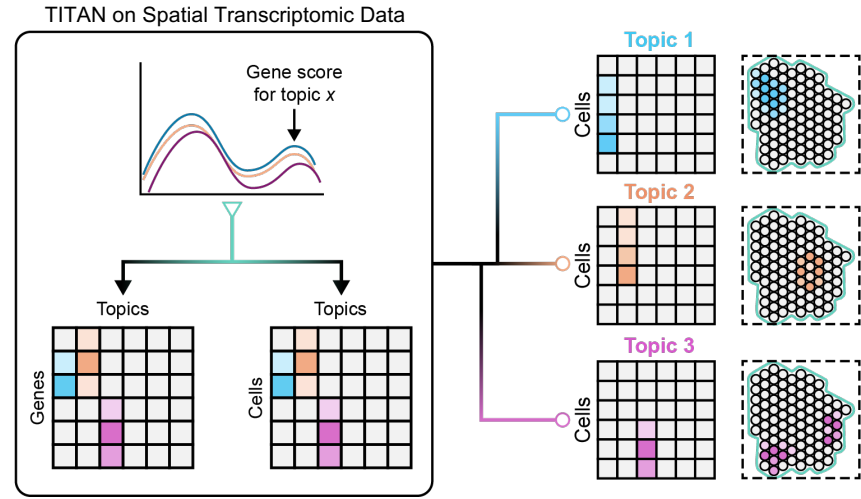
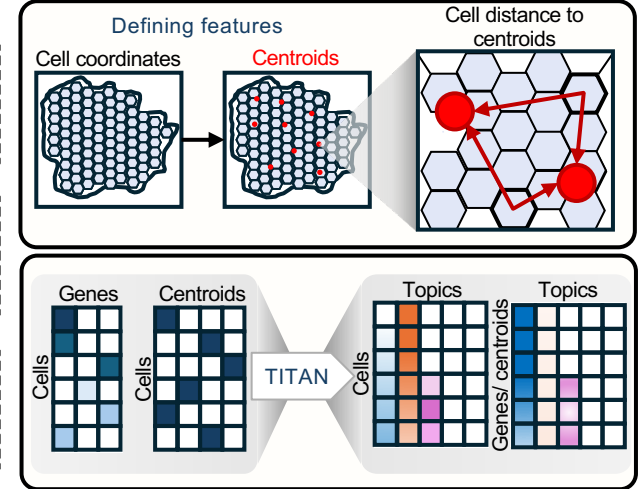


