Supplementary Methods

We benchmarked SigFormer against nine existing computational methods, including four widely used tools for intercellular ligand–receptor inference (CellChat[1], CellPhoneDB[2], and ICELLNET[3], LRLoop[4]) and five methods capable of inferring both ligand–receptor interactions and downstream intracellular signaling (NicheNet[5], exFINDER[6], scSeqComm[7], CellCall[8], and scMLnet[9]). Benchmarking was performed using scRNA-seq datasets from eight tumor types—glioblastoma, leukemia, pancreatic cancer, breast cancer, colorectal cancer, lung cancer, melanoma, and ovarian cancer—obtained from the Curated Cancer Cell Atlas (3CA)[10]. Preprocessed scRNA-seq datasets (in h5ad format) are provided with the SigFormer toolkit. Detailed parameter settings for each method are described below.

SigFormer

SigFormer was applied to each cancer scRNA-seq dataset with the following parameters:

- scRNAseq_path = "./test data/scRNA-seq";
- scProteomics_path = None when benchmarking against existing methods, "./test data/ scProteomics" when constructing CCPA;
- scATACseq_path = None when benchmarking against existing methods, "./test data/ scATAC-seq" when constructing CCPA;
- pathway_file = "./reference library/Intracellular signaling.txt";
- ligand_file = "./reference library/Ligand_secreted&membrane.txt";
- index_cell = "Malignant" (index cell type);

- min_cell = 0.01 (genes expressed in <1% of cells removed);
- min_gene = 0.01 (cells with <1% of genes detected removed);
- normalize = TRUE (library-size normalization);
- log_trans = TRUE (natural logarithm transformation);
- hvg_top_gene = 5000 (top highly variable genes);
- cell_top_gene = 500 (top expressed genes per cell);
- spatial = FALSE when benchmarking against existing methods, TRUE when evaluating the impact of spatial information;
- knn = 10 (number of nearest neighbors per index cell when spatial = TRUE);
- classification_accuracy = 0.8 (cell classification threshold);
- edge_threshold = 0.8 (edge reconstruction threshold);
- min_cell_count = 5 (minimum cells per cell type);
- max_length = 5 (maximum pathway length);
- num_epochs = 20 (number of training epochs);
- learning_rate = 0.001 (optimizer learning rate);
- block_size = 5000 (segment block size for large datasets);
- random_seed = 43

NicheNet

NicheNet is a framework for predicting intercellular communication by integrating gene expression data with prior knowledge of ligand–target signaling networks. We used the R implementation *nichenetr* (https://github.com/saeyslab/nichenetr) and the default precomputed networks provided by the authors, including lr_network.RData,

ligand_target_matrix.RData, sig_network.RData, and gr_network.RData. For each pair of interacting cell types A and B, genes significantly upregulated in each cell type were identified using the Wilcoxon signed-rank test (adjusted $P < 0.05$ and $\log_2$ fold change > 1). Default NicheNet functions were then applied to infer ligand–target regulatory networks between malignant cells and each non-malignant cell type.

LRLoop

LRLoop infers bidirectional cell–cell communication by identifying reciprocal ligand–receptor pairs forming closed feedback loops. We used the default prior networks provided by the authors, including lr_network.RData, ligand_target_matrix.RData, and receptor_target_matrix.RData. Differentially expressed genes were identified using the Wilcoxon signed-rank test (adjusted $P < 0.05$ and $\log_2$ fold change > 1). The PrepareBasics() function was then applied with default parameters to construct LRLoop network matrices between interacting cell types.

CellCall

CellCall (https://github.com/ShellyCoder/cellcall) infers intercellular communication and associated transcriptional regulatory networks. TPM expression matrices were used to construct expression objects via the CreateNichConObject() function. The TransCommuProfile() function was applied to compute ligand–receptor interaction scores and identify associated signaling pathways, with parameters set as follows: correlation threshold = 0.05, *P*-value threshold = 0.1, multiple testing correction threshold = 0.05, and weighted average signal aggregation. The LR2TF() function was used to infer ligand–receptor–TF relationships, and trans2tripleScore() was applied to

identify connected ligand–receptor–TF triplets. TF–target gene relationships were extracted from GSEA results to characterize downstream regulatory effects.

exFINDER

exFINDER identifies external signaling inputs in scRNA-seq data using prior pathway knowledge. We used the R implementation available at https://github.com/ChanghanGitHub/exFINDER. Human regulatory databases provided by the authors (LR_layer1_human.rda, RTF_layer2_human.rda, and TFT_layer3_human.rda) were used to construct multi-layer signaling networks. The get_potentialex() function was applied to identify candidate ligands across cell types (correlation threshold = −0.5), followed by the get_exSigNet() function to infer intercellular signaling networks between specific cell pairs (significance filtering threshold = 0.6).

