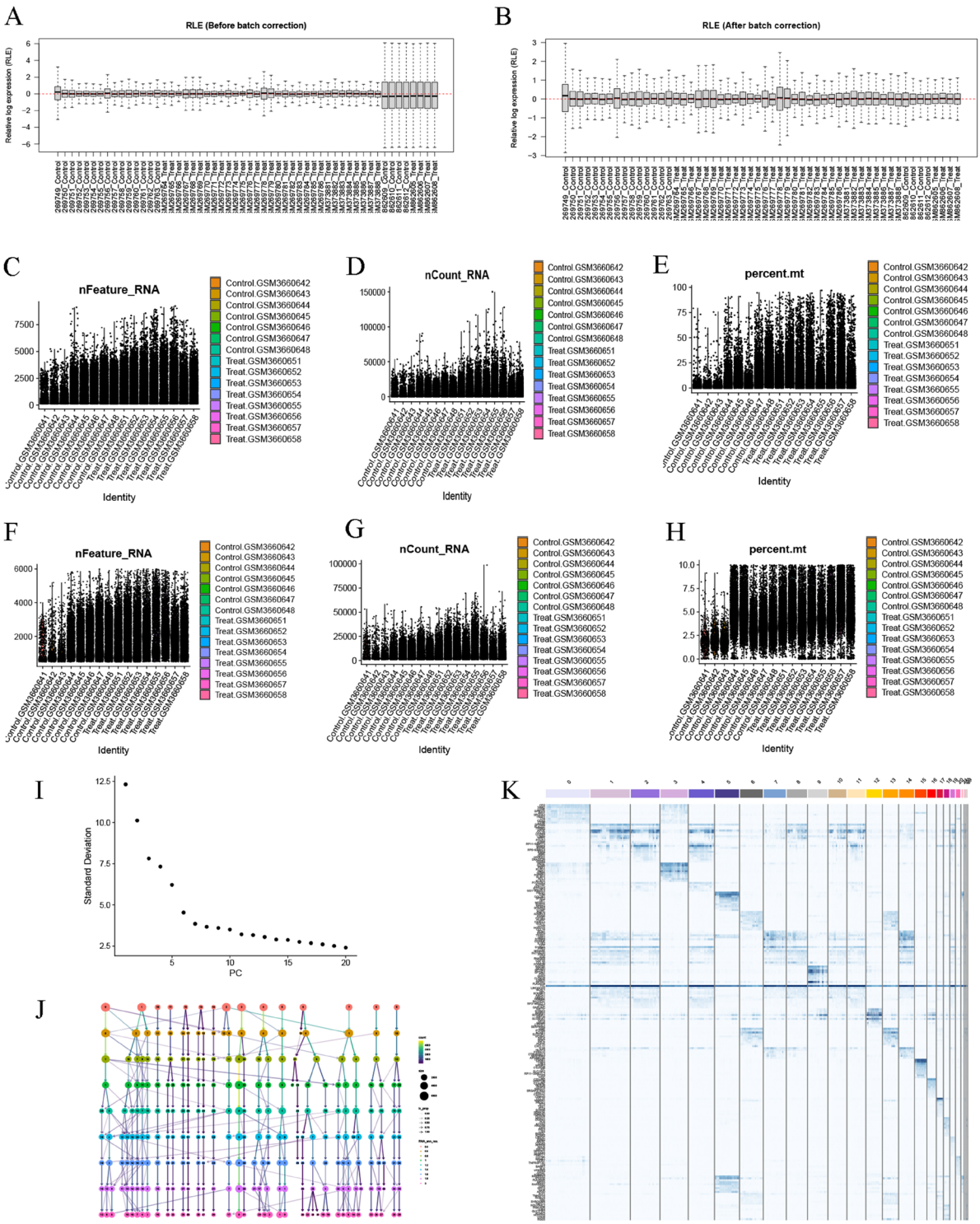

Figure S1 (A,B) RLE plots before and after batch correction of the bulk training datasets GSE35145 and GSE10667; (C-E) distributions of nFeature_RNA, nCount_RNA, and percent.mt across samples before single-cell quality control; (F-H) distributions of nFeature_RNA, nCount_RNA, and percent.mt across samples after single-cell quality control; (I) elbow plot used to determine the number of principal components for downstream dimensionality reduction; (J) clustree plot used to evaluate clustering resolution during single-cell preprocessing; (K) heatmap of representative marker genes across the initial 24 single-cell clusters.

https://dx.doi.org/10.21037/jtd-2026-0896

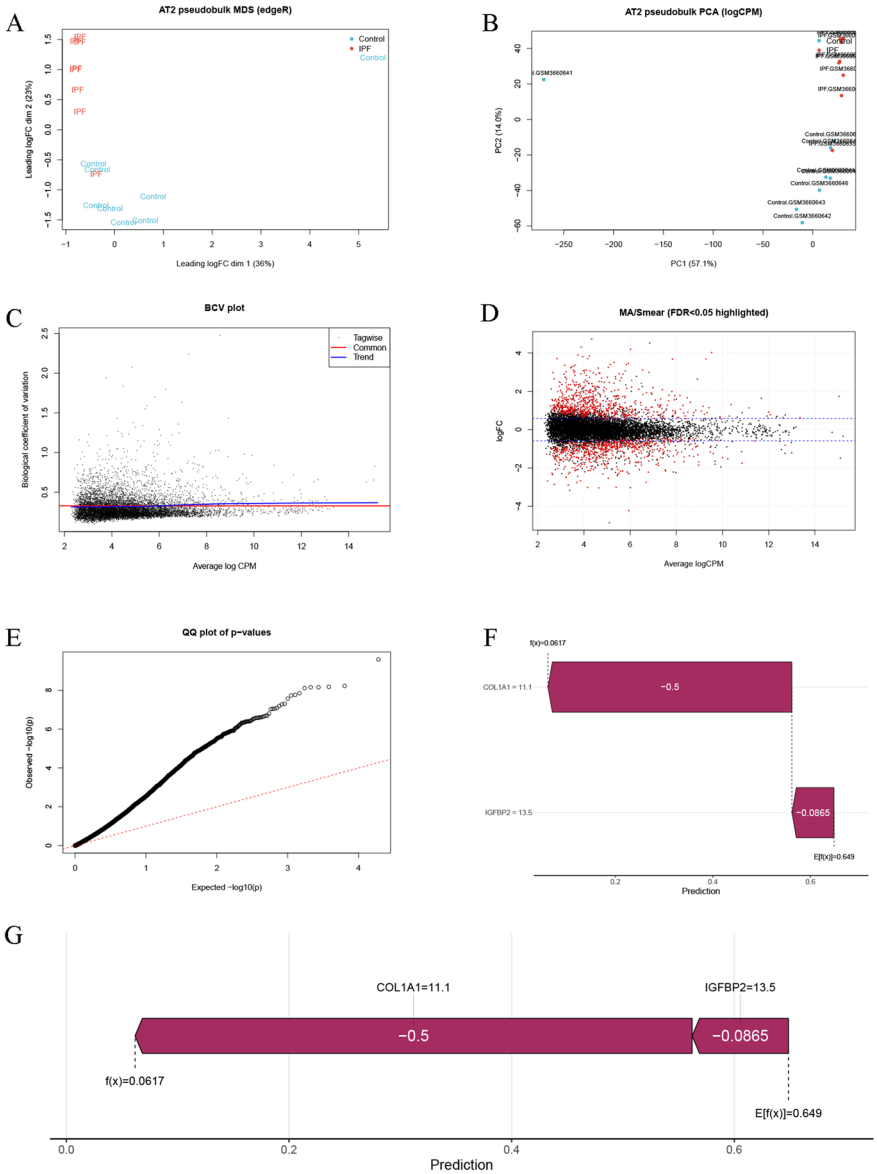

Figure S2 (A) Multidimensional scaling (MDS) plot of AT2 pseudo-bulk samples; (B) Principal component analysis (PCA) plot of AT2 pseudo-bulk samples; (C) Biological coefficient of variation (BCV) plot; (D) MA/smear plot of differential expression analysis, with significantly differentially expressed genes highlighted; (E) Quantile-quantile (QQ) plot of P values; (F-G) Representative individual-sample prediction interpretation plots showing feature contributions to model output.

