Supplementary Material

S1 Additional Data

In addition to the data utilized in the main body of our paper, we employed several publicly available datasets in the supplementary material. These datasets can be accessed via the following websites or using the provided accession numbers: (1) Stereo-seq data[48].

S2 Comparison baselines

Our experimental results demonstrate that our approach achieves very high accuracy. However, it is important to note that previous literature on these methods has some limitations:

1. Previous spatial transcriptomics analysis methods have not been able to enhance gene expression to single-cell resolution without using single-cell RNA-seq data.

2. Certain deconvolution methods use public single-cell RNA-seq references, including the Human Cell Atlas, the BRAIN Initiative Cell Census Network (BICCN), and the Human BioMolecular Atlas, to address the problem of low resolution in spatial transcriptomics. However, these methods are prone to incomplete identification of cell types due to batch effects and tissue heterogeneity in samples. Additionally, the accuracy of deconvolution may be impacted by the different perturbations that affect single-cell references and spatial transcriptomics.

To overcome these limitations, we propose a novel framework called TransformerST, which utilizes a Transformer-based approach to associate the heterogeneity of local gene expression properties and reveal the structural relationships at nearly single-cell resolution. We have also included a table that compares the features of various spatial transcriptomics analysis methods, highlighting their respective strengths and limitations (Supplementary Table 1).

S3 Tissue identification with Stereo-seq technology

We use the proposed method to a Stereo-seq dataset derived from mouse olfactory bulb tissue. Nowadays, stereo-seq is one of the most promising new spatial holography methods available. DNA nanosphere-patterned array chips provide spatial resolution down to the subcellular level. The information in this study was split down to a cellular level (14 m) for clarity. The rostral migratory stream (RMS), granular cell layer (GCL), endothelial cell layer (GCL), internal plexiform layer (IPL), filamentous cell layer (MCL), exterior plexiform layer (EPL), and olfactory nerve layer were all labeled in DAPI-stained pictures of

Supplementary Table 1 Comparison between TransformerST with baselines.

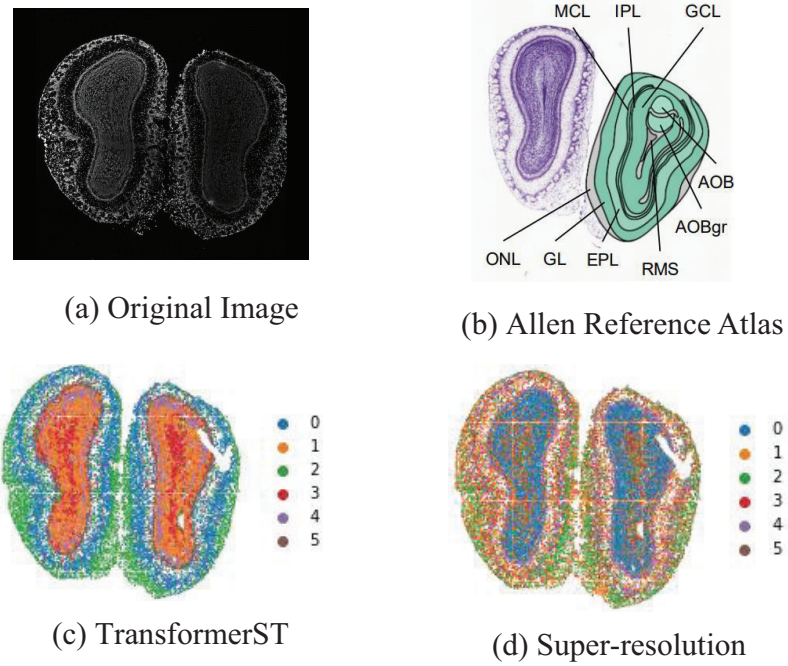
Methods	Objective	Super-resolution	Reference-free	Histology image
TransformerST	Clustering, super-resolution	Single-cell	Yes	Yes
SpaGCN	Clustering	Original	No	Yes
BayesSpace	Clustering, super-resolution	Multi-cellular	Yes	No
CCST	Clustering	Original	No	No
STAGATE	Clustering	Original	No	No
DeepST	Clustering	Original	No	Yes
stLearn	Clustering	Original	No	No
STdeconvolve	Deconvolution	Multi-cellular	Yes	No

the coronal mouse olfactory bulb by Fu et al.[48]. The results are shown in Supplementary Supplementary Fig. 1. The results derived from TransformerST align with the Allen Reference Atlas, as depicted in Supplementary Supplementary Fig. 1b, demonstrating the method's precision and consistency. In Supplementary Supplementary Fig. 1c, TransformerST, utilizing the original spot resolution, yielded results that closely mirrored those of the Allen Reference Atlas. This showcases its capacity to capture intricate gene expression patterns and cellular structures.