E1

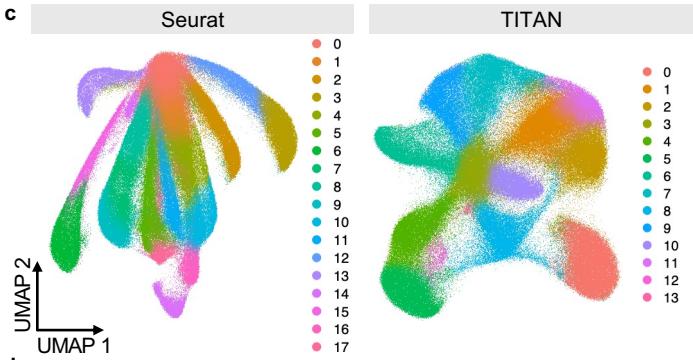
a



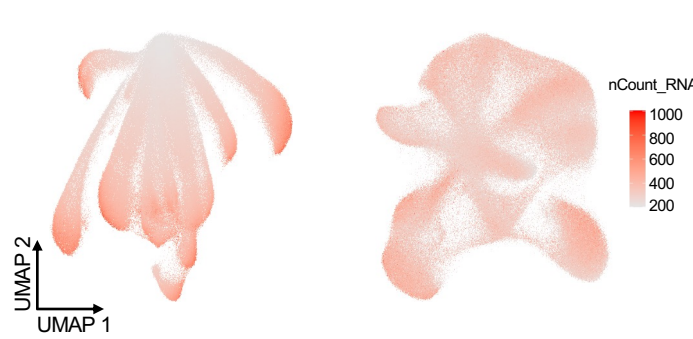
b Incorporating spatial data to TITAN



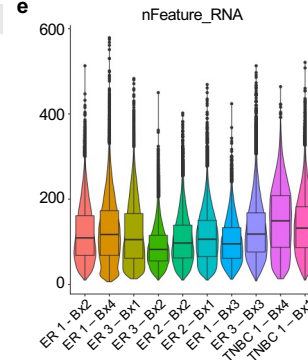
c



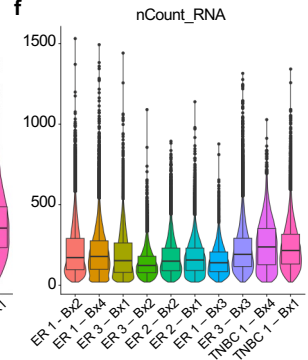
d



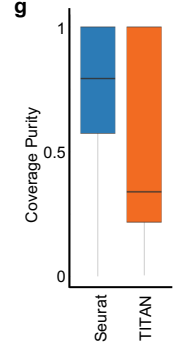
e



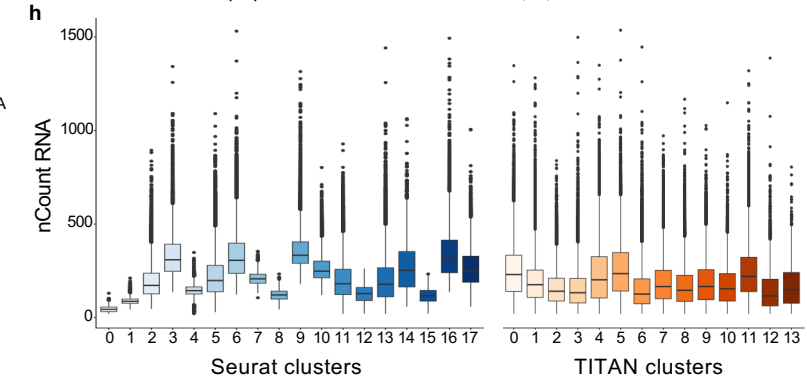
f



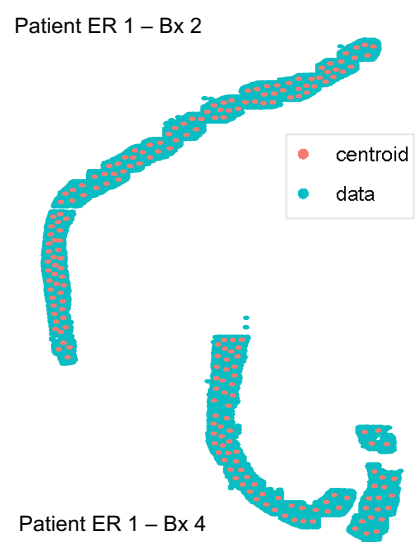
g



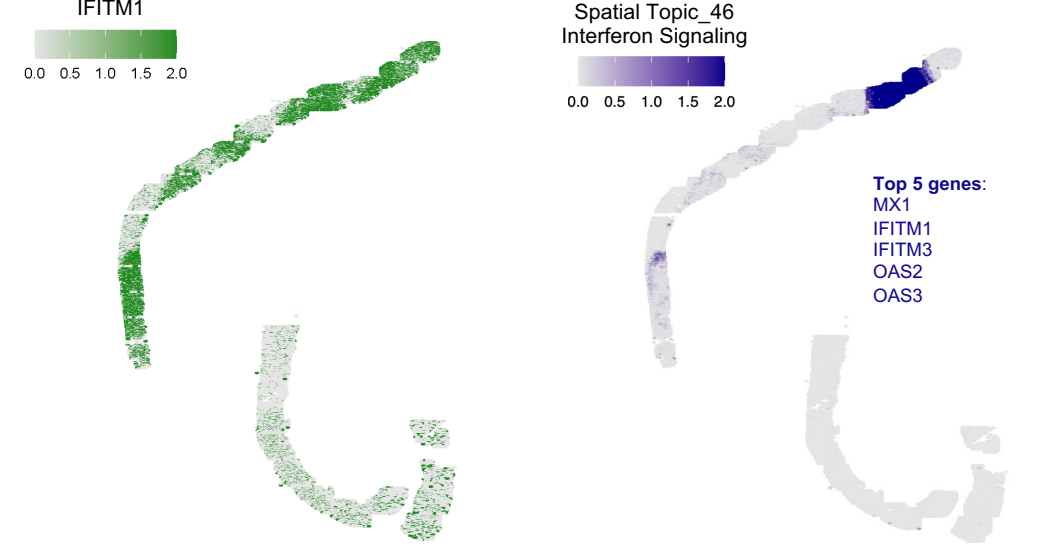
h



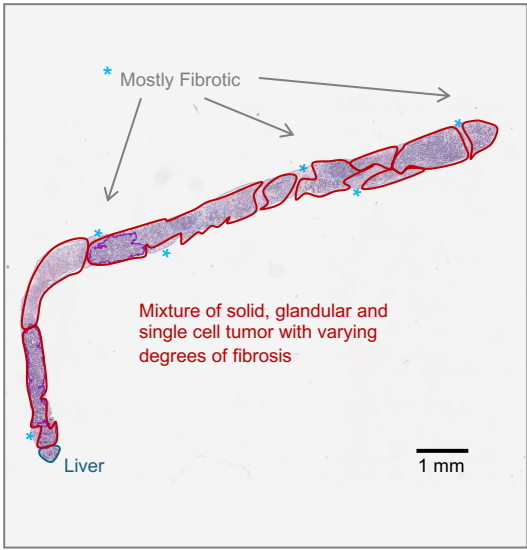
i



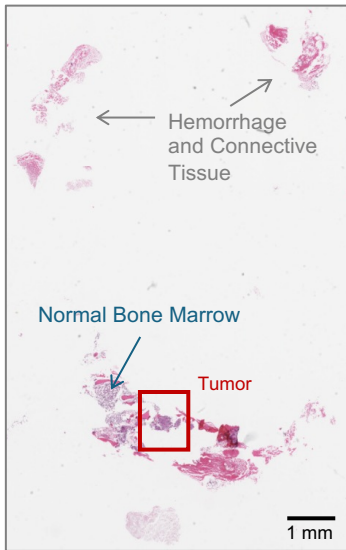
j



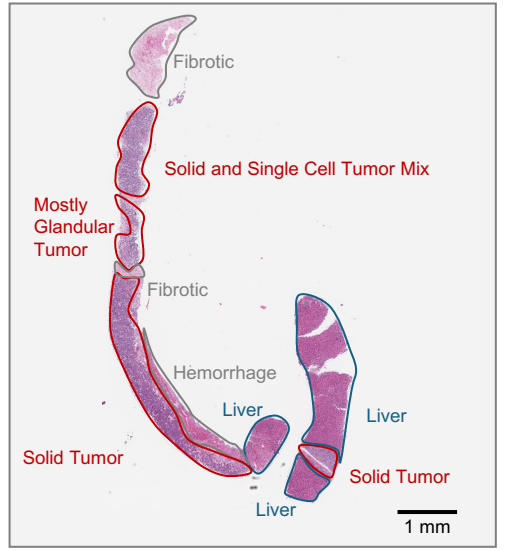
Patient ER 1 – Biopsy 2
Liver Metastasis



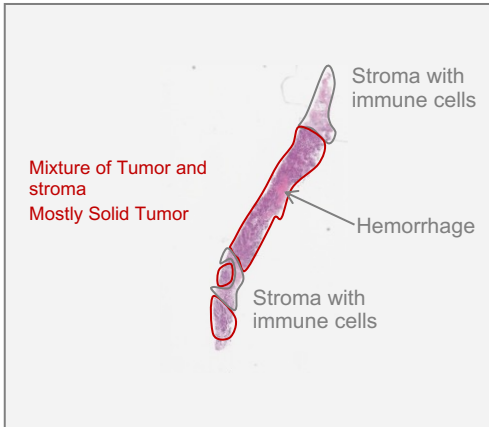
Patient ER 1 – Biopsy 3
Lumbar Spine Metastasis