https://dx.doi.org/10.21037/jtd-2026-0896

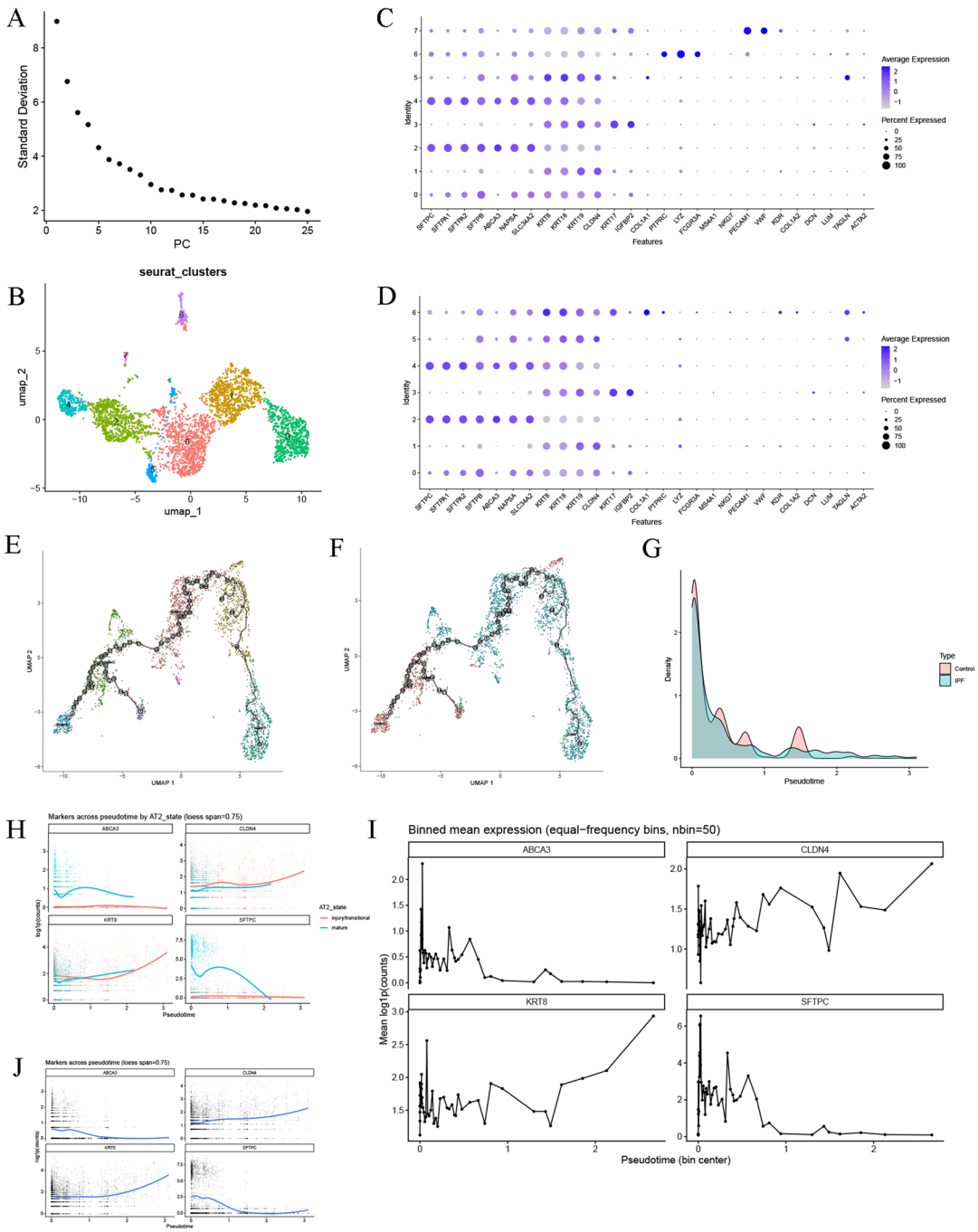

Figure S3 (A) Elbow plot used to determine the number of principal components for AT2 subclustering; (B) initial UMAP showing subclusters identified after reclustering the annotated AT2 population; (C) dot plot of representative marker genes showing expression of non-AT2 lineage markers in contaminating subclusters; (D) dot plot of marker-gene expression after removal of contaminating clusters, showing improved AT2 specificity; (E) pseudotime trajectory of clean-AT2 cells colored by cluster identity; (F) pseudotime trajectory of clean-AT2 cells colored by sample group; (G) density distribution of pseudotime across control and IPF samples; (H) LOESS-smoothed expression patterns of representative mature and injury-associated markers across pseudotime, stratified by AT2 state; (I) binned mean expression of representative mature and injury-associated markers along pseudotime; (J) overall LOESS-smoothed expression patterns of representative mature and injury-associated markers across pseudotime.

https://dx.doi.org/10.21037/jtd-2026-0896

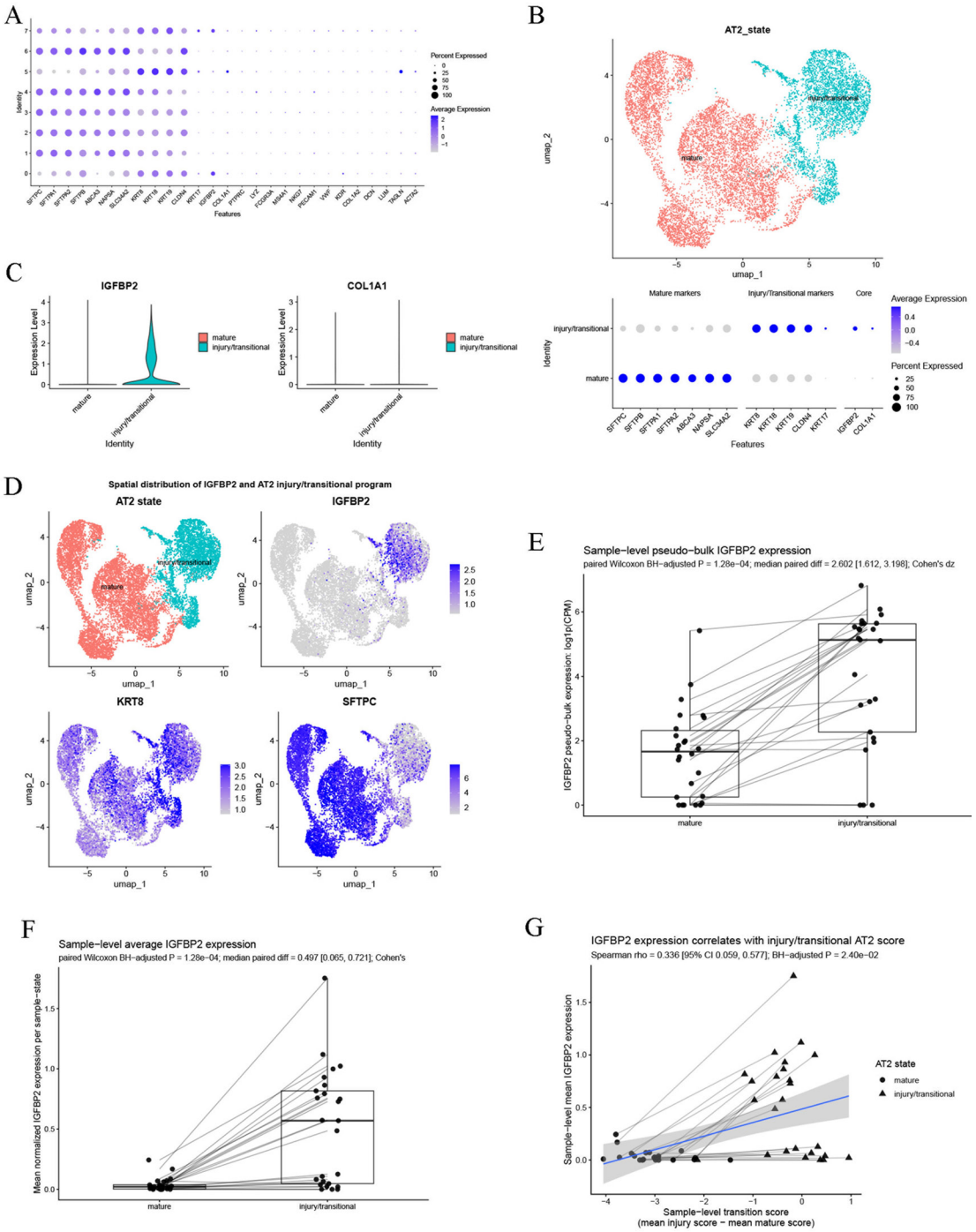

Figure S4 (A) Dot plot of representative marker genes supporting cell annotation and AT2-related marker patterns in the validation cohort. (B) UMAP and marker-gene dot plot supporting mature and injury/transitional AT2 state assignment. (C) Expression of IGFBP2 and COL1A1 across mature and injury/transitional AT2 states. (D) Spatial distribution of AT2 state, IGFBP2, KRT8, and SFTPC in UMAP space. (E) Sample-level pseudo-bulk IGFBP2 expression measured as log1p(CPM). (F) Sample-level average normalized IGFBP2 expression. (G) Correlation between sample-level IGFBP2 expression and the AT2 transition score.