Furthermore, in Supplementary Supplementary Fig. 1d, TransformerST effectively reconstructed the original resolution of spatial transcriptomics data using downsampled inputs. The outcomes closely resembled those in the Allen Reference Atlas, underscoring TransformerST's robustness and flexibility in managing various data resolutions and ensuring accurate results irrespective of input quality.

The compatibility of TransformerST with the Allen Reference Atlas can also be attributed to its sophisticated algorithms. These enable the precise identification of spatial gene expression patterns and facilitate the extraction of biologically significant information. Additionally, the method's robust feature extraction and pattern recognition capabilities aid in effectively distinguishing between different cell types, tissue structures, and gene expression patterns. Moreover, TransformerST's versatility allows it to be applied across multiple platforms and datasets, maintaining consistent performance under a wide array of experimental conditions. This adaptability ensures the method's reliability and relevance, further reinforcing its alignment with the established Allen Reference Atlas.

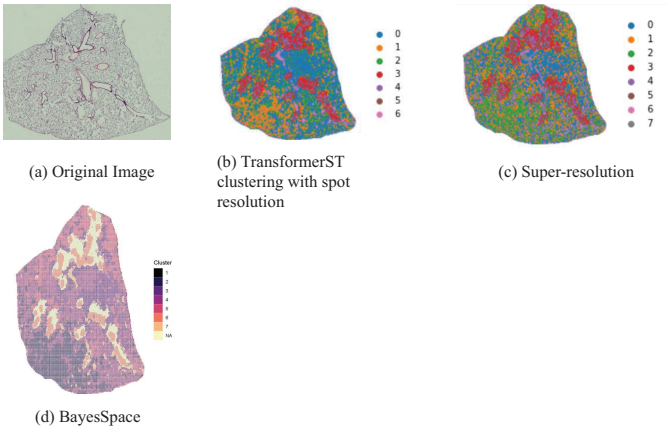
In conclusion, the concordance between the results obtained from TransformerST and the Allen Reference Atlas can be attributed to the method's



Supplementary Figure 1 TransformerST improves the identification of known tissue structures in the olfactory bulb tissue

precision, consistency, adaptability, and its ability to effectively manage various data resolutions. These attributes position TransformerST as a valuable tool in spatial transcriptomics analysis, offering researchers reliable and biologically pertinent insights.

We applied the proposed methods to a segment of mouse lung tissue profiled by Stereo-seq and found that the spatial regions identified by the method showed strong concordance with the patterns of the original image (Supplementary Supplementary Fig. 2). This underscores the effectiveness of our approach in accurately delineating tissue structures and gene expression patterns, a critical aspect for understanding the complex biological processes within the sample. The method's ability to yield results consistent with expert annotations underscores its potential as a valuable tool for spatial transcriptomics analysis across diverse tissue types and experimental conditions. We also showcase the results of the BayesSpace method in Supplementary Supplementary Fig. 2d, which uses the original data for tissue type clustering. Our simulation outcome aligns with the results of BayesSpace. Importantly, our simulation results for super-resolution with downsampled data are capable of producing more refined patterns that closely mirror the original image.



Supplementary Figure 2 TransformerST improves the identification of known tissue structures in the mouse lung tissue

This further attests to the effectiveness of our proposed approach in handling various resolutions and providing detailed spatial transcriptomics analysis.

For both datasets, our proposed method produced visually superior tissue identification and super-resolution results when compared to other methods. In the two figures, 'TransformerST' denotes the clustering result at the original resolution, while 'Super-resolution' refers to the super-resolution outcome achieved by TransformerST. BayesSpace denotes the result achieved when data at the original resolution is employed for tissue type clustering.

S4 Enhanced Gene Expression Prediction at Sub-cellular Resolution

Enhanced Gene expression prediction at sub-cellular resolution in breast cancer data HER2+. To evaluate the tissue identification and super-resolution performance in predicting gene expression at sub-cellular resolution using histology images, we employed the leave-one-out method (36 fold) using HER2+ breast cancer data, which includes 36 tissue sections from 8 patients. This approach underscores the effectiveness in gene expression prediction and super-resolution. For the leave-one-out evaluation, 32 sections were utilized to train the tissue identification and super-resolution model, with the remaining sections reserved for evaluation. The results of this approach are represented as TransformerST. We also assessed the super-resolution performance at the sub-cellular level, referred to as Super-resolution.

