE50081	GEO	GEO	Publicly available
External validation cohort	GSE72094	GEO	GEO	Publicly available
Immunotherapy cohort	GSE135222	GEO	GEO	Publicly available
Bulk RNA-seq and clinical data	TCGA-LUAD	TCGA	The Cancer Genome Atlas	Publicly available (via data portal)
Spatial transcriptomics	E-MTAB-13530	ArrayExpress	EMBL-EBI ArrayExpress	Publicly available
Immunotherapy dataset	POPLAR	Roche / Data provider	Clinical trial data platform	Controlled access
Immunotherapy dataset	OAK	Roche / Data provider	Clinical trial data platform	Controlled access

Supplementary Table S3. Summary of datasets used in this study.

This table summarizes all datasets included in the analysis, including single-cell RNA sequencing data, bulk RNA-seq cohorts, spatial transcriptomics data, and immunotherapy-treated cohorts. Information on dataset accession numbers, platforms, data sources, and availability is provided.

Supplementary Table S4. Numbers of major cell types identified in each LUAD and normal lung sample.

Sample	B_cells	Endothelial	Epithelial	Fibroblast	Myeloid	Plasma	T_cells	Total_cells	Group
GSM9237901	169	257	381	537	462	78	445	2329	Normal
GSM9237904	1906	2548	7033	4080	4584	869	8625	29645	Normal
GSM9237905	694	1220	3205	3087	1723	202	1648	11779	Normal
GSM9237909	1055	3660	3911	5065	2369	68	986	17114	Normal
GSM9237912	128	321	353	308	220	26	188	1544	Normal
GSM9237915	888	1918	3225	1954	662	113	1070	9830	Normal
GSM9237918	4	156	1230	320	103	8	27	1848	Normal
GSM9237921	304	519	823	542	224	36	309	2757	Normal
GSM9237925	78	337	742	348	215	39	157	1916	Normal
GSM9237928	149	446	550	535	195	22	223	2120	Normal
GSM9237931	64	255	401	339	189	19	150	1417	Normal
GSM9237934	66	166	199	197	78	18	114	838	Normal
GSM9237937	100	253	480	236	222	21	140	1452	Normal
GSM9237940	291	461	470	569	182	3	170	2146	Normal
GSM9237943	176	526	512	505	304	26	313	2362	Normal
GSM9237946	751	1518	3329	2312	920	294	1455	10579	Normal
GSM9237949	715	2142	2544	2975	1035	39	889	10339	Normal
GSM9237903	270	455	976	503	578	66	681	3529	LUAD
GSM9237908	1894	2597	10867	5176	4257	1195	6136	32122	LUAD
GSM9237911	1392	525	3915	1952	2446	976	5958	17164	LUAD
GSM9237914	507	768	1408	1152	782	443	1621	6681	LUAD
GSM9237916	738	1386	2948	2546	1072	174	1130	9994	LUAD
GSM9237917	757	710	2761	1607	882	388	1951	9056	LUAD
GSM9237920	53	399	4389	1260	768	101	253	7223	LUAD
GSM9237924	655	1178	2207	1667	707	258	1258	7930	LUAD
GSM9237927	1343	1126	6599	1613	1463	1720	4300	18164	LUAD
GSM9237930	893	615	1588	1844	1340	1256	2908	10444	LUAD
GSM9237933	161	484	1163	873	406	141	458	3686	LUAD
GSM9237936	383	310	409	724	410	214	815	3265	LUAD
GSM9237939	254	605	1397	637	421	57	643	4014	LUAD
GSM9237942	764	583	2113	1465	711	225	1578	7439	LUAD
GSM9237945	374	580	1249	736	422	179	726	4266	LUAD
GSM9237948	1127	471	2188	2662	2120	875	4037	13480	LUAD

GSM9237951	867	2823	3138	4283	1503	59	2244	14917	LUAD
ALL(N=34)	45821	31599	51041	57939	34957	10198	51834	283389	

Supplementary Table S4. Numbers of major cell types identified in each LUAD and normal lung sample.

This table summarizes the numbers of major cell types identified in each paired normal lung and LUAD sample from the single-cell RNA-seq dataset.

Supplementary Table S5. Distribution of epithelial subclusters in each LUAD and normal sample