https://dx.doi.org/10.21037/jtd-2026-0896

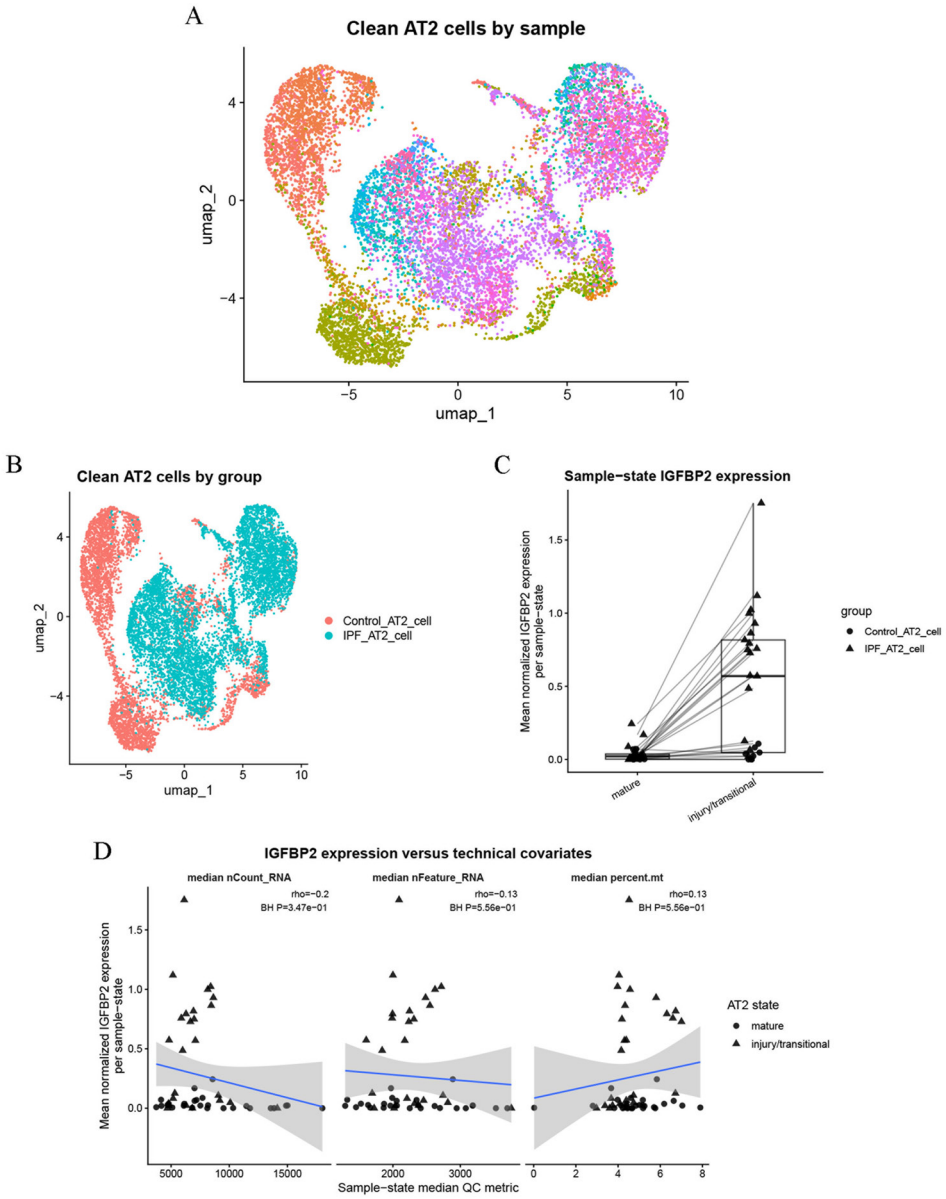

Figure S5 (A) UMAP visualization of the clean AT2 subset colored by individual sample. (B) UMAP visualization of the clean AT2 subset colored by disease group. (C) Sample-state mean normalized IGFBP2 expression in mature and injury/transitional AT2 cells. Gray lines connect paired sample-state values from the same sample. (D) Correlations between sample-state mean normalized IGFBP2 expression and technical covariates, including median nCount_RNA, median nFeature_RNA, and median percent.mt. Spearman correlation coefficients and Benjamini-Hochberg-adjusted P values are indicated. No significant correlations were detected, supporting that IGFBP2 enrichment was not primarily driven by technical covariates.

https://dx.doi.org/10.21037/jtd-2026-0896

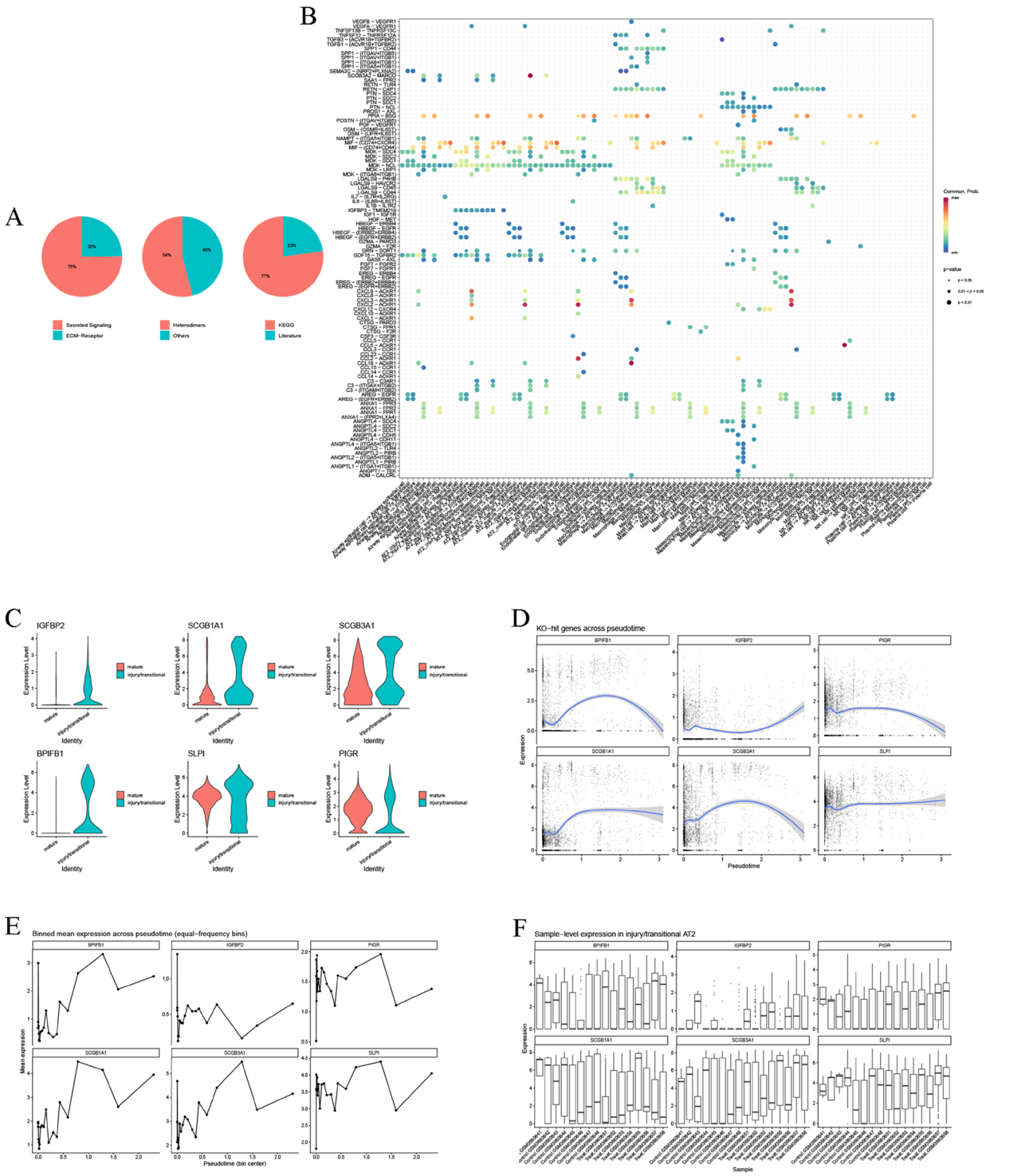

Figure S6 (A) Composition of signaling interaction categories included in the CellChat analysis; (B) global ligand-receptor interaction landscape inferred across major cell populations; (C) expression of IGFBP2 and representative knockout-inferred genes across mature and injury/transitional AT2 states; (D) LOESS-smoothed expression trajectories of IGFBP2 and representative knockout-inferred genes across pseudotime; (E) binned mean expression of IGFBP2 and representative knockout-inferred genes across pseudotime; (F) sample-level expression of representative knockout-inferred genes within injury/transitional AT2 cells.

