Supplementary Figures for "CoxFormer enables spatial omics inference with multimodal generative modeling"

Yiyang Yang[1,3†], Xu Liao[1†], Haoyu Zhang[1†], Yida Wu[1], Xiaobo Sun[2], Yao Wang[3*], Tianshu Yu[1*], and Jin Liu[1*]

[1]School of Data Science, The Chinese University of Hong Kong-Shenzhen, Shenzhen, China,

[2]Department of Human Genetics, Emory University School of Medicine, Atlanta, USA.

[3]School of Management, Xi'an Jiaotong University, Xi'an, China,

[†]Equal contributions.

[*]Corresponding author. Email: yao.s.wang@gmail.com, yutianshu@cuhk.edu.cn, liujinlab@cuhk.edu.cn


a
ADAMTS6
ADAM15
ADAMTS3
FAM171B
ADAMTS9-AS1
ADAMTS1
FAM171A1
FAM124A
FAM13C
ADAMTS18
FAM162B
FAM167B
ADAMTS9
FAM107A
FAM124B
Blood Vessel Morphogenesis
Regulation of Endothelial Cell-Proliferation
Regulation of Cell Migration
Positive Regulation of Endothelial Cell-Proliferation
Vasculogenesis
Sprouting Angiogenesis
Positive Regulation of Angiogenesis
0
5
10
15
-log10 Adjusted P-value
b
IFIT1
IFIT2
IFI6
IFI44L
IFIH1
IFI27
HLA-B
HLA-F
HLA-E
IFI35
HLA-G
HLA-C
HLA-A
Cellular Response to Type I Interferon
Type I Interferon Signaling Pathway
Defense Response to Virus
Defense Response to Symbiont
Cytokine-mediated Signaling Pathway
Negative Regulation of Viral Process
Cellular Response to Interferon-gamma
0
20
40
60
-log10 Adjusted P-value
c
ANKRD37
ANKRD35
ANKK1
ANKRD18A
ANKRD26
BBS2
BBS1
BBS4
ANKMY1
BBS12
ANKEF1
ANKRD54
ANKRD45
ANKRD66
ANKFN1
Cilium Assembly
Cilium Organization
Plasma Membrane Bounded Cell Projection-Assembly
Organelle Assembly
Intraciliary Transport Involved in-Cilium Assembly
Cilium Movement
Axoneme Assembly
0
20
40
-log10 Adjusted P-value
d
KLK13
KLK10
KRT15
KLK5
KRT4
KRT5
KLK7
KRT17
KRT13
KLK6
KRT14
KRT16
KLK11
KRT6A
KLK8
Epidermis Development
Skin Development
Epidermal Cell Differentiation
Keratinocyte Differentiation
Defense Response to Bacterium
Peptide Cross-linking
Establishment of Skin Barrier
0
5
10
15
20
25
-log10 Adjusted P-value

Figure S1: **Gene Ontology (GO) enrichment analysis highlights major functional modules encoded by CoxFormer gene embeddings**. Shown here are representative functional categories recovered from the embedding manifold, where each panel pairs a gene cluster in the low-dimensional projection (left; representative genes annotated) with its top enriched GO terms (right; bars indicate $-\log_{10}$ adjusted P values). **(a)** Blood-vessel and endothelial programs. **(b)** Interferon and antiviral immune response. **(c)** Cilium assembly and movement. **(d)** Epidermis and keratinocyte differentiation.

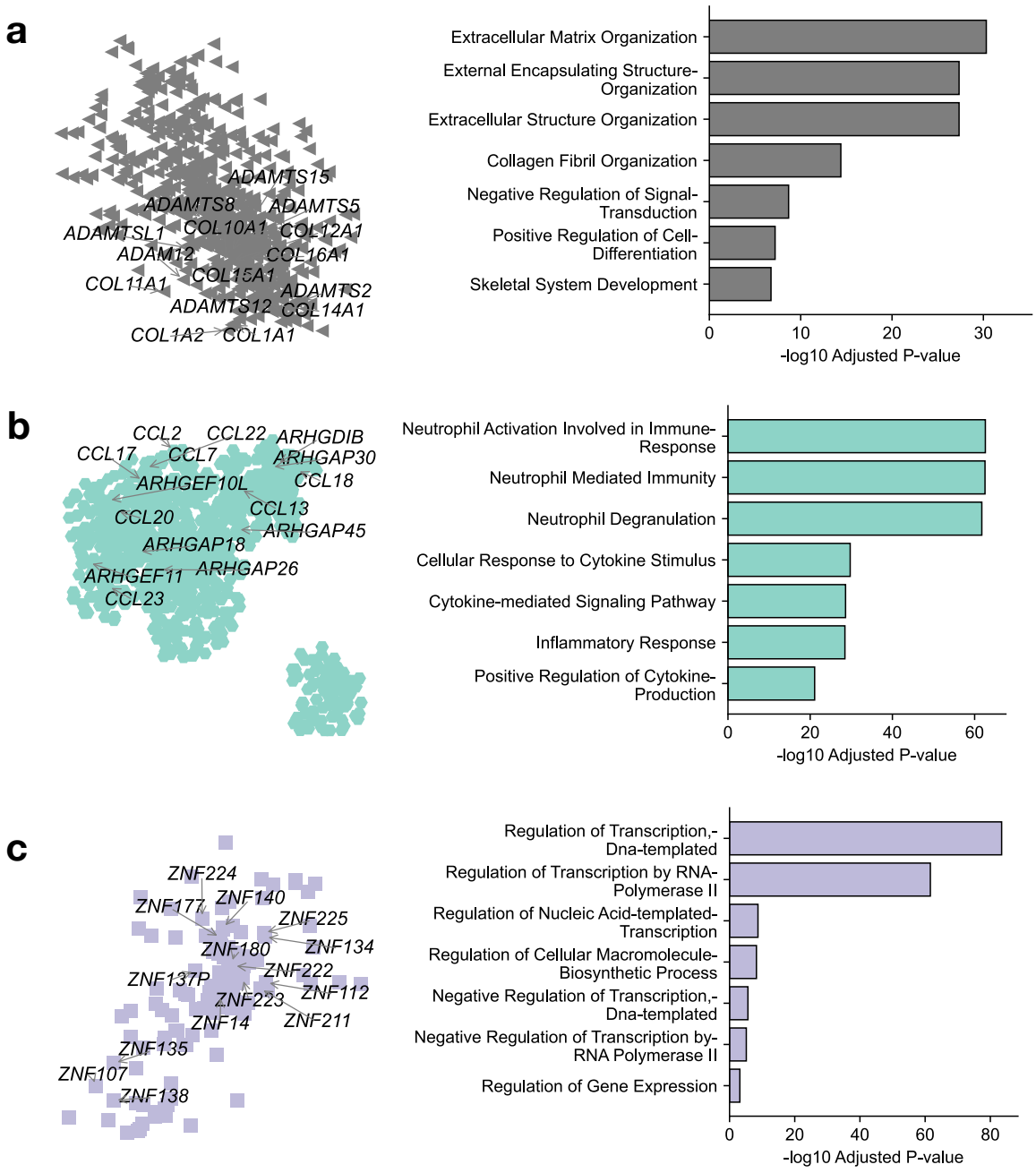
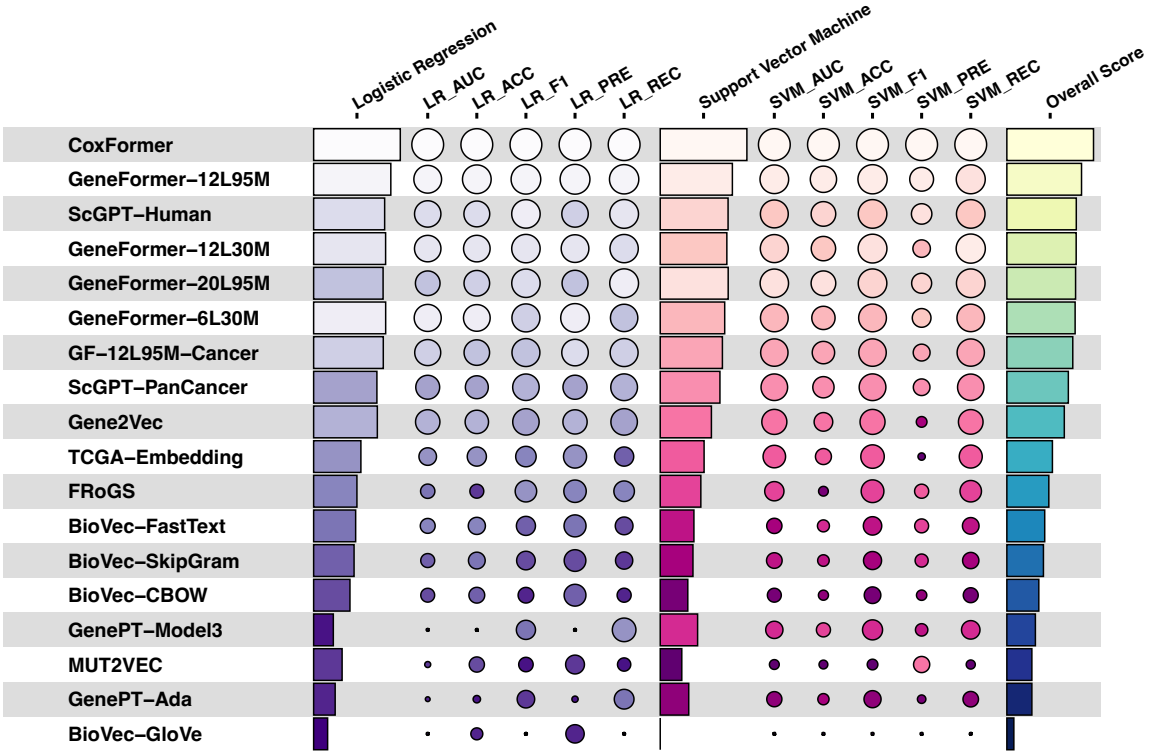


Figure S2: **Additional functional modules identified from the CoxFormer embedding manifold.** Shown here are representative functional categories recovered from the embedding manifold, where each panel pairs a gene cluster in the low-dimensional projection (left; representative genes annotated) with its top enriched GO terms (right; bars indicate $-\log_{10}$ adjusted P values). **(a)** Extracellular-matrix organization and collagen fibril organization. **(b)** Neutrophil activation and cytokine-mediated inflammatory signaling. **(c)** Transcriptional regulation and RNA polymerase–associated processes.

a

Distinguishes between bivalent and Lys4-only-methylated genes (whole genome)



b

Distinguishes between dosage-sensitive and dosage-insensitive transcription factors

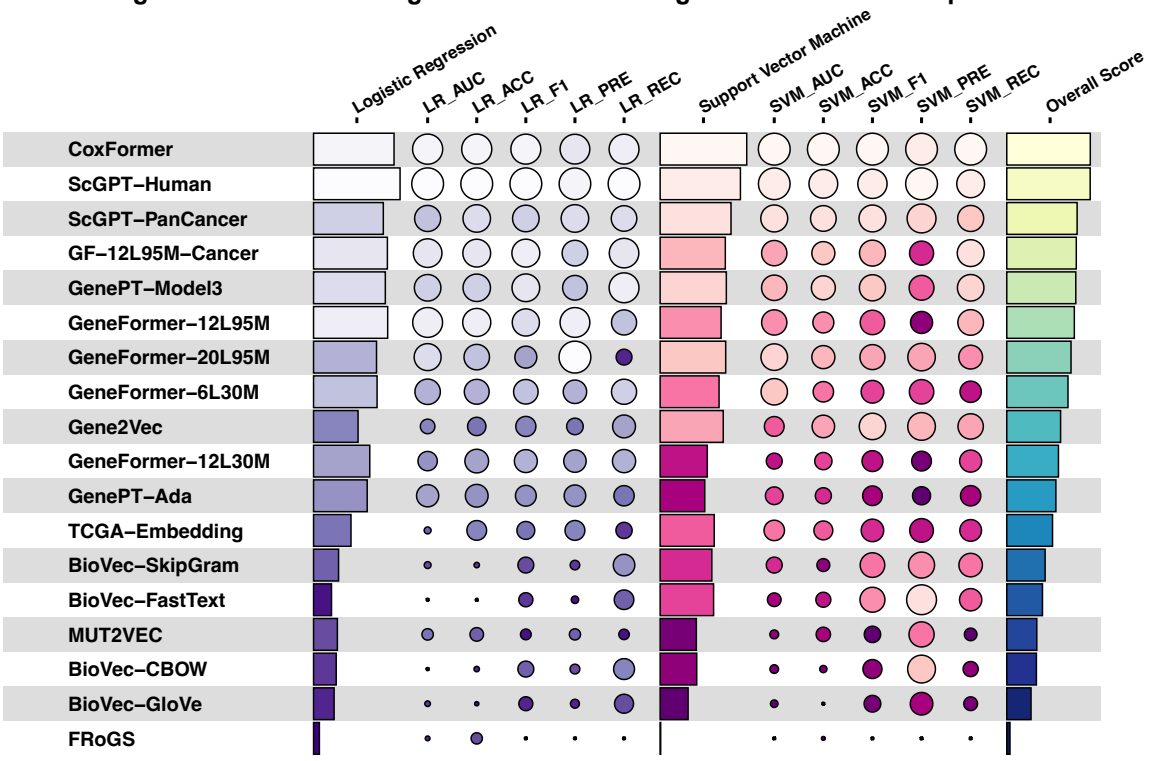
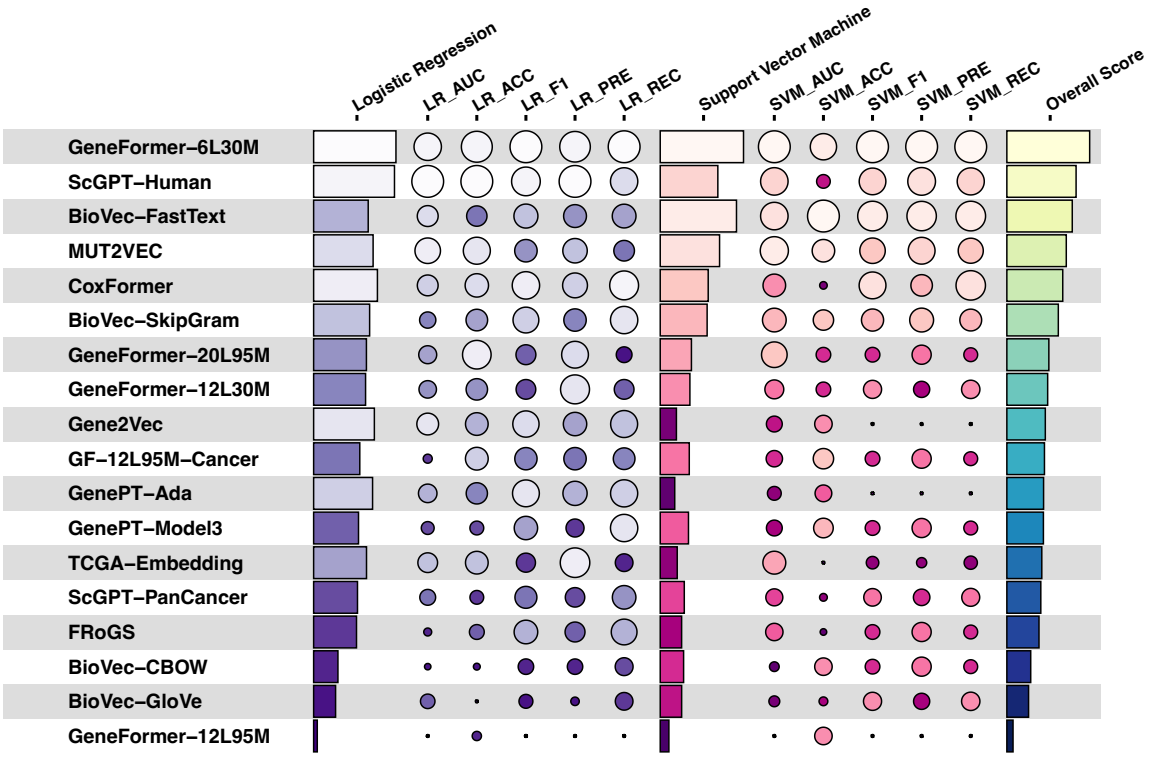