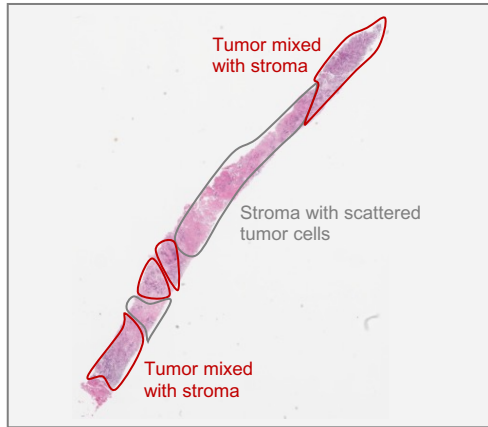
Patient ER 1 – Biopsy 4
Liver Metastasis



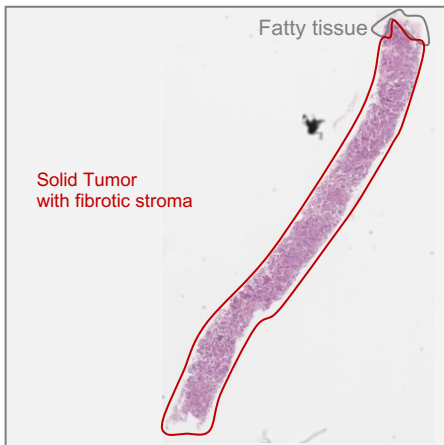
Patient ER 2 – Biopsy 1
Lymph Node Metastasis



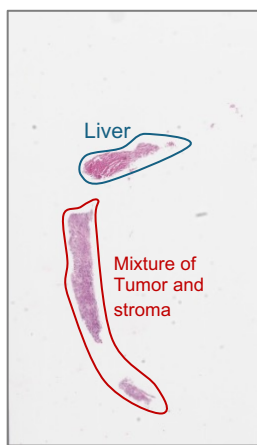
Patient ER 2 – Biopsy 2
Lymph Node Metastasis



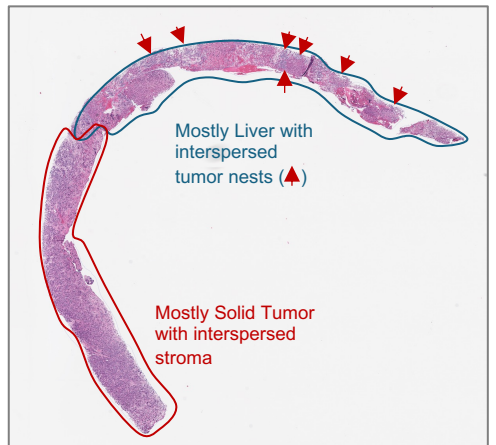
Patient ER 3 – Biopsy 1
Lymph Node Metastasis



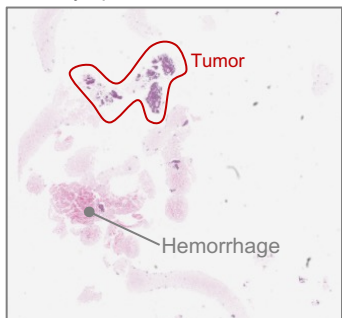
Patient ER 3 – Biopsy 2
Liver Metastasis



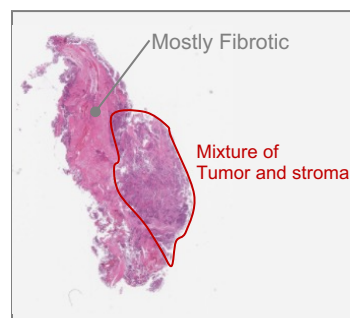
Patient ER 3 – Biopsy 3
Liver Metastasis