https://dx.doi.org/10.21037/jtd-2026-0896

Table S1 Marker genes for cell type annotation in single-cell RNA sequencing

Cell type	Maker
Macrophage	MRC1, C1QA, MARCO
Monocyte	CD14, FCN1, FCGR3A
Endothelial cell	VWF, PECAM1, KDR
Mast cell	MS4A2, CPA3, TPSAB1
Mesenchymal cell	PDGFRA, COL1A1, ACTA2
NK cell	NKG7, GNLY
CD4+ T cell	CD4, TRAC, IL7R
CD8+ T cell	CD8A, TRAC, GZMK
Plasma cell	MZB1, JCHAIN, XBP1
AT1 cell	AGER, CAV1
AT2 cell	SFTPB, SFTPC, ABCA3
Airway epithelial cell	FOXJ1, SCGB1A1, MUC5B

https://dx.doi.org/10.21037/jtd-2026-0896

Table S2 PSRGs

No.	Gene
1	ABHD12
2	ACTA1
3	ADGRV1
4	ANGPT2
5	ANKRD1
6	ANKRD23
7	ANO3
8	AQP1
9	ASIC2
10	ASIC3
11	ATAT1
12	ATOH7
13	ATP1A2
14	ATP8A2
15	BACE1
16	BAD
17	BAG3
18	BAK1
19	BCL10
20	BDKRB1
21	BGLAP
22	BMP6
23	BNIP3
24	BTG2
25	CAPN2
26	CASP1
27	CASP2
28	CASP5
29	CASP8
30	CASP8AP2
31	CAV3
32	CD40
33	CDH2
34	CHEK1
35	CHI3L1
36	CHRNA10
37	CHRNA9
38	CITED2
39	CLCN6
40	CNN2
41	CNTNAP2
42	COL11A1
43	COL1A1
44	CRADD
45	CSRP3
46	CTNNB1
47	CXCL10
48	CXCL12
49	CXCR4
50	DAG1
51	DCANP1
52	DDR2
53	DMD
54	DRD2
55	EDN1
56	ENG
57	ETV1
58	F11R
59	FADD
60	FAS
61	FGF2
62	FOS
63	FOSB
64	FOSL1
65	FYN
66	GADD45A
67	GATA4
68	GCLC
69	GDF5
70	GPI
71	GSN
72	HABP4
73	HPN
74	HTR2A
75	HTT
76	IGF1R
77	IGFBP2
78	IHH
79	IL13
80	IL1B
81	IL33
82	IRF1
83	ITGA2
84	ITGB3
85	JUP
86	KCNA1
87	KCNA5
88	KCNC1
89	KCNJ2
90	KCNK2
91	KCNK4
92	KCNQ1
93	KCNQ3
94	KIAA0319
95	KIT
96	KRT5
97	LARGE1
98	LHFPL5
99	LRP11
100	LTBR
101	MAG
102	MAP1B
103	MAP2K4
104	MAP3K1
105	MAP3K14
106	MAP3K2
107	MAPK14
108	MAPK3
109	MAPK8
110	MBD2
111	MDK
112	MEIS2
113	MKKS
114	MMP14
115	MMP2
116	MPO
117	MTPN
118	MYD88
119	NEUROG1
120	NFKB1
121	NFKBIA
122	NPPA
123	NRXN1
124	NRXN2
125	NTRK1
126	P2RX3
127	P2RX7
128	P2RY1
129	PDE2A
130	PDZD7
131	PHF24
132	PIEZO1
133	PIEZO2
134	PIK3CA
135	PJVK
136	PKD1
137	PKD1L1
138	PKD1L3
139	PKD2
140	PKD2L1
141	PKD2L2
142	PKDREJ
143	PLEC
144	POSTN
145	PPL
146	PSPH
147	PTCH1
148	PTGER4
149	PTGS2
150	PTK2
151	PTK2B
152	PTN
153	RAF1
154	RELA
155	RETN
156	RPS6KB1
157	RYR2
158	SCEL
159	SCN11A
160	SCN1A
161	SCN9A
162	SCX
163	SERPINE2
164	SHANK3
165	SLC1A3
166	SLC26A5
167	SLC2A1
168	SLC38A2
169	SLC8A1
170	SLC9A1
171	SLITRK6
172	SOST
173	SOX9
174	SRC
175	STAT1
176	STRA6
177	STRBP
178	STRC
179	SUN1
180	TACR1
181	TCAP
182	THBS1
183	TIFAB
184	TLR3
185	TLR4
186	TLR5
187	TLR7
188	TLR8
189	TMC1
190	TMC2
191	TMEM120A
192	TMEM150C
193	TMEM87A
194	TNC
195	TNF
196	TNFRSF10A
197	TNFRSF10B
198	TNFRSF11A
199	TNFRSF1A
200	TNFRSF8
201	TNFSF14
202	TRPA1
203	TRPV4
204	TTN
205	TUBA1A
206	TXNIP
207	UCN
208	USP53
209	WHRN
210	WNT11
211	XPA
212	XPC

Table S2 (*continued*)

https://dx.doi.org/10.21037/jtd-2026-0896

Table S3 Cross-cohort reproducibility of IGFBP2 enrichment and transition-score association

Dataset	Cohort role	Analysis	Unit	n	Statistical test	Effect size/ correlation	95% CI	BH-adjusted P	Direction/ interpretation
GSE128033	Discovery scRNA cohort	IGFBP2, injury/transitional *vs.* mature AT2	Sample-level pseudo-bulk log1p(CPM)	16 paired samples	Paired Wilcoxon	Cohen's dz =0.87; log_2FC =6.189	Median diff 0.849 [0.288, 1.970]; log2FC [1.438, 11.491]	1.09e-03	Higher in injury/ transitional AT2
GSE135893	Independent validation scRNA cohort	IGFBP2, injury/transitional *vs.* mature AT2	Sample-level pseudo-bulk log1p(CPM)	24 paired samples	Paired Wilcoxon	Cohen's dz =1.38; log_2FC =3.385	Median diff 2.602 [1.612, 3.198]; log_2FC [2.200, 4.827]	1.28e-04	Higher in injury/ transitional AT2
GSE128033	Discovery scRNA cohort	IGFBP2, injury/transitional *vs.* mature AT2	Sample-level mean normalized expression	16 paired samples	Paired Wilcoxon	Cohen's dz =0.83	Median diff 0.161 [0.086, 0.339]	1.33e-03	Higher in injury/ transitional AT2
GSE135893	Independent validation scRNA cohort	IGFBP2, injury/transitional *vs.* mature AT2	Sample-level mean normalized expression	24 paired samples	Paired Wilcoxon	Cohen's dz =1.01	Median diff 0.497 [0.065, 0.721]	1.28e-04	Higher in injury/ transitional AT2
GSE128033	Discovery scRNA cohort	IGFBP2 *vs.* AT2 transition score	Sample-state level	32 sample-state points	Spearman	rho =0.695	0.464-0.822	3.07e-05	Positive association
GSE135893	Independent validation scRNA cohort	IGFBP2 *vs.* AT2 transition score	Sample-state level	51 sample-state points	Spearman	rho =0.336	0.059-0.577	2.40e-02	Positive association

The same sample-level pseudo-bulk, mean-expression, and transition-score framework was applied to GSE128033 and GSE135893. BH, Benjamini-Hochberg adjustment; CI, confidence interval; CPM, counts per million. Transition score was calculated as mean injury marker score minus mean mature AT2 marker score.

Table S4 Technical negative-control correlations in GSE135893

Dataset	Cohort role	Analysis	Unit	n	Statistical test	Effect size/ correlation	95% CI	BH-adjusted P	Direction/ interpretation
GSE135893	Independent validation scRNA cohort	IGFBP2 *vs.* nCount_RNA	Sample-state level	51 sample-state points	Spearman	rho =-0.20	Not estimated	0.347	Not QC-driven
GSE135893	Independent validation scRNA cohort	IGFBP2 *vs.* nFeature_RNA	Sample-state level	51 sample-state points	Spearman	rho =-0.13	Not estimated	0.556	Not QC-driven
GSE135893	Independent validation scRNA cohort	IGFBP2 *vs.* percent.mt	sample-state level	51 sample-state points	Spearman	rho =0.13	Not estimated	0.556	Not QC-driven

QC-metric correlations were included as negative-control checks for technical confounding. Not estimated indicates that a confidence interval was not calculated for these QC negative-control correlations.

https://dx.doi.org/10.21037/jtd-2026-0896

Table S5 Top mesenchymal-to-AT2 incoming ligand-receptor pairs inferred by CellChat.