Figure S3: **Detailed per-task results for gene-level binary classification benchmarks (I).** Results are reported for two classifiers (logistic regression and support vector machine) and five complementary evaluation metrics (AUC, Accuracy, F1, Precision, Recall), together with an aggregated overall score (rightmost column). Bubble size denotes the normalized score (higher is better) and bubble color encodes the ranking within each metric (lighter indicates better performance). **(a)** Distinguishes between bivalent and Lys4-only-methylated genes (whole genome). **(b)** Distinguishes between dosage-sensitive and dosage-insensitive transcription factors.

a

Differentiates between long-range and short-range transcription factors (TFs)



b

Identifies differences between bivalent and non-methylated genes

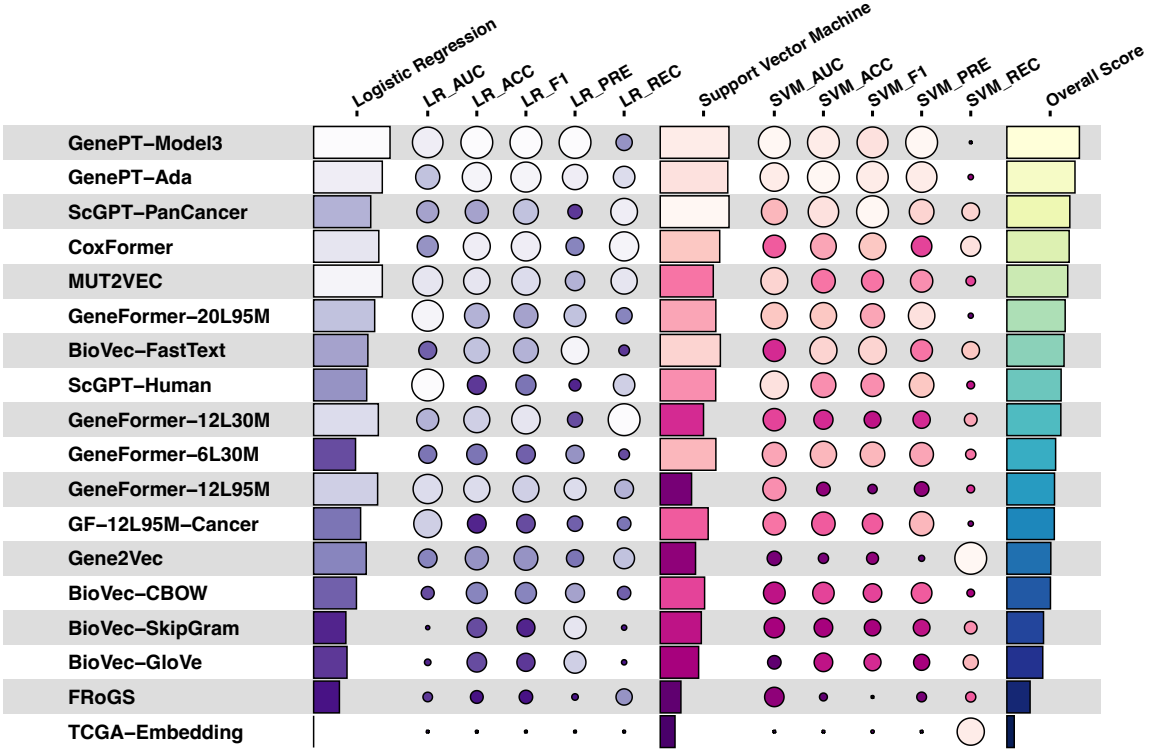
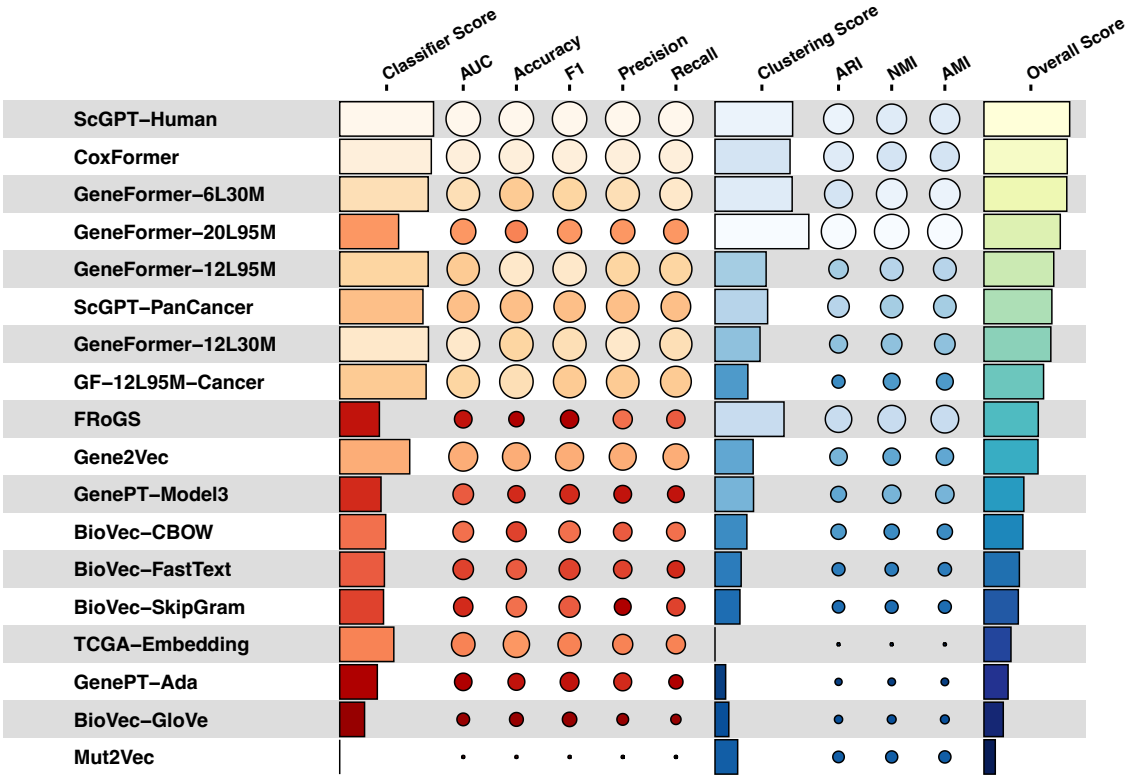


Figure S4: **Detailed per-task results for additional gene-level binary classification benchmarks (II).** Results are reported for two classifiers (logistic regression and support vector machine) and five complementary evaluation metrics (AUC, Accuracy, F1, Precision, Recall), together with an aggregated overall score (rightmost column). Bubble size denotes the normalized score (higher is better) and bubble color encodes the ranking within each metric (lighter indicates better performance). **(a)** Differentiates between long-range and short-range transcription factors (TFs). **(b)** Identifies differences between bivalent and non-methylated genes.

a


Human CD34+ Bone Marrow (BM)
Classifier Score
AUC
Accuracy
F1
Precision
Recall
Clustering Score
ARI
NMI
AMI
Overall Score
ScGPT–Human
CoxFormer
GeneFormer–6L30M
GeneFormer–20L95M
GeneFormer–12L95M
ScGPT–PanCancer
GeneFormer–12L30M
GF–12L95M–Cancer
FRoGS
Gene2Vec
GenePT–Model3
BioVec–CBOW
BioVec–FastText
BioVec–SkipGram
TCGA–Embedding
GenePT–Ada
BioVec–GloVe
Mut2Vec

b

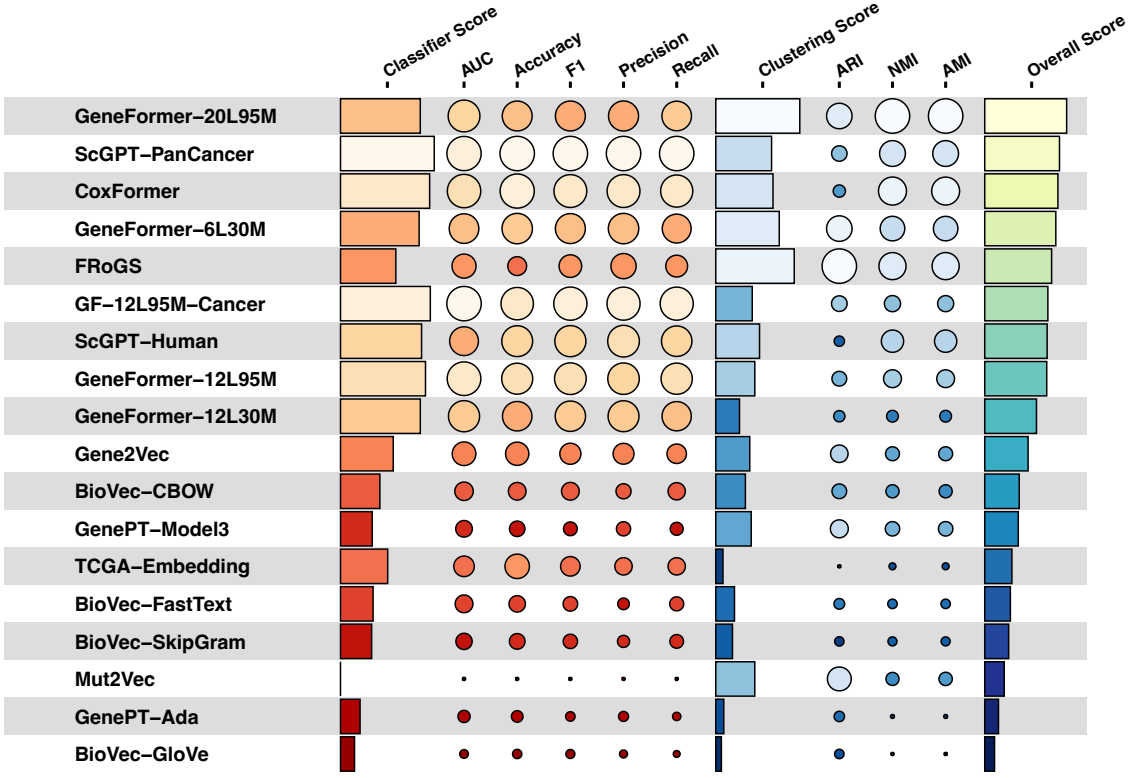
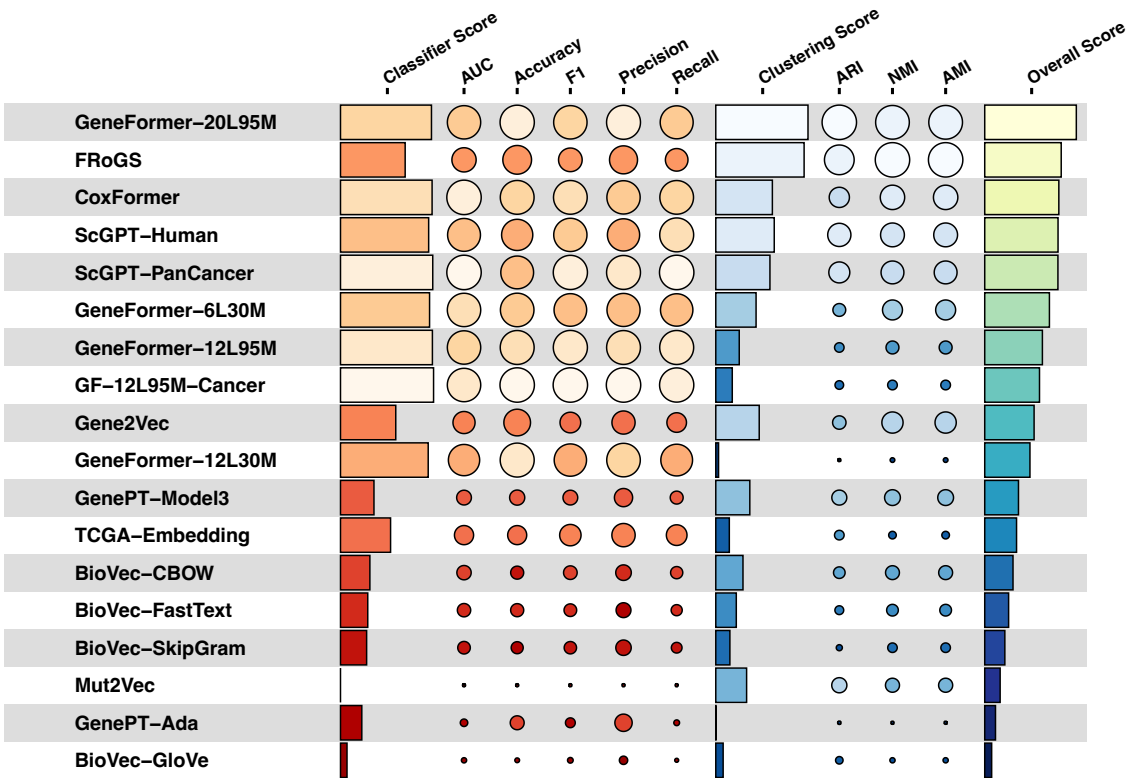