Manual annotation of three tissue sections was included for evaluating clustering accuracy [49]. We compared our proposed method, TransformerST,

Supplementary Table 2 Average correlation of predicted expression for top 50 most highly variable genes (HVG) compared to ground truth expressions on held out dataset

Methods	HVG
TransformerST	0.226
BLEEP	0.168
TCGN	0.151

with TCGN [37] and BLEEP [36] for gene expression prediction using three tissue sections, as detailed in Supplementary Fig. 3. TCGN and BLEEP, in comparison, demonstrated less effective gene prediction performance. The leave-one-out evaluation showed a higher correlation with biological interpretation.

TransformerST markedly improved clustering accuracy (ARI) for the evaluated sections out of the 32 samples, significantly outperforming TCGN and BLEEP. This advancement in clustering accuracy by TransformerST is evident in the results for sample B1 (Supplementary Fig. 3a), where TransformerST achieved the highest clustering accuracy (ARI of 0.308 for TransformerST compared to lower ARIs for TCGN and BLEEP). Similarly, Supplementary Fig. 3b displays the superiority of TransformerST (ARI=0.325) over TCGN and BLEEP. These results might be due to substantial gene expression differences among patients, which led to the leave-one-out evaluation by TransformerST attaining superior tissue identification performance, as shown in Supplementary Fig. 3). Supplementary Table 2 displays the performance comparison of BLEEP, TCGN, and TransformerST in predicting the top 50 most highly variable genes (HVG). In this assessment, the expression profiles predicted by TransformerST showed the strongest correlation with the actual data among all gene sets.

The enhanced single-cell resolution results further demonstrate TransformerST could predict the biological meaningful patterns as in the manual annotations. While it is hard to estimate the ARI for the super-resolution result, the study is visually consistent with the manual annotations by pathologists in the spatial domain (Supplementary Fig. 3).

S5 Meta Gene and SVGs Analysis with DLPFC and IDC Samples

To further demonstrate that TransformerST could explore the biological relevance, we detected the spatial variable genes and meta genes for LIBD human dorsolateral prefrontal cortex (DLPFC) data and IDC sample. As shown in Supplementary Fig. 4a, SVGs and their corresponding meta gene show similar spatial patterns for human DLPFC samples at spot resolution. For example, TMSB10 is enriched in cluster 0 of tissue sample 151508. The combination of meta genes (TMSB10+MBP-MT-CO2) shows the strengthened spatial patterns in the neighboring regions. GFAP is enriched in cluster

2 of tissue sample 151508, its corresponding meta gene is GFAP+SNORC-TMSB10+CDT3-MBP, which is also spatially correlated with the SVGs of cluster 2 in the histology image.

As illustrated in Supplementary Fig. 4b, boxplots display Moran's I and Geary's C values for spatially variable genes (SVGs) identified by TransformerST, SpatialPCA [38], and Vesalius[39], offering a comparative analysis of the performance of these methods in detecting spatial patterns in brain tissue slice 151508. Notably, the Moran's I and Geary's C values for SVGs detected by SpatialPCA and Vesalius are lower than those identified by TransformerST, indicating a weaker presence of spatial patterns in their results. The experimental results with different tissue samples and different cluster domains demonstrate TransformerST could mark specific gene-expressed regions for different cluster domains.

To illustrate how TransformerST works for different tissue samples, We detected the spatial variable genes and meta gene for IDC samples at nearly single-cell resolution. As shown in Supplementary Fig. 4c, TransformerST detected single SVGs (ACADSB) for cluster 2. Its corresponding meta gene was defined as ACADSB+NME2-MUC1+ATP5MPL-CD74+LAPTM4B-CRIP1. TransformerST detected DEGS1 SVG for cluster 3, which accords with its meta gene DEGS1+RPS18-CXCL14+AGR2-MGP+CSTA-NEAT1 visually. TTLL12 is enriched in cluster 4 with its corresponding meta gene as TTLL12+HMGN2-MALAT1+KRT8-SLC9A3R1.