Direction	Rank	Ligand	Receptor	Pair	Pathway	Prob.	P value
Mesenchymal -> AT2 mature	1	COL1A2	SDC4	COL1A2-SDC4	COLLAGEN	0.1160	<0.001
Mesenchymal -> AT2 mature	2	COL6A2	SDC4	COL6A2-SDC4	COLLAGEN	0.0924	<0.001
Mesenchymal -> AT2 mature	3	FN1	SDC4	FN1-SDC4	FN1	0.0877	<0.001
Mesenchymal -> AT2 mature	4	COL6A1	SDC4	COL6A1-SDC4	COLLAGEN	0.0781	<0.001
Mesenchymal -> AT2 mature	5	COL1A1	SDC4	COL1A1-SDC4	COLLAGEN	0.0769	<0.001
Mesenchymal -> AT2 mature	6	COL1A2	CD44	COL1A2-CD44	COLLAGEN	0.0688	<0.001
Mesenchymal -> AT2 mature	7	COL6A3	SDC4	COL6A3-SDC4	COLLAGEN	0.0590	<0.001
Mesenchymal -> AT2 mature	8	COL6A2	CD44	COL6A2-CD44	COLLAGEN	0.0542	<0.001
Mesenchymal -> AT2 mature	9	FN1	CD44	FN1-CD44	FN1	0.0514	<0.001
Mesenchymal -> AT2 mature	10	COL1A2	SDC1	COL1A2-SDC1	COLLAGEN	0.0488	<0.001
Mesenchymal -> AT2 injury/ transitional	1	COL1A2	SDC4	COL1A2-SDC4	COLLAGEN	0.0776	<0.001
Mesenchymal -> AT2 injury/ transitional	2	COL6A2	SDC4	COL6A2-SDC4	COLLAGEN	0.0612	<0.001
Mesenchymal -> AT2 injury/ transitional	3	COL1A2	SDC1	COL1A2-SDC1	COLLAGEN	0.0585	<0.001
Mesenchymal -> AT2 injury/ transitional	4	FN1	SDC4	FN1-SDC4	FN1	0.0581	<0.001
Mesenchymal -> AT2 injury/ transitional	5	COL6A1	SDC4	COL6A1-SDC4	COLLAGEN	0.0515	<0.001
Mesenchymal -> AT2 injury/ transitional	6	COL1A1	SDC4	COL1A1-SDC4	COLLAGEN	0.0507	<0.001
Mesenchymal -> AT2 injury/ transitional	7	COL6A2	SDC1	COL6A2-SDC1	COLLAGEN	0.0460	<0.001
Mesenchymal -> AT2 injury/ transitional	8	FN1	SDC1	FN1-SDC1	FN1	0.0436	<0.001
Mesenchymal -> AT2 injury/ transitional	9	COL1A2	CD44	COL1A2-CD44	COLLAGEN	0.0428	<0.001
Mesenchymal -> AT2 injury/ transitional	10	COL6A3	SDC4	COL6A3-SDC4	COLLAGEN	0.0386	<0.001

Pairs are ranked by CellChat communication probability within each source-target direction. Prob., communication probability.

https://dx.doi.org/10.21037/jtd-2026-0896

Table S6 Top outgoing ligand-receptor pairs from injury/transitional AT2 cells

Direction	Rank	Ligand	Receptor	Pair	Pathway	Prob.	P value
AT2 injury/transitional -> Endothelial	1	CXCL8	ACKR1	CXCL8-ACKR1	CXCL	0.0885	<0.001
AT2 injury/transitional -> Endothelial	2	CXCL2	ACKR1	CXCL2-ACKR1	CXCL	0.0683	<0.001
AT2 injury/transitional -> Endothelial	3	CXCL1	ACKR1	CXCL1-ACKR1	CXCL	0.0590	<0.001
AT2 injury/transitional -> Endothelial	4	MDK	NCL	MDK-NCL	MK	0.0233	<0.001
AT2 injury/transitional -> Endothelial	5	MDK	ITGA6_ITGB1	MDK-(ITGA6+ITGB1)	MK	0.0132	<0.001
AT2 injury/transitional -> Endothelial	6	CXCL3	ACKR1	CXCL3-ACKR1	CXCL	0.0118	<0.001
AT2 injury/transitional -> Endothelial	7	GDF15	TGFBR2	GDF15-TGFBR2	GDF	0.0118	<0.001
AT2 injury/transitional -> Endothelial	8	NAMPT	ITGA5_ITGB1	NAMPT-(ITGA5+ITGB1)	VISFATIN	0.0097	<0.001
AT2 injury/transitional -> Endothelial	9	CCL18	ACKR1	CCL18-ACKR1	CCL	0.0064	<0.001
AT2 injury/transitional -> Endothelial	10	LAMB3	ITGA6_ITGB1	LAMB3-(ITGA6+ITGB1)	LAMININ	0.0061	<0.001
AT2 injury/transitional -> Macrophage	1	MIF	CD74_CD44	MIF-(CD74+CD44)	MIF	0.0570	<0.001
AT2 injury/transitional -> Macrophage	2	PPIA	BSG	PPIA-BSG	CypA	0.0548	<0.001
AT2 injury/transitional -> Macrophage	3	MIF	CD74_CXCR4	MIF-(CD74+CXCR4)	MIF	0.0377	<0.001
AT2 injury/transitional -> Macrophage	4	LAMB3	CD44	LAMB3-CD44	LAMININ	0.0211	<0.001
AT2 injury/transitional -> Macrophage	5	MDK	NCL	MDK-NCL	MK	0.0182	<0.001
AT2 injury/transitional -> Macrophage	6	ANXA1	FPR1	ANXA1-FPR1	ANNEXIN	0.0136	<0.001
AT2 injury/transitional -> Macrophage	7	ANXA1	FPR3	ANXA1-FPR3	ANNEXIN	0.0131	<0.001
AT2 injury/transitional -> Macrophage	8	SCGB3A2	MARCO	SCGB3A2-MARCO	UGRP1	0.0116	<0.001
AT2 injury/transitional -> Macrophage	9	ANXA1	FPR2_LXA4	ANXA1-(FPR2+LXA4)	ANNEXIN	0.0109	<0.001
AT2 injury/transitional -> Macrophage	10	MDK	SDC4	MDK-SDC4	MK	0.0092	<0.001
AT2 injury/transitional -> Mesenchymal	1	PPIA	BSG	PPIA-BSG	CypA	0.0681	<0.001
AT2 injury/transitional -> Mesenchymal	2	MDK	LRP1	MDK-LRP1	MK	0.0235	<0.001
AT2 injury/transitional -> Mesenchymal	3	MDK	SDC2	MDK-SDC2	MK	0.0232	<0.001
AT2 injury/transitional -> Mesenchymal	4	MDK	NCL	MDK-NCL	MK	0.0191	<0.001
AT2 injury/transitional -> Mesenchymal	5	LAMB3	CD44	LAMB3-CD44	LAMININ	0.0114	<0.001
AT2 injury/transitional -> Mesenchymal	6	LAMB3	ITGA1_ITGB1	LAMB3-(ITGA1+ITGB1)	LAMININ	0.0076	<0.001
AT2 injury/transitional -> Mesenchymal	7	GDF15	TGFBR2	GDF15-TGFBR2	GDF	0.0031	<0.001
AT2 injury/transitional -> Mesenchymal	8	LAMA5	CD44	LAMA5-CD44	LAMININ	0.0020	<0.001
AT2 injury/transitional -> Mesenchymal	9	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0016	<0.001
AT2 injury/transitional -> Mesenchymal	10	GAS6	AXL	GAS6-AXL	GAS	0.0014	<0.001

Outgoing interactions from injury/transitional AT2 cells to endothelial, macrophage, and mesenchymal cells are shown.

Table S7 Top outgoing ligand-receptor pairs from mature AT2 cells

Direction	Rank	Ligand	Receptor	Pair	Pathway	Prob.	P value
AT2 mature -> Endothelial	1	CXCL2	ACKR1	CXCL2-ACKR1	CXCL	0.1082	<0.001
AT2 mature -> Endothelial	2	CXCL8	ACKR1	CXCL8-ACKR1	CXCL	0.0313	<0.001
AT2 mature -> Endothelial	3	CXCL1	ACKR1	CXCL1-ACKR1	CXCL	0.0169	<0.001
AT2 mature -> Endothelial	4	CXCL3	ACKR1	CXCL3-ACKR1	CXCL	0.0149	<0.001
AT2 mature -> Endothelial	5	GDF15	TGFBR2	GDF15-TGFBR2	GDF	0.0143	<0.001
AT2 mature -> Endothelial	6	NAMPT	ITGA5_ITGB1	NAMPT-(ITGA5+ITGB1)	VISFATIN	0.0104	<0.001
AT2 mature -> Endothelial	7	CCL18	ACKR1	CCL18-ACKR1	CCL	0.0069	<0.001
AT2 mature -> Endothelial	8	MDK	NCL	MDK-NCL	MK	0.0066	<0.001
AT2 mature -> Endothelial	9	MDK	ITGA6_ITGB1	MDK-(ITGA6+ITGB1)	MK	0.0037	<0.001
AT2 mature -> Endothelial	10	VEGFA	FLT1	VEGFA-VEGFR1	VEGF	0.0017	<0.001
AT2 mature -> Macrophage	1	SCGB3A2	MARCO	SCGB3A2-MARCO	UGRP1	0.1245	<0.001
AT2 mature -> Macrophage	2	MIF	CD74_CD44	MIF-(CD74+CD44)	MIF	0.0569	<0.001
AT2 mature -> Macrophage	3	PPIA	BSG	PPIA-BSG	CypA	0.0539	<0.001
AT2 mature -> Macrophage	4	MIF	CD74_CXCR4	MIF-(CD74+CXCR4)	MIF	0.0376	<0.001
AT2 mature -> Macrophage	5	ANXA1	FPR1	ANXA1-FPR1	ANNEXIN	0.0098	<0.001
AT2 mature -> Macrophage	6	ANXA1	FPR3	ANXA1-FPR3	ANNEXIN	0.0095	<0.001
AT2 mature -> Macrophage	7	C3	ITGAX_ITGB2	C3-(ITGAX+ITGB2)	COMPLEMENT	0.0080	<0.001
AT2 mature -> Macrophage	8	ANXA1	FPR2_LXA4	ANXA1-(FPR2+LXA4)	ANNEXIN	0.0079	<0.001
AT2 mature -> Macrophage	9	C3	ITGAM_ITGB2	C3-(ITGAM+ITGB2)	COMPLEMENT	0.0066	<0.001
AT2 mature -> Macrophage	10	ANXA1	FPR2	ANXA1-FPR2	ANNEXIN	0.0065	<0.001
AT2 mature -> Mesenchymal	1	PPIA	BSG	PPIA-BSG	CypA	0.0670	<0.001
AT2 mature -> Mesenchymal	2	MDK	LRP1	MDK-LRP1	MK	0.0066	<0.001
AT2 mature -> Mesenchymal	3	MDK	SDC2	MDK-SDC2	MK	0.0066	<0.001
AT2 mature -> Mesenchymal	4	MDK	NCL	MDK-NCL	MK	0.0054	<0.001
AT2 mature -> Mesenchymal	5	GDF15	TGFBR2	GDF15-TGFBR2	GDF	0.0037	<0.001
AT2 mature -> Mesenchymal	6	LAMB3	CD44	LAMB3-CD44	LAMININ	0.0030	<0.001
AT2 mature -> Mesenchymal	7	LAMB3	ITGA1_ITGB1	LAMB3-(ITGA1+ITGB1)	LAMININ	0.0020	<0.001
AT2 mature -> Mesenchymal	8	GAS6	AXL	GAS6-AXL	GAS	0.0006	<0.001