Breast Cancer FFPE (BC)
Classifier Score
AUC
Accuracy
F1
Precision
Recall
Clustering Score
ARI
NMI
AMI
Overall Score
GeneFormer–20L95M
ScGPT–PanCancer
CoxFormer
GeneFormer–6L30M
FRoGS
GF–12L95M–Cancer
ScGPT–Human
GeneFormer–12L95M
GeneFormer–12L30M
Gene2Vec
BioVec–CBOW
GenePT–Model3
TCGA–Embedding
BioVec–FastText
BioVec–SkipGram
Mut2Vec
GenePT–Ada
BioVec–GloVe

Figure S5: **Detailed per-dataset results for cell-type clustering and classification benchmarks (I).** Results are reported for four scRNA-seq datasets and summarized into an aggregated overall cell-task score (rightmost column; higher indicates better performance; Methods). Two complementary evaluations are included: unsupervised clustering (ARI, NMI, AMI) and supervised cell-type classification (AUC, Accuracy, F1, Precision, Recall). Bubble size denotes the normalized score (higher is better) and bubble color encodes the ranking within each metric (lighter indicates better performance). **(a)** Human $CD34^+$ Bone Marrow (BM). **(b)** Breast Cancer FFPE (BC).

a


Diffuse Large B-cell Lymphoma FFPE (DLBL)
Classifier Score
AUC
Accuracy
F1
Precision
Recall
Clustering Score
ARI
NMI
AMI
Overall Score
GeneFormer–20L95M
FRoGS
CoxFormer
ScGPT–Human
ScGPT–PanCancer
GeneFormer–6L30M
GeneFormer–12L95M
GF–12L95M–Cancer
Gene2Vec
GeneFormer–12L30M
GenePT–Model3
TCGA–Embedding
BioVec–CBOW
BioVec–FastText
BioVec–SkipGram
Mut2Vec
GenePT–Ada
BioVec–GloVe

b

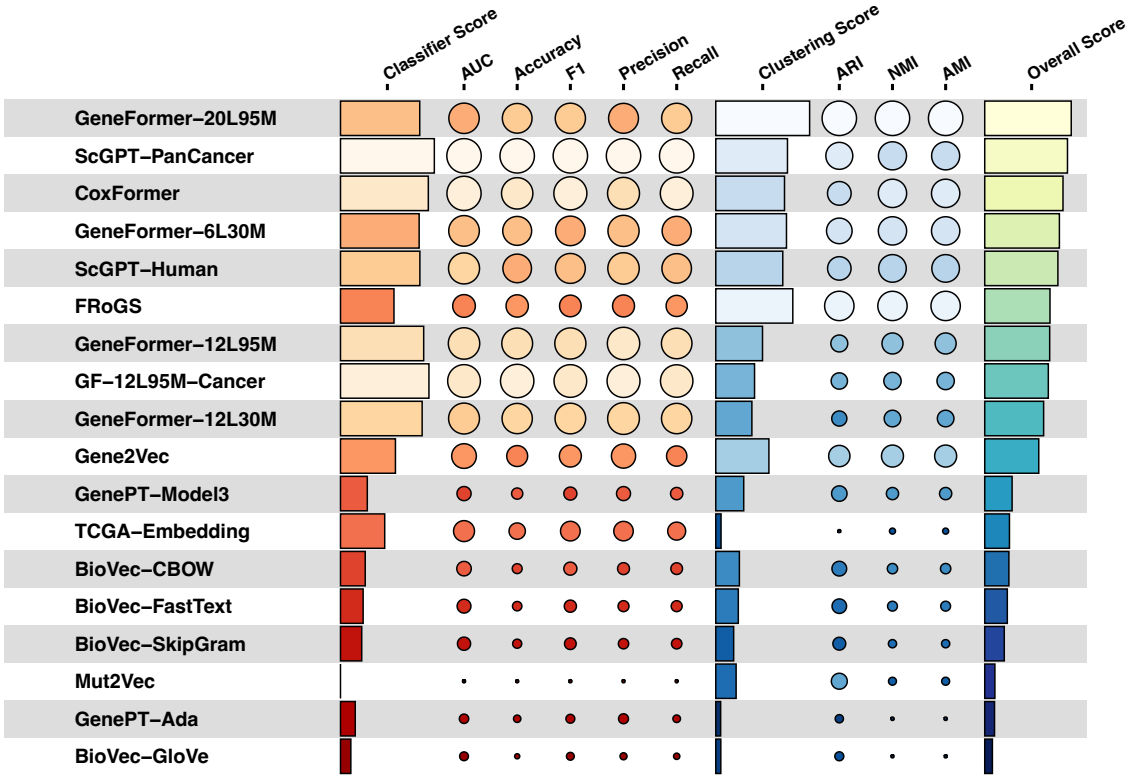

Lung Cancer FFPE (LC)
Classifier Score
AUC
Accuracy
F1
Precision
Recall
Clustering Score
ARI
NMI
AMI
Overall Score
GeneFormer–20L95M
ScGPT–PanCancer
CoxFormer
GeneFormer–6L30M
ScGPT–Human
FRoGS
GeneFormer–12L95M
GF–12L95M–Cancer
GeneFormer–12L30M
Gene2Vec
GenePT–Model3
TCGA–Embedding
BioVec–CBOW
BioVec–FastText
BioVec–SkipGram
Mut2Vec
GenePT–Ada
BioVec–GloVe

Figure S6: **Detailed per-dataset results for additional cell-type clustering and classification benchmarks (II).** Results are reported for four scRNA-seq datasets and summarized into an aggregated overall cell-task score (rightmost column; higher indicates better performance; Methods). Two complementary evaluations are included: unsupervised clustering (ARI, NMI, AMI) and supervised cell-type classification (AUC, Accuracy, F1, Precision, Recall). Bubble size denotes the normalized score (higher is better) and bubble color encodes the ranking within each metric (lighter indicates better performance). **(a)** Diffuse Large B-cell Lymphoma FFPE (DLBL). **(b)** Lung Cancer FFPE (LC).

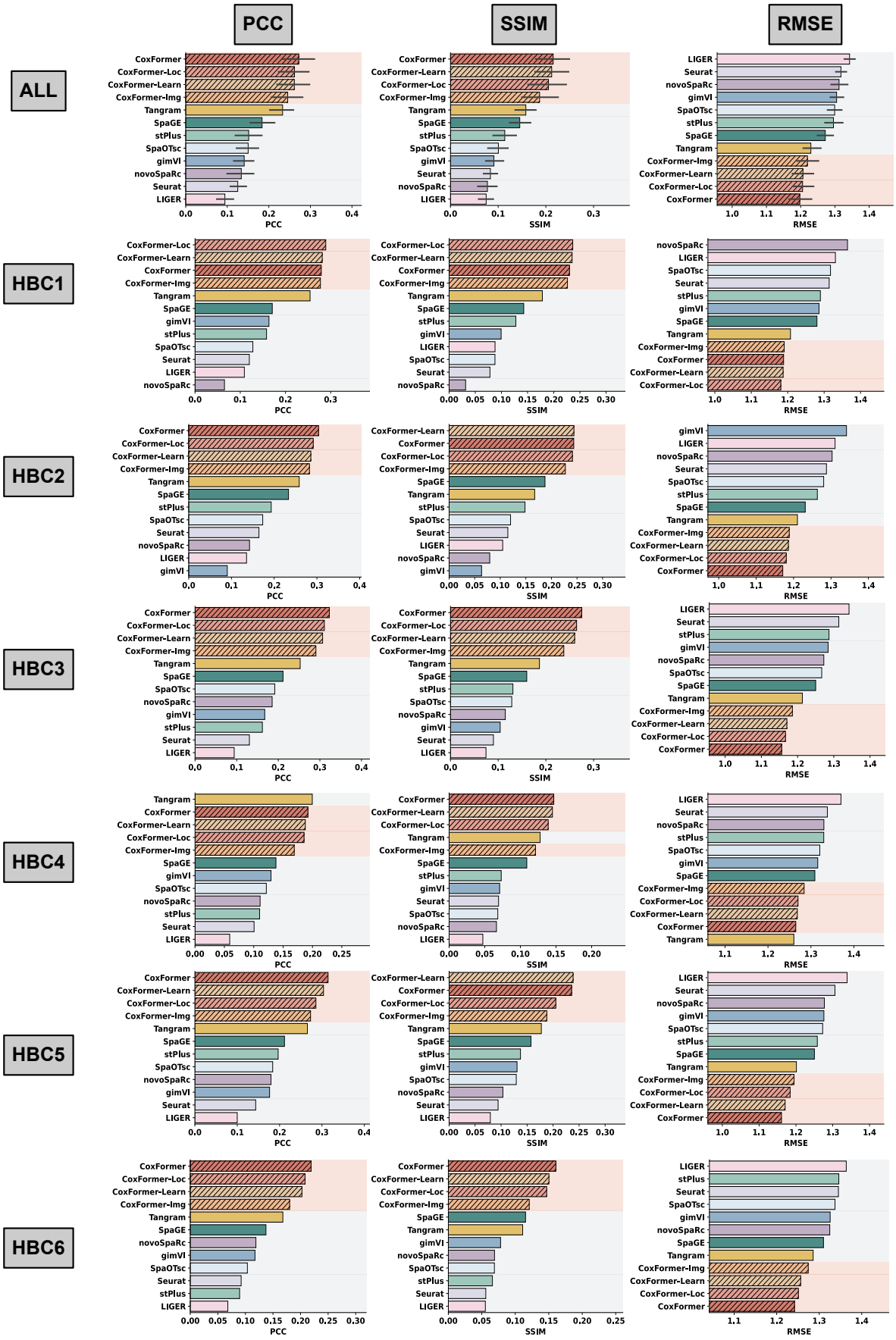

PCC
SSIM
RMSE
ALL
CoxFormer
CoxFormer-Loc
CoxFormer-Learn
CoxFormer-Img
Tangram
SpaGE
stPlus
SpaOTsc
gimVI
novoSpaRc
Seurat
LIGER
PCC
SSIM
RMSE
HBC1
HBC2
HBC3
HBC4
HBC5
HBC6

Figure S7: **Detailed results for pseudo-panel gene prediction on six human breast cancer (HBC) Visium datasets.** We compare CoxFormer with scRNA-seq reference-based baselines across the aggregated set (ALL) and each individual dataset (HBC1-HBC6). Prediction quality is assessed by three complementary metrics: Pearson correlation coefficient (PCC), structural similarity index (SSIM), and root mean squared error (RMSE). Ablation variants (CoxFormer-Loc, CoxFormer-Img, CoxFormer-Learn) are included to quantify the contribution of different modalities; incorporating comprehensive spatial context yields the best overall performance.

