small methods
www.small-methods.com

UCS: A Unified Approach to Cell Segmentation for Subcellular Spatial Transcriptomics

Yuheng Chen, Xin Xu, Xiaomeng Wan, Jiashun Xiao, and Can Yang**

Subcellular Spatial Transcriptomics (SST) represents an innovative technology enabling researchers to investigate gene expression at the subcellular level within tissues. To comprehend the spatial architecture of a given tissue, cell segmentation plays a crucial role in attributing the measured transcripts to individual cells. However, existing cell segmentation methods for SST datasets still face challenges in accurately distinguishing cell boundaries due to the varying characteristics of SST technologies. In this study, a unified approach is proposed to cell segmentation (UCS) specifically designed for SST data obtained from diverse platforms, including 10X Xenium, NanoString CosMx, MERSCOPE, and Stereo-seq. UCS leverages deep learning techniques to achieve high accuracy in cell segmentation by integrating nuclei segmentation from nuclei staining and transcript data. Compared to current methods, UCS not only provides more precise transcript assignment to individual cells but also offers computational advantages for large-scale SST data analysis. The analysis output of UCS further supports versatile downstream analyses, such as subcellular gene classification and missing cell detection. By employing UCS, researchers gain the ability to characterize gene expression patterns at both the cellular and subcellular levels, leading to a deeper understanding of tissue architecture and function.

1. Introduction

Understanding how different cell types assemble into tissues and organs, how cells interact to transmit and receive biological signals, and how cellular communities make decisions in their spatial context are fundamental questions in biomedical and biological research. Subcellular Spatial Transcriptomics (SST) offers a powerful method to measure gene expression at the subcellular level within intact tissue samples, offering unprecedented opportunities to explore the spatial organization and dynamic functions of tissues.[1–4] To deepen our understanding of cellular behavior using SST technologies, it is essential to accurately attribute the measured transcripts to individual cells through the segmentation of the spatial domain. The precision of cell segmentation directly impacts various downstream tasks in SST data analysis, including gene expression quantification, cell type annotation, and quantitative assessment of the cellular environment. For example, if cell segmentation is oversized, it may include transcripts from neighboring cells or extracellular regions, introducing noise that can distort the spatial gene expression patterns and subcellular gene classification. Conversely, fragmented or underestimated segmentation can result in the loss of significant numbers of transcripts, leading to incomplete or inaccurate gene expression profiles and a subsequent loss of valuable information about cellular behavior. Therefore, reliable and robust cell segmentation is crucial for extracting valuable biological insights from SST datasets.

Currently, SST technologies exhibit notable differences in their capabilities across key parameters such as resolution, capture efficiency, and gene profiling capacity, which are critical considerations for their application in various biological research areas.[5] It poses a significant challenge to develop cell segmentation methods that are applicable to multiple SST platforms. Among these platforms, imaging methods like Vizgen's MERSCOPE[6] and 10X Genomics' Xenium[7] achieve high resolutions of approximately 100 nm while maintaining capture rates exceeding 90%. This level of resolution is crucial for applications requiring detailed spatial analysis at the cellular or subcellular level. Alternatively, sequencing-based methods offer distinct capabilities. For example, Stereo-seq[8] enables mRNA collection on an array with a fixed spatial resolution of about 0.5 μm and a specific capture efficiency of 12 661 transcripts per 100 μm^2.[9] The gene profiling capacity also varies among these technologies. Platforms like Xenium and MERSCOPE generally focus on

Y. Chen, X. Xu, X. Wan, C. Yang
Department of Mathematics
The Hong Kong University of Science and Technology
Hong Kong SAR 999077, China
E-mail: macyang@ust.hk

J. Xiao
Shenzhen Research Institute of Big Data
ShenZhen 518100, China
E-mail: xiaojiashun@sribd.cn

C. Yang
State Key Laboratory of Molecular Neuroscience
The Hong Kong University of Science and Technology
Hong Kong 999077, P.R. China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smtd.202400975

DOI: 10.1002/smtd.202400975

profiling hundreds of genes, making them suitable for targeted studies. In contrast, NanoString's[10] technology enables the profiling of thousands of genes, providing broader coverage. Meanwhile, Stereo-seq stands out by offering whole transcriptome analysis, capturing over 10 000 genes, which is ideal for applications requiring extensive gene expression profiling. Adaptability to the diverse characteristics of these technologies is crucial in order to achieve accurate and reliable cell segmentation.

Considerable efforts have been dedicated to the development of cell segmentation methods for various SST platforms. To facilitate discussion, we can broadly categorize these methods into two groups: transcript-based methods and image-based methods. Transcript-based methods primarily leverage the abundant transcript information from SST datasets. For example, BIDCell[11] achieves highly accurate segmentation by utilizing extensive datasets that include single-cell RNA sequencing reference data, positive marker genes, and even elusive negative marker genes. This method is particularly effective in accurately delineating boundaries for elongated cell types like fibroblasts. However, the extensive information requirements of BIDCell make it inconvenient to use in practical settings. JSTA[12] employs a novel cell-level classifier that assigns cell (sub)types to pixels, facilitating segmentation. Nonetheless, JSTA is often limited to small tissue sections and heavily relies on single-cell RNA information, limiting its scalability and broad applicability. SCS[13] assigns spots to cells by adaptively learning their position relative to the cell center using a transformer neural network, achieving high accuracy. However, SCS may encounter potentially unstable training phases and exhibit some cross-nuclei segmentation. Baysor[14] offers a unique approach by spatially clustering observed molecules to cells and modeling each cell as an ellipsoid using a Gaussian distribution of transcript composition. Although innovative, Baysor is often sensitive to parameter settings and may segment single cells into multiple smaller cells due to varying gene clusters inside, making its practical use challenging compared to faster deep learning methods. On the other hand, image-based methods rely solely on stain images, particularly nuclei staining, which are often the only staining available in SST datasets. Commercial platforms like 10X and Vizgen employ these methods, but they often face challenges such as inaccurately large cells resulting from nuclei dilation in Xenium's default segmentation or loss of some transcript data due to smaller segmented cells in Vizgen's approach. Another example is Cellpose,[15] a general method for cell segmentation that primarily utilizes various staining images to delineate cell boundaries. When applied to SST datasets, this approach fails to incorporate SST data and only has nuclei staining as the input, potentially leading to less informative and accurate segmentation outcomes.

In this paper, we introduce a unified approach to cell segmentation (UCS) for SST datasets. The advantages of UCS over existing methods are threefold. First, UCS integrates accurate nuclei segmentation results from nuclei staining with the transcript data to predict precise cell boundaries, thereby significantly improving the segmentation accuracy. The combination of nuclei segmentation and transcript data offers a comprehensive perspective that enhances cell segmentation. Nuclei segmentation provides a precise identification of cell centers, which is critical for accurately locating individual cells within a complex tissue structure. On the other hand, transcript data delineates the spatial distribution of gene expression, which helps in determining the cell boundaries more accurately. Second, UCS is applicable to a wide range of SST platforms, including 10X Xenium, NanoString CosMx, MERSCOPE, and Stereo-seq, and shows robust performance across diverse datasets. Third, UCS stands out for its computational efficiency and user-friendliness, in contrast to methods like BIDCell or JSTA, which require extensive inputs such as single-cell RNA data and specific marker gene information. Through real data examples and downstream analysis, UCS demonstrates its power by enhancing the accuracy of cell segmentation. It produces better visualization results, matching the hematoxylin and eosin (H and E) staining. Moreover, UCS improves downstream analysis by enabling more accurate spatial gene analysis, facilitating subcellular gene classification, and accommodating diverse cell morphologies. The UCS software is now publicly available at: https://github.com/YangLabHKUST/UCS.

2. Results

2.1. Method Overview

UCS is a deep learning-based approach designed to enhance cell segmentation in diverse SST datasets. By integrating both nuclei segmentation from nuclei staining and transcript data, UCS can distinguish the cell boundaries with high precision. UCS is composed of two strategically designed efficient convolutional neural networks—one for predicting the foreground, the region with higher transcript density that likely corresponds to cellular areas and the other for distinguishing the exact cell boundaries. More specifically, UCS comprises two steps in real data applications. Initially, it uses the transcript data to identify the foreground region of candidate cells using the first network according to the density of transcripts (**Figure 1**a). Subsequently, UCS transforms the nuclei segmentation mask into a softmask. The nuclei segmentation mask is derived from nuclei staining and serves as an initial identification of where cell nuclei are located. Each pixel's value in the softmask is determined by its distance to the nearby nucleus, representing the probability that the pixel belongs to a particular cell. Then the softmask is combined with the segmented foreground regions to generate cell boundaries for each nucleus utilizing the second network (Figure 1b). Intuitively, the nuclei segmentation mask can serve as an anchor for each cell, while the softmask acts as a probabilistic guidance like a prior, guiding the cell segmentation process using the foreground regions detected from the first step. UCS significantly improved the cell segmentations by associating nuclei with their corresponding cytoplasmic regions, and achieves high consistency with cell segmentations from H and E staining image (Figure 1c), facilitating downstream analyses.

UCS is a versatile approach designed to address multiple challenges inherent in cell segmentation task, offering a range of optional utilities that enhance its applicability (Figure 1d). First, UCS exhibits robust detection capabilities, identifying cells that are frequently missed by other segmentation methods. Newly identified cells exhibit sufficient gene expression and compatibility with previously identified cells, ensuring reliable detection. Second, for elongated cell types, UCS employs a scaled softmask to maintain shape consistency with the nuclei, thereby