scMLnet

scMLnet is a multilayer network method that integrates intercellular ligand–receptor interactions with intracellular receptor–TF and TF–target gene regulatory networks. We used the prior network files provided by the authors (LigRec.txt, RecTF.txt, and TFTargetGene.txt) from the scMLnet database. Differential expression thresholds were set to $P < 0.05$ and log fold change > 0.15.

scSeqComm

scSeqComm (https://gitlab.com/sysbiobig/scseqcomm) identifies and characterizes cellular communication within a single biological condition. Curated ligand–receptor pairs from LR_pairs_Kumar_2018 were used to infer intercellular interactions. Intracellular signaling was modeled using the integrated transcriptional regulatory

network TF_TG_TRRUSTv2_HTRIdb_RegNetwork_High together with KEGG pathway annotations.

ICELLNET

ICELLNET (https://github.com/soumelis-lab/ICELLNET) quantifies intercellular communication using curated ligand–receptor databases. Expression matrices were standardized for each cell type, scaled using the gene.scaling() function, and interaction scores were computed using icellnet.score(). Ligand–receptor pairs with interaction scores greater than zero were retained as biologically meaningful interactions.

CellChat

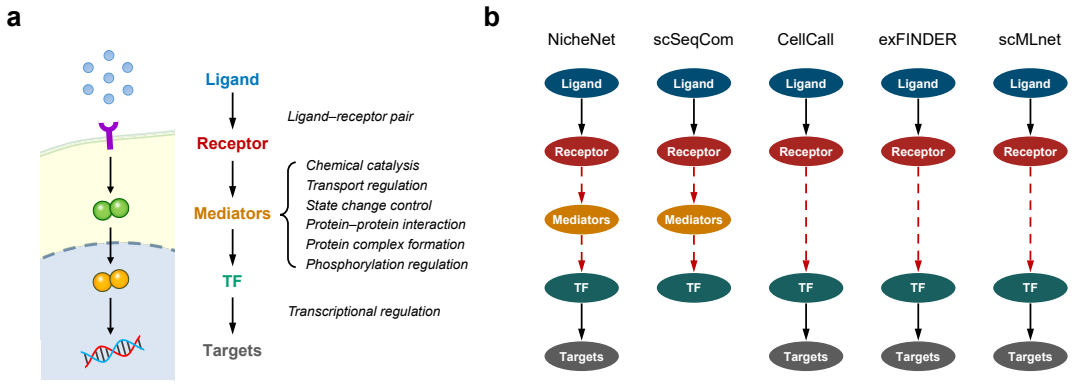
CellChat (https://github.com/sqjin/CellChat) infers intercellular communication using curated ligand–receptor signaling databases. We used the human CellChat database (CellChatDB.human) and included all three signal categories (secreted signaling, extracellular matrix receptor signaling, and cell–cell contact). Intercellular communication probabilities were computed using computeCommunProb() with the truncated mean method (trim = 0.1), and interactions were filtered using min.cells = 5.

CellPhoneDB

CellPhoneDB (https://github.com/ventolab/CellPhoneDB) is a Python-based framework for inferring ligand–receptor interactions between cell types. We used the built-in ligand–receptor database and performed statistical analysis using cpdb_statistical_analysis_method.call, with a cell expression ratio threshold of 0.1 and four parallel threads. Functional ligand–receptor pairs were retained, and bidirectional communication relationships between cell types were extracted along with their