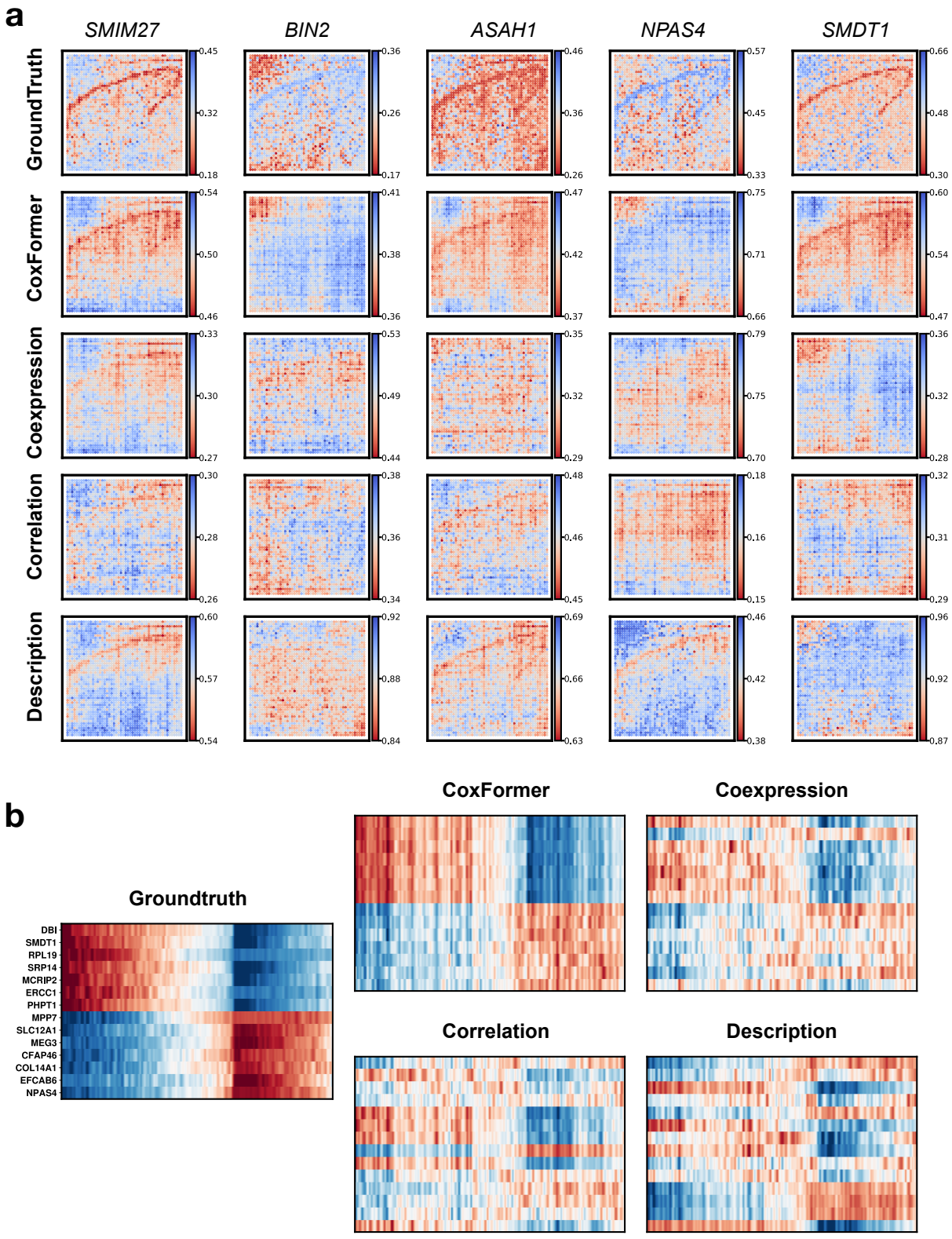


Figure S8: **ATAC-derived gene activity score (GAS) prediction.** We compare CoxFormer with three ablated variants that isolate individual information sources: gene co-expression, gene correlation, and gene description. **(a)** Spatial GAS prediction for five representative genes (*SMIM27*, *BIN2*, *ASAH1*, *NPAS4*, *SMDT1*), showing that CoxFormer better matches the ground-truth spatial contrast and localization than individual-source variants. **(b)** Region-by-gene GAS heatmaps based on top differentially expressed genes (DEGs) between DG GCL and CA3 Pyr, where CoxFormer more faithfully preserves the underlying block-wise structure.

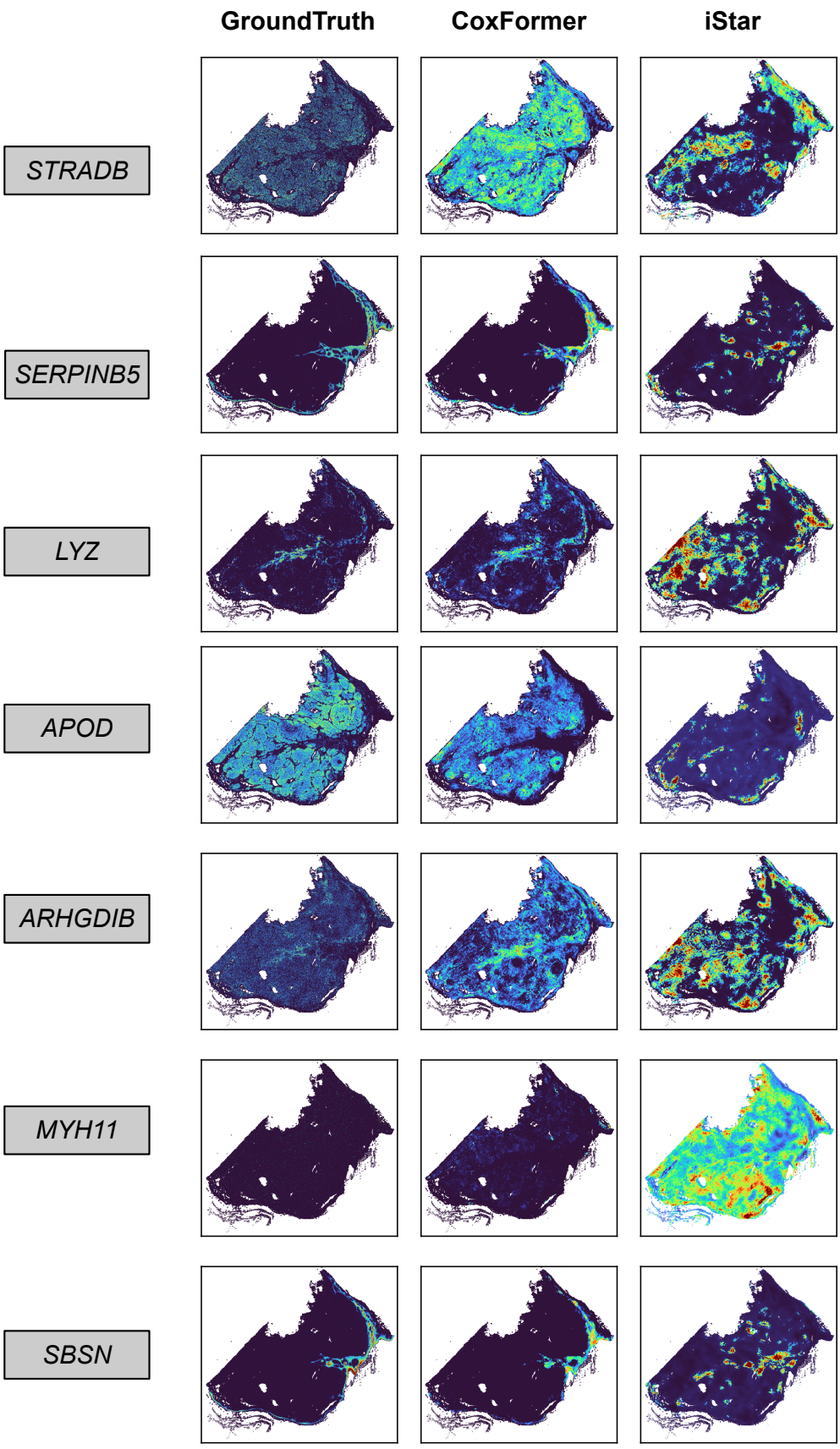

GroundTruth
CoxFormer
iStar
STRADB
SERPINB5
LYZ
APOD
ARHGDIB
MYH11
SBSN

Figure S9: **Visual comparison of super-resolution gene maps between CoxFormer and iStar.** We visualize enhanced-resolution predictions for representative genes by comparing ground-truth, CoxFormer, and iStar. CoxFormer produces smoother and more spatially coherent expression patterns that better aligned with tissue structure, whereas iStar often yields fragmented and patchy maps.

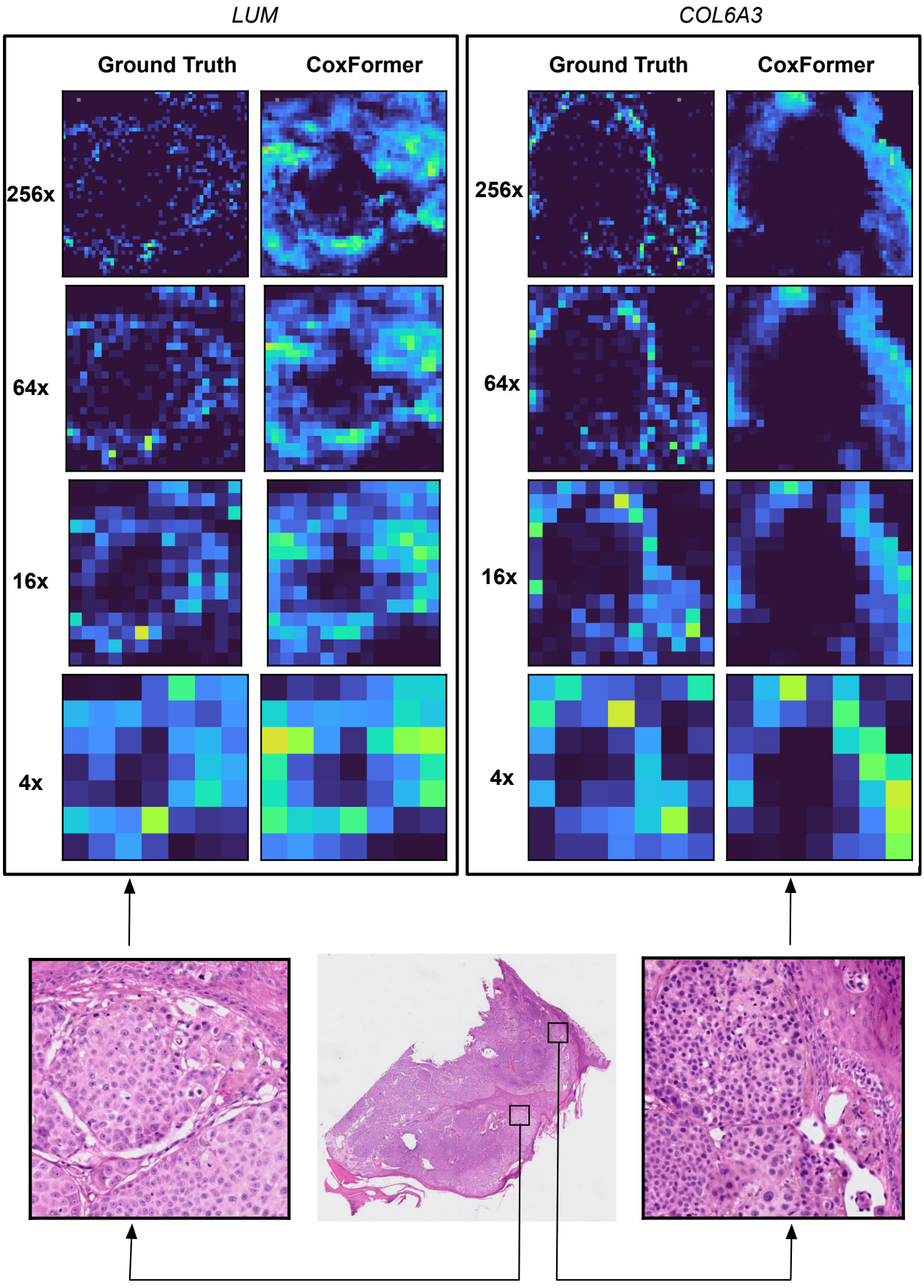

LUM
COL6A3
Ground Truth
CoxFormer
Ground Truth
CoxFormer
256x
64x
16x
4x
256x
64x
16x
4x

Figure S10: **Zero-shot marker gene super-resolution prediction and multi-scale Xenium validation: extracellular-matrix–associated stromal programs.** Using the 56 genes unobserved in both sections, we evaluate CoxFormer in a stringent zero-shot regime and compare its subcellular-scale predictions with Xenium ground truth for representative marker genes that delineate distinct tissue compartments. Multi-scale visualizations are shown at four resolutions (4×, 16×, 64×, and 256×), together with matched H&E context and the corresponding zoom-in locations on the whole-slide image (*LUM*, *COL6A3*).