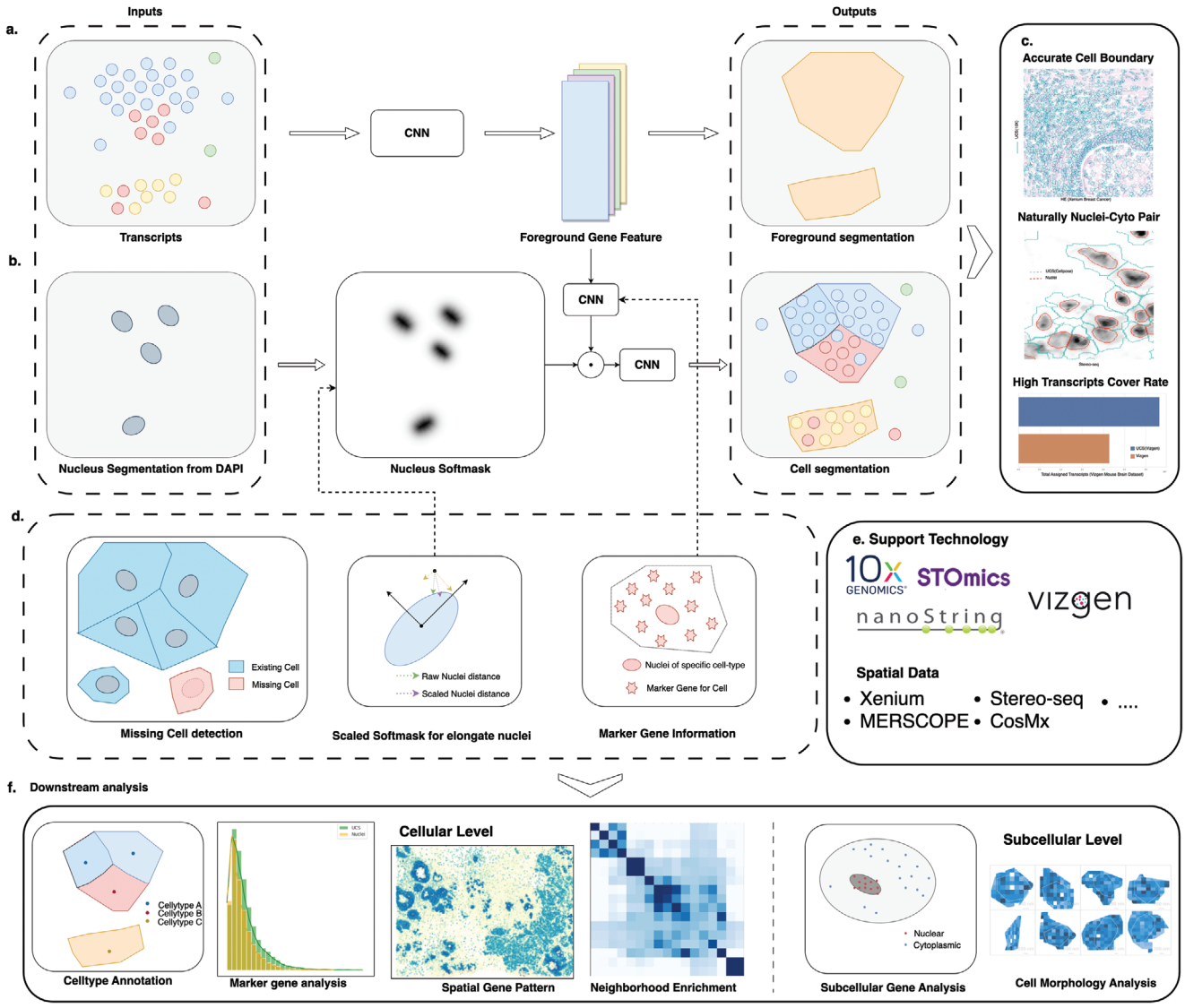


Figure 1. Overview of UCS. a) UCS first uses the transcript data as the input and output the potential foreground where cells exist. b) Then the nuclei segmentation mask will be processed into a softmask, and integrated with foreground information to get the final cell segmentation mask. c) UCS can give accurate cell boundaries and naturally nuclei-cyto pair with a relatively high transcript cover rate. d) Other versatile utility of UCS. e) UCS can be used on SST datasets from different platforms. f) UCS also improves downstream analysis at both cellular level and subcellar level.

preserving the morphological integrity of these cells. Additionally, UCS can integrate marker gene information to enhance segmentation, ensuring that each nucleus is associated with the correct cell-type specific markers. In summary, UCS is a unified approach that can be applied across various SST datasets generated by different technologies, including 10X Xenium, Vizgen, Nanostring CosMx, and Stereo-seq (Figure 1e). The superior performance of UCS lays the foundation for a range of downstream analyses (Figure 1f). At the cellular level, UCS facilitates cell type annotation, marker gene analysis, spatial pattern recognition, and neighborhood enrichment. At the subcellular level, it supports subcellular gene classification and cell morphology analysis. These capabilities demonstrate the extensive utility of UCS in analyzing SST datasets.

2.2. UCS Shows Robust Performance Across Diverse SST Datasets

As a general framework that can accommodate various SST datasets through a well-structured training process, UCS can handle the diverse characteristics and varying quality of data across different platforms, ensuring accurate and reliable cell segmentation. Variability within SST datasets, particularly in relation to the tissue type and dataset quality, significantly complicates the process of cell segmentation. For instance, tissues such as breast cancer contain elongated fibroblast cells, which pose unique segmentation challenges compared to other cell types due to their distinct morphology (**Figure 2a**, I). The quality of SST data also varies across platforms; for example, Vizgen and Nanostring tend to produce cleaner data, whereas Stereo-seq data

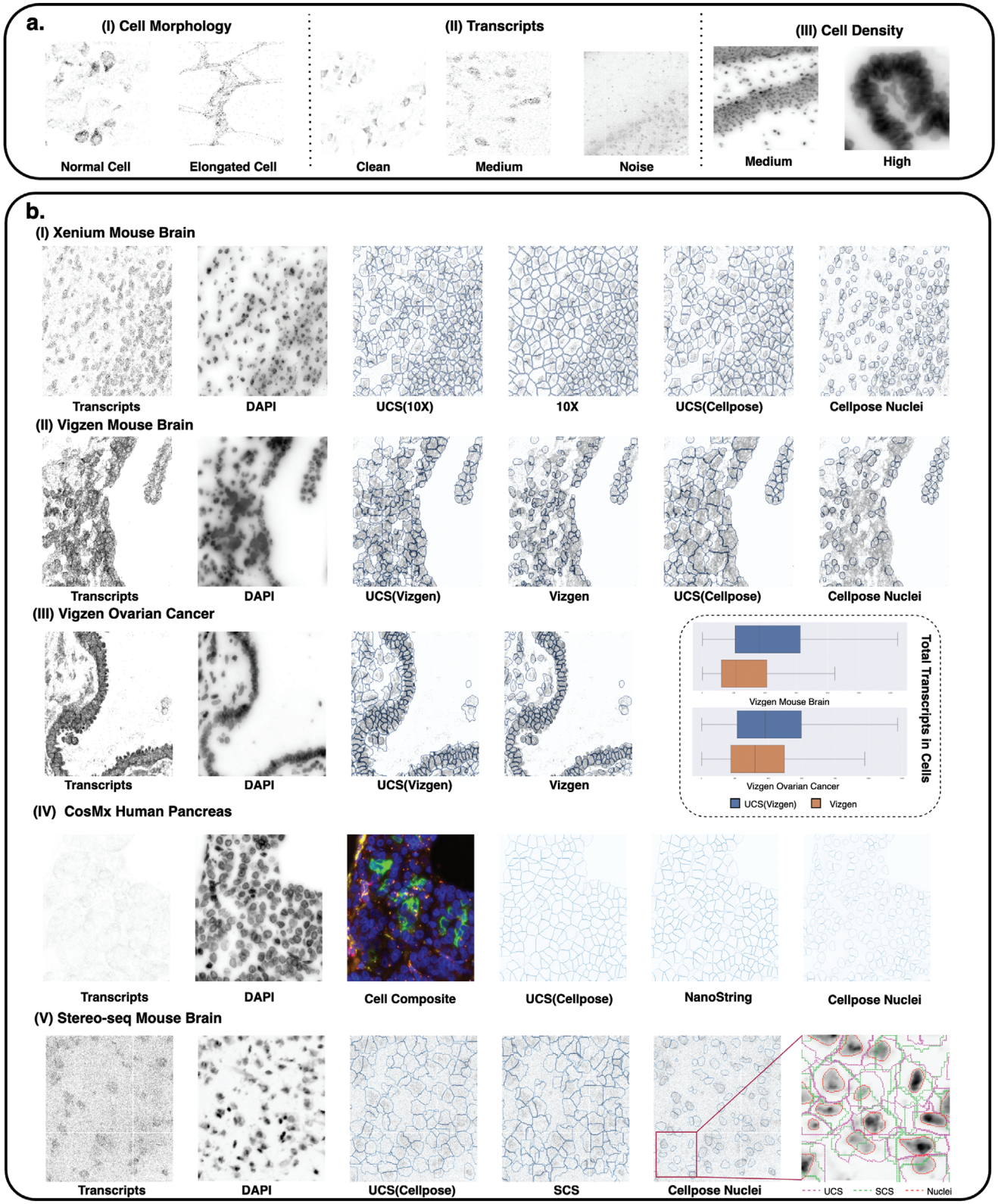


Figure 2. UCS shows robust performance across diverse SST Datasets. UCS (10X) means the nuclei segmentation input for UCS model is produced by 10X, and same notation for UCS (Cellpose) and UCS (Vizgen). a) SST datasets vary in cell morphology, transcript noise level and cell density. b) Comparative visualization of cell segmentation across diverse SST datasets. UCS shows robust performance across datasets from different platforms. Here, the cell composite image of CosMx Huamn Pancreas dataset is the mixed fluorescence image of nuclei, cytoplasm and cell boundaries.

is often noisier (Figure 2a, II). Additionally, cell density can differ markedly between datasets or even within different regions of the same dataset (Figure 2a, III), necessitating adaptable segmentation methods capable of handling such variability. These challenges underscore the need for robust and versatile cell segmentation techniques that can accommodate the diverse and complex nature of SST data.

The proposed UCS method demonstrates superior performance in cell segmentation across diverse SST datasets. Comparative analysis on Xenium Mouse Brain indicates that UCS achieves higher accuracy and produces more compact segmentation results compared to the 10X method (Figure 2b, I), which uses dilation of the nuclei segmentation that often results in oversized cells with low morphological diversity. By leveraging the underestimated segmentation results from Vizgen as "nuclei," UCS improves transcript coverage rates and includes areas with high gene expression levels (Figure 2b, II, III), thus increasing transcript counts per cell for downstream analyses. Vizgen, on the other hand, adopts overly conservative segmentation strategies, resulting in low transcript coverage rates. In the CosMx Human Pancreas, UCS produces segmentation boundaries comparable to the latest method of NanoString based on the staining of the whole cell (Figure 2b, IV), despite using only nuclei fluorescence information and transcripts. This is important since most SST datasets only provide nuclei staining, but most of the existing image-based cell segmentation methods solely rely on nuclei staining, making it infeasible to precisely segment the cytoplasm of cells based solely on image data. Although recent advancements in platforms like 10X Genomics Xenium 2.0 and NanoString have integrated more sophisticated fluorescence-based imaging techniques (like the newly released CosMx Human Pancreas dataset in the figure), enabling a richer and more detailed capture of both cellular and subcellular structures, the availability of datasets generated from these new technologies remains relatively limited. In the Stereo-seq Mouse Brain, UCS effectively avoids the problematic cross-nuclei segmentation observed with SCS (Figure 2b, V). The stability of UCS is particularly noteworthy in noisy datasets like Stereo-seq, where methods like SCS exhibit unstable training and segmentation results. UCS consistently provides stable and reliable segmentation across all tested datasets, underscoring its robustness and adaptability to various SST scenarios (Figure S2).

We also conducted an experiment on the Xenium Breast Cancer dataset by subsampling the transcripts at different ratios (80%, 40%, and 20%) randomly to simulate datasets with varying signal levels (**Figure 3**a). This approach intentionally introduces noise and incompleteness to evaluate how UCS performs under these challenging conditions. Using the original dataset with 100% of transcripts as the ground truth, we compared the performance of UCS and BIDCell across the subsampled datasets (Figure 3). UCS maintained high performance, with F1-scores above 0.9 across all subsampling levels, demonstrating its robustness in cell segmentation even as the signal decreased. Furthermore, the cell boundaries produced by UCS remained consistent with the HE image, indicating its ability to preserve spatial accuracy under noisy conditions. In contrast, BIDCell's performance showed a more noticeable decline as the subsampling ratio decreased. The F1-scores for BIDCell were more affected, and the segmentation results exhibited greater inconsistencies, highlighting that UCS is better equipped to handle noisy and incomplete datasets. These findings reinforce the claim that UCS is well-suited for cell segmentation in various spatial transcriptomics datasets.