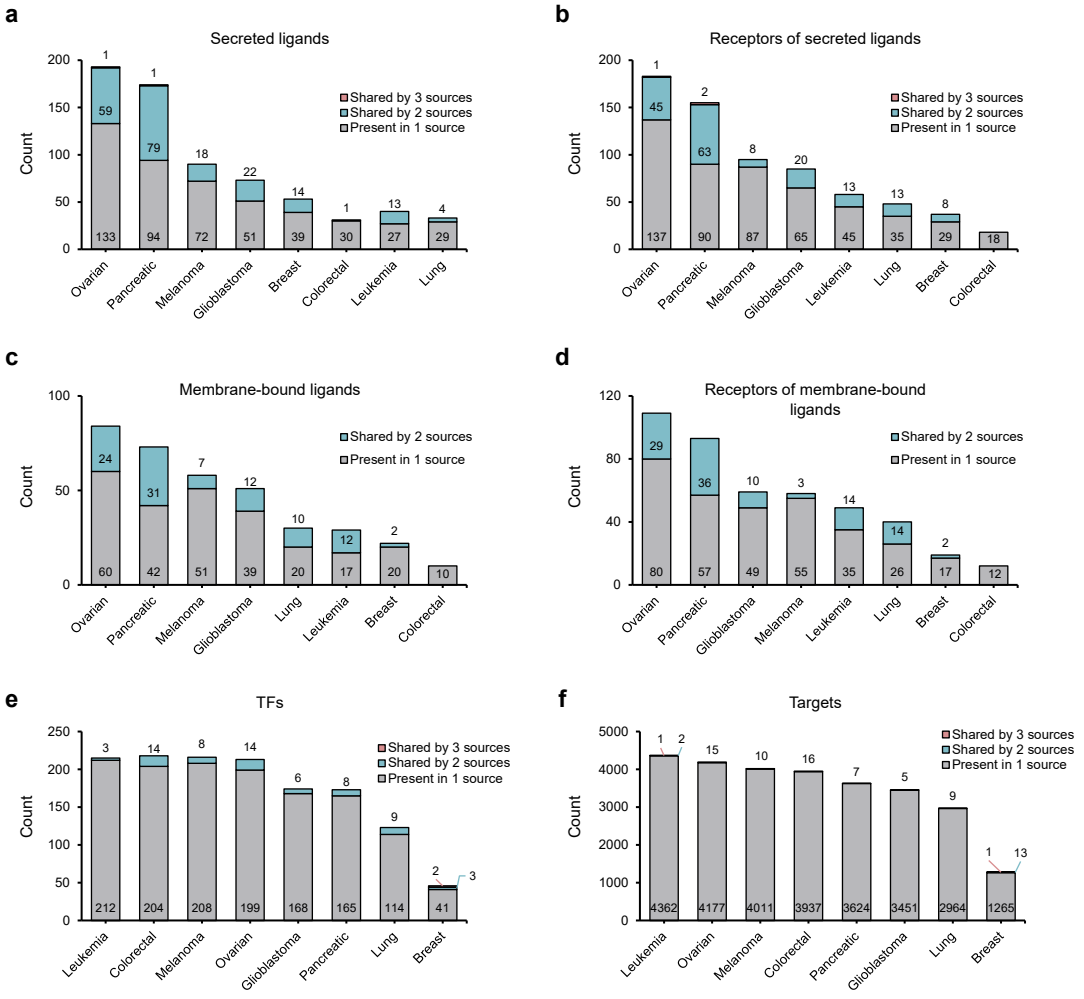
average expression levels.

Supplementary Figures

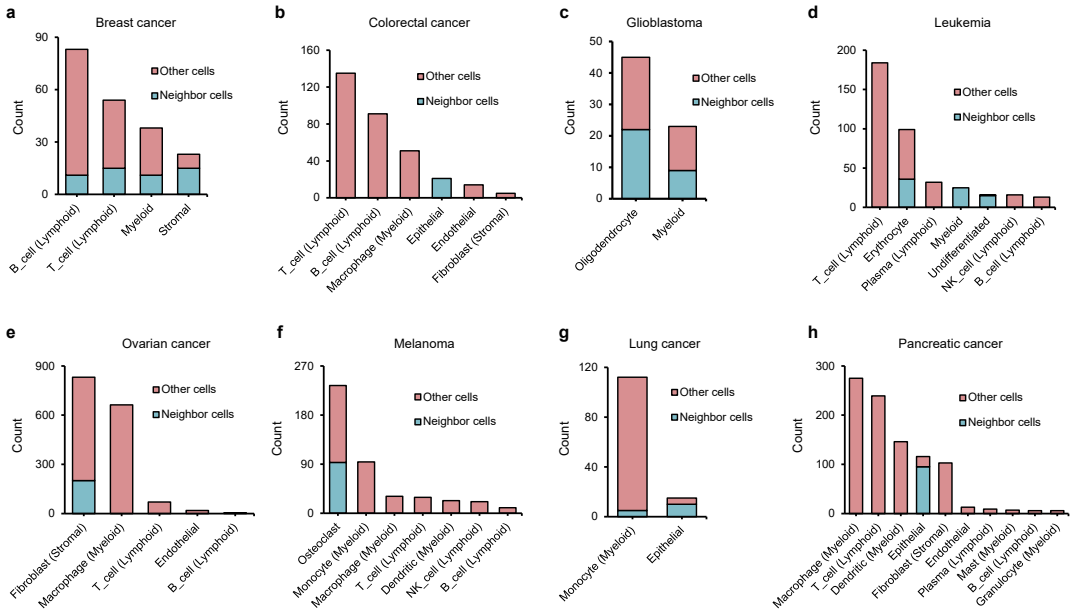


Extended Data Fig. 1 | Overview of transcellular signaling pathway reconstruction.