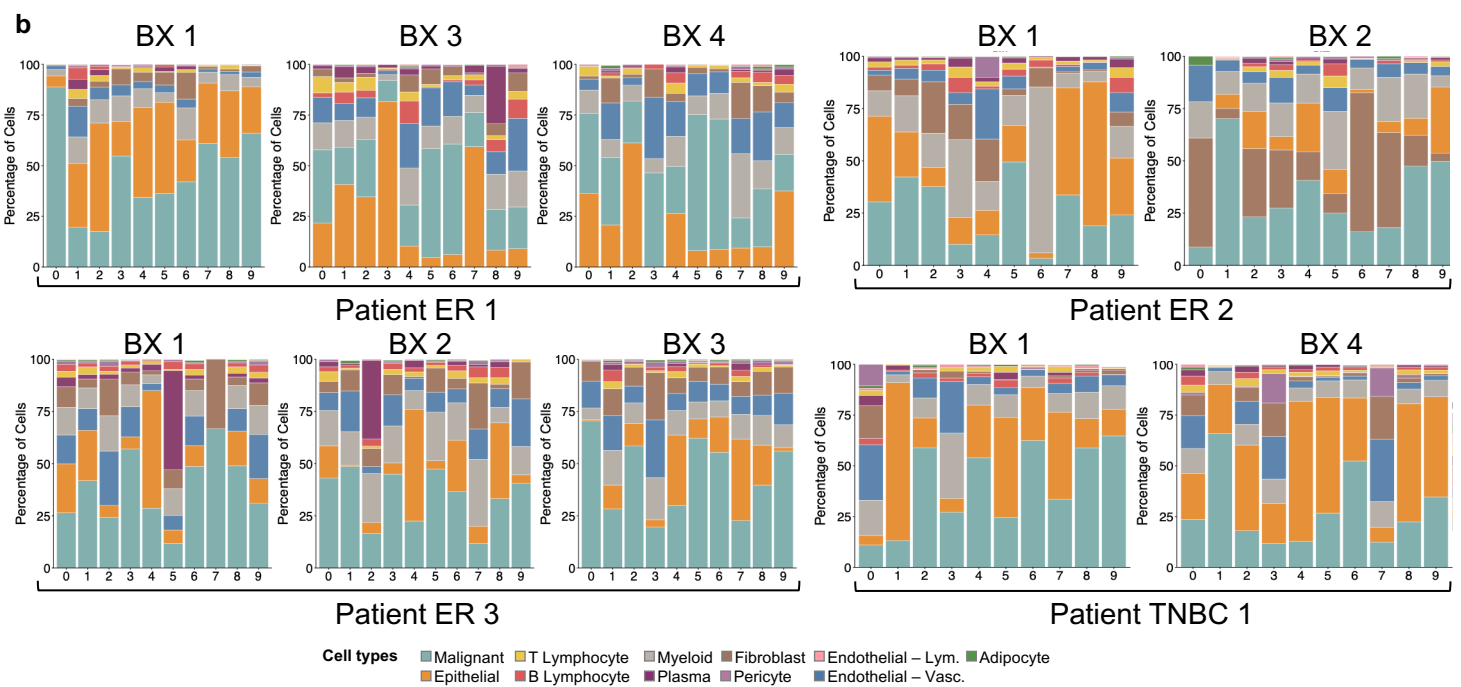
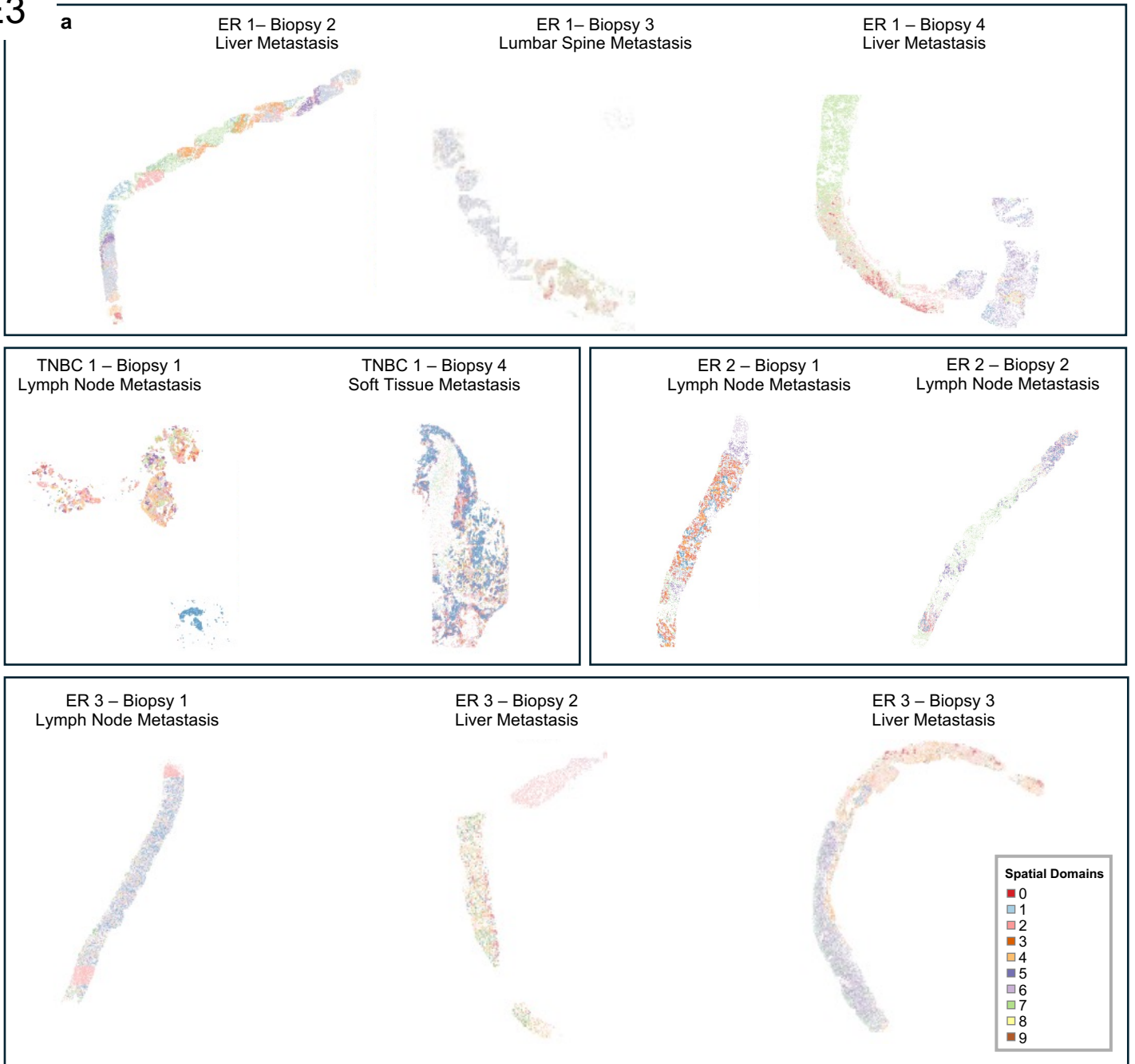
Patient TNBC 1 – Biopsy 1
Lymph Node Metastasis