By addressing the limitations of both transcripts-based and image-based segmentation methods, UCS offers a unified approach that enhances cell segmentation accuracy and reliability across a variety of SST data from different platforms, tissues, and sections, paving the way for more precise and insightful SST analyses.

2.3. Benchmarking Studies Show the Superior Performance of UCS in Cell Segmentation on Xenium Datasets

To quantitatively compare the cell segmentation performance of UCS and the other methods, we conducted experiments with UCS on various Xenium datasets, including the Xenium Human Breast Cancer, Xenium Mouse Brain, and the recently introduced Xenium 2.0 Human Lung Cancer datasets. UCS excels in both normal and elongated cell regions within the Xenium Human Breast Cancer dataset (**Figure** 4a). By leveraging the scaled soft-mask, marker gene information, and nuclei segmentation provided by Cellpose, UCS achieves segmentation results comparable to BIDCell—a method specifically designed for elongated cell types, which utilizes prior knowledge about the elongation of cells. In comparison, Baysor, which depends on gene clustering, often yields small or fragmented cells. This fragmentation is a consequence of clustering-based approaches that may not fully consider the spatial continuity and morphological context of cells. On the other hand, BIDCell employs extensive additional information to assist in cell segmentation, such as single-cell RNA-seq data and other contextual markers. This limits BIDCell's applicability. Furthermore, UCS exhibits greater consistency with H and E staining (Figure 4b). The regions with deep color staining, which indicate the presence of cells, are more accurately covered by UCS segmentation than by BIDCell, while blank regions are not erroneously segmented—a problem seen in the segmentation results of 10X. UCS also ensures that the boundaries of cell clusters are distinctly clear and demonstrates superior accuracy in cell annotation.

To establish baseline characteristics, we computed several metrics across different methods, including cell number, total transcripts, cell area, total transcripts in one cell, transcripts per cell pixel, and total genes in one cell (Figure 4c). When comparing methods, there is often a tradeoff between cell size and the total transcripts included. For example, methods like Baysor segment cells with smaller areas, prioritizing high transcript density. While this approach reduces noise, it risks excluding transcripts located near the cell boundaries, potentially underrepresenting the full expression profile of the cell. On the other hand, approaches such as the nuclei dilation method used in the 10X official pipeline segment cells with larger areas, which may include transcripts from surrounding extracellular regions. This results in a higher total transcript count per cell but a lower transcript density, as many of the included genes may not truly belong to the cell. Our method, UCS, aims to strike a balance between these extremes. It achieves a trade-off by adaptively adjusting the segmentation to include high-density transcript regions

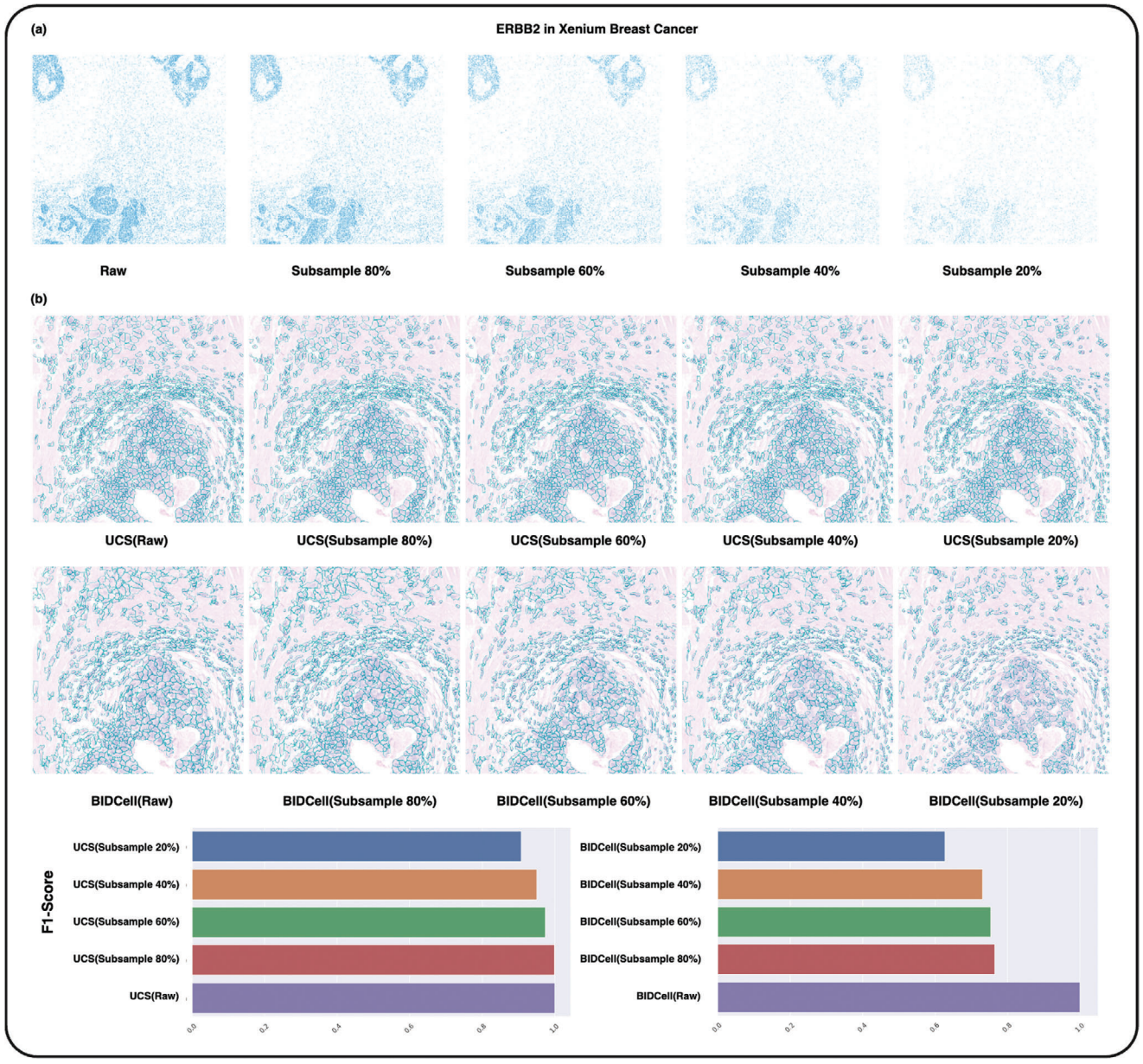


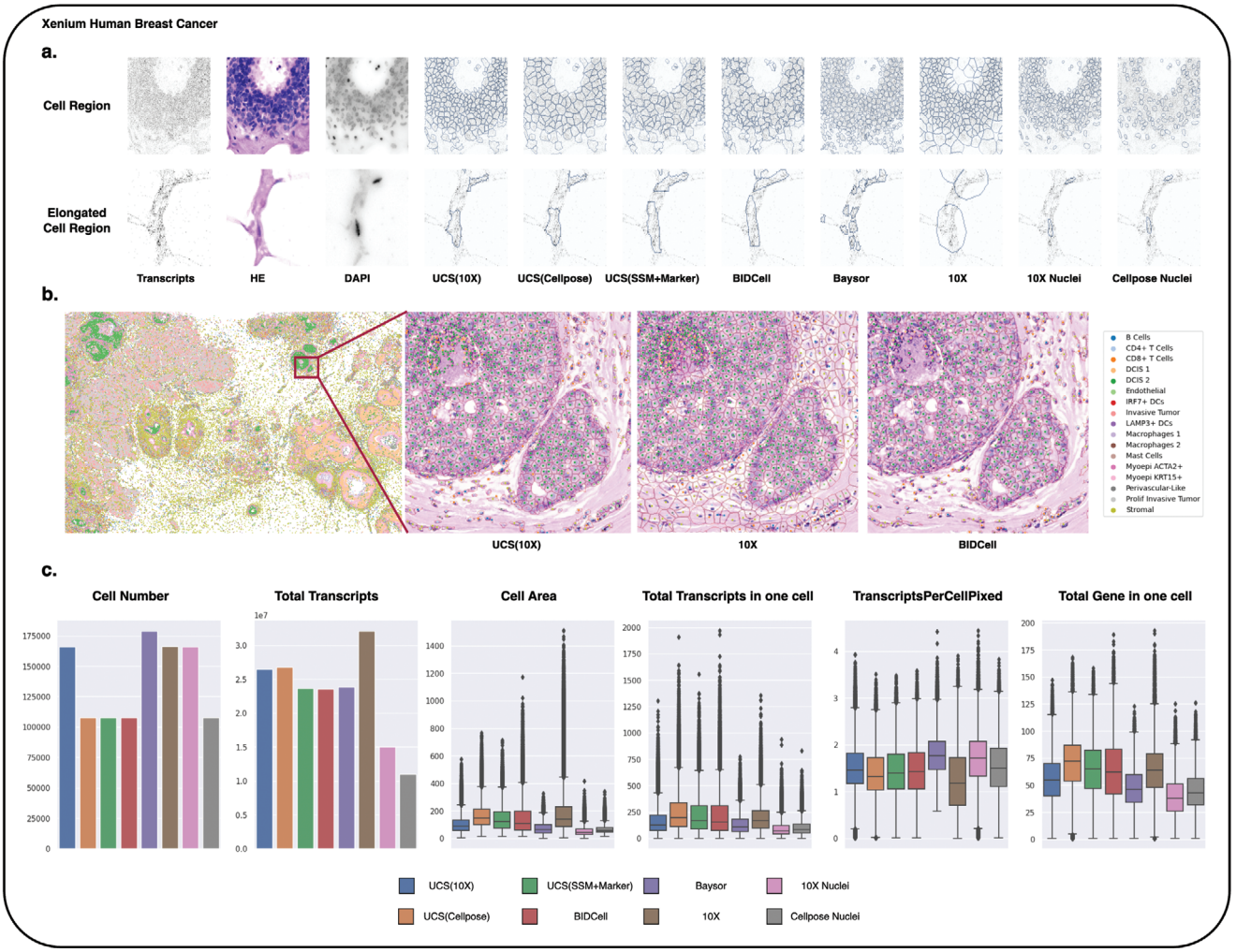
Figure 3. Experiment on the Xenium Breast Cancer dataset by subsampling the transcripts at different ratios (80%, 40%, and 20%) to simulate datasets with varying signal levels. a) Transcript density of *ERBB2* at different subsample ratios. b) Performance comparison of UCS and BIDCell across the subsampled datasets. UCS consistently maintained high performance, achieving F1 scores above 0.9 across all subsampling levels, while BIDCell's performance was more affected, showing cell boundary shrinkage and even missing cells in some cases. UCS also maintained better consistency with the HE image.

while avoiding the inclusion of noise from sparse or distant areas. UCS demonstrates a balanced performance in terms of cell size and transcript inclusion, ensuring both spatial precision and transcriptional completeness.