Matched mature AT2 outgoing interactions are shown as the comparator for injury/transitional AT2 cells.

Table S8 Comparative outgoing communication to Endothelial cells: injury/transitional AT2 versus mature AT2

Interaction	Ligand	Receptor	Pathway	Injury prob.	Mature prob.	Delta prob.	Fold change	Injury P	Mature P	Directionality
CXCL8-ACKR1	CXCL8	ACKR1	CXCL	0.0885	0.0313	0.0573	2.83	<0.001	<0.001	Higher in injury AT2
CXCL1-ACKR1	CXCL1	ACKR1	CXCL	0.0590	0.0169	0.0420	3.48	<0.001	<0.001	Higher in injury AT2
CXCL2-ACKR1	CXCL2	ACKR1	CXCL	0.0683	0.1082	-0.0399	0.63	<0.001	<0.001	Lower in injury AT2
MDK-NCL	MDK	NCL	MK	0.0233	0.0066	0.0167	3.53	<0.001	<0.001	Higher in injury AT2
MDK-(ITGA6+ITGB1)	MDK	ITGA6_ITGB1	MK	0.0132	0.0037	0.0095	3.56	<0.001	<0.001	Higher in injury AT2
CXCL6-ACKR1	CXCL6	ACKR1	CXCL	0.0060	0.0000	0.0060	6036.26	<0.001	NA	Injury AT2 only
LAMB3-(ITGA6+ITGB1)	LAMB3	ITGA6_ITGB1	LAMININ	0.0061	0.0016	0.0045	3.77	<0.001	<0.001	Higher in injury AT2
CXCL3-ACKR1	CXCL3	ACKR1	CXCL	0.0118	0.0149	-0.0030	0.80	<0.001	<0.001	Lower in injury AT2
GDF15-TGFBR2	GDF15	TGFBR2	GDF	0.0118	0.0143	-0.0025	0.83	<0.001	<0.001	Lower in injury AT2
IGFBP3-TMEM219	IGFBP3	TMEM219	IGFBP	0.0018	0.0000	0.0018	1830.36	<0.001	NA	Injury AT2 only
LAMA5-(ITGA6+ITGB1)	LAMA5	ITGA6_ITGB1	LAMININ	0.0011	0.0000	0.0011	1076.36	<0.001	NA	Injury AT2 only
NAMPT-(ITGA5+ITGB1)	NAMPT	ITGA5_ITGB1	VISFATIN	0.0097	0.0104	-0.0007	0.94	<0.001	<0.001	Lower in injury AT2
CCL18-ACKR1	CCL18	ACKR1	CCL	0.0064	0.0069	-0.0005	0.93	<0.001	<0.001	Lower in injury AT2
VEGFA-VEGFR1	VEGFA	FLT1	VEGF	0.0017	0.0017	0.0000	1.03	<0.001	<0.001	Higher in injury AT2

Delta prob. was calculated as injury/transitional AT2 probability minus mature AT2 probability. Fold change was calculated as (injury probability + 1e-6)/(mature probability + 1e-6).

https://dx.doi.org/10.21037/jtd-2026-0896

Table S9 Comparative outgoing communication to Macrophage cells: injury/transitional AT2 versus mature AT2

Interaction	Ligand	Receptor	Pathway	Injury prob.	Mature prob.	Delta prob.	Fold change	Injury P	Mature P	Directionality
SCGB3A2-MARCO	SCGB3A2	MARCO	UGRP1	0.0116	0.1245	-0.1129	0.09	<0.001	<0.001	Lower in injury AT2
LAMB3-CD44	LAMB3	CD44	LAMININ	0.0211	0.0056	0.0154	3.73	<0.001	<0.001	Higher in injury AT2
MDK-NCL	MDK	NCL	MK	0.0182	0.0051	0.0131	3.55	<0.001	<0.001	Higher in injury AT2
MDK-SDC4	MDK	SDC4	MK	0.0092	0.0026	0.0066	3.57	<0.001	<0.001	Higher in injury AT2
MDK-SDC2	MDK	SDC2	MK	0.0078	0.0022	0.0056	3.57	<0.001	<0.001	Higher in injury AT2
C3-(ITGAX+ITGB2)	C3	ITGAX_ITGB2	COMPLEMENT	0.0032	0.0080	-0.0049	0.40	<0.001	<0.001	Lower in injury AT2
MDK-LRP1	MDK	LRP1	MK	0.0059	0.0016	0.0042	3.58	<0.001	<0.001	Higher in injury AT2
C3-(ITGAM+ITGB2)	C3	ITGAM_ITGB2	COMPLEMENT	0.0026	0.0066	-0.0040	0.40	<0.001	<0.001	Lower in injury AT2
ANXA1-FPR1	ANXA1	FPR1	ANNEXIN	0.0136	0.0098	0.0038	1.38	<0.001	<0.001	Higher in injury AT2
LAMA5-CD44	LAMA5	CD44	LAMININ	0.0038	0.0000	0.0038	3766.05	<0.001	NA	Injury AT2 only
ANXA1-FPR3	ANXA1	FPR3	ANNEXIN	0.0131	0.0095	0.0036	1.38	<0.001	<0.001	Higher in injury AT2
ANXA1-(FPR2+LXA4)	ANXA1	FPR2_LXA4	ANNEXIN	0.0109	0.0079	0.0030	1.38	<0.001	<0.001	Higher in injury AT2
IGFBP3-TMEM219	IGFBP3	TMEM219	IGFBP	0.0027	0.0000	0.0027	2669.21	<0.001	NA	Injury AT2 only
ANXA1-FPR2	ANXA1	FPR2	ANNEXIN	0.0090	0.0065	0.0025	1.38	<0.001	<0.001	Higher in injury AT2
C3-C3AR1	C3	C3AR1	COMPLEMENT	0.0016	0.0040	-0.0024	0.39	<0.001	<0.001	Lower in injury AT2
SAA1-FPR2	SAA1	FPR2	SAA	0.0013	0.0000	0.0013	1251.32	<0.001	NA	Injury AT2 only
PPIA-BSG	PPIA	BSG	CypA	0.0548	0.0539	0.0009	1.02	<0.001	<0.001	Higher in injury AT2
GRN-SORT1	GRN	SORT1	GRN	0.0030	0.0021	0.0008	1.40	<0.001	<0.001	Higher in injury AT2
GAS6-AXL	GAS6	AXL	GAS	0.0009	0.0004	0.0005	2.34	<0.001	<0.001	Higher in injury AT2
GDF15-TGFBR2	GDF15	TGFBR2	GDF	0.0019	0.0023	-0.0004	0.83	<0.001	<0.001	Lower in injury AT2
NAMPT-(ITGA5+ITGB1)	NAMPT	ITGA5_ITGB1	VISFATIN	0.0037	0.0039	-0.0003	0.93	<0.001	<0.001	Lower in injury AT2
MIF-(CD74+CD44)	MIF	CD74_CD44	MIF	0.0570	0.0569	0.0001	1.00	<0.001	<0.001	Higher in injury AT2
MIF-(CD74+CXCR4)	MIF	CD74_CXCR4	MIF	0.0377	0.0376	0.0001	1.00	<0.001	<0.001	Higher in injury AT2

Delta prob. was calculated as injury/transitional AT2 probability minus mature AT2 probability. Fold change was calculated as (injury probability + 1e-6)/(mature probability + 1e-6).