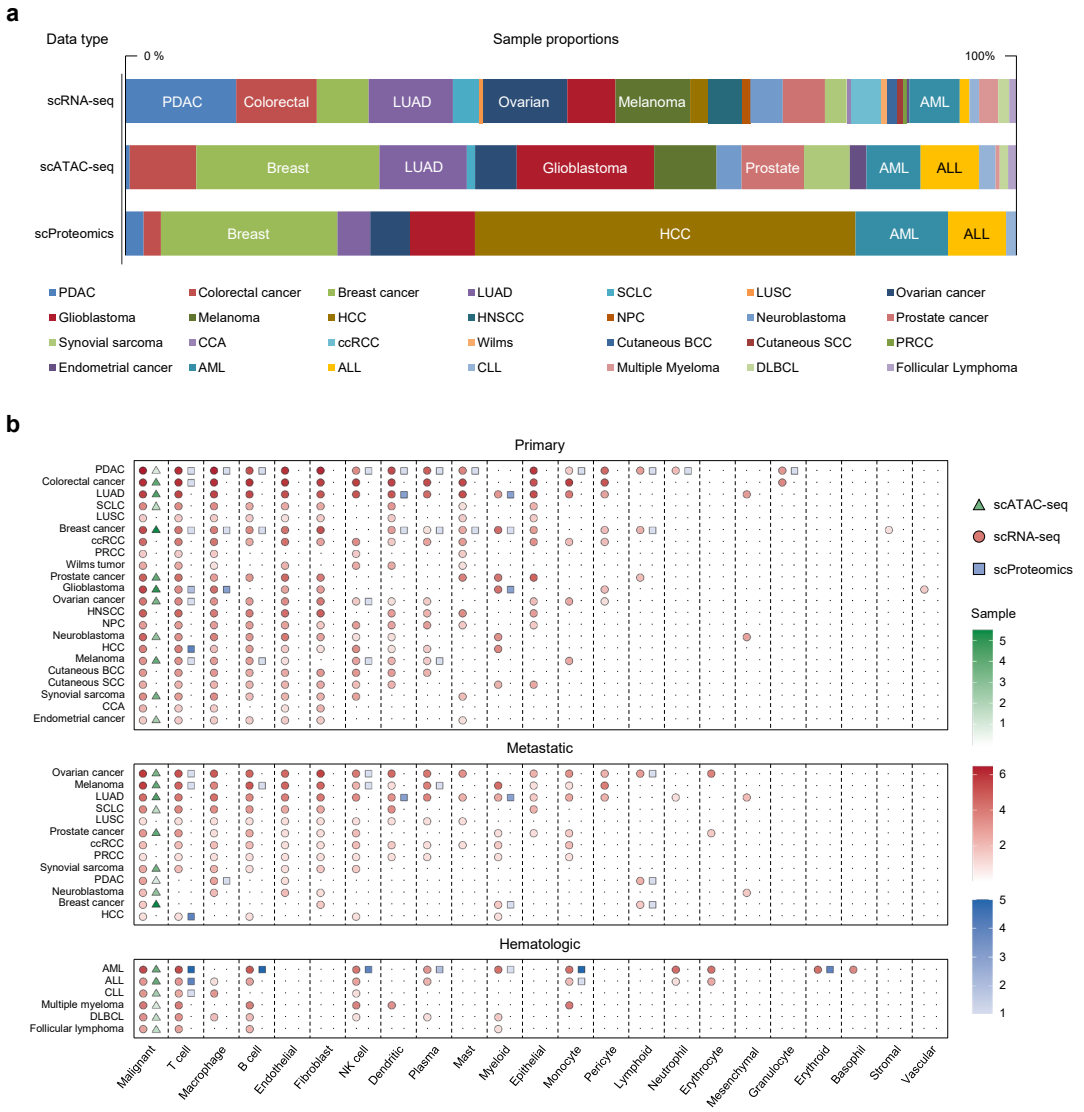
a, Representative stages in transcellular signaling pathways: ligand, receptor, intracellular signaling proteins, TF, and target genes. **b,** Stages supported by representative signaling pathway inference methods. Solid and dashed arrows represent direct and indirect relationships, respectively.



Extended Data Fig. 2 | Positive sets for pathway inference benchmark. a,b, Numbers of positive secreted ligands and their receptors across eight cancers. **c,d,** Numbers of positive membrane-bound ligands and their receptors. **e,f,** Numbers of positive TFs and their target genes. Signaling components are grouped by the number of independent cancer-related evidence sources supporting them: present in one source, shared by two sources, or shared by three sources.