In addition to the Human Breast Cancer dataset, we further evaluated UCS on the Xenium Mouse Brain and the new Xenium 2.0 Human Lung Cancer datasets (Figure 4d). The Xenium 2.0 dataset employs advanced fluorescence technology to stain both nuclei and cell boundaries, providing a more convincing ground truth for cell segmentation evaluation. Using the 10X Multimodel result for benchmarking, we calculated the F1 Score for various methods (Figure 4d).[16] UCS achieved the highest F1 Score of 0.84, indicating its segmentation results are closely aligned with the ground truth provided by the 10X 2.0 dataset.

2.4. UCS Enhances the Consistency of Cellular Expression Profiles with scRNA-seq Data

An accurate cell segmentation method will generate cellular gene expression profiles that closely match the corresponding scRNA-seq reference dataset. Given this fact, we integrated the spatial



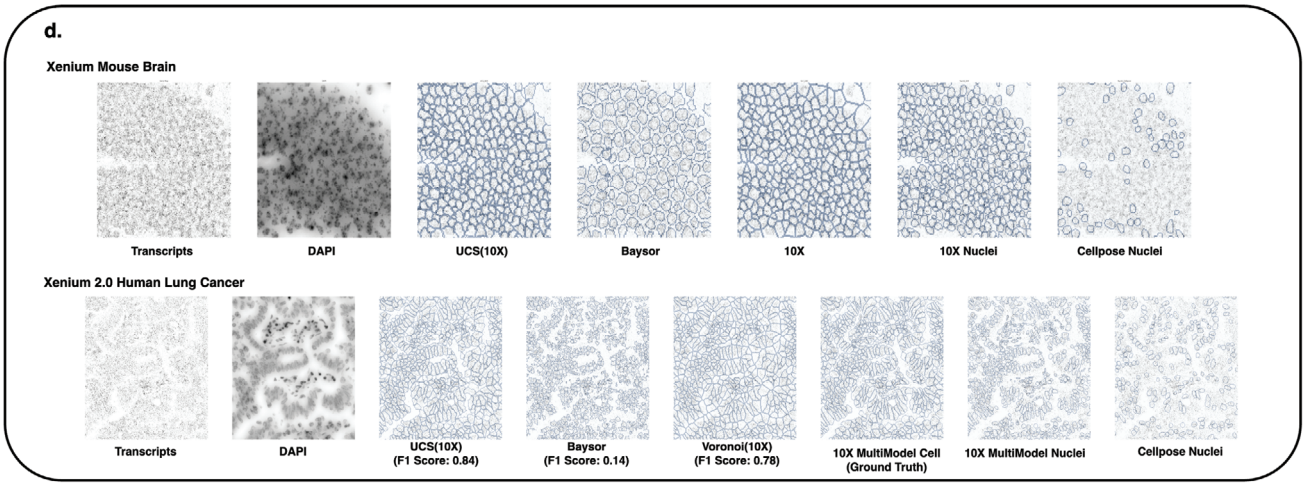


Figure 4. Benchmark using Xenium datasets. UCS(SSM+Marker) means UCS using scaled softmask, additional marker gene information(discussed in detail in the Method section) and Cellpose nuclei segmentation as the input. a) Comparison of normal cell region and elongated cell region with different methods on Xenium Human Breast Cancer dataset. The H&E images are provided for visual comparison. b) Cell annotation result on Xenium Human Breast Cancer dataset. UCS shows greater consistency with the H and E image which is not used for building the segmentation model compared to 10X or BIDCell. c) Cell segmentation characteristics of different methods. d) More comparisons on Xenium Mouse Brain dataset and Xenium 2.0 Human Lung Cancer dataset. UCS has the largest F1 Score with the 10X MultiModel Cell which uses additional fluorescence information.

single cells segmented by the compared methods with a scRNA-seq dataset from the same tissue. We evaluated the integration result with two aspects. From the cell type labeling aspect, we transferred the cell type labels from scRNA-seq to the segmented cells by different methods and then assessed their spatial organizations. As shown in **Figure 6**a, UCS enabled a more precise identification of the spatial distribution of cell types in human breast cancer tissue, especially accurately distinguishing the two subtypes (Figure 6a) (Green and Light orange regions) in areas of ductal carcinoma in situ (DCIS). While different segmentation methods generally reflected similar proportions of each cell type, the 10X segmentation approach often resulted in oversized cell boundaries. This oversizing tends to incorporate more extraneous genetic noise, thereby affecting cell type annotations.

For gene expression aspect, we compared the Pearson correlation (PCC) between the gene expression of segmented cells of each cell type and corresponding cells from the scRNA-seq reference. To calculate the correlation, we first averaged the cell expression grouped by cell type for each segmentation method's results and the single-cell reference dataset. Then, we computed the Pearson correlation for each cell type between the segmentation results and the single-cell reference. Finally, the average correlation across all cell types was calculated to obtain the overall correlation for each segmentation method. This approach provides a quantitative measure of how well the segmentation results align with the single-cell reference in terms of gene expression profiles. As shown in right panel of Figure 6b and Figure 6c II, the cell expression profiles obtained using UCS demonstrated an improvement in similarity with scRNA-seq reference by maintaining appropriate cell sizes and accurate cell boundaries. The enhanced segmentation accuracy of UCS resulted in gene expression profiles that more closely matched those observed in scRNA-seq datasets, yielding cleaner and clearer spatial gene patterns such as *CHD1* and *TACSTD2* (Figure 6c I), which are highly expressed in tumor cells. Additionally, UCS enabled the identification of highly expressed or marker genes in cells, such as *ERBB2* in invasive tumors and *LYZ* in macrophages (Figure 6e). The precise boundary delineation ensured that the number of genes identified was almost the same as with the oversized 10X segmentation results (Figure 6d).

Furthermore, UCS's segmentation results facilitated the discovery of neighborhood enrichment patterns in breast cancer tissues (Figure 6f). For instance, our analysis revealed that tumor cells inhibit the growth of neighboring cells and suggested potential interactions between immune cells and IRF7+ DCs. These findings underscore UCS's utility in elucidating complex cellular interactions within the tissue microenvironment, thereby advancing our understanding of spatially resolved transcriptomics.

2.5. UCS Enhances Subcellular Gene Localization and Morphological Correlation

To further illustrate the utility and advantage of UCS in subcellular gene classification, we employed Bento[17] to determine the primary expression locations of genes–nuclei, cytoplasm, or no preference. We specifically analyzed an invasive tumor region with a cell count comparable to the dataset used in the original Bento paper (**Figure 6**a, left). Notably, in both UCS and 10X segmentation results, the top three genes expressed in the nuclei were *ANKRD30A*, *POLR2J3*, and *MLPH*. However, discrepancies emerged in the cytoplasmic gene classification: UCS identified *SMS*, *HMGA1*, and *SEC11C*, whereas 10X identified *LLM*, *SEC11C*, and *FAM107B*, with *LLM* showing significantly higher expression levels (Figure 6a, right). Upon further investigation using scRNA-seq references, *LLM* was found to be highly expressed in stromal cells but rarely in tumor cells, indicating a misclassification by the 10X method due to incorrect cell boundary assignments. In contrast, *SMS*, which UCS identified, is moderately expressed in tumor cells and predominantly localized to the cytoplasm, corroborating the accuracy of UCS segmentation.

Moreover, we compared cell morphology by calculating the eccentricity for nuclei and cells across different cell types (Figure 6b, upper). Both UCS and BIDCell demonstrate the ability to preserve diverse and accurate cell morphology, whereas other methods failed. A concrete example of stromal cells with notably elongated morphology is shown in the bottom panel of Figure 6b, where UCS successfully segmented these elongated cells with the aid of scaled soft masks and marker gene information, matching the performance of BIDCell, which requires extensive prior knowledge about cell elongation and other marker gene information.

2.6. UCS Enables Detection of Missing Cells in SST Datasets

One of the significant advancements introduced by UCS is its ability to identify missing cells that traditional image-based segmentation methods often overlook. This capability stems from UCS's two-stage strategy of cell segmentation, which leverages both transcriptomic and imaging data. In the first stage, UCS predicts the confidence foreground region for cells solely based on the SST data. This transcript map provides a detailed view of gene expression across the tissue sample, independent of the staining intensity or quality. By focusing on the gene expression data, UCS can identify potential cell regions that might not be apparent in imaging data alone. In the second stage, UCS refines the initial predictions using prior segmentation results. Specifically, the unassigned potential cell regions are defined as the difference between the confidence foreground and the final cell segmentation result (**Figure 7**a). This difference highlights areas where traditional image-based methods, which often rely heavily on nuclei staining, might fail due to extremely dense nuclei fluorescence, as illustrated by the red region in the Vizgen Mouse Brain dataset (Figure 7b).

Next, we applied UCS to detect missing cells in less dense spatial data of Xenium Breast Cancer (Figure 7c). We observed that the de nova cells are characterized by transcript counts per pixel that are smaller than those of existing nuclei but larger than those of fully segmented cells (Figure 7d, I). This nuanced detection captures cells that might be underrepresented or entirely missed due to low staining signal intensity in traditional methods. The accuracy of UCS in detecting these novel cells is further validated by integrating their expression profiles into the PCA embedding space (Figure 7d, II). These novel cells' expression profiles align well with those in corresponding scRNA datasets, indicating the correctness of the UCS-based detections.From Figure 7e,