Table S10 Comparative outgoing communication to Mesenchymal cells: injury/transitional AT2 versus mature AT2

Interaction	Ligand	Receptor	Pathway	Injury prob.	Mature prob.	Delta prob.	Fold change	Injury P	Mature P	Directionality
MDK-LRP1	MDK	LRP1	MK	0.0235	0.0066	0.0168	3.53	<0.001	<0.001	Higher in injury AT2
MDK-SDC2	MDK	SDC2	MK	0.0232	0.0066	0.0167	3.53	<0.001	<0.001	Higher in injury AT2
MDK-NCL	MDK	NCL	MK	0.0191	0.0054	0.0137	3.54	<0.001	<0.001	Higher in injury AT2
LAMB3-CD44	LAMB3	CD44	LAMININ	0.0114	0.0030	0.0084	3.75	<0.001	<0.001	Higher in injury AT2
LAMB3-(ITGA1+ITGB1)	LAMB3	ITGA1_ITGB1	LAMININ	0.0076	0.0020	0.0056	3.76	<0.001	<0.001	Higher in injury AT2
LAMA5-CD44	LAMA5	CD44	LAMININ	0.0020	0.0000	0.0020	2021.94	<0.001	NA	Injury AT2 only
IGFBP3-TMEM219	IGFBP3	TMEM219	IGFBP	0.0016	0.0000	0.0016	1568.67	<0.001	NA	Injury AT2 only
LAMA5-(ITGA1+ITGB1)	LAMA5	ITGA1_ITGB1	LAMININ	0.0013	0.0000	0.0013	1339.78	<0.001	NA	Injury AT2 only
PPIA-BSG	PPIA	BSG	CypA	0.0681	0.0670	0.0011	1.02	<0.001	<0.001	Higher in injury AT2
GAS6-AXL	GAS6	AXL	GAS	0.0014	0.0006	0.0008	2.34	<0.001	<0.001	Higher in injury AT2
GDF15-TGFBR2	GDF15	TGFBR2	GDF	0.0031	0.0037	-0.0006	0.83	<0.001	<0.001	Lower in injury AT2

Delta prob. was calculated as injury/transitional AT2 probability minus mature AT2 probability. Fold change was calculated as (injury probability + 1e-6)/(mature probability + 1e-6).

Table S11 IGFBP pathway ligand-receptor pairs detected in the CellChat network

Source	Target	Ligand	Receptor	Pair	Pathway	Prob.	P value
AT2 injury/transitional	Airway epithelial	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0029	<0.001
AT2 injury/transitional	Macrophage	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0027	<0.001
AT2 injury/transitional	Endothelial	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0018	<0.001
AT2 injury/transitional	AT2 mature	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0018	<0.001
AT2 injury/transitional	AT2 injury/transitional	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0017	<0.001
AT2 injury/transitional	Mesenchymal	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0016	<0.001
AT2 injury/transitional	Mast	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0009	<0.001
AT2 injury/transitional	Monocyte	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0008	<0.001
AT2 injury/transitional	Plasma	IGFBP3	TMEM219	IGFBP3-TMEM219	IGFBP	0.0003	<0.001

No robust IGFBP2-specific ligand-receptor pair was detected in the provided CellChat network. The IGFBP pathway pairs detected here involved IGFBP3-TMEM219 and should be interpreted at the pathway level rather than as direct IGFBP2-mediated intercellular signaling.

Table S12 Parameter settings for scTenifoldKnk-based virtual knockout analysis

Item	Setting used in this study	Rationale/note
Analysis purpose	In-silico virtual knockout of IGFBP2	Used to infer candidate regulatory changes associated with IGFBP2 perturbation in injury/transitional AT2 cells
Software / package	scTenifoldKnk	Gene-regulatory-network-based single-cell virtual knockout framework
Input object	AT2_subcluster.rds	Refined AT2 subcluster object used for perturbation analysis
Target cell population	Injury/transitional AT2 cells	Cells intended to be selected using AT2_state == injury/transitional
Metadata column for cell-state selection	AT2_state	Used to define mature and injury/transitional AT2 states
State retained for analysis	injury/transitional	Only injury/transitional AT2 cells were included in the virtual knockout analysis
Target gene for virtual knockout	IGFBP2	Core candidate prioritized by the integrative analysis
Random seed	123	Set before analysis for reproducibility
Input assay	RNA assay	Expression data were extracted from the Seurat RNA assay
Input matrix	Raw UMI count matrix	Extracted using GetAssayData
Normalization before HVG selection	Seurat NormalizeData; LogNormalize default	Used before highly variable gene selection
HVG selection method	vst	Performed using Seurat FindVariableFeatures
Number of HVGs	5,000	HVGs were combined with IGFBP2 for the input gene set
Target gene retention	Forced retention of IGFBP2	IGFBP2 was retained in the input matrix even if excluded by filters
Low-expression gene filter	Genes expressed in at least 30 cells	min_cells_expr =30
Zero-variance gene filter	SD > 1e-12	min_sd_eps =1e-12 to avoid unstable correlations
Mitochondrial QC threshold	0.1	qc_mtThreshold =0.1 in scTenifoldKnk
Minimum library size	1,000	qc_minLSize =1,000 in scTenifoldKnk
Number of reconstructed networks	25	nc_nNet =25
Cells sampled per network	Up to 1,000 cells	nc_nCells_use = min(1000, number of available cells)
Virtual knockout argument	gKO = IGFBP2	IGFBP2 was passed as the computational knockout target
Differential regulation output	res$diffRegulation	Source table for allDiff.txt and sigDiff.txt
All tested genes after virtual knockout	3,168	Number of genes in allDiff.txt after excluding the target gene
Significant genes	5	Genes with adjusted P value <0.05 in sigDiff.txt
Significance cutoff	Adjusted P value < 0.05	Used to define significant knockout-inferred genes
Main output files	allDiff.txt; sigDiff.txt; barplot.pdf; vol.pdf; pie.pdf	Outputs used for downstream summary and visualization
Interpretation boundary	Network-inferred perturbation; not experimental knockout	Results indicate computational regulatory associations rather than causal wet-lab validation

The parameter values are based on the provided scTenifoldKnk script. The virtual knockout analysis should be interpreted as gene-regulatory-network-inferred perturbation, not as experimental knockout evidence.

https://dx.doi.org/10.21037/jtd-2026-0896

Table S13 Significant genes inferred after IGFBP2 virtual knockout

Gene	Distance	Z score	FC	P value	Adjusted P value	Functional interpretation
SCGB3A1	2.259e-05	4.793	2408.127	3.424e-95	5.387e-89	Secretory epithelial program; secretoglobin family member associated with airway/epithelial secretory differentiation
SCGB1A1	9.037e-06	3.961	385.398	8.312e-86	8.780e-83	Secretory epithelial program; canonical club/secretory epithelial marker
BPIFB1	3.649e-06	3.229	62.851	2.229e-15	1.766e-12	Mucosal defense and epithelial barrier-associated gene
SLPI	2.649e-06	2.991	33.117	8.678e-09	5.500e-06	Anti-protease, epithelial barrier, and inflammatory-control related gene
PIGR	2.236e-06	2.869	23.601	1.185e-06	6.259e-04	Mucosal immunity and epithelial immunoglobulin transport

Significant genes were defined using adjusted P value <0.05 from res$diffRegulation. Only five genes met this threshold in sigDiff.txt. FC and Z score are scTenifoldKnk output statistics and should be interpreted as network-inferred regulatory perturbation metrics.

https://dx.doi.org/10.21037/jtd-2026-0896

Table S14 ADMET prediction results of parent molecules

Compound-related parameters/ names	Values/information
smiles	O=C(c1ccc(-c2cccc3cccnc23) cc1)N1C[C@H](CO)O[C@@H] (CO)C1
LogS	-2.608
LogD	1.298
LogP	1.415
Pgp-inh	0.347
Pgp-sub	0.005
HIA	0.025
F(20%)	0.948
F(30%)	0.465
Caco-2	-5.119
MDCK	6.1736313909932505e-06
BBB	0.445
PPB	85.19%
VDss	0.969
Fu	7.02%
CYP1A2-inh	0.323
CYP1A2-sub	0.072
CYP2C19-inh	0.487
CYP2C19-sub	0.099
CYP2C9-inh	0.528
CYP2C9-sub	0.236
CYP2D6-inh	0.04
CYP2D6-sub	0.507
CYP3A4-inh	0.52
CYP3A4-sub	0.484
CL	1.803
T12	0.312
hERG	0.469
H-HT	0.989
DILI	0.988
Ames	0.631
ROA	0.199
FDAMDD	0.312
SkinSen	0.274
Carcinogenicity	0.884
EC	0.003
EI	0.025
Respiratory	0.028
BCF	0.446
IGC50	3.616
LC50	4.565
LC50DM	4.42
NR-AR	0.056
NR-AR-LBD	0.212
NR-AhR	0.783
NR-Aromatase	0.72
NR-ER	0.166
NR-ER-LBD	0.008
NR-PPAR-gamma	0.789
SR-ARE	0.828
SR-ATAD5	0.058
SR-HSE	0.926
SR-MMP	0.432
SR-p53	0.783
MW	378.16
Vol	388.269
Dense	0.974
nHA	6
nHD	2
TPSA	82.89
nRot	5
nRing	4
MaxRing	10
nHet	6
fChar	0
nRig	24
Flex	0.208
nStereo	2
NonBiodegradable	2
NonGenotoxic_Carcinogenicity	0
SureChEMBL	0
LD50_oral	0
Skin_Sensitization	0
Acute_Aquatic_Toxicity	0
Toxicophores	1
Genotoxic_Carcinogenicity_ Mutagenicity	0
QED	0.727
Synth	3.051
Fsp3	0.273
MCE-18	69
Natural Product-likeness	-0.589
Alarm_NMR	0
BMS	0
Chelating	0
PAINS	0
Lipinski	Accepted
Pfizer	Accepted
GSK	Accepted
GoldenTriangle	Accepted