The detection outcomes for metagenes and SVGs demonstrate that TransformerST can effectively identify heterogeneity among spatial domains and predict boundaries not recognized by current state-of-the-art methods. These findings highlight TransformerST's ability to better uncover spatial patterns utilizing a graph transformer network.The detection outcomes for metagenes and SVGs showcase that TransformerST not only effectively identifies heterogeneity among spatial domains, but also accurately predicts boundaries that remain undetected by current state-of-the-art methods. These findings not only highlight TransformerST's superior performance in uncovering spatial patterns but also emphasize its ability to leverage a graph transformer network for a more comprehensive understanding of spatial transcriptomics data. Consequently, TransformerST emerges as a valuable tool for researchers aiming to analyze complex spatial relationships within biological samples.

S6 Tissue identification and super-resolution in breast cancer block at nearly single-cell resolution with Xenium technology.

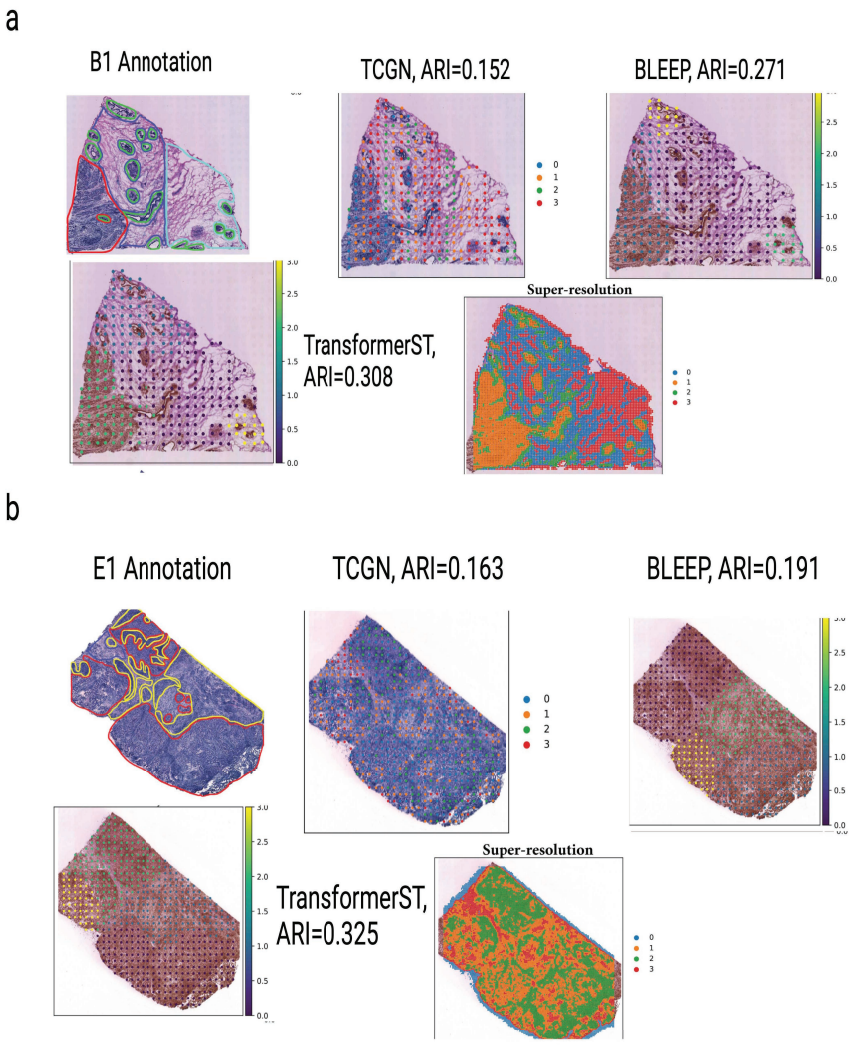
We introduce the cutting-edge Xenium In Situ technology, which has a wide field of view and the unique ability to combine gene expression and histological pictures (H&E and IF staining) in a single tissue segment. Using this method, Chromium scFFPE-seq data were generated from 2 x 25 μm FFPE sections obtained from a breast cancer block (stage II-B, ER+/PR-/HER2+).

These sections subjected to scFFPE-seq were situated across the tissue sections employed for Visium and Xenium procedures.