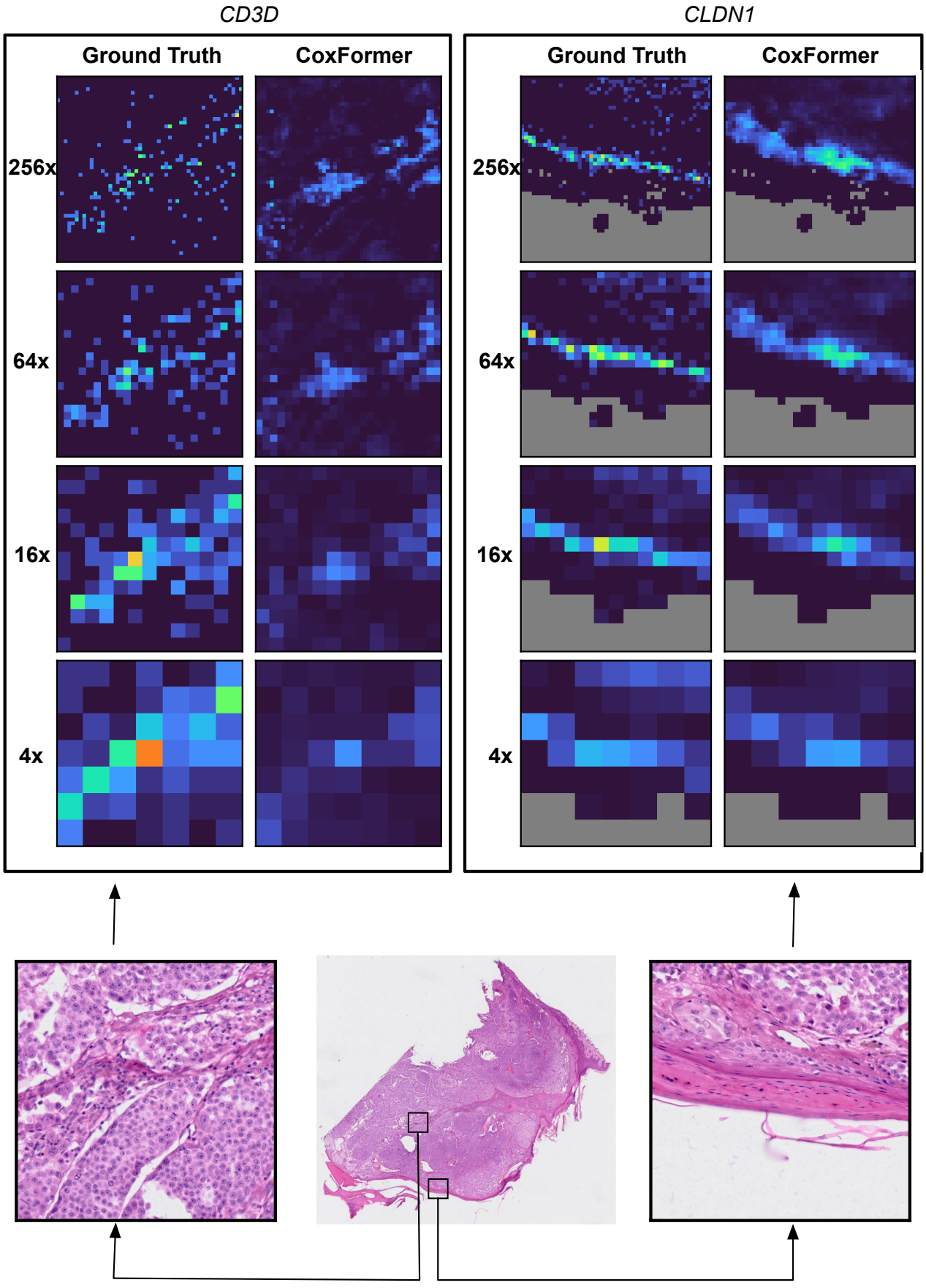

CD3D
CLDN1
Ground Truth
CoxFormer
Ground Truth
CoxFormer
256x
64x
16x
4x
256x
64x
16x
4x

Figure S11: **Zero-shot marker gene super-resolution prediction and multi-scale Xenium validation: epidermis/immune compartment markers.** Using the 56 genes unobserved in both sections, we evaluate CoxFormer in a stringent zero-shot regime and compare its subcellular-scale predictions with Xenium ground truth for representative marker genes that delineate distinct tissue compartments. Multi-scale visualizations are shown at four resolutions (4×, 16×, 64×, and 256×), together with matched H&E context and the corresponding zoom-in locations on the whole-slide image (*CD3D*, *CLDN1*).

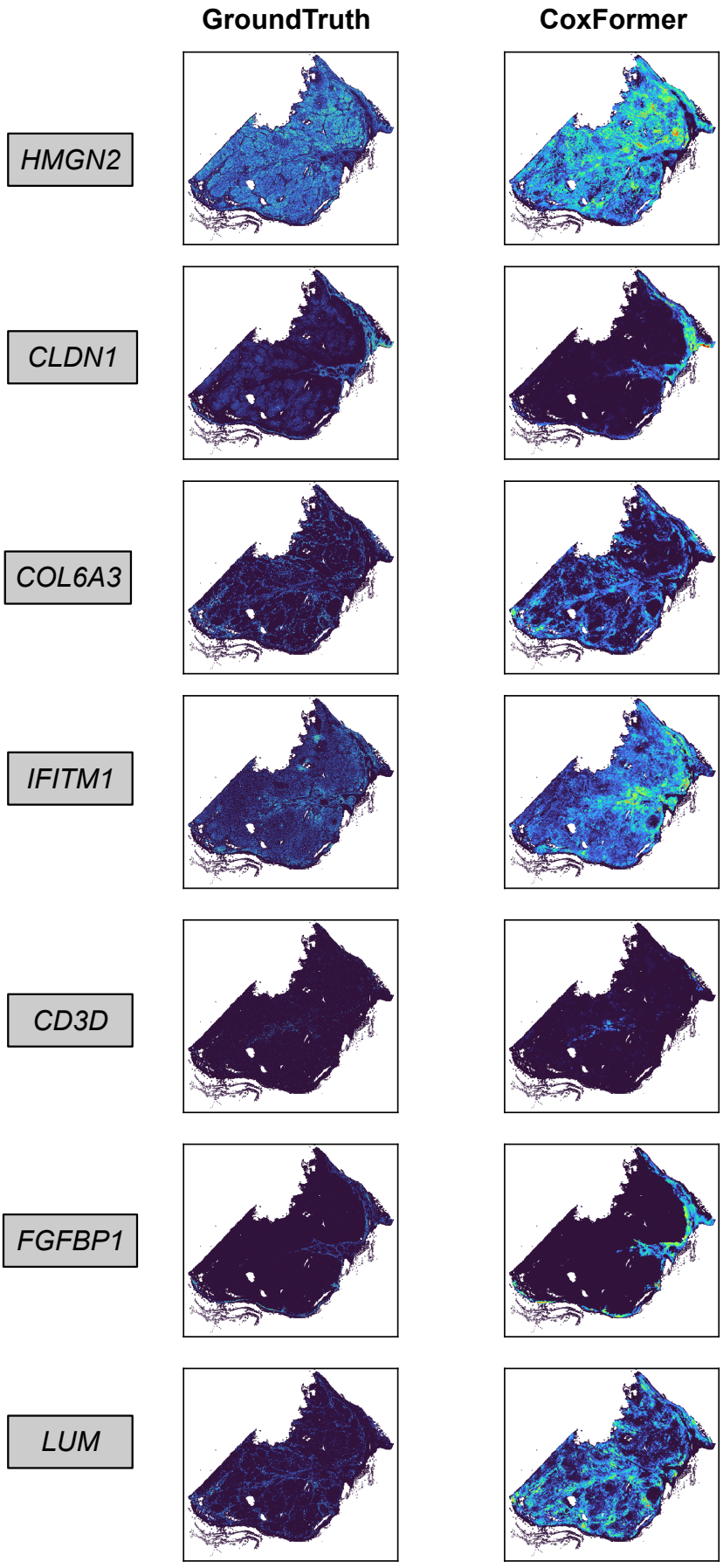
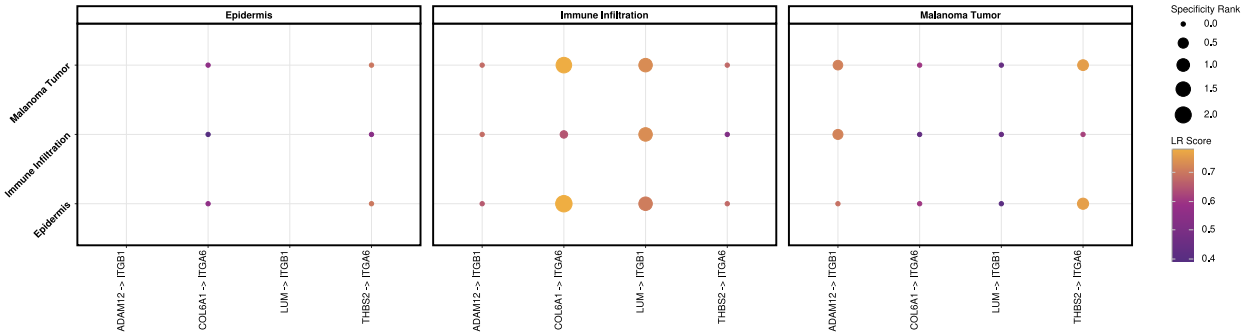

GroundTruth
CoxFormer
HMGN2
CLDN1
COL6A3
IFITM1
CD3D
FGFBP1
LUM

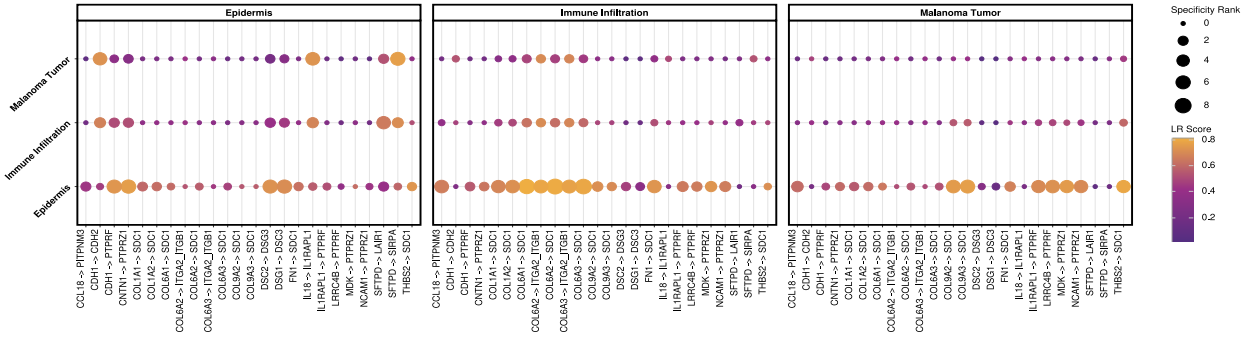
Figure S12: **Full-section zero-shot super-resolution predictions with Xenium ground truth.** In a stringent zero-shot setting where target genes are entirely absent from the measured panel, we visualize full-section predictions for representative unobserved genes by comparing Xenium ground truth with CoxFormer outputs.

a

Dotplots for ligand–receptor interactions before imputation

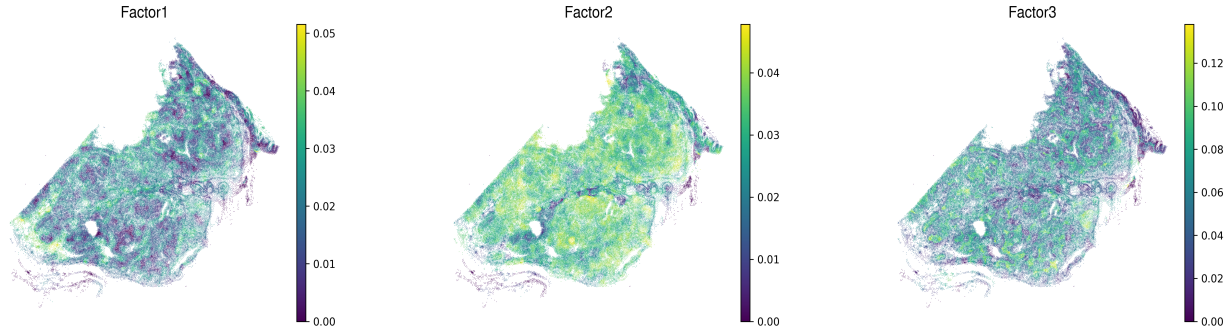


Dotplots for ligand–receptor interactions after imputation by CoxFormer



b

Spatial NMF visualization before imputation



Spatial NMF visualization after imputation by COPT

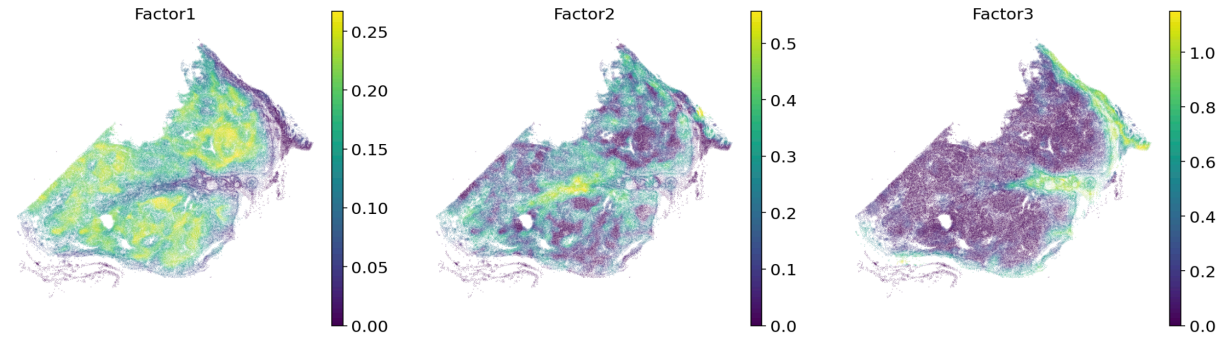


Figure S13: **CCC analysis before and after transcriptome-wide imputation in Section B.**. **(a)** Spatial NMF factors are visualized on Section B before imputation and after imputation by CoxFormer, illustrating clearer region-resolved programs after transcriptome-wide expansion. **(b)** Ligand-receptor dotplots are inferred from the original 280-gene panel (top), where CCC inference is underpowered and yields only four robust ligand-receptor pairs, versus the transcriptome-expanded data after CoxFormer imputation (bottom), where the inferred CCC landscape become substantially richer.