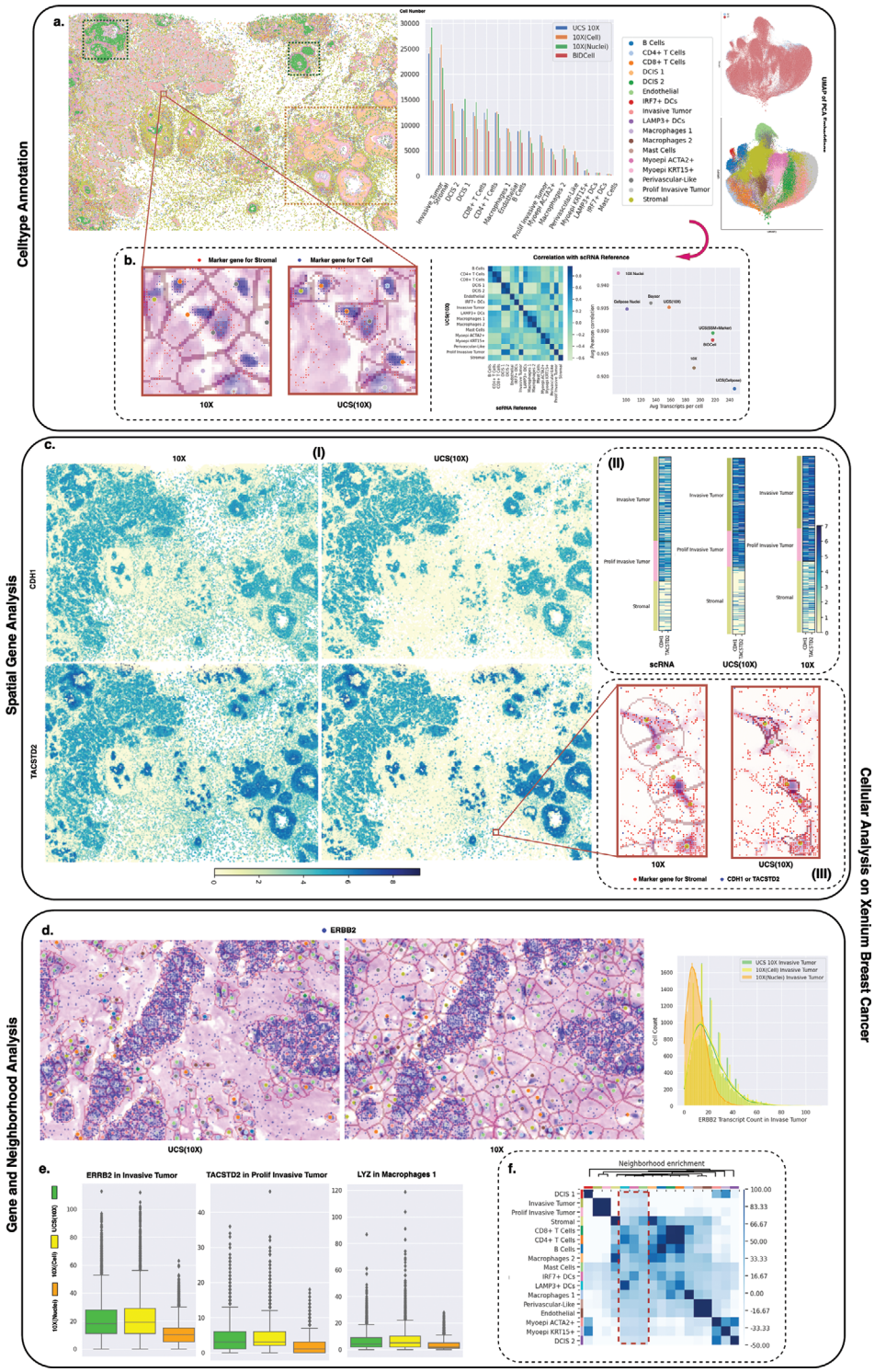


Figure 5. (Previous page.) Cellular level analysis. a) Cell type annotation results of UCS for Xenium Human Breast Cancer, alongside a joint UMAP visualization of PCA embeddings of both the scRNA dataset and the spatial dataset, demonstrate well-integrated cell expression profiles from the UCS segmentation and the paired scRNA dataset. b) Left: A region demonstrates the superior alignment of UCS segmentation with H&E images compared to 10X. UCS includes fewer irrelevant genes, resulting in more accurate cell type annotation. Right: Correlation with scRNA reference. UCS with 10X nuclei achieves high average correlation, maintaining relatively large cell sizes. c) The spatial patterns of CDH1 and TACSTD2, predominantly expressed in tumor cells, are clearer and cleaner with UCS compared to 10X. This improvement is due to UCS excluding noise genes that are distant from cell nuclei and located in sparse transcript regions. UCS also shows better consistency in cell expression with the scRNA dataset. d) UCS identifies nearly the same highly expressed genes or marker genes in different cells, albeit with relatively smaller but more refined cell sizes compared to 10X. Additionally, noticeable spatial neighborhood enrichment patterns are evident in UCS's segmentation results.

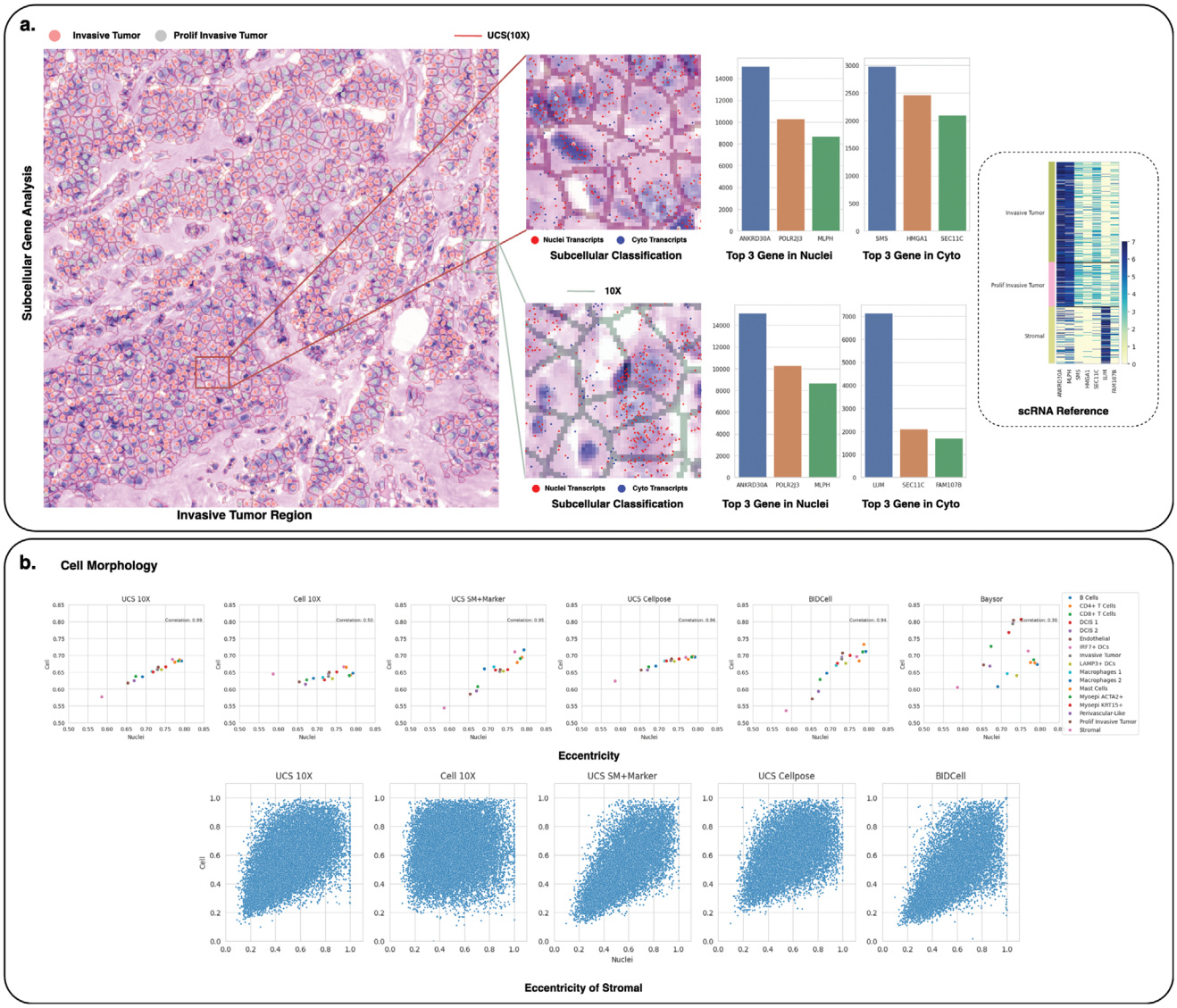


Figure 6. Subcellular level analysis. a) Subcellular classification results in an invasive tumor region. Genes are classified based on their primary expression location–nuclei, cytoplasm, or no preference–and the top three genes expressed in the cytoplasm and nuclei are listed for both UCS and 10X segmentation results. UCS correctly identifies the relevant genes, whereas 10X misclassifies genes from other cell types as cytoplasmic genes of the tumor, primarily due to its inaccurate boundary definitions. b) Eccentricity of the nuclei (x-axis) versus cells (y-axis) to compare cell morphology across methods for each cell type, highlighting the elongated morphology of stromal cells.

the nuclei staining for these segmented new cells are missing but the density of transcripts is relatively high. The newly detected cells exhibit similar characteristics to the existing ones, including comparable cell sizes and total transcript counts. This integration ensures that the newly identified cells are not artifacts but genuine parts of the cellular landscape.

2.7. Computational Time

UCS demonstrates its advantages over existing methods in terms of computational efficiency, making it suitable for large-scale SST data analysis. For the Xenium Breast Cancer dataset, UCS completes the segmentation in approximately 1.5 h when utilizing nuclei segmentation obtained by Cellpose, processing 107 848 cells. When using 10X nuclei segmentation, the computational time increases slightly to 2 h for 165 752 cells. Additionally, UCS requires only 12 min to segment a 1200 × 1200 patch of Stereo-seq and 15 min of one 1400 × 1400 Field of view (FOV) of the CosMx Human Pancreas dataset. In contrast, BIDCell, another prominent method, takes around 2 h to process the Xenium Breast Cancer dataset containing 107 829 cells. Baysor demonstrates a considerably longer computational time. For the Xenium Breast Cancer dataset, Baysor requires approximately 48 h to complete the segmentation process. The SCS method, when applied to a 1200 × 1200 patch of Stereo-seq data, takes about 14 h to complete the segmentation. All the experiments were conducted on a Single NVIDIA Tesla V100 GPU with a 44-core CPU.