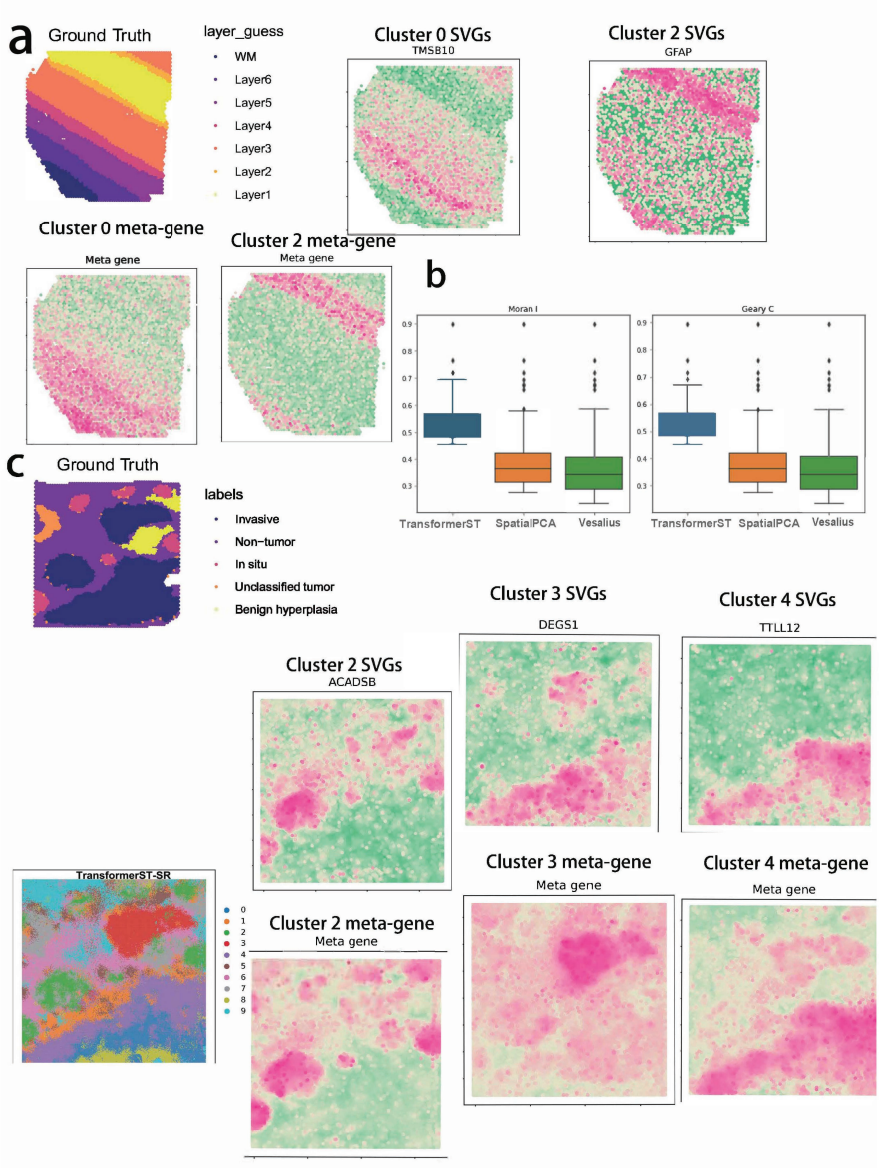
The Visium CytAssist and Xenium platforms were employed to analyze 5 μm tissue sections adjacent to those used for scFFPE-seq. Prior to imaging, sections subjected to Visium were deparaffinized and stained with H&E. The glass slides containing tissue sections were then processed using a Visium CytAssist instrument to transfer analytes to Visium slides. Using the same probe set as in scFFPE-seq, 18,536 genes targeted by 54,018 probes were quantified for sections subjected to Visium. The median number of genes in the Visium data set was 5,712, and dimensionality reduction revealed 17 spatial clusters (the same number of clusters as the scFFPE-seq data). The molecular subtypes of ductal carcinoma in situ (DCIS) (hence referred to as DCIS #1 and DCIS #2) and invasive tumors were each identified as separate clusters by scFFPE-seq, and Visium was able to determine their exact locations.

We use the following pipeline to show the performance of the proposed method. Firstly, we choose the region of interest using the Visium image. Then we obtain the corresponding higher-resolution image and gene expression using Xenium technology. We perform cell segmentation. To allocate mRNAs to cells, facilitating subsequent analysis and integration with Chromium and Visium data, the spatial boundaries between cells and mRNA transcripts need to be determined. Initially, DAPI images were employed to identify nuclei using a neural network. Subsequently, each nucleus was expanded outward up to a maximum distance of 15 μm or until encountering the border of another cell. With the original Visium image, we perform cell-type clustering. Next, we employ the proposed method to perform super-resolution, utilizing the clustering and cell segmentation outcomes. The processed steps are shown in Supplementary Fig. 5a.