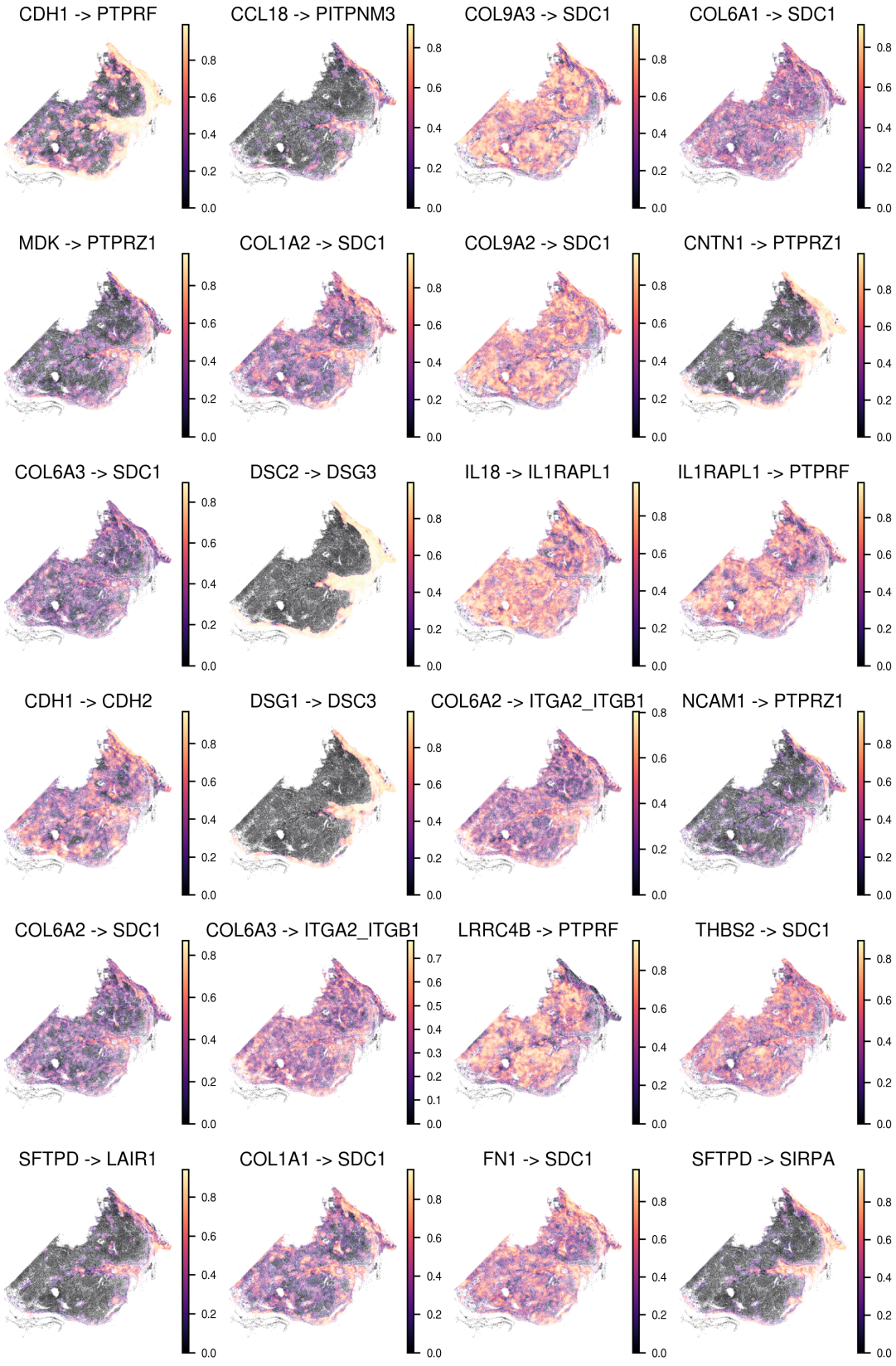

CDH1 -> PTPRF
CCL18 -> PITPNM3
COL9A3 -> SDC1
COL6A1 -> SDC1
MDK -> PTPRZ1
COL1A2 -> SDC1
COL9A2 -> SDC1
CNTN1 -> PTPRZ1
COL6A3 -> SDC1
DSC2 -> DSG3
IL18 -> IL1RAPL1
IL1RAPL1 -> PTPRF
CDH1 -> CDH2
DSG1 -> DSC3
COL6A2 -> ITGA2_ITGB1
NCAM1 -> PTPRZ1
COL6A2 -> SDC1
COL6A3 -> ITGA2_ITGB1
LRRC4B -> PTPRF
THBS2 -> SDC1
SFTPD -> LAIR1
COL1A1 -> SDC1
FN1 -> SDC1
SFTPD -> SIRPA

Figure S14: **Spatial patterns of ligand-receptor pairs identified from dotplot CCC analysis after transcriptome-wide imputation.** The spatial maps reveal region-aligned signaling programs across tissue compartments, including extracellular-matrix and adhesion interactions (e.g., *COL1A1→SDC1*, *THBS2→SDC1*, *FN1→SDC1*), epidermal junction programs with strong epidermis-associated specificity (e.g., *DSG1→DSC3* and *DSC2→DSG3*), and immune recognition signals enriched in the immune-infiltrated region (e.g., *SFTPD→LAIR1* and *SFTPD→SIRPA*).

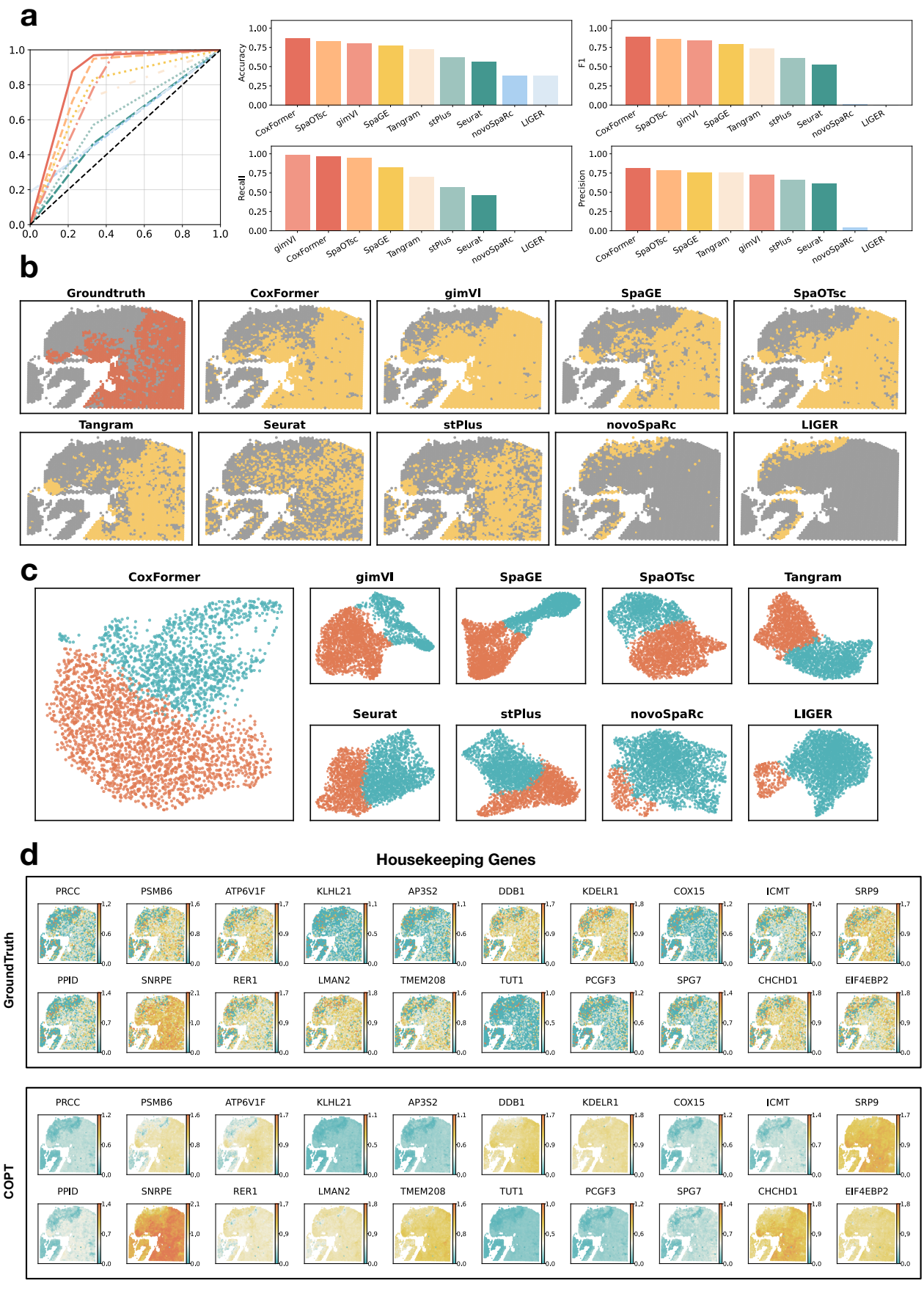

a
Accuracy
F1
Recall
Precision
CoxFormer
SpaOTsc
gimVI
SpaGE
Tangram
stPlus
Seurat
novoSpaRc
LIGER
b
Groundtruth
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
c
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
d
Housekeeping Genes
GroundTruth
COPT
PRCC
PSMB6
ATP6V1F
KLHL21
AP3S2
DDB1
KDELR1
COX15
ICMT
SRP9
PPID
SNRPE
RER1
LMAN2
TMEM208
TUT1
PCGF3
SPG7
CHCHD1
EIF4EBP2

Figure S15: **Example results of abnormal-region annotation on the untreated colorectal cancer (CRC) section.** We compare CoxFormer with representative reference-required baselines (SpaGE, SpaOTsc, Tangram, gimVI, Seurat, stPlus, novoSpaRc, and LIGER). **(a)** Quantitative performance on this section, reporting AUROC (left) and complementary classification metrics (Accuracy, F1, Recall, Precision). **(b)** Spatial segmentation results are overlaid on the tissue, comparing predicted abnormal regions from each method with the ground-truth annotation. **(c)** UMAP visualization of the predicted labels from panel (b), illustrating separation between abnormal and normal spots under different methods. **(d)** Representative housekeeping-gene expression patterns for ground-truth versus CoxFormer predictions.

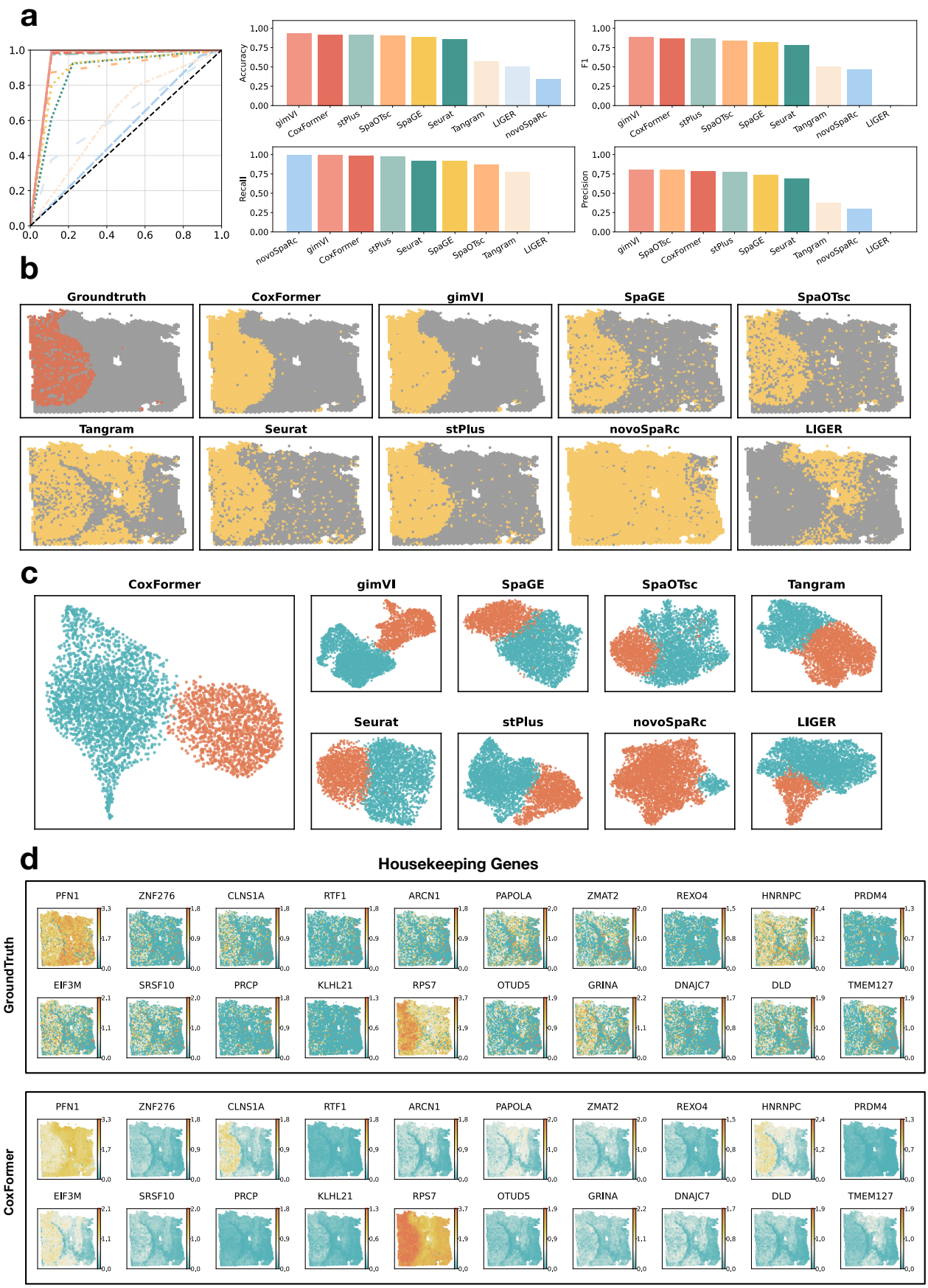

a
Accuracy
F1
Recall
Precision
gimVI
CoxFormer
stPlus
SpaOTsc
SpaGE
Seurat
Tangram
LIGER
novoSpaRc
b
Groundtruth
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
c
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
d
Housekeeping Genes
GroundTruth
CoxFormer
PFN1
ZNF276
CLNS1A
RTF1
ARCN1
PAPOLA
ZMAT2
REXO4
HNRNPC
PRDM4
EIF3M
SRSF10
PRCP
KLHL21
RPS7
OTUD5
GRINA
DNAJC7
DLD
TMEM127

Figure S16: **Example results of abnormal-region annotation on the therapy-perturbed colorectal cancer (CRC) section.** We compare CoxFormer with representative reference-required baselines (SpaGE, SpaOTsc, Tangram, gimVI, Seurat, stPlus, novoSpaRc, and LIGER). **(a)** Quantitative performance on this section, reporting AUROC (left) and complementary classification metrics (Accuracy, F1, Recall, Precision). **(b)** Spatial segmentation results are overlaid on the tissue, comparing predicted abnormal regions from each method with the ground-truth annotation. **(c)** UMAP visualization of the predicted labels from panel (b), illustrating separation between abnormal and normal spots under different methods. **(d)** Representative housekeeping-gene expression patterns for ground-truth versus CoxFormer predictions.