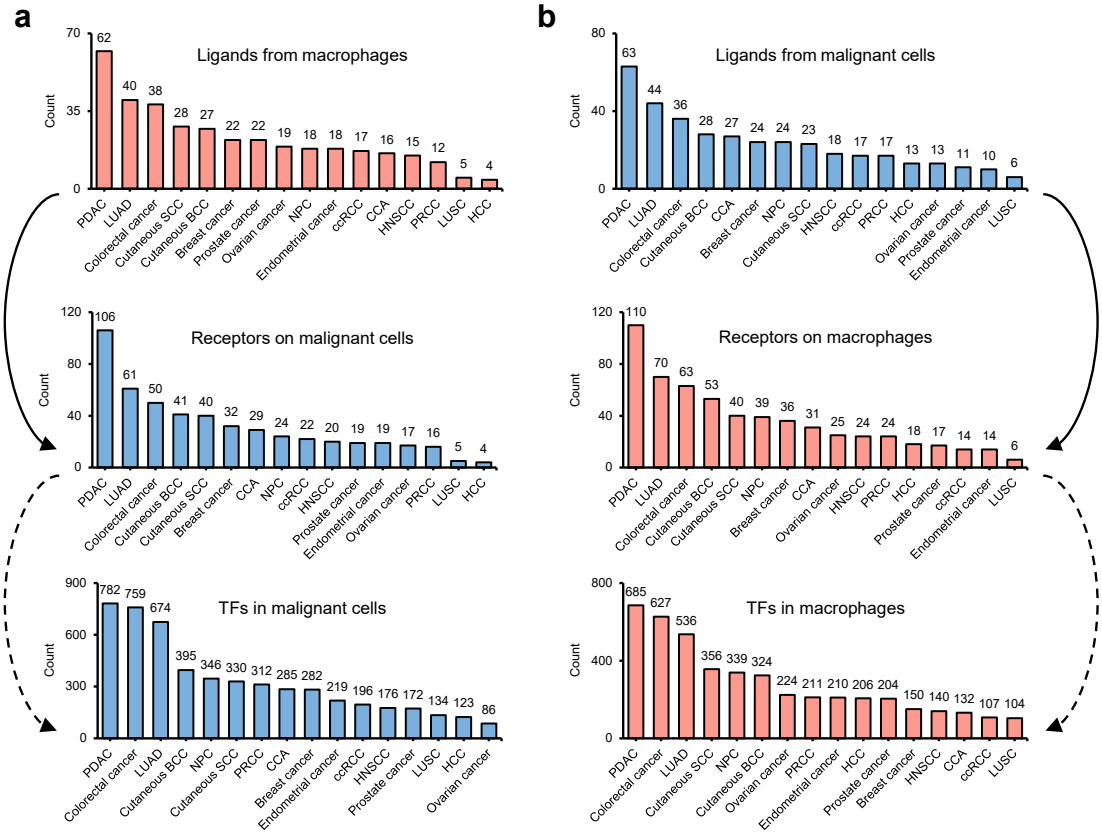


Extended Data Fig. 3 | Distribution of non-malignant cells in malignant cell neighborhoods. Proportion of each non-malignant cell type located within malignant cell neighborhoods relative to its total abundance across eight cancer types. Cyan indicates cells within malignant cell neighborhoods, whereas magenta indicates remaining cells.