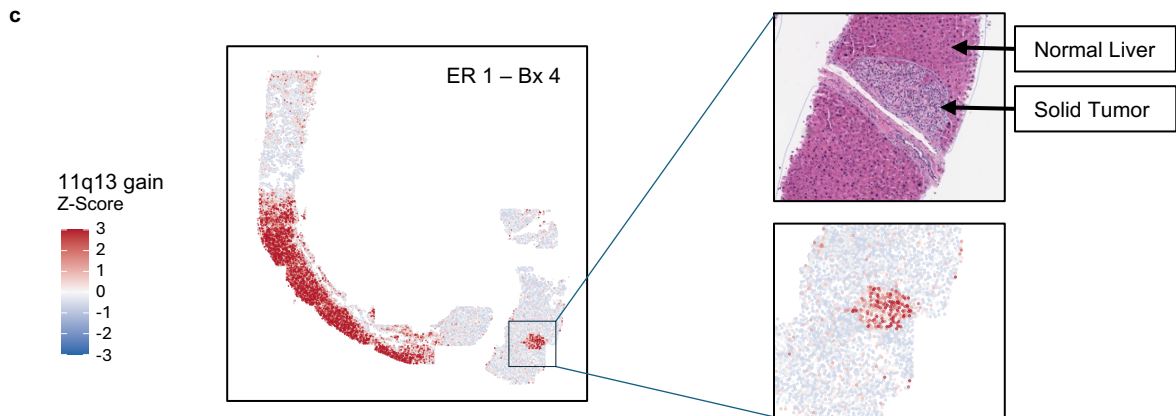
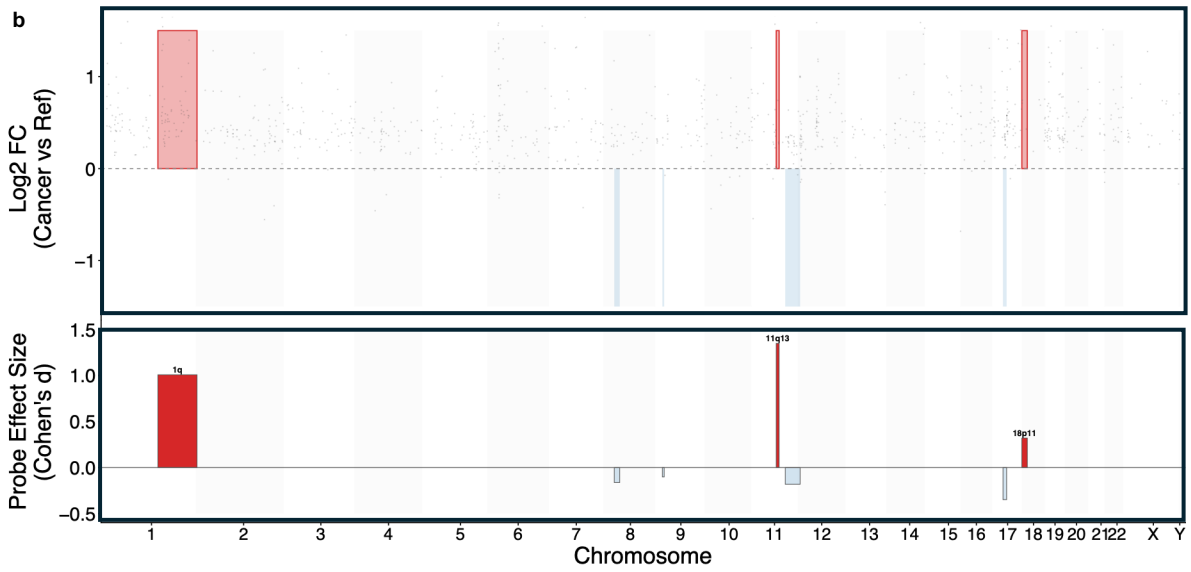
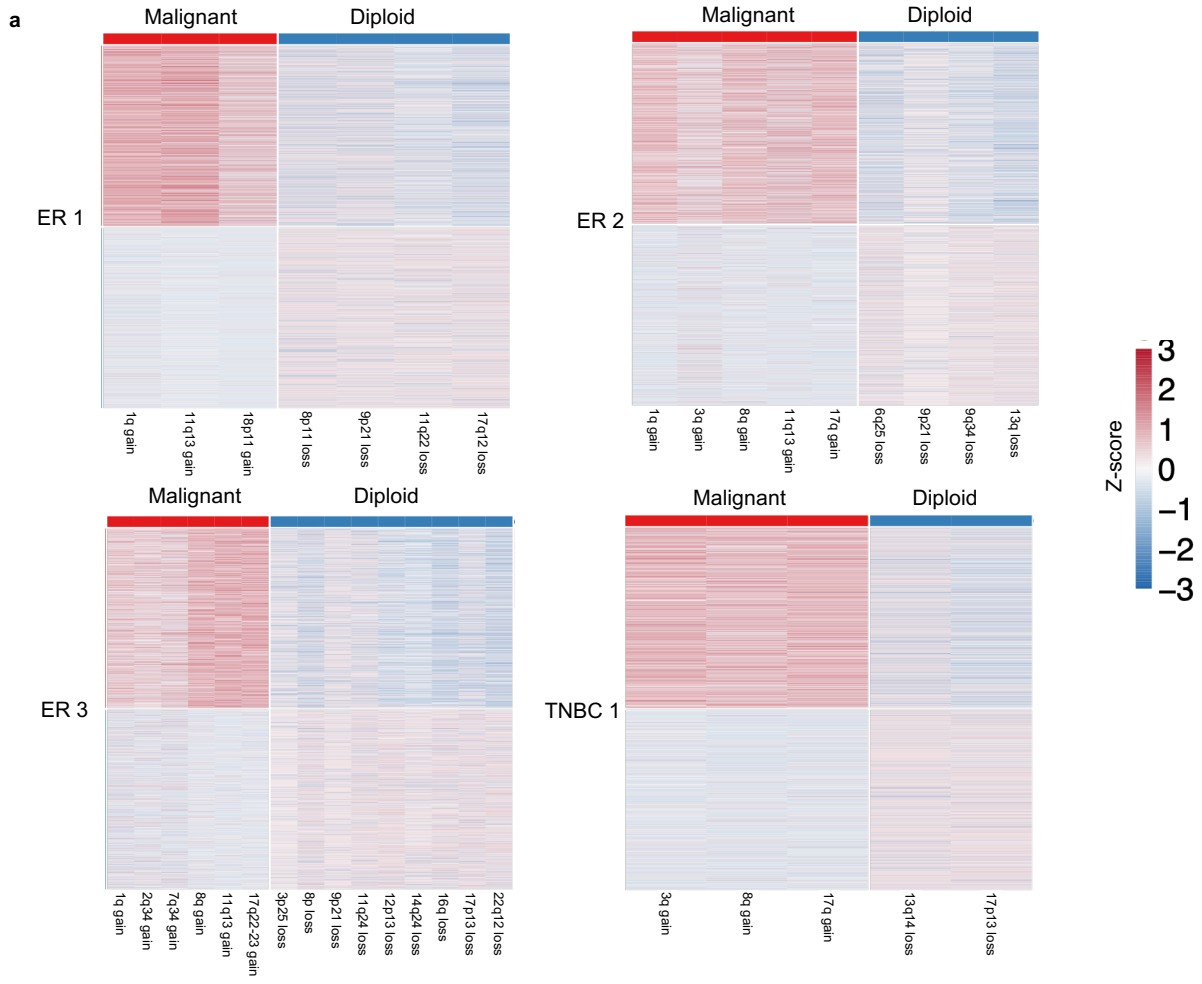
Patient TNBC 1 – Biopsy 4
Soft Tissue Metastasis