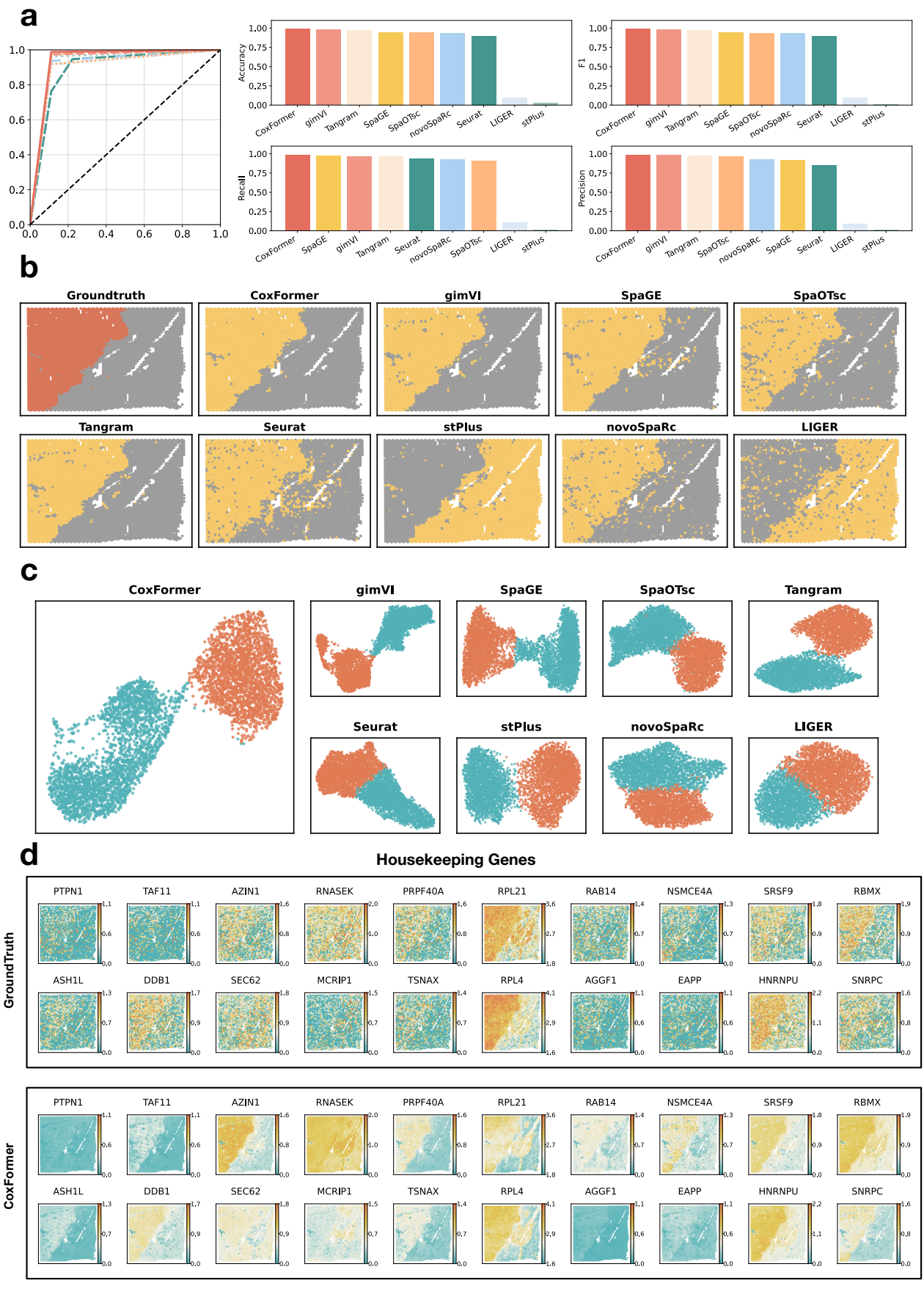

a
Accuracy
F1
Recall
Precision
CoxFormer
gimVI
Tangram
SpaGE
SpaOTsc
novoSpaRc
Seurat
LIGER
stPlus
b
Groundtruth
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
c
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
d
Housekeeping Genes
GroundTruth
CoxFormer
PTPN1
TAF11
AZIN1
RNASEK
PRPF40A
RPL21
RAB14
NSMCE4A
SRSF9
RBMX
ASH1L
DDB1
SEC62
MCRIP1
TSNAX
RPL4
AGGF1
EAPP
HNRNPU
SNRPC

Figure S17: **Example results of abnormal-region annotation on the untreated liver metastasis (LM) section.** We compare CoxFormer with representative reference-required baselines (SpaGE, SpaOTsc, Tangram, gimVI, Seurat, stPlus, novoSpaRc, and LIGER). **(a)** Quantitative performance on this section, reporting AUROC (left) and complementary classification metrics (Accuracy, F1, Recall, Precision). **(b)** Spatial segmentation results are overlaid on the tissue, comparing predicted abnormal regions from each method with the ground-truth annotation. **(c)** UMAP visualization of the predicted labels from panel (b), illustrating separation between abnormal and normal spots under different methods. **(d)** Representative housekeeping-gene expression patterns for ground-truth versus CoxFormer predictions.

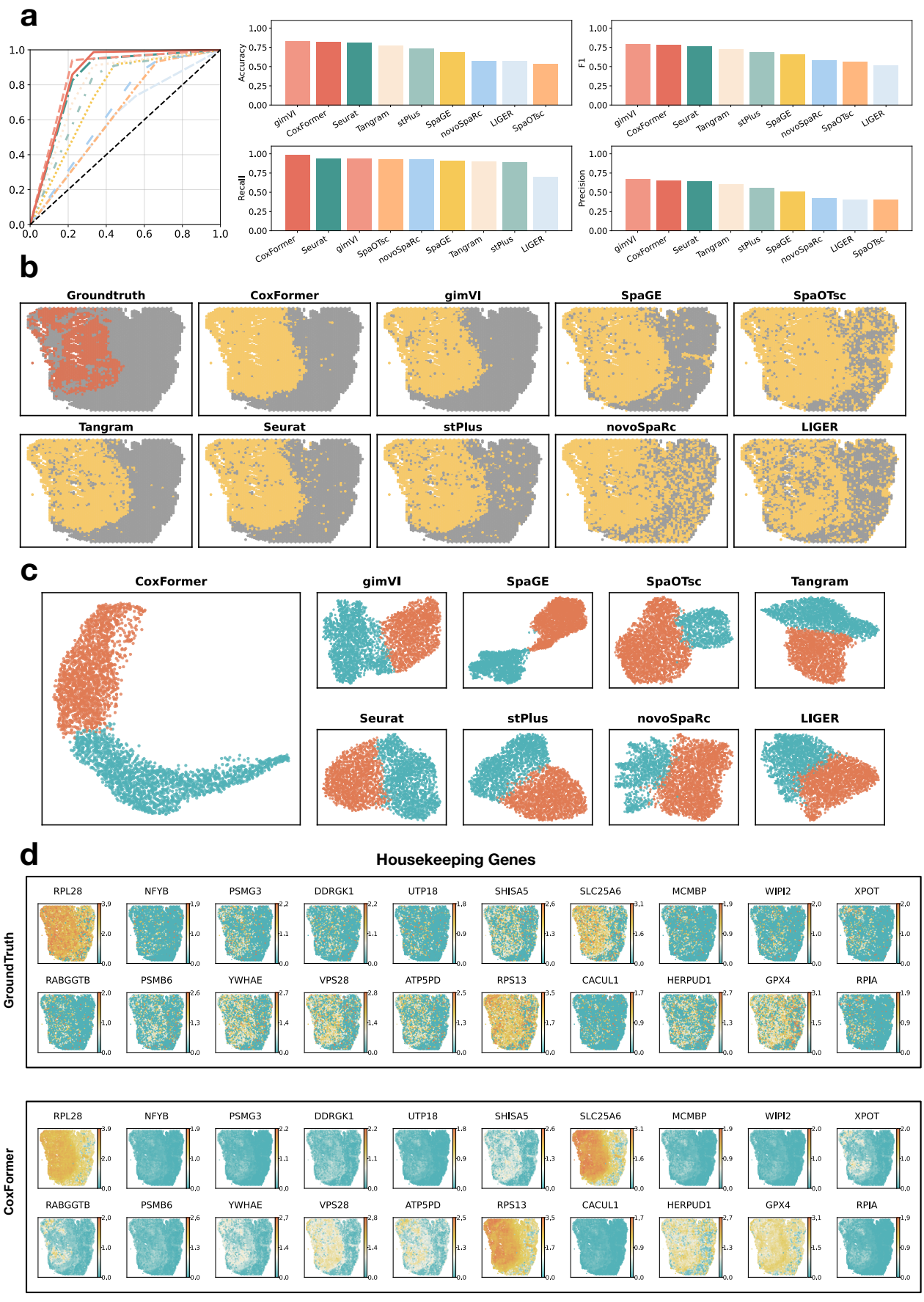

a
Accuracy
F1
Recall
Precision
gimVI
CoxFormer
Seurat
Tangram
stPlus
SpaGE
novoSpaRc
SpaOTsc
LIGER
b
Groundtruth
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
c
CoxFormer
gimVI
SpaGE
SpaOTsc
Tangram
Seurat
stPlus
novoSpaRc
LIGER
d
Housekeeping Genes
GroundTruth
CoxFormer
RPL28
NFYB
PSMG3
DDRGK1
UTP18
SHISA5
SLC25A6
MCMBP
WIPI2
XPOT
RABGGTB
PSMB6
YWHAE
VPS28
ATP5PD
RPS13
CACUL1
HERPUD1
GPX4
RPIA

Figure S18: **Example results of abnormal-region annotation on the therapy-perturbed liver metastasis (LM) section.** We compare CoxFormer with representative reference-required baselines (SpaGE, SpaOTsc, Tangram, gimVI, Seurat, stPlus, novoSpaRc, and LIGER). **(a)** Quantitative performance on this section, reporting AUROC (left) and complementary classification metrics (Accuracy, F1, Recall, Precision). **(b)** Spatial segmentation results are overlaid on the tissue, comparing predicted abnormal regions from each method with the ground-truth annotation. **(c)** UMAP visualization of the predicted labels from panel (b), illustrating separation between abnormal and normal spots under different methods. **(d)** Representative housekeeping-gene expression patterns for ground-truth versus CoxFormer predictions.