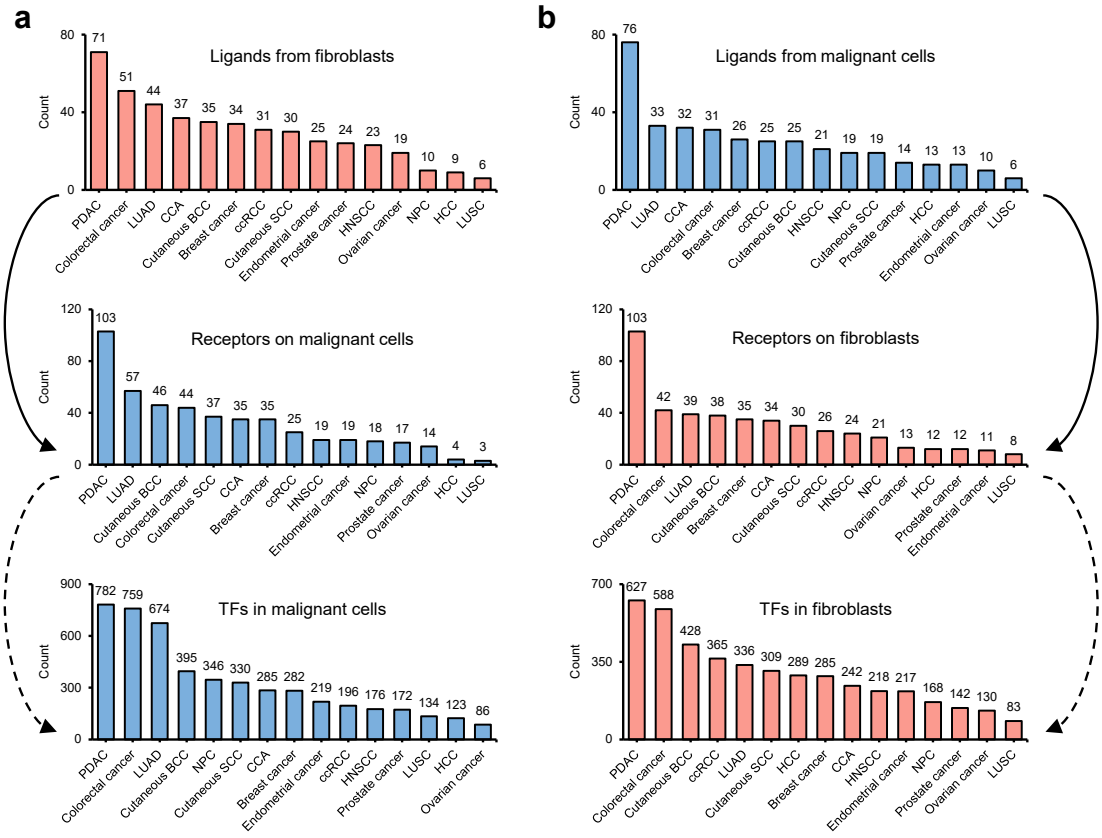


Extended Data Fig. 4 | Overview of single-cell multi-omics data. a, Proportions of tumor samples across different omics layers, including pancreatic ductal adenocarcinoma (PDAC), colorectal cancer, breast cancer, lung adenocarcinoma (LUAD), small cell lung cancer (SCLC), lung squamous cell carcinoma (LUSC), clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (PRCC), Wilms tumor, prostate cancer, glioblastoma, ovarian cancer, head and neck squamous cell carcinoma (HNSCC), nasopharyngeal carcinoma (NPC), neuroblastoma, hepatocellular carcinoma (HCC), melanoma, cutaneous basal cell carcinoma (BCC), cutaneous

squamous cell carcinoma (SCC), synovial sarcoma, cholangiocarcinoma (CCA), endometrial cancer, acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), multiple myeloma, diffuse large B-cell lymphoma (DLBCL), and follicular lymphoma. **b,** Tumor and cell types represented in scRNA-seq (red circles), scATAC-seq (green triangles), and scProteomics (blue squares) datasets. Samples were grouped into 22 primary tumors, 13 metastatic tumors, and 6 hematologic tumors. Color intensity indicates the log2-transformed number of samples in which each cell type was detected.