Sample	Group	Epi_0	Epi_1	Epi_2	Epi_3	Epi_4	Epi_5	Epi_6	Epi_7	Epi_8
GSM9237901	Normal	143	0	13	2	27	0	0	0	1
GSM9237904	Normal	1727	107	573	94	161	0	0	3	174
GSM9237905	Normal	854	17	31	8	19	0	0	1	2
GSM9237909	Normal	1706	12	290	2	196	0	1	2	61
GSM9237912	Normal	94	0	54	0	38	0	0	0	14
GSM9237915	Normal	856	3	392	7	477	0	0	0	102
GSM9237918	Normal	290	3	64	2	95	0	0	0	11
GSM9237921	Normal	139	0	96	0	161	0	0	1	130
GSM9237925	Normal	188	9	201	10	120	0	0	0	23
GSM9237928	Normal	151	3	94	0	121	0	0	3	34
GSM9237931	Normal	117	2	24	1	42	0	0	1	16
GSM9237934	Normal	66	0	19	1	10	0	0	0	6
GSM9237937	Normal	188	0	37	0	22	0	0	0	17
GSM9237940	Normal	231	1	33	0	23	0	0	0	3
GSM9237943	Normal	234	1	16	1	30	0	0	1	3
GSM9237946	Normal	813	18	538	28	441	0	0	0	385
GSM9237949	Normal	890	3	103	3	59	0	0	0	19
GSM9237903	LUAD	234	7	164	204	154	0	0	4	19
GSM9237908	LUAD	1587	4405	342	83	261	0	0	5	285
GSM9237911	LUAD	56	134	35	20	17	0	3525	1	28
GSM9237914	LUAD	234	221	119	258	88	0	0	0	47
GSM9237916	LUAD	772	10	250	40	228	0	0	8	74
GSM9237917	LUAD	291	198	221	1364	78	0	0	12	46
GSM9237920	LUAD	484	119	197	3	185	0	0	2946	26
GSM9237924	LUAD	311	134	592	77	264	1	0	1	165
GSM9237927	LUAD	417	17	350	32	233	3815	0	2	204
GSM9237930	LUAD	197	511	131	45	87	0	5	1	66
GSM9237933	LUAD	319	12	36	117	34	0	0	0	12
GSM9237936	LUAD	54	84	24	54	18	0	0	0	14
GSM9237939	LUAD	346	37	29	467	48	0	0	1	11
GSM9237942	LUAD	314	904	256	199	126	0	0	13	17

GSM9237945	LUAD	262	32	545	6	14	6	0	1	3
GSM9237948	LUAD	95	134	353	1209	97	0	1	0	101
GSM9237951	LUAD	802	18	375	739	256	0	0	1	39
ALL(N=34)		15462	7156	6597	5076	4230	3822	3532	3008	2158

Supplementary Table S5. Distribution of epithelial subclusters in each LUAD and normal sample.

This table summarizes the distribution of epithelial subclusters (Epi_0–Epi_8) across individual LUAD and normal lung samples.

Supplementary Table S6. Clinical and transcriptomic datasets used in this study

Cohort	Platform	Sample size	High-risk	Low-risk	Endpoint	Source	Usage
TCGA-LUAD	RNA-seq	515	257	258	OS	TCGA	Training
GSE31210	Microarray	226	105	121	OS	GEO	Validation
GSE50081	Microarray	127	61	66	OS	GEO	Validation
GSE72094	RNA-seq	398	206	192	OS	GEO	Validation

Supplementary Table S6. Clinical and transcriptomic datasets used in this study.

This table summarizes the bulk transcriptomic cohorts used for model training and external validation, including platform information, sample sizes, clinical endpoints, and data sources.

Supplementary Table S7. Main MR Results

Exposure	Outcome	Method	SNPs	Beta	SE	OR (95% CI)	P value
SLC2A1	Lung adenocarcinoma	IVW	6	0.349	0.154	1.418 (1.049–1.918)	0.023
SLC2A1	Lung adenocarcinoma	Weighted median	6	0.315	0.191	1.371 (0.943–1.992)	0.098
SLC2A1	Lung adenocarcinoma	MR Egger	6	1.454	0.782	4.280 (0.92–19.77)	0.063

SLC2A1	Lung adenocarcinoma	Simple mode	6	0.221	0.164	1.247 (0.902–1.725)	0.178
SLC2A1	Lung adenocarcinoma	Weighted mode	6	0.228	0.138	1.256 (0.943–1.687)	0.123

Supplementary Table S7. Main MR results.

This table summarizes the Mendelian randomization (MR) analyses evaluating the association between genetically predicted SLC2A1 expression and lung adenocarcinoma risk.

Supplementary Table S8. Forest Plot Data

Exposure	Outcome	SNP / Overall	Method	Beta	SE	P value
SLC2A1	Lung adenocarcinoma	rs1770813	Wald ratio	0.52	0.337	0.122
SLC2A1	Lung adenocarcinoma	rs4660691	Wald ratio	0.213	0.401	0.594
SLC2A1	Lung adenocarcinoma	rs4660697	Wald ratio	0.505	0.381	0.185
SLC2A1	Lung adenocarcinoma	rs710221	Wald ratio	0.187	0.343	0.586
SLC2A1	Lung adenocarcinoma	rs859511	Wald ratio	0.241	0.413	0.56
SLC2A1	Lung adenocarcinoma	rs9438964	Wald ratio	0.398	0.413	0.336
SLC2A1	Lung adenocarcinoma	All-IVW	IVW	0.349	0.154	0.023

Supplementary Table S8. Forest plot data.

This table provides the detailed Mendelian randomization effect estimates for individual SNPs and the overall inverse variance weighted (IVW) analysis.