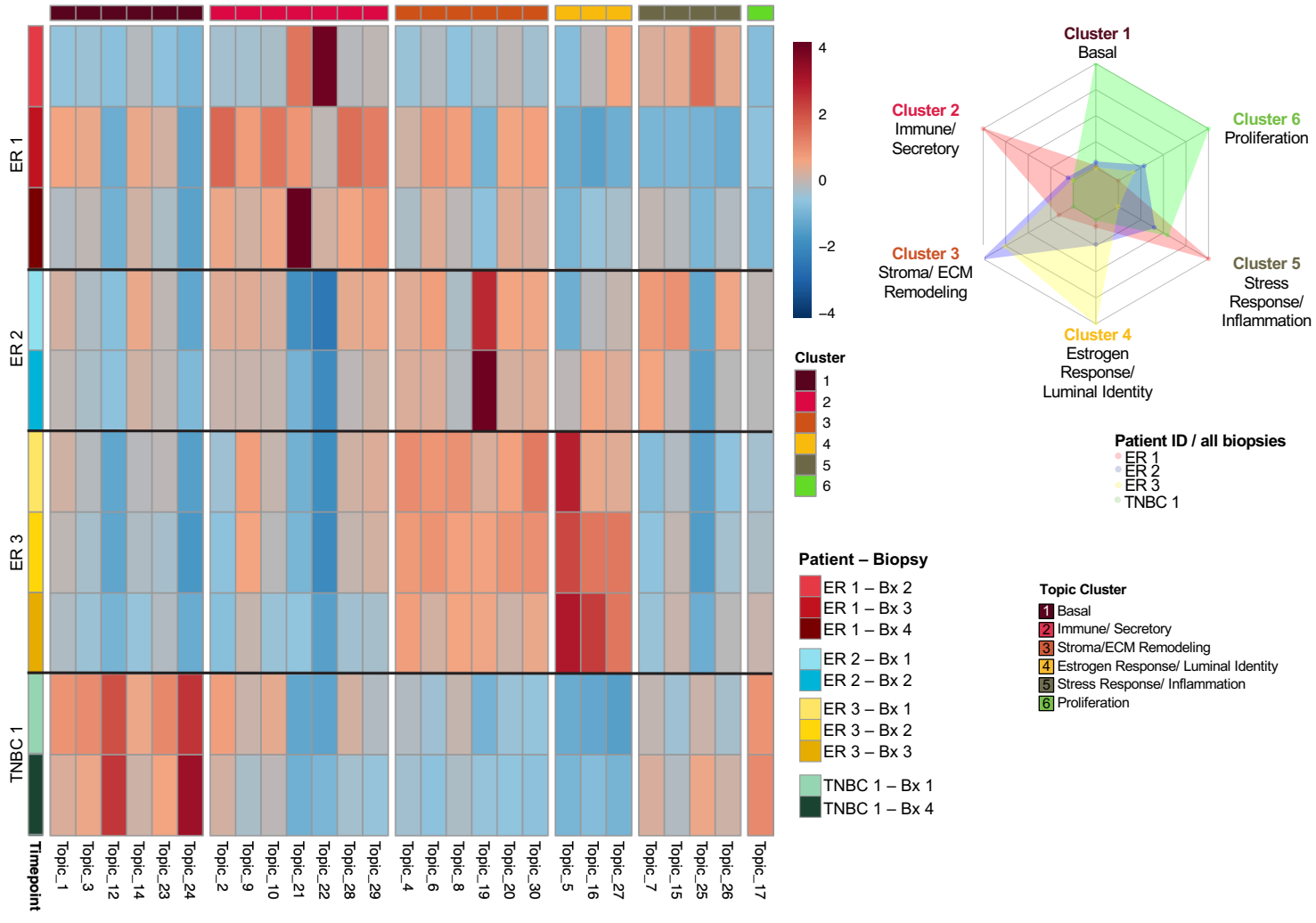


E3



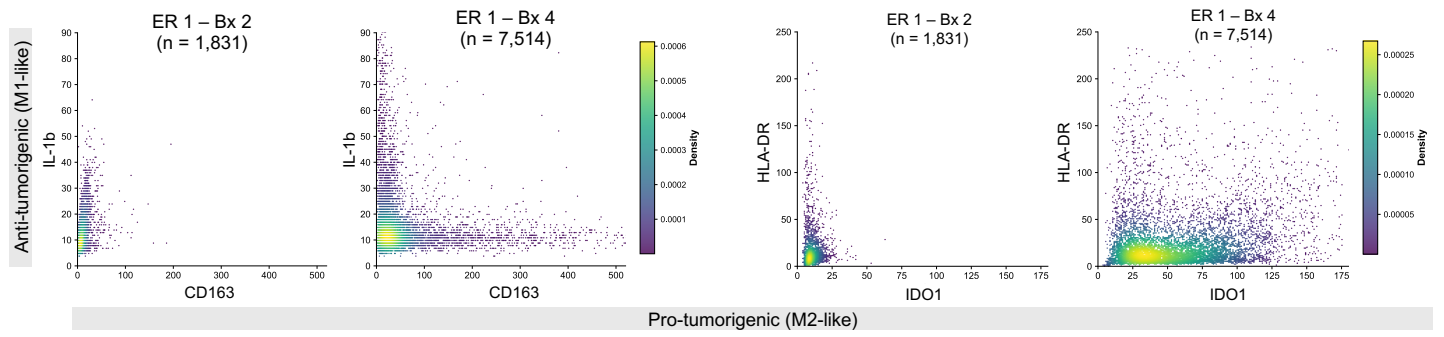
E4





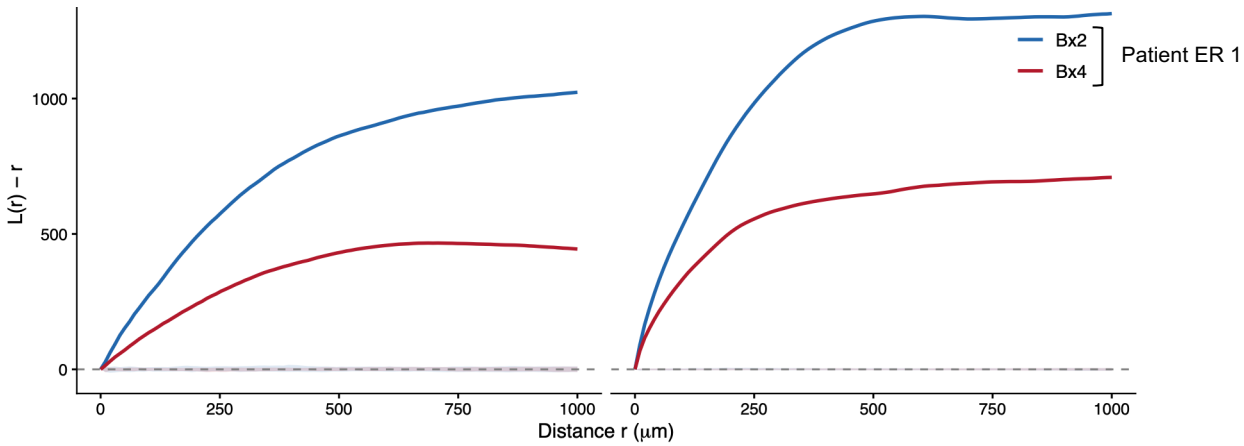
a

Macrophages State Shifts



b