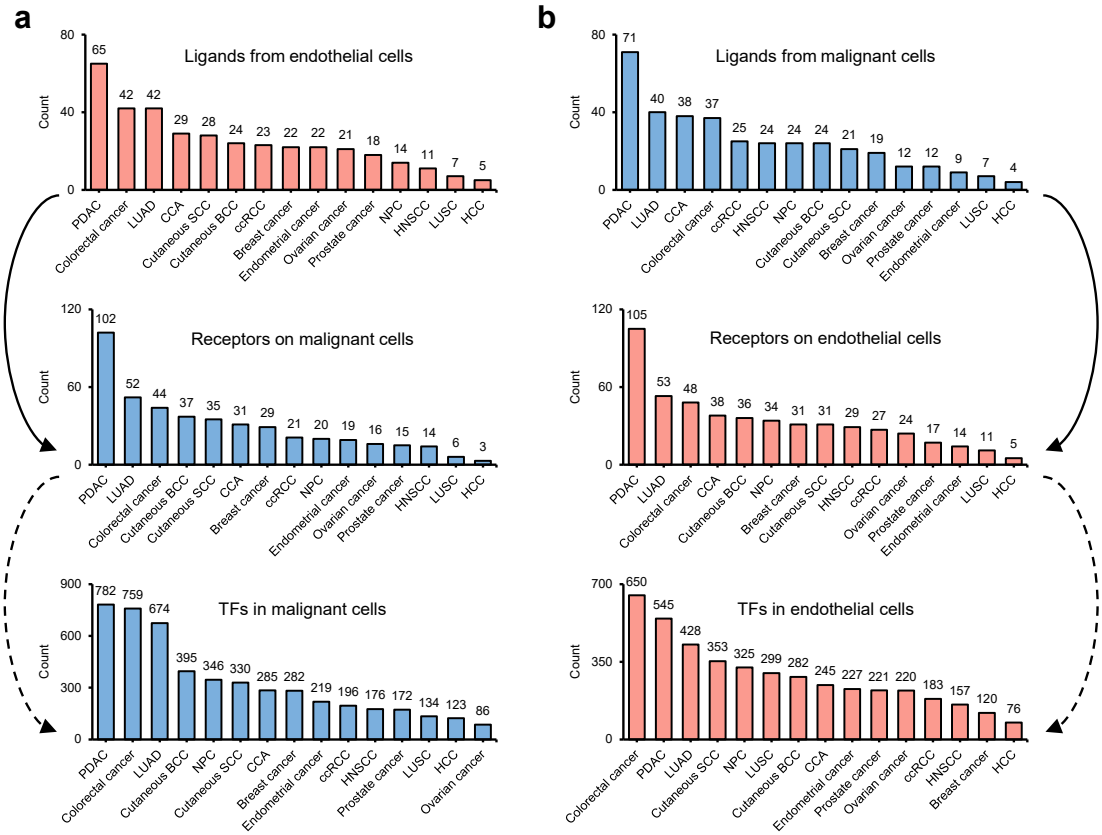


Extended Data Fig. 5 | Macrophage–malignant cell signaling across epithelial tumors. a, Numbers of macrophage-derived ligands (pink) and malignant cell receptors and TFs (blue) across 16 epithelial tumors. **b,** Numbers of malignant cell-derived ligands (blue) and macrophage receptors and TFs (pink) across 16 epithelial tumors.

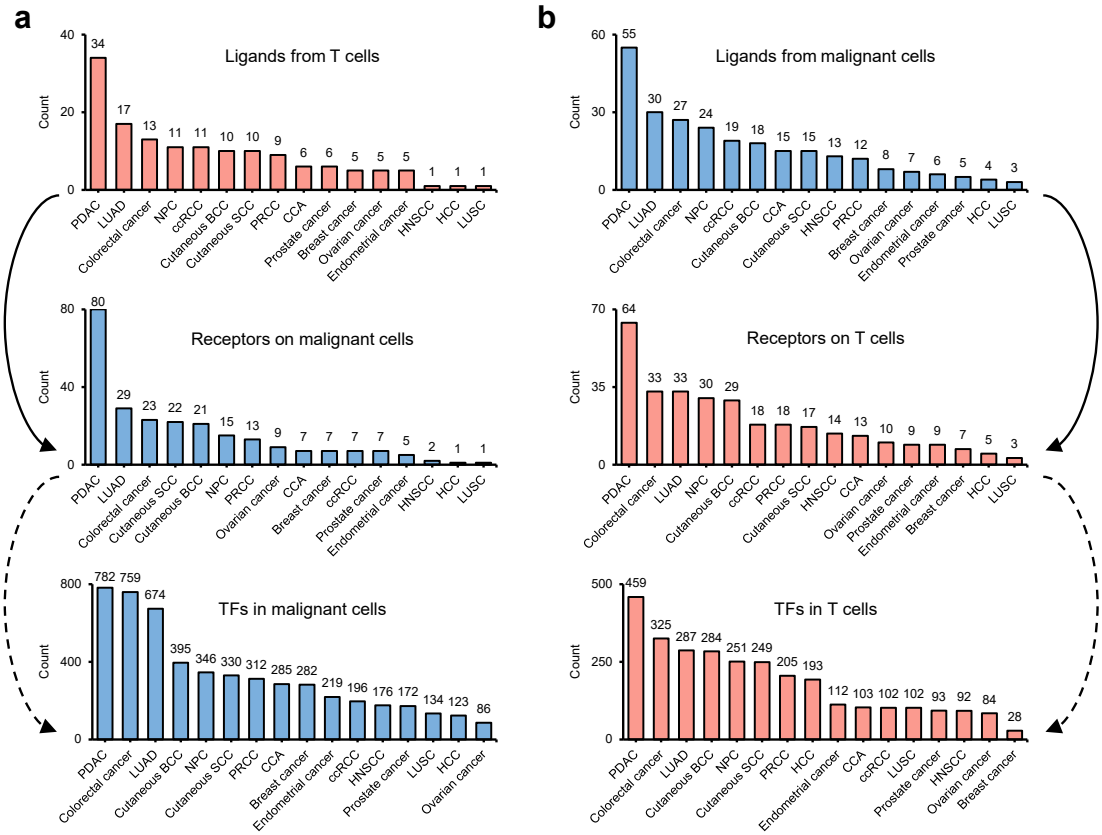


Extended Data Fig. 6 | Fibroblast–malignant cell signaling across epithelial tumors.