Table S15 ADMET prediction results of improved molecules

Compound-related parameters/ names	Values/Information
SMILES	CCOC(=O)OC1CCC(O) CC1C1CCC(C)(C(=O)N2C[C@H] (CO)O[C@@H](CO)C2)CC1
LogS	-2.214
LogD	1.04
LogP	1.069
Pgp-inh	0.025
Pgp-sub	0.818
HIA	0.005
F(20%)	0.007
F(30%)	0.058
Caco-2	-5.016
MDCK	0.0001496721088187769
BBB	0.082
PPB	52.34%
VDss	0.949
Fu	41.71%
CYP1A2-inh	0.006
CYP1A2-sub	0.076
CYP2C19-inh	0.024
CYP2C19-sub	0.788
CYP2C9-inh	0.01
CYP2C9-sub	0.069
CYP2D6-inh	0.01
CYP2D6-sub	0.163
CYP3A4-inh	0.43
CYP3A4-sub	0.761
CL	6.618
T12	0.433
hERG	0.02
H-HT	0.886
DILI	0.141
Ames	0.506
ROA	0.007
FDAMDD	0.946
SkinSen	0.233
Carcinogenicity	0.163
EC	0.003
EI	0.009
Respiratory	0.032
BCF	0.344
IGC50	1.934
LC50	3.601
LC50DM	4.123
NR-AR	0.816
NR-AR-LBD	0.188
NR-AhR	0.022
NR-Aromatase	0.02
NR-ER	0.063
NR-ER-LBD	0.015
NR-PPAR-gamma	0.18
SR-ARE	0.651
SR-ATAD5	0.006
SR-HSE	0.876
SR-MMP	0.237
SR-p53	0.619
MW	457.27
Vol	456.74
Dense	1.001
nHA	9
nHD	3
TPSA	125.76
nRot	9
nRing	3
MaxRing	6
nHet	9
fChar	0
nRig	20
Flex	0.45
nStereo	5
NonBiodegradable	1
NonGenotoxic_Carcinogenicity	0
SureChEMBL	0
LD50_oral	0
Skin_Sensitization	0
Acute_Aquatic_Toxicity	0
Toxicophores	0
Genotoxic_Carcinogenicity_ Mutagenicity	0
QED	0.512
Synth	4.457
Fsp3	0.913
MCE-18	57.273
Natural Product-likeness	1.005
Alarm_NMR	0
BMS	0
Chelating	0
PAINS	0
Lipinski	Accepted
Pfizer	Accepted
GSK	Rejected
GoldenTriangle	Accepted

https://dx.doi.org/10.21037/jtd-2026-0896

Table S16 DrugCLIP web-screening and compound prioritization settings for IGFBP2

Item	Setting	Rationale/note
Analysis aim	Exploratory virtual screening and structural assessment of potential small-molecule accessibility of IGFBP2	Used to provide preliminary computational clues, not therapeutic validation
Screening platform	DrugCLIP web platform	The platform was used as a hosted screening service; local model retraining was not performed
DrugCLIP core principle	Contrastive protein-molecule representation learning; virtual screening formulated as dense retrieval between a protein pocket query and candidate molecules	Summarized from the original DrugCLIP publication
Target protein	Human IGFBP2	Core candidate gene prioritized for state-focused downstream analyses
Protein structure	AlphaFold AF-P18065-F1-v6	Used as the IGFBP2 structural input
Protein input used for DrugCLIP	Predicted IGFBP2 3D structure / pocket representation	DrugCLIP encodes 3D pocket information for similarity-based retrieval
Compound library	FCHGroup_SC_202007; Enamine_screening_collection_sdf_202504; Princeton_BioMolecular_Research	The top candidate compounds reported in Table 1 were derived from these DrugCLIP-accessible compound sources. Library names were retained as returned by the DrugCLIP screening output
Binding-pocket selection	Highest-scoring DrugCLIP-predicted protein binding pocket	Selected as the pocket for subsequent docking-center definition
Primary ranking metric	DrugCLIP score	Candidate compounds were first ranked by the DrugCLIP score
Initial compound output	Top candidate compounds ranked by DrugCLIP score	Top candidates are reported in the main manuscript table.
Secondary refinement	Molecular docking score	Top candidates were further evaluated by AutoDock Vina docking
Top candidates for docking	Top 10 compounds ranked by DrugCLIP score	Used for subsequent docking evaluation.
ADMET / drug-likeness assessment	ADMETlab 2.0	Used for predicted toxicity and drug-likeness assessment before and after structural optimization
Structural optimization	Manual structure-guided optimization of the top scaffold	Performed to reduce predicted toxicity and retain drug-likeness; should be interpreted as exploratory

Table S17 DrugCLIP model principle and published validation information

Item	Setting	Rationale/note
Model formulation	Dense retrieval framework	DrugCLIP treats a protein pocket as the query and retrieves high-similarity molecules from a compound library
Input modalities in DrugCLIP	3D protein pocket and 3D molecule representations	The original model tokenizes pocket and molecule atoms and uses atom types and coordinates
Encoder architecture	UniMol-style 3D encoder/SE(3)-aware transformer architecture	Summarized from the DrugCLIP publication; not re-trained in this study
Similarity scoring	Dot product or cosine similarity between pocket and molecule embeddings	The web-screening output was used as a DrugCLIP score-based ranking
Training objective	Bidirectional contrastive learning using pocket-to-molecule and molecule-to-pocket losses	Model-level information from the original DrugCLIP publication
Negative-pair construction	In-batch negative sampling	Model-level information from the original DrugCLIP publication
Training datasets reported for DrugCLIP	PDBBind 2019, BioLip, and ChEMBL	Training data reported in the original DrugCLIP publication
Data augmentation	HomoAug and RDKit-generated molecule conformations	Model-level information from the original DrugCLIP publication
Validation benchmarks reported for DrugCLIP	DUD-E and LIT-PCBA	Used to summarize the method's reported benchmarking, not to revalidate this study's IGFBP2 candidates
Reported validation metrics	AUROC, BEDROC, enrichment factor, and ROC enrichment	Metrics reported in the original DrugCLIP publication
DUD-E zero-shot DrugCLIP performance	AUROC =80.93%; BEDROC =50.52%; EF 0.5% =38.07; EF 1% =31.89; EF 5% =10.66	Reported benchmark performance; not generated in this study
LIT-PCBA DrugCLIP performance	AUROC 57.17%; BEDROC =6.23%; EF 0.5% =8.56; EF 1% =5.51; EF 5% =2.27	Reported benchmark performance; not generated in this study
Local model training parameters	Not applicable	The hosted DrugCLIP web platform was used; no local model training, hyperparameter tuning, or re-training was performed
Model validation in this study	Not independently re-trained or benchmarked	The study used DrugCLIP output for exploratory candidate prioritization only

Table S18 Docking, ADMET, and molecular dynamics settings

Item	Setting	Rationale/note
Docking software	AutoDock Vina	Used for docking of the top DrugCLIP-ranked candidates
Docking grid center	x = -6.99, y = -2.38, z = 12.33	Defined according to the predicted center of the optimal DrugCLIP binding pocket
Docking grid size	47 x 47 x 47 Angstrom	Used for AutoDock Vina docking
Exhaustiveness	32	AutoDock Vina search parameter
Random seed	20250101	Used for docking reproducibility
Docking pose selection	Pose with the lowest Vina score	Selected for downstream molecular dynamics simulation
Molecular dynamics software	GROMACS 2024.4	Used for the 200-ns production simulation
Production simulation length	200 ns	Exploratory MD assessment of the IGFBP2-ligand complex
Force field	Amber99-ILDN	Used for system parameterization
Water model	TIP3P	Used for solvent modeling
Ion treatment	Na^+/Cl^- ions added for charge neutralization	Used before MD production simulation
Equilibration workflow	Energy minimization followed by NVT and NPT equilibration	Used before production MD
Trajectory saving interval	10 ps	Used for downstream trajectory analysis
MD readouts	RMSD, RMSF, radius of gyration, and solvent-accessible surface area	Used to evaluate structural stability and compactness
Interpretation boundary	Exploratory structural assessment only	Does not constitute biochemical validation or therapeutic evidence

DrugCLIP was used through the web platform as a hosted screening service. Therefore, local model-training hyperparameters were not modified or re-estimated in this study. Model principles and reported benchmark metrics are summarized from the original DrugCLIP publication. The structural analysis was exploratory and should not be interpreted as biochemical or therapeutic validation.

https://dx.doi.org/10.21037/jtd-2026-0896