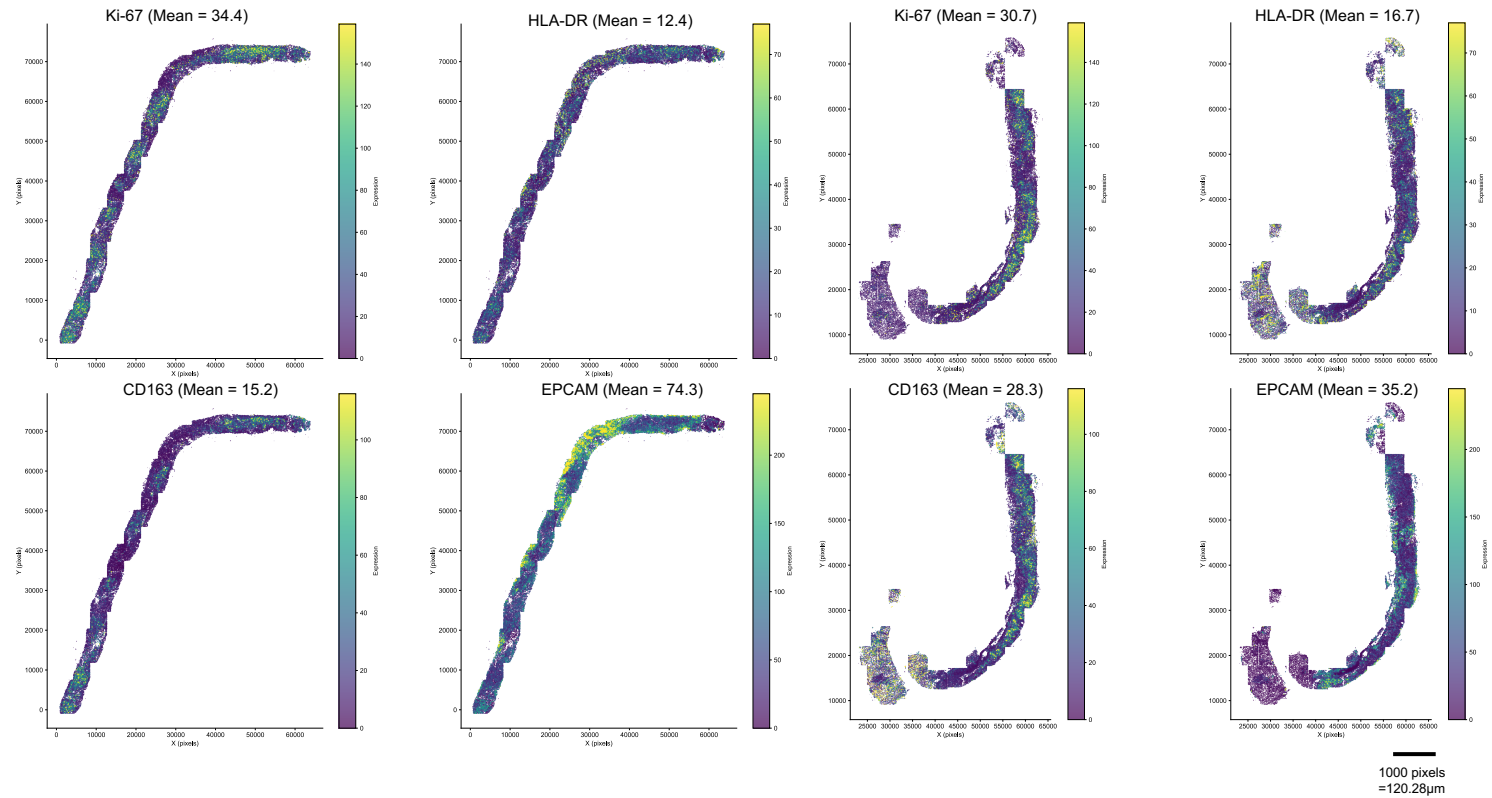
T-cell Spatial Aggregation Strength – Ripley's L (RNA) T-cell Spatial Aggregation Strength – Ripley's L (Protein)



c

Patient ER 1 – Bx 2 CosMx (Protein Expression)