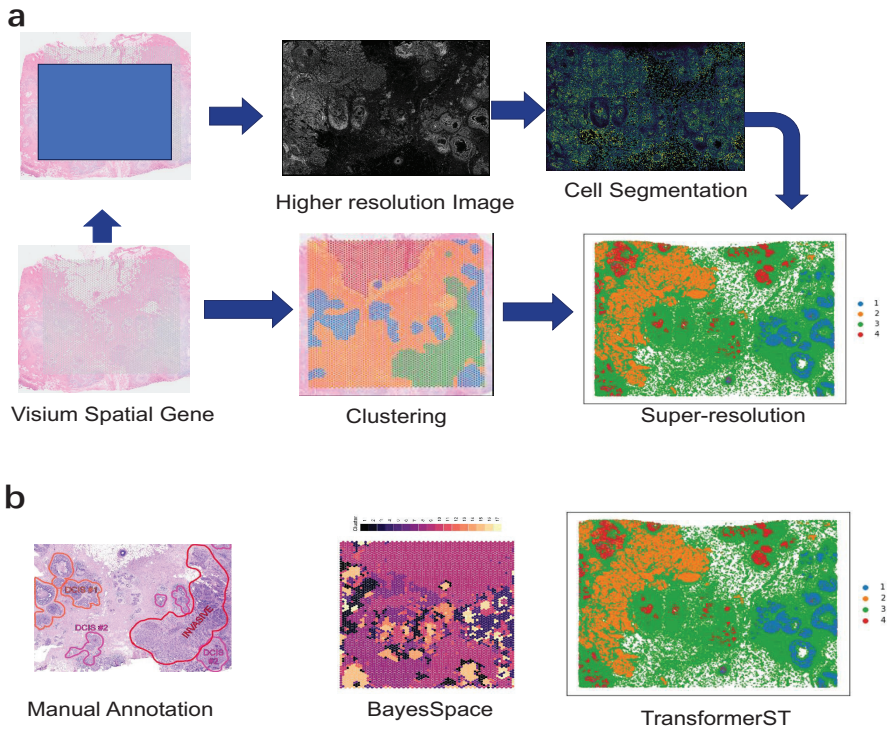
As shown in Supplementary Fig. 5b, similar to Xenium, the proposed method is able to pinpoint precisely where adipocyte markers are located, but it shows much more detail in cases when adipocyte transcripts escape the boundaries of the cells. The examination of TransformerST and BayesSpace demonstrates that the methods can be replicated with high precision. Notably, strong correlations were observed between the quantities of transcripts and the proportions of cell types in these replicates, further validating the effectiveness and reproducibility of the approaches.



Supplementary Figure 3 Super-resolved gene expression prediction with breast cancer data. The first column displays the histology image accompanied by pathologists' annotations, in which the red lines represent invasive cancer, green lines signify breast glands, yellow lines indicate immune infiltration and blue lines denote connective tissue. a, Tissue type assignments and nearly single cell super-resolution using B1 section. b, Tissue type assignments and nearly single cell super-resolution using E1 section. $TransformerST$ refers to the clustering results obtained from TransformerST. Super-resolution, on the other hand, represents the ability of the method to achieve sub-cellular resolution performance in enhancing the spatial transcriptomics data.



Supplementary Figure 4 Spatial variable genes (SVGs) and meta gene detection.a, SVGs and corresponding meta genes for cluster 0 (TMSB10, TMSB10+MBP-MT-CO2), cluster 2 (GFAP, GFAP+SNORC-TMSB10+CDT3-MBP) in brain tissue slice 151508 at spot resolution. b, Boxplots of Moran's I and Geary's C values for SVGs detected by TransformerST, SpatialPCA, and Vesalius using brain tissue slice 151508. c, SVGs and corresponding meta gene for cluster 2 (ACADSB, ACADSB+NME2-MUC1+ATP5MPL-CD74+LAPTM4B-CRIP1), cluster 3 (DEGS1, DEGS1+RPS18-CXCL14+AGR2-MGP+CSTA-NEAT1), and cluster 4 (TTLL12, TTLL12+HMGN2-MALAT1+KRT8-SLC9A3R1) in IDC sample at nearly single cell resolution. Pathologist annotated different regions in different colors (IC outlined in red, carcinoma in yellow, benign hyperplasia in green, unclassified tumor in grey).



Supplementary Figure 5 Tissue identification and super-resolution in breast cancer block at nearly single-cell resolution with Xenium technology. a, Schematic illustration of TransformerST with Xenium data. b, Tissue identification in breast cancer tissue with Xenium and Visium data.