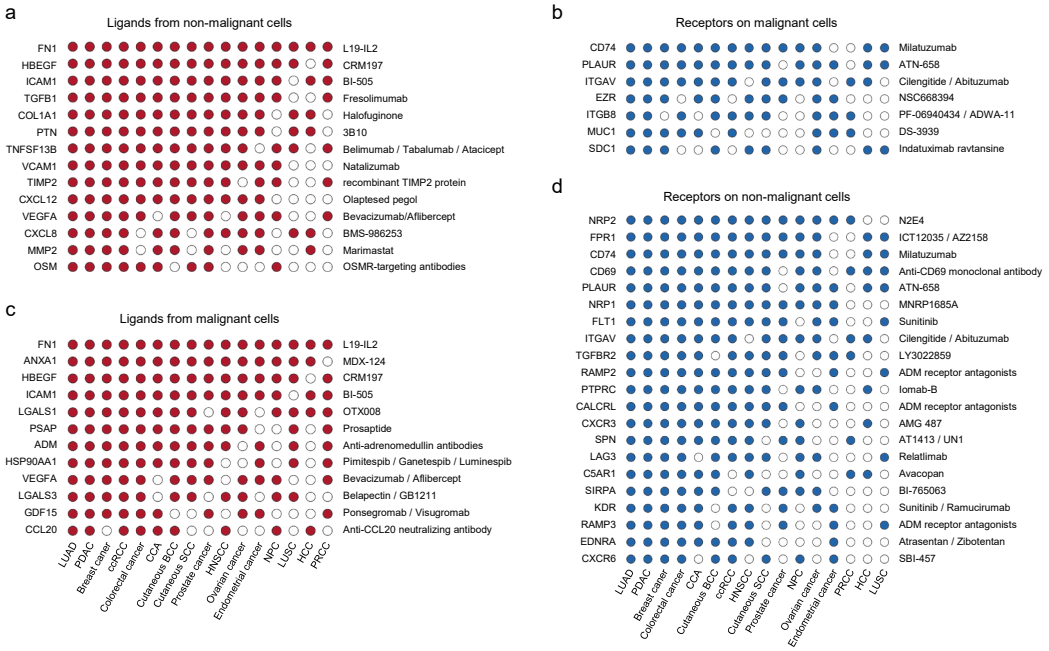
a, Numbers of fibroblast-derived ligands (pink) and malignant cell receptors and TFs (blue) across 16 epithelial tumors. **b,** Numbers of malignant cell-derived ligands (blue) and fibroblast receptors and TFs (pink) across 16 epithelial tumors.



Extended Data Fig. 7 | Endothelial–malignant cell signaling across epithelial tumors. a, Numbers of endothelial cell-derived ligands (pink) and malignant cell receptors and TFs (blue) across 16 epithelial tumors. **b,** Numbers of malignant cell-derived ligands (blue) and endothelial cell receptors and TFs (pink) across 16 epithelial tumors.



Extended Data Fig. 8 | T cell–malignant cell signaling across epithelial tumors. a, Numbers of T cell-derived ligands (pink) and malignant cell receptors and TFs (blue) across 16 epithelial tumors. **b,** Numbers of malignant cell-derived ligands (blue) and T cell receptors and TFs (pink) across 16 epithelial tumors.



Extended Data Fig. 9 | Drug repurposing opportunities in epithelial tumors. Drugs targeting **a,** non-malignant cell-derived ligands; **b,** malignant cell receptors; **c,** malignant cell-derived ligands; and **d,** non-malignant cell receptors shared by more than half of the 16 epithelial tumors. Targets are shown on the left and drugs on the right. Each column represents a tumor type, and hotspots indicate the presence of targets across tumors.

Reference

1. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).
2. Efremova, M., Vento-Tormo, M., Teichmann, S.A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506 (2020).
3. Noel, F. et al. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat Commun* **12**, 1089 (2021).
4. Xin, Y. et al. LRLoop: a method to predict feedback loops in cell-cell communication. *Bioinformatics* **38**, 4117-4126 (2022).
5. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**, 159-162 (2020).
6. He, C., Zhou, P. & Nie, Q. exFINDER: identify external communication signals using single-cell transcriptomics data. *Nucleic Acids Res* **51**, e58 (2023).

7. Baruzzo, G., Cesaro, G. & Di Camillo, B. Identify, quantify and characterize cellular communication from single-cell RNA sequencing data with scSeqComm. *Bioinformatics* **38**, 1920-1929 (2022).
8. Zhang, Y. et al. CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res* **49**, 8520-8534 (2021).
9. Cheng, J., Zhang, J., Wu, Z. & Sun, X. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Brief Bioinform* **22**, 988-1005 (2021).
10. Tyler, M. et al. The Curated Cancer Cell Atlas provides a comprehensive characterization of tumors at single-cell resolution. *Nat Cancer* **6**, 1088-1101 (2025).