Patient ER 1 – Bx 4 CosMx (Protein Expression)



Extended Data Figure 1. TITAN topic modeling framework, spatial feature integration, and benchmarking against standard clustering.

- a.** Overview of the TITAN topic modeling framework. TITAN decomposes single-cell gene expression into topic-cell and gene-topic matrices to identify latent transcriptional programs. Inferred topic scores can be projected onto tissue space to enable spatial visualization of transcriptional states.
- b.** Schematic illustrating the incorporation of spatial information into the TITAN framework. Spatial domain centroids are defined from cell coordinates, and per-cell distances to these centroids are computed and added to TITAN's input matrices as additional features for topic inference.
- c.** Uniform Manifold Approximation and Projection (UMAP) visualizations of spatial transcriptomics data of Seurat (left) and TITAN (right) clustering in all cells (n=345,207).
- d.** UMAP visualizations of spatial transcriptomics data as show in (c) colored by total RNA counts per cell (nCount_RNA).
- e.** Distribution of detected features (nFeature_RNA) across all samples shown as violin plots.
- f.** Distribution of RNA counts (nCount_RNA) across all samples shown as violin plots.
- g.** Clustering quality comparison between Seurat and TITAN shown as a bar chart of coverage purity scores; higher relative values indicate greater influence of sequencing depth on cluster identity.
- h.** Gene count distributions (nCount RNA) across Seurat (left) and TITAN (right) clusters shown as box plots of log-transformed RNA counts.
- i.** Spatial mapping in Patient ER 1 Biopsy 2 (Bx2) and Bx4 of domain centroids as defined in methods and panel (b).
- j.** Spatial mapping in Patient ER 1 Biopsy 2 (Bx2) and Bx4 of gene expression for a representative interferon-stimulated gene (IFITM1), and topic score for a spatial topic representative of localized interferon signaling (Topic 46).

Extended Data Figure 2. Histopathological characterization of all biopsies.

Representative H&E-stained sections from all biopsies across the four patients with pathological annotations. ER 1: Bx2 (liver), Bx3 (lumbar spine), Bx4 (liver). ER 2: Bx1 and Bx2 (lymph node). ER 3: Bx1 (lymph node), Bx2 and Bx3 (liver). TNBC 1: Bx1 (lymph node), Bx4 (soft tissue). Scale bars: 1 mm.

Extended Data Figure 3. STAGATE-defined spatial domain organization and cell type annotation across all biopsies.

- a.** STAGATE-defined spatial domain organization across all biopsies. Each cell is colored by its assigned domain identity (0–9).
- b.** Cell type distribution across STAGATE-defined spatial domains for all biopsies.

Extended Data Figure 4. Single-cell copy number alteration profiles and clone identification.

- a.** Single-cell copy number alteration (CNA) profiles for all four patients shown as heatmaps. Rows represent individual cells ordered by clone assignment (Malignant or Diploid); columns represent chromosomal regions. Key recurrent alterations are annotated for each patient.
- b.** Visualization of patient ER 1 specific CNA at a genome wide level. (Top) Log₂ fold change (Log₂FC) of gene expression in malignant cells relative to a diploid baseline for key copy number events; gains shown in red and losses in blue. (Bottom) Robustness of each inferred event quantified by probe effect size (Cohen's d).
- c.** Spatial overlay of CNA-based clone assignments in ER 1, Bx4, distinguishing malignant from diploid epithelial populations across the tissue section, with 11q13 copy number amplification Z-score overlaid on tissue morphology. A high-resolution inset shows a tumor nest within the host liver tissue and its corresponding H&E stain; adjacent normal liver and solid tumor regions are annotated.

Extended Data Figure 5. Transcriptional topic landscape across all patients and biopsies.

Transcriptional program composition across all patients and biopsies. (Left) Relative enrichment of all 30 TITAN-identified topics across all patient biopsies shown as a heatmap, grouped into six functional clusters: Basal (Cluster 1), Immune/Secretory (Cluster 2), Stroma/ECM Remodeling (Cluster 3), Estrogen Response/Luminal Identity (Cluster 4), Stress Response/Inflammation (Cluster 5), and Proliferation (Cluster 6). (Right) Relative contribution of each functional cluster per patient across all biopsies shown as a radar plot.

Extended Data Figure 6. T cell spatial organization and protein marker validation in ER 1.

- a.** Macrophage phenotype marker co-expression in ER 1 Bx2 (n=1,831) and Bx4 (n=7,514) shown as scatter plots comparing anti-tumorigenic M1-like markers (IL-1 β , HLA-DR) and pro-tumorigenic M2-like markers (CD163, IDO1), supporting the macrophage polarization shift toward an M2-like state in Bx4 shown in Fig. 2h.
- b.** T cell spatial aggregation in ER 1 quantified using Ripley's L function, comparing Bx2 and Bx4 based on RNA-based (left) and protein-based (right) T cell identification. Decreased L(r) values in Bx4 indicate dispersal of spatially organized T cell clusters.
- c.** Spatial distribution of key protein marker (Ki-67, HLA-DR, CD163, EPCAM) expression in ER 1 Bx2 and Bx4 from the CosMx protein panel, with mean expression values indicated per biopsy. 1 pixel=0.12028 μ m.