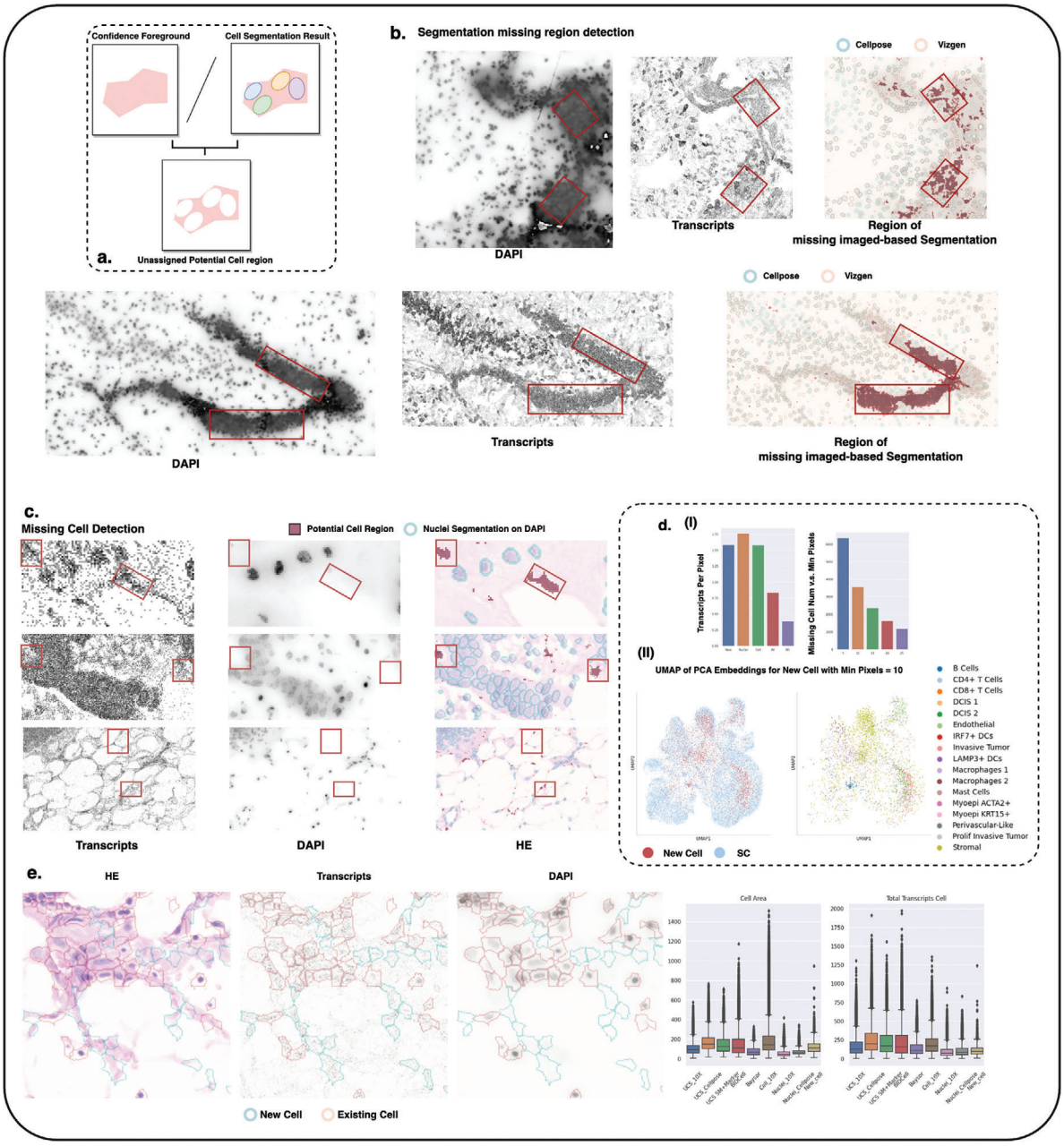


Figure 7. Missing Cell detection. a) The unassigned potential cell region is defined as the difference between the confidence foreground and the final cell segmentation result by UCS. b) UCS can identify regions where image-based segmentation methods fail, likely due to the extremely dense distribution of different nuclei fluorescence. c) Missing cell detection on Xenium Breast Cancer dataset. d) UCS can detect missing cells with high gene expression, and the joint UMAP of PCA embeddings of both the scRNA dataset and the detected cells shows that these novel cells integrate well with cells in the scRNA dataset, demonstrating their correctness. e) Visualization shows that these detected cells are located in regions without DAPI staining but exhibit high gene transcript density. These cells also share similar sizes and total transcripts per cell compared to other cells.

2.8. Influence of Nuclei Segmentation Quality on UCS Performance

UCS relies on nuclei segmentation as it uses the nuclei as the foundational anchor for segmenting the entire cell. Variations in the quality of nuclei segmentation directly influence the accuracy of UCS's final cell segmentation outcomes.

To investigate this, we conducted an ablation study using the 10X Xenium Breast Cancer dataset,[7] which provides highly accurate official nuclei segmentation results that served as the ground truth for our analysis. We compared the performance of UCS using nuclei segmentation results derived from multiple methods, including the Cyto models of Cellpose version 1,[15] version 2,[18] and version 3,[19] Nuclei model of Cellpose version 1, StarDist,[20] and Watershed.[21] The study first evaluated the

quality of nuclei segmentation for these methods and confirmed that the 10X official nuclei segmentation delivered the best performance. Using these as the benchmark, F1-scores were calculated to assess the accuracy of UCS results generated from alternative nuclei segmentation inputs. The results are shown in Figure S3 (Supporting Information).

Our findings indicate that nuclei segmentation methods with lower accuracy, such as Watershed, led to reduced F1-scores for UCS, negatively affecting the final segmentation outcomes. This is likely due to the inability of such methods to adequately delineate nuclei boundaries, which in turn impacts the proper identification of cell boundaries by UCS. In contrast, methods like Cellpose and StarDist, which provided more accurate nuclei segmentation, resulted in UCS segmentations that maintained high consistency with HE images. These results demonstrate the robustness of UCS in preserving spatial accuracy when reliable nuclei segmentation inputs are used. Furthermore, while the choice of nuclei segmentation method affects the final results, UCS demonstrates a degree of resilience by producing relatively similar outcomes when high-performing methods like Cellpose or StarDist are employed.

These results emphasize the critical importance of accurate nuclei segmentation for achieving reliable UCS performance. Fortunately, most spatial omics technologies, such as 10X Genomics, MERFISH, and NanoString CosMx, provide official nuclei segmentation derived from staining images, which are typically of high accuracy. Additionally, third-party algorithms such as Cellpose and StarDist offer robust alternatives, allowing UCS to maintain reliable performance across diverse datasets and experimental conditions.

2.9. Discussion

We introduce UCS, a deep learning-based approach designed to enhance cell segmentation within diverse SST datasets. UCS integrates nuclei segmentation from nuclei staining with transcriptomic data, distinguishing cell boundaries with high precision. The method comprises two convolutional neural networks: the first network predicts the foreground regions of candidate cells, while the second network delineates the exact cell boundaries. This two-step process ensures that UCS achieves superior segmentation accuracy, effectively associating nuclei with their corresponding cytoplasmic regions.

The results of employing UCS demonstrate significant advancements in various aspects of SST data analysis. The cell segmentation results of UCS exhibit high accuracy in visual observations and show remarkable consistency with H&E staining images. One of the salient advantages of UCS is its ability to support a broad range of downstream analyses. For example, UCS substantially enhances the precision of cell type annotation, facilitating more accurate identification and classification of cell types within tissue samples. Moreover, UCS significantly improves the fidelity of gene expression profiles by maintaining appropriate cell sizes and accurate boundaries. This enhancement results in gene expression profiles that closely mirror those obtained from scRNA-seq references, thereby enabling more precise spatial gene pattern recognition and more reliable subcellular gene classification.

Additionally, the improved segmentation accuracy afforded by UCS facilitates the identification of highly expressed or marker genes in specific cell types. UCS's ability to maintain cell morphology is particularly noteworthy for elongated cell types, such as fibroblasts, which are often challenging to segment accurately with other methods. By preserving the morphological information of these cells, UCS ensures that the resultant gene expression data is both reliable and biologically meaningful.

In conclusion, UCS signifies an advancement in the field of spatial transcriptomics by offering a robust and efficient solution for precise cell segmentation. Its well integration of multimodal data, coupled with its computational efficiency and applicability across a variety of SST platforms, positions UCS as a versatile and powerful tool for researchers. The enhanced segmentation accuracy and downstream analytical capabilities provided by UCS promise to yield more accurate and reliable biological insights, thereby contributing to a deeper understanding of tissue structure and function.

3. Experimental Section

Preprocess SST Dataset from Different Platform: To preprocess SST datasets using the UCS model, we start by converting all detected transcripts with spatial locations into a transcript map. This conversion results in a 3D array denoted as $G \in \mathbb{R}^{N\times H\times W}$, where N represents the number of genes in the dataset and H, W is the height and the width of transcript map. Each channel in this array corresponds to a specific gene, with each pixel within a channel reflecting the transcript count of that gene at a given location.

For nuclei segmentation, two distinct approaches based on the availability of data from the platform were employed. One approach utilizes the nuclei segmentation data provided directly by platforms such as 10X Genomics Xenium or Vizgen MERSCOPE. These platforms offer pre-segmented nuclei data, which can be seamlessly integrated into the preprocessing workflow. The alternative approach involves using the Cellpose[15] algorithm to segment the nuclei staining. Cellpose is a versatile tool that can handle diverse staining conditions and image qualities, making it ideal for datasets from different platforms. The comparison of different nuclei segmentation methods is shown in Figure S3 (Supporting Information).

Following the segmentation, the segmentation mask was adjusted to ensure uniform spatial resolution with the transcript map. This step involves resizing the segmentation mask to 0.5 μm for datasets from the Stereo-seq and CosMx platforms and to 1 μm for datasets from 10X Genomics Xenium and Vizgen MERSCOPE. This resizing ensures the segmentation mask pairs accurately with the transcriptomic data, maintaining the integrity of the size $M \in \mathbb{R}^{h\times w}$. The final input to the UCS model includes both the transcript map $\mathbf{G}$ and the resized segmentation mask $\mathbf{M}$. By integrating these elements, UCS leverages the combined strengths of transcriptomic data and pre-existing segmentation information. This dual-input strategy enhances the accuracy and robustness of cell segmentation, effectively delineating cell boundaries and identifying individual cells within the tissue.

The Model of UCS: UCS is a deep learning-based method designed for the precise segmentation of cells in diverse spatially resolved transcriptomics data. It comprises two primary neural networks: the Foreground Predict Net and the Cell Predict Net (Figure S7, Supporting Information).

The Model of UCS—Foreground Predict Net: The Foreground Predict Net focuses on identifying regions within the transcriptomic data that likely correspond to cellular areas. To achieve this, the whole transcripts map $\mathbf{G}$ was divided into smaller patches $\{\mathbf{G}^{(p)}\}$ with size $h \times w$ to facilitate processing (default patch size is 48x48). The input to this network is the transcripts map of a patch $G^{(p)} \in \mathbb{R}^{N\times h\times w}$. The Foreground Predict Net consists of:

- A feature extractor $F(\cdot)$ with three 1×1 kernel convolutional layers.
- An encoder with four 3×3 kernel convolutional layers $\{e_i\}$, $i = 1, 2, 3, 4$. The output of each layer is concatenated to provide a rich and comprehensive feature set. The concatenated output of the encoder is denoted by $E(\cdot)$.
- An classifier $C(\cdot)$ with four 1×1 kernel convolutional layers.

By employing these convolutional layers without any pooling,[22] the network ensures that critical gene expression information is not lost, resulting in a robust prediction. Both stride and padding are adjusted to maintain the original dimensions of the patches, thereby preserving the full spatial context of the gene expression data. The prediction can be derived as follows:

$$P_{\text{fg}} = Softmax(C(E(F(G^{(p)})))) \in \mathbb{R}^{2\times h\times w} \tag{1}$$

This output has two channels: the first channel represents the probability that a given pixel is part of the background, and the second channel represents the probability that a pixel is part of any cell. $\hat{p}_{ij}$ was denoted as the predicted probability of the pixel (i, j) being foreground.

The Model of UCS—Nucleus Softmask Generation: An integral aspect of our UCS model is the generation and utilization of a probability softmask derived from the nuclei segmentation mask. This softmask serves as a crucial input to the Cell Predict Net. To begin, the nuclei segmentation mask $\{\mathbf{M}\}$ is preprocessed into smaller patches $\{\mathbf{M}^{(p)}\}$, akin to the division of the transcript map. The nuclei mask for each cell is then expanded into the channel dimension, resulting in an input tensor $M^{(p)} \in \mathbb{R}^{n_{\text{cells}}\times h\times w}$. Here, each channel represents a binary mask corresponding to an individual cell's nuclei.