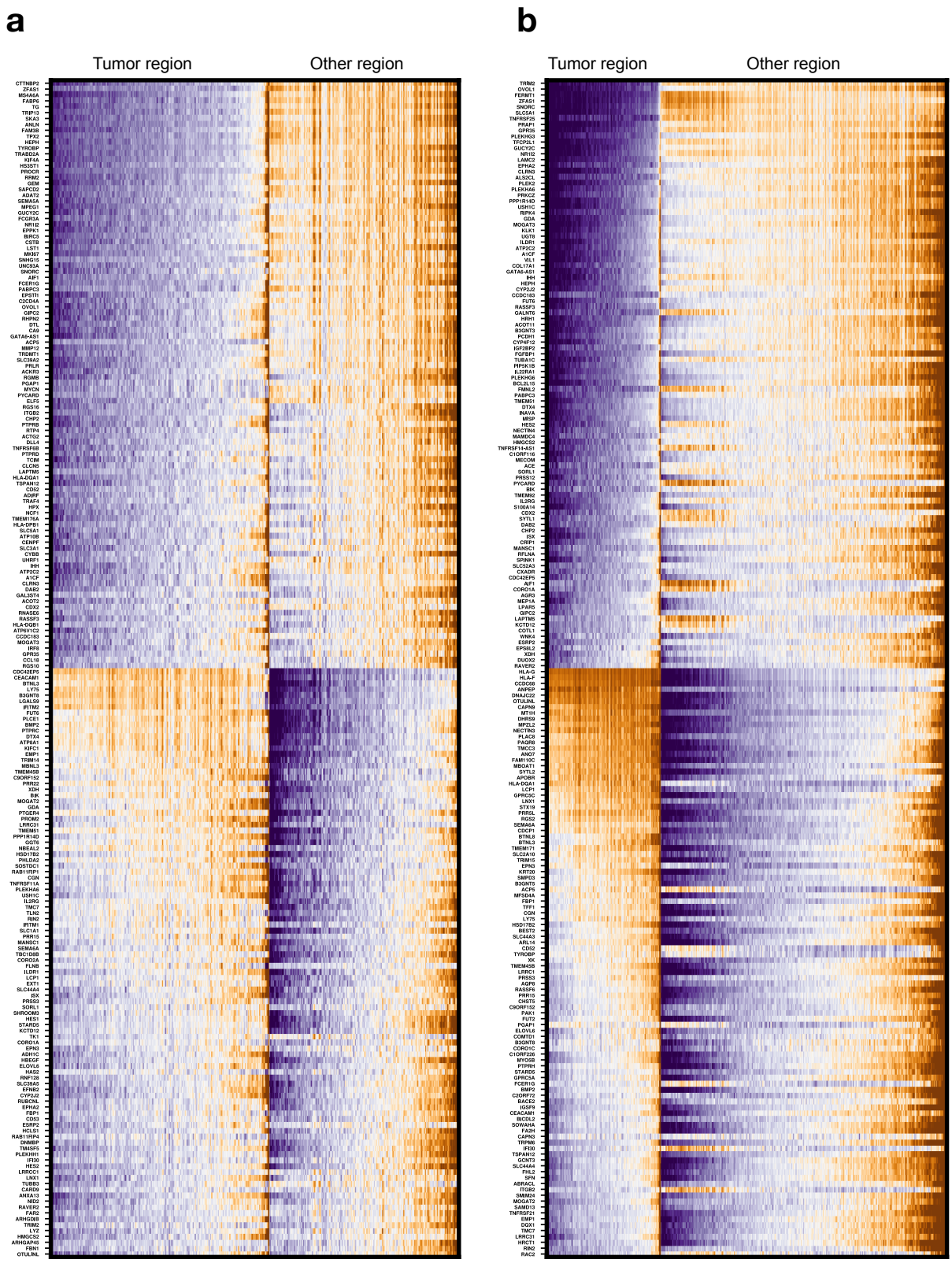

a
Tumor region
Other region
b
Tumor region
Other region

Figure S19: **Differential expression patterns of top marker genes between CoxFormer-segmented normal and abnormal regions in colorectal cancer (CRC) sections.** For each CRC section, we visualize the expression of the top 200 differentially expressed genes (DEGs) identified between the normal and abnormal regions segmented by CoxFormer, highlighting coherent region-wise contrasts associated with pathological states. **(a)** Untreated CRC section. **(b)** Therapy-perturbed (PR) CRC section.

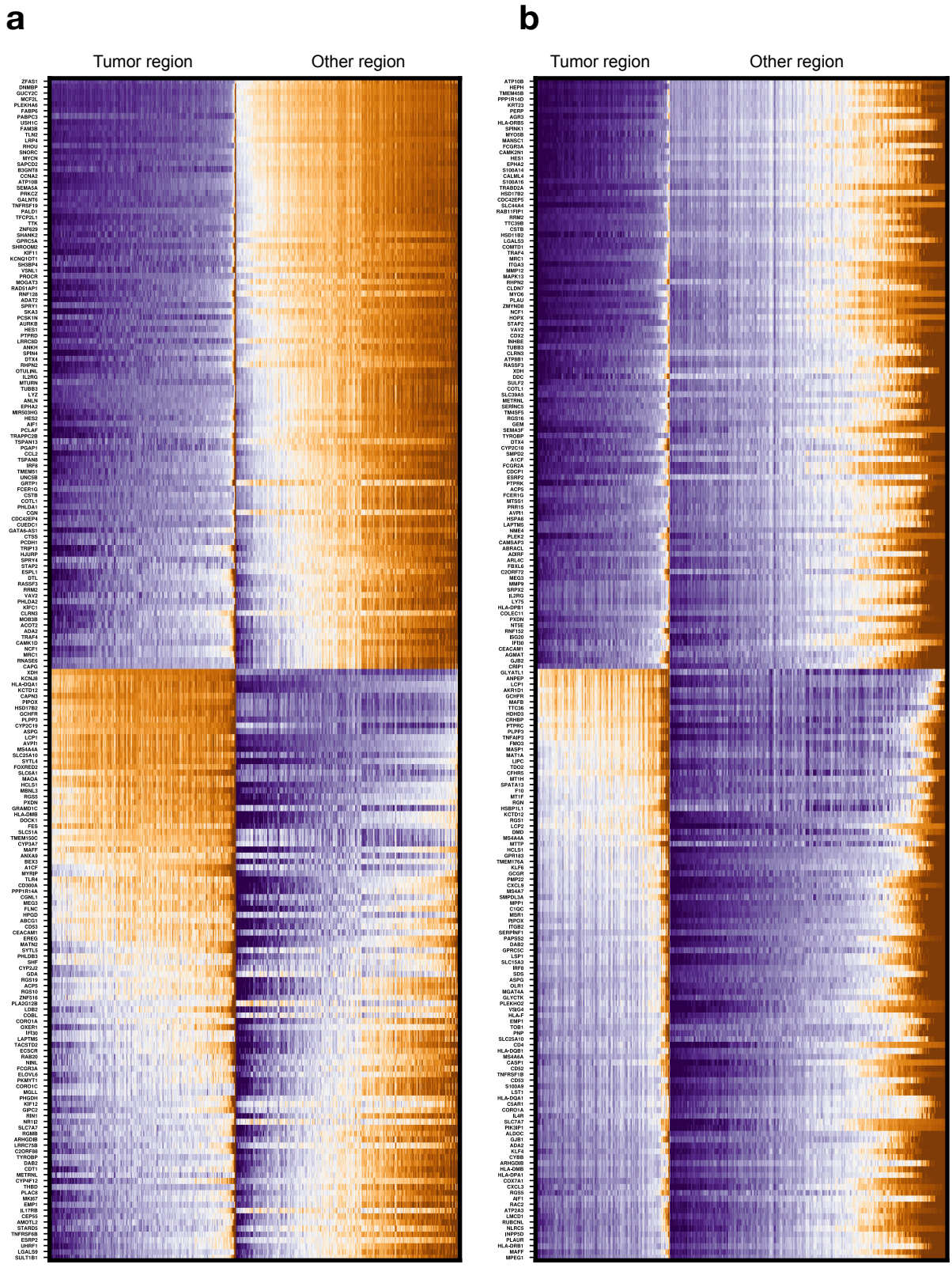
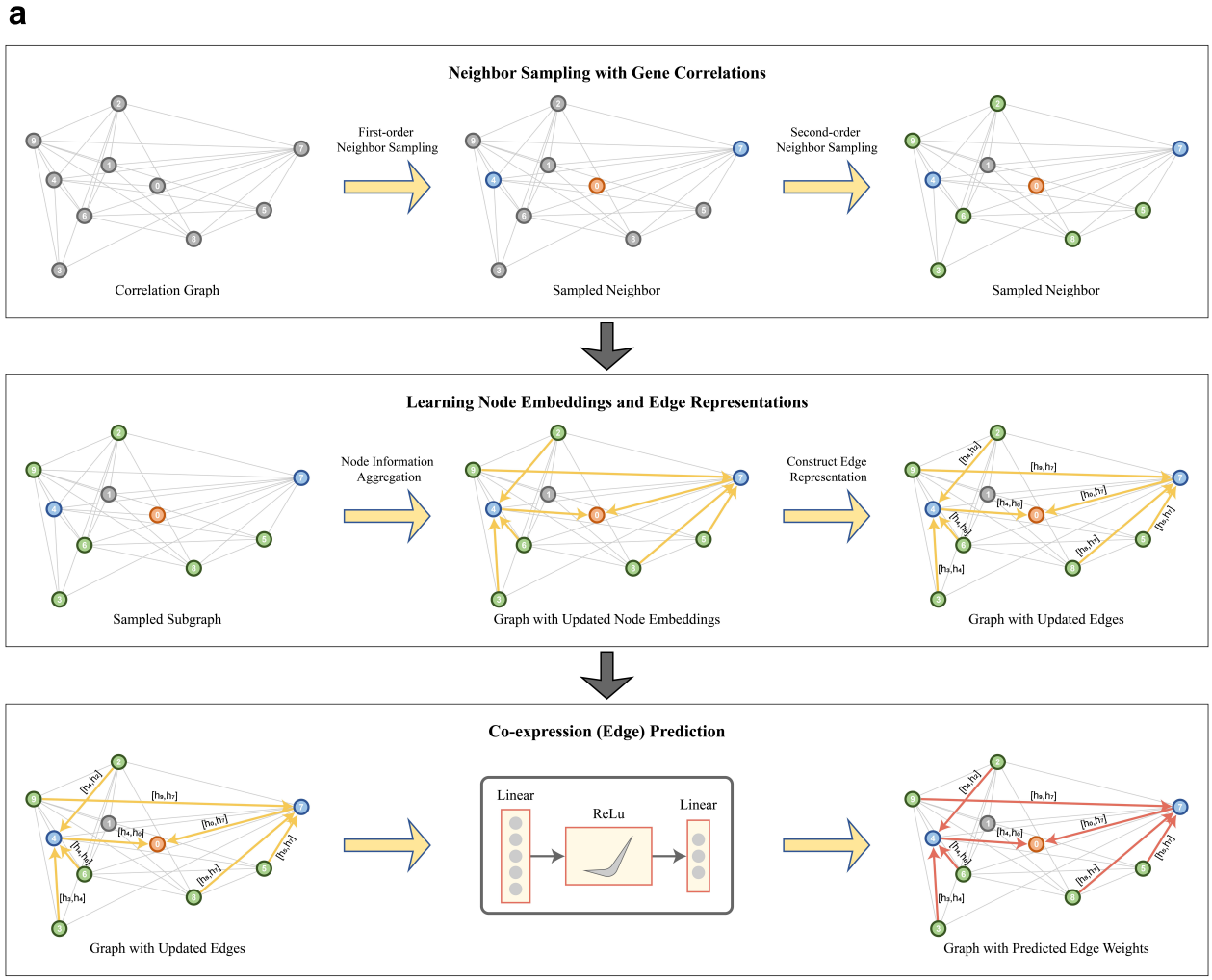

a
Tumor region
Other region
b
Tumor region
Other region

Figure S20: **Differential expression patterns of top marker genes between CoxFormer-segmented normal and abnormal regions in liver metastasis (LM) sections.** For each LM section, we visualize the expression of the top 200 differentially expressed genes (DEGs) identified between the normal and abnormal regions segmented by CoxFormer, highlighting coherent region-wise contrasts associated with pathological states. **(a)** Untreated LM section. **(b)** Therapy-perturbed (PR) LM section.


a
Neighbor Sampling with Gene Correlations
First-order
Neighbor Sampling
Second-order
Neighbor Sampling
Correlation Graph
Sampled Neighbor
Sampled Neighbor
Learning Node Embeddings and Edge Representations
Node Information
Aggregation
Construct Edge
Representation
Sampled Subgraph
Graph with Updated Node Embeddings
Graph with Updated Edges
Co-expression (Edge) Prediction
Linear
ReLu
Linear
Graph with Updated Edges
Graph with Predicted Edge Weights

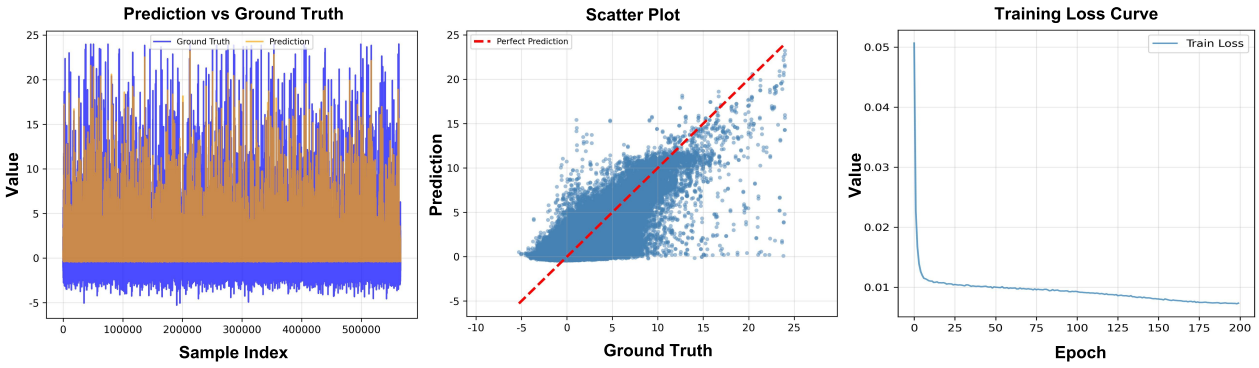

b
Prediction vs Ground Truth
Ground Truth
Prediction
Value
Sample Index
Scatter Plot
Perfect Prediction
Prediction
Ground Truth
Training Loss Curve
Train Loss
Value
Epoch

Figure S21: **Training Pipeline and Training Analysis. (a)** Coxformer predicts gene co-expression by learning node embeddings and edge representations on a gene correlation graph. Although training starts from the global graph, optimisation is performed on correlation-guided sampled subgraphs for efficiency. For each target node (or candidate gene pair), we construct a compact computational neighbourhood by sampling first-order neighbours that provide direct co-expression context and second-order neighbours that add higher-order relational evidence via shared correlation partners. Node embeddings are updated via neighbourhood aggregation, edge representations are constructed from pairs of updated node embeddings, and an edge-value regressor outputs predicted co-expression (edge) weights for sampled gene pairs, yielding a graph with updated (predicted) edge weights. **(b)** Training analysis and prediction performance: (left) prediction and ground-truth values across sampled indices, (middle) scatter plot of prediction vs. ground truth (with the diagonal indicating perfect prediction), and (right) training loss curve over epochs.