For each cell's nuclei segmentation mask $\{M_c^{(p)} \in \mathbb{R}^{h\times w}, c = 1, \ldots, n_{\text{cells}}\}$, we model the weight of each pixel (i, j) using a sigmoid function $\sigma(\cdot)$, which accounts for the pixel's association with the nuclei.[23] The formula for this transformation is given by:

$$\hat{M}_{cij}^{(p)} = \sigma\left(\frac{C(M_c^{(p)}, i, j) \cdot D(M_c^{(p)}, i, j)}{\tau}\right) \tag{2}$$

In this equation, $C(M_c^{(p)}, i, j)$ is a value that equals 1 if the pixel (i, j) lies within the nuclei of cell c and -1 otherwise. $D(M_c^{(p)}, i, j)$ denotes the distance from the pixel to the closest point on the nuclei segmentation boundary. The parameter τ, controls the sharpness of the sigmoid function. The resulting softmask $\hat{M}$ provides a probability distribution over the spatial domain, which is then fed into the Cell Predict Net. The softmask effectively bridges the SST data with the nuclei segmentation information.

The Model of UCS—Cell Predict Net: The Cell Predict Net, inspired by the FiLM framework,[24] is designed to seamlessly integrate the foreground features derived from the Foreground Predict Net with the softmask to achieve precise cell segmentation. The input to this network is the feature from the Foreground Predict Net

$$I^{(p)} = E(F(G^{(p)})) \in \mathbb{R}^{d\times h\times w} \tag{3}$$

and the softmask $\hat{M}^{(p)}$. The Cell Predict Net consists of:

- A network $T(\cdot)$ with three 1×1 kernel convolutional layers designed to process the foreground feature.
- A single 1×1 kernel convolutional layer $H(\cdot)$ as the predicted head.

The final output for a particular cell c in the patch will be:

$$P_{\text{cell}} = Softmax(H(T(I^{(p)}) \odot \hat{M}_c^{(p)})) \in \mathbb{R}^{2\times h\times w} \tag{4}$$

where $\odot$ is the element-wise product, effectively weighting the feature map according to the probabilistic information provided by the softmask. This operation ensures that the network focuses on regions with higher probabilities of being part of a cell. The network generates a probability distribution over a cell and background for each pixel. Note that in practice the first dimension (n_{cells}) of $\hat{M}^{(p)}$ is placed in the batch size dimension, and cells from the same patch use the same foreground feature. $\hat{q}_{ij}^{c}$ was denoted as the predicted probability of the pixel (i, j) being cell c in the corresponding patch.

To make the final prediction for each patch, consider that there are C nuclei of cells within the patch. Given that the batch size is equal to the number of cells, the outputs of the Cell Predict Net for each pixel include the predicted probability for each cell, $\hat{q}_{ij}^{c}$, where $c = 1, \ldots, C$. Since there are multiple background probabilities, they were averaged per pixel to form the predicted background probability:

$$\hat{q}_{ij}^{\text{bg}} = \frac{1}{C}\sum_{c=1}^{C}\left(1 - \hat{q}_{ij}^{c}\right) \tag{5}$$

The final segmentation result for a pixel is determined by selecting the highest probability among the background and each cell in the patch:

$$\hat{\gamma}_{ij} = \arg\max_{c}\left(\hat{q}_{ij}^{\text{bg}}, \hat{q}_{ij}^{1}, \hat{q}_{ij}^{2}, \ldots, \hat{q}_{ij}^{C}\right) \tag{6}$$

Training Details—Training Objective for the Foreground Predict Net: The training objective for the Foreground Predict Net begins with the definition of foreground and background samples. $M_{Binary}^{(p)} \in \mathbb{R}^{h\times w}$ is defined to be the binary mask for all nucleus in the patch. For the foreground, pixels located within the nuclei regions ($M_{Binary}^{(p)} > 0$) are utilized as ground truth positive samples. This ensures that the model accurately identifies areas that are indeed part of the cells.

To define the negative samples, the nuclei segmentation mask is manipulated to generate $\delta(M_{Binary}^{(p)})$, where δ represents a dilation operator. This operator dilates the nuclei masks using a 10×10 kernel over four iterations by default. Subsequently, the average expression on the zero-area of the dilation mask is calculated using the formula:

$$\bar{g} = \frac{1}{\#\{(i,j) : \delta_{ij}(M_{Binary}^{(p)}) = 0\}}\sum_{(i,j):\delta_{ij}(M_{Binary}^{(p)})=0}\sum_{k=1}^{N} G_{kij}^{(p)} \tag{7}$$

The background is then defined as $\mathbf{B}^{(p)}$:

$$B_{ij}^{(p)} = \begin{cases} 1 & \text{if } \delta_{ij}(M_{Binary}^{(p)}) = 0 \text{ and } \sum_{k=1}^{N} G_{kij}^{(p)} < \bar{g} \\ 0 & \text{Otherwise.} \end{cases} \tag{8}$$

This formulation ensures that regions without significant transcriptional activity are correctly identified as non-cell areas. Pixels that do not fall within the nuclei or background regions are excluded from the training process. Recall that $\hat{p}_{ij}$ is the predicted probability that pixel (i, j) belongs to the foreground. The weighted Cross-Entropy (CE) loss for training the Foreground Predict Net L_{fg} on a single patch is then formulated as:

$$L_{fg} = -\sum_{(i,j):B_{ij}^{(p)}>0 \text{ or } M_{Binary,ij}^{(p)}>0}\left[w_1 M_{Binary,ij}^{(p)}\log(\hat{p}_{ij}) + w_2(1 - M_{Binary,ij}^{(p)})\log(1 - \hat{p}_{ij})\right] \tag{9}$$

where w_1 and w_2 are the weighted parameters of the foreground region and background region, respectively. This selective training approach enhances the model's ability to distinguish between cellular and non-cellular regions.

Training Details—Training Objective for the Cell Predict Net: The training for the Cell Predict Net relies on the predictions of the Foreground Predict Net. Similar to the training process above, for each cell the nuclei region was designated as the positive sample, while the background predicted by the Foreground Predict Net ($\hat{p}_{ij} < 0.5$) is treated as the negative

sample. The other region will be dropped at the training stage. Suppose C represents the number of nuclei in the nuclei segmentation mask $\mathbf{M}^{(p)}$ and $\hat{q}_{ij}^{c}$ denotes the predicted segmentation probability for cell c in the patch, The final weighted CE loss for training the Cell Predict Net, L_{cell} on a single patch is then formulated as follows:

$$L_{cell} = -\frac{1}{C}\sum_{c=1}^{C}\sum_{(i,j):M_{cij}^{(p)}>0 \text{ or } \hat{p}_{ij}<0.5}[w_3 M_{cij}^{(p)}\log(\hat{q}_{ij}^{c}) + w_4(1-M_{cij}^{(p)})\log(1-\hat{q}_{ij}^{c})] \quad (10)$$

In this formulation, $\mathbf{M}_{p,c}(i,j)$ is the binary mask indicating whether the spot (i,j) belongs to the nuclei of cell c. The weights w_3 and w_4 are balancing factors that adjust the contribution of positive and negative samples to the loss function. This weighted CE loss ensures that the model learns to accurately distinguish between cell interior and predicted background regions, thus refining the segmentation boundaries.

Training Details—Other Details: An essential aspect of this training process is the influence of the dilation parameters for defining the background area. Adjusting this parameter allows to balance cell size and the number of included transcripts per cell, ensuring that the model can adapt to various tissue types and experimental conditions. The training process is expedited by the model's relatively simple architecture and the limited number of parameters, allowing for rapid convergence. For optimization, the Adam optimizer was employed with a learning rate of 10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 10^{-4} across different datasets. The weighted parameters w_1, w_2, w_3, w_4 are generally set to 1. For the Stereo-seq dataset, w_1 and w_3 are set to 5 as their default values. The postprocessing procedure for UCS adheres closely to the strategies employed by BIDCell that combine with several morphological operations.

Identifying Missing Cells: To segment missing cells, UCS follows a systematic approach that leverages its Foreground Prediction Network, which is trained to identify potential cellular regions based on the spatialdistribution of transcriptomic signals. The steps to achieve this are as follows:

1. *Standard UCS Training*: Initially, UCS is trained in the usual manner, integrating staining and transcriptomic data to learn segmentation patterns.
2. *Foreground Probability Prediction*: Using the trained Foreground Prediction Network, UCS generates a foreground probability map for each pixel in the region of interest. This map indicates the likelihood of a pixel belonging to a cell, based on transcriptomic signals.
3. *Thresholding for Potential Nuclei*: A threshold is applied to the foreground probability map to identify potential nuclei regions. A recommended threshold is greater than 0.95, as pixels with high probability values are more likely to correspond to nuclei. This step ensures that only highly probable nuclei regions are considered for further processing.
4. *Morphological Refinement*: To clean up the potential nuclei mask, morphological operations including dilation and erosion are applied. These operations help eliminate small, discontinuous regions or artifacts that may not correspond to actual cells. This step enhances the reliability of the potential nuclei mask.
5. *Exclusion of Overlapping Regions*: Any regions that overlap with existing nuclei are excluded to avoid duplicate segmentation or misidentification.
6. *Final Prediction*: With the refined potential nuclei mask, UCS predicts the newly segmented cells, incorporating the transcriptomic data to delineate their boundaries accurately.

Scaled Softmask and Integrating the Marker Gene Information: To enhance the segmentation of elongated cells without compromising the accuracy of normal cells, we propose a refined approach to scaling the softmask. This process begins with fitting a minimal enclosing ellipse around the nuclei of each cell. The major and minor axes of the ellipse are denoted as a and b, respectively. The eccentricity e of the ellipse is then calculated using the formula

$$e = \sqrt{1-(\frac{b}{a})} \quad (11)$$

where a and b represent the lengths of the major and minor axes. Following this, the distance between each pixel and the closest point on the nucleus is decomposed into two components along the major and minor axes, denoted as d_a and d_b. These distances are then scaled by the square root of the eccentricity. Specifically, the distance along the major axis d_a is multiplied by $\sqrt{e}$, and the distance along the minor d_b is divided by $\sqrt{e}$. This recalibration results in the weight of each pixel along the major axis becoming larger and along the minor axis becoming smaller (Figure S4, Supporting Information). If the nucleus is nearly circular (i.e., $\sqrt{e} \approx 0$), the distances remain largely unchanged, thus preserving the segmentation accuracy for normal cells.

In addition to adjusting the softmask, we integrate marker gene information into the model. This integration starts with annotating the nuclei to obtain cell types. For each cell, a mask is generated where each pixel is assigned a value based on the presence of marker genes in the transcript map: if a marker gene is present at a given pixel, it is assigned a value of 1; otherwise, it is assigned a value of 0. This marker gene mask is then used as an additional input to the Cell Predict Net. An additional cross-entropy loss function is introduced. This loss function penalizes discrepancies between the predicted segmentation and the marker gene mask, thereby guiding the model to incorporate the marker gene data into its segmentation predictions. Note that scaling the softmask for elongated cells and integrating marker gene information are designed to complement the normal segmentation procedure without significantly altering it.

Addressing Gene Variability and Tuning the Hyperparameters: For datasets encompassing the whole transcriptome, such as Stereo-seq and NanoString CosMx, which include more than 10 000 genes, the top 1000 highly variable genes (HVGs) were identified after filtering out control genes. This step is crucial for focusing on the most informative genes and reducing computational complexity. Given that the Stereo-seq Mouse Brain dataset captures an entire adult mouse brain slice (an area of approximately 5.3×7.0 mm^2) with a resolution of 0.5 µm, the transcript map is significantly larger compared to datasets such as 10X Xenium and Vizgen MERSCOPE.

It is essential to note that different regions of the Stereo-seq dataset exhibit distinct sets of HVGs due to the diversity in cell types (Figure S5, Supporting Information). This observation underscores a potential flaw in previous transcriptomics methodologies that apply the same set of HVGs across the entire tissue, potentially overlooking regional variations in gene expression. To address this, the Stereo-seq dataset was segmented into small datasets of 1200×1200 pixels and train a separate model on each small dataset. Similarly, for the NanoString CosMx dataset, models were trained on each field of view (FOV) to accommodate the variability in gene expression across different regions of the tissue.

The UCS workflow is detailed in the supplementary materials (Figure S6, Supporting Information). Regarding hyperparameters, primarily the number of training epochs was adjusted based on the size of the SST dataset and the parameters for the dilation kernel and iterations, which influence the perceived cell size.

Downstream Analysis: For the downstream analysis of our SST data, we employed several specialized software tools. Single-cell Variational Inference (SCVI)[25,26] (https://scvi-tools.org/) was used to transfer labels from the scRNA-seq dataset to the cells in the SST dataset based on segmentation. For cellular level downstream analysis, Scanpy[27] (https://scanpy.readthedocs.io/en/stable/) and Squidpy[28] (https://squidpy.readthedocs.io/en/stable/) were utilized, which are popular tools for single-cell gene expression analysis and spatial pattern detection, respectively. Additionally, Bento[17] (https://github.com/ckmah/bento-tools) was used for subcellular gene classification. These tools collectively enhance the analysis and interpretation of high-resolution spatial gene expression data.

Parameter Setting for Method Comparison: Several methods were used to benchmark against the proposed method. We employed SCS on Stereo-seq using the default parameters. For BIDCell, due to its requirement for extensive additional data, such as negative and positive marker genes, its application was limited to the Xenium Breast Cancer dataset. The official instructions (https://www.10xgenomics.com/jp/analysis-guides/using-baysor-to-perform-xenium-cell-segmentation) detailed in the 10x Genomics analysis guide were followed to run Baysor. Cellpose was used for segmenting the nuclei staining using the "cyto" model. Last, Voronoi diagrams were generated using the py-clesperanto Python package (https://github.com/clEsperanto/pyclesperanto_prototype).

Data Availability: All datasets used in this work are publicly available through online sources.

- Xenium Human Breast Cancer and the scRNA Reference using the adjacent tissues (https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast)
- Xenium Mouse Brain (https://www.10xgenomics.com/datasets/fresh-frozen-mouse-brain-replicates-1-standard)
- Xenium Human Lung Cancer (https://www.10xgenomics.com/datasets/preview-data-ffpe-human-lung-cancer-with-xenium-multimodal-cell-segmentation-1-standard)
- Stereo-seq Mouse Brain (https://db.cngb.org/stomics/mosta/download/)
- NanoString CosMx Human Pancreas (https://nanostring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/cosmx-smi-human-pancreas-ffpe-dataset/)
- Vizgen MERFISH Mouse Brain S2R1 (https://info.vizgen.com/mouse-brain-map)
- Vizgen MERSCOPE Ovarian Cancer 3 (https://info.vizgen.com/ffpe-showcase)

Code Availability: The UCS software is available at https://github.com/YangLabHKUST/UCS.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge the following grants: the Innovation and Technology Commission (ITCPD/17-9), Hong Kong Research Grant Council grants nos. 16301419, 16308120, 16307221, 16307322, and 16302823, The Hong Kong University of Science and Technology Startup Grants R9405 and Z0428 from the Big Data Institute to C.Y.; NSFC (Grant No. 12401384) and Shenzhen Science and Technology Program (Grant No. RCBS20231211090613024) to J.X.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords

cell segmentation, deep learning, subcellular spatial transcriptomics

Received: June 29, 2024
Revised: December 14, 2024
Published online: January 7, 2025

[1] V. Marx, *Nat. Methods* **2021**, *18*, 9.
[2] J. L. Close, B. R. Long, H. Zeng, *Nat. Methods* **2021**, *18*, 23.
[3] G. Wang, J. Zhao, Y. Yan, Y. Wang, A. R. Wu, C. Yang, *Nat. Mach. Intell.* **2023**, *5*, 1200.
[4] X. Wan, J. Xiao, S. S. T. Tam, M. Cai, R. Sugimura, Y. Wang, X. Wan, Z. Lin, A. R. Wu, C. Yang, *Nat. Commun.* **2023**, *14*, 7848.
[5] C. G. Williams, H. J. Lee, T. Asatsuma, R. Vento-Tormo, A. Haque, *Genome Med.* **2022**, *14*, 68.
[6] M. Zhang, S. W. Eichhorn, B. Zingg, Z. Yao, K. Cotter, H. Zeng, H. Dong, X. Zhuang, *Nature* **2021**, *598*, 137.
[7] A. Janesick, R. Shelansky, A. D. Gottscho, F. Wagner, S. R. Williams, M. Rouault, G. Beliakoff, C. A. Morrison, M. F. Oliveira, J. T. Sicherman, A. Kohlway, J. Abousoud, T. Y. Drennon, S. H. Mohabbat, 10x Development Teams, S. E. B. Taylor, *Nat. Commun.* **2023**, *14*, 8353.
[8] A. Chen, S. Liao, M. Cheng, K. Ma, L. Wu, Y. Lai, X. Qiu, J. Yang, J. Xu, S. Hao, X. Wang, H. Lu, X. Chen, X. Liu, X. Huang, Z. Li, Y. Hong, Y. Jiang, J. Peng, S. Liu, M. Shen, C. Liu, Q. Li, Y. Yuan, X. Wei, H. Zheng, W. Feng, Z. Wang, Y. Liu, Z. Wang, et al., *Cell* **2022**, *185*, 1777.
[9] Y. Wang, B. Liu, G. Zhao, Y. Lee, A. Buzdin, X. Mu, J. Zhao, H. Chen, X. Li, *Genomics* **2023**, 110671.
[10] S. He, R. Bhatt, C. Brown, E. A. Brown, D. L. Buhr, K. Chantranuvatana, P. Danaher, D. Dunaway, R. G. Garrison, G. Geiss, M. T. Gregory, M. L. Hoang, R. Khafizov, E. E. Killingbeck, D. Kim, T. K. Kim, Y. Kim, A. Klock, M. Korukonda, A. Kutchma, Z. R. Lewis, Y. Liang, J. S. Nelson, G. T. Ong, E. P. Perillo, J. C. Phan, T. Phan-Everson, E. Piazza, T. Rane, Z. Reitz, et al., *Nat. Biotechnol.* **2022**, *40*, 1794.
[11] X. Fu, Y. Lin, D. M. Lin, D. Mechtersheimer, C. Wang, F. Ameen, S. Ghazanfar, E. Patrick, J. Kim, J. Y. Yang, *Nat. Commun.* **2024**, *15*, 509.
[12] R. Littman, Z. Hemminger, R. Foreman, D. Arneson, G. Zhang, F. Gómez-Pinilla, X. Yang, R. Wollman, *bioRxiv* **2020**, 2020.
[13] H. Chen, D. Li, Z. Bar-Joseph, *Nat. Methods* **2023**, *20*, 1237.
[14] V. Petukhov, R. J. Xu, R. A. Soldatov, P. Cadinu, K. Khodosevich, J. R. Moffitt, P. V. Kharchenko, *Nat. Biotechnol.* **2022**, *40*, 345.
[15] C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, *Nat. Methods* **2021**, *18*, 100.
[16] G. Lee, S. Kim, J. Kim, S.-Y. Yun, *arXiv preprint arXiv:2212.03465* **2022**.
[17] C. K. Mah, N. Ahmed, N. A. Lopez, D. C. Lam, A. Pong, A. Monell, C. Kern, Y. Han, G. Prasad, A. J. Cesnik, E. Lundberg, Q. Zhu, H. Carter, G. W. Yeo, *Genome Biol.* **2024**, *25*, 82.
[18] C. Stringer, M. Pachitariu, *Nat. Methods* **2022**, *19*, 1634.
[19] C. Stringer, M. Pachitariu, *bioRxiv* **2024**.
[20] M. Weigert, U. Schmidt, in *The IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, **2022**.
[21] X. Qiu, D. Y. Zhu, Y. Lu, J. Yao, Z. Jing, K. H. Min, M. Cheng, H. Pan, L. Zuo, S. King, Q. Fang, H. Zheng, M. Wang, S. Wang, Q. Zhang, S. Yu, S. Liao, C. Liu, X. Wu, Y. L. Hao, Z. Zhang, L. Wu, Y. Zhang, M. Li, Z. Tu, J. Lin, Z. Yang, Y. Li, Y. Gu, D. Ellison, et al., *Cell* **2024**.
[22] A. Azulay, Y. Weiss, *J Mach Learn Res* **2019**, *20*, 1.
[23] J. Lazarow, W. Xu, Z. Tu, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. **2022**, 4382.
[24] E. Perez, F. Strub, H. De Vries, V. Dumoulin, A. Courville, in *Proceedings of the AAAI conference on artificial intelligence*, volume *32*, **2018**.
[25] A. Gayoso, R. Lopez, G. Xing, P. Boyeau, V. Valiollah Pour Amiri, J. Hong, K. Wu, M. Jayasuriya, E. Mehlman, M. Langevin, Y. Liu, J. Samaran, G. Misrachi, A. Nazaret, O. Clivio, C. Xu, T. Ashuach, M. Gabitto, M. Lotfollahi, V. Svensson, E. da Veiga Beltrame, V.

Kleshchevnikov, C. Talavera-López, L. Pachter, F. J. Theis, A. Streets, M. I. Jordan, J. Regier, N. Yosef, *Nat. Biotechnol.* **2022**.
[26] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, *Nat. Methods* **2018**, *15*, 1053.
[27] F. A. Wolf, P. Angerer, F. J. Theis, *Genome Biol.* **2018**, *19*, 1.
[28] G. Palla, H. Spitzer, M. Klein, D. Fischer, A. C. Schaar, L. B. Kuemmerle, S. Rybakov, I. L. Ibarra, O. Holmberg, I. Virshup, M. Lotfollahi, S. Richter, F. J. Theis, *Nat. Methods* **2022**, *19*, 171.

