COMMENTARY

Spatial transcriptomics in cancer research: Opportunities and challenges

Zhihua Ou,[1,2] Jianhua Yin,[1] Liang Wu,[1] Florent Ginhoux,[3,4,5,6] and Xin Jin[1,7,8,9,*]

[1]BGI-Shenzhen, Shenzhen 518103, China
[2]Shenzhen Key Laboratory of Unknown Pathogen Identification, BGI-Shenzhen, Shenzhen 518083, China
[3]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore 138648, Republic of Singapore
[4]Shanghai Institute of Immunology, Department of Immunology and Microbiology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
[5]Inserm U1015, Gustave Roussy, Villejuif 94800, France
[6]SingHealth Duke-NUS Academic Medical Centre, Translational Immunology Institute, Singapore 169856, Singapore
[7]School of Medicine, South China University of Technology, Guangzhou 510006, China
[8]Shenzhen Key Laboratory of Transomics Biotechnologies, BGI-Shenzhen, Shenzhen 518083, China
[9]BGI Asia-Pacific, BGI-Shenzhen, Singapore 138567, Singapore
*Correspondence: jinxin@genomics.cn (X. J.)

Received: March 3, 2023; Accepted: May 8, 2023; Published Online: May 31, 2023; https://doi.org/10.59717/j.xinn-life.2023.100006

Citation: Ou Z., Yin J., Wu L., et al., (2023). Spatial transcriptomics in cancer research: Opportunities and challenges. The Innovation Life **1(1)**, 100006.

Spatial transcriptomics (ST) technologies can be divided into two categories, depending on the strategies used to measure transcripts: imaging-based and sequencing-based technologies. Imaging-based ST technologies employ microscopy to detect transcripts using fluorescence *in situ* hybridization (FISH). On the other hand, sequencing-based technologies use sequencing to capture the spatial expression patterns of genes. These sequencing-based technologies can be further divided into three types based on how spatial information is obtained: laser capture microdissection (LCM), *in situ* sequencing (ISS), and *in situ* barcoding.[1] In this context, we aim to identify the opportunities and challenges of spatial transcriptomics in cancer research.

OPPORTUNITIES

ST technologies facilitate *in situ* single-cell-level observations of cell composition, cell states, and cell interactions. These observations enable in-depth exploration of the underlying mechanisms involved in tumor origin, metastasis, and tumor recurrence, thus offering great opportunities to unravel the core hallmarks and enabling characteristics of cancer.[2] Potential application research include the identification of the cellular and molecular factors that drive tumorigenesis and immune evasion, the characterization of tumor subclones to understand heterogeneous immune response, metabolic preference, and metastatic potential, as well as the investigation of the factors contributing to resistance against different therapies. When combined with histopathological data, ST can accelerate the discovery of biological factors with diagnostic, prognostic, and therapeutic potential for various types of cancers.

CHALLENGES IN ST TECHNOLOGIES

Although an increasing number of ST technologies can detect transcript signals at the subcellular level, several limitations still exist (Figure 1). 1) Sample compatibility: The most commonly used sample type in ST studies is freshly frozen samples. However, Some ST technologies such as Visium and Xenium from 10x Genomics, GeoMx DSP, CosMx SMI from NanoString Technologies, and DBiT-seq can also work with formalin-fixed paraffin-embedded (FFPE) samples.[1] However, the RNA integrity of FFPE samples is generally lower as compared to frozen fresh samples. Improving the capacity to detect mRNA from FFPE samples is an urgent need for ST research involving archived tumor samples. 2) Capture size: Only a few ST technologies such as Stereo-seq, 10x Genomics Xenium, NanoString CosMx SMI, Matrix-seq, sci-Space, PIXEL-seq, and MERFISH can capture tissue sections larger than 100 mm^2.[1] It needs delicate arrangements to prepare very thin or small tissues for ST experiments, such as mouse lymph nodes or small needle biopsies. Additionally, small chips hinder the investigation of the spatial atlas of large mammals. 3) Target restriction: Imaging-based ST technologies that use *in situ* hybridization require prior knowledge to design detection probes. Most of these technologies can only detect a limited number of genes, ranging from dozens to hundreds. However, MERFISH and seqFISH+ can detect over 1000 genes simultaneously with high efficiency.[1] In sequencing-based ST technologies, the LCM technologies can detect the whole transcriptome, the ISS technologies mainly detect targeted RNAs, and the technologies using *in situ* barcoding mostly capture RNAs with a polyA tail.[1] However, bacteria and some viruses lack a polyA tail in their mRNAs, and the genes of these organisms are generally not covered by widely used ST technologies. Further customization of ST detection targets is needed to explore the correlation between the microbiome and cancers, such as cervical cancer and colorectal cancer, etc. 4) Low detection efficiency: Most ST technologies capture fewer genes than single-cell RNA sequencing (scRNA-seq).[3] In cancer tissues, the stromal cells, immune cells, and cancer stem cells typically produce a lower amount of RNA transcripts compared to cancer cells and may be dispersed sporadically in the tumor microenvironment (TME). This makes it challenging to identify the expression signals of these cells. 5) mRNA diffusion: The diffusion of mRNA from adjacent active cells can obscure the true activity of relatively inactive cells in the tissue. This diffusion effect also results in significant background noise in tissues containing cavities such as the vasculature and alveoli. The combined effects of low detection efficiency and mRNA diffusion hinder accurate cell segmentation and annotation of minor or rare cell types in the TME, which may play a fundamental role in cancer development. 6) Cost: Although the cost per cell of ST is much lower than that of scRNA-seq, the overall cost per sample usually exceeds thousands of US dollars. This cost limitation hinders its application in research projects involving large cohorts with multiple organs and time points.

CHALLENGES IN EXPERIMENTAL DESIGN

To maximize the scientific value of the ST data, several aspects of the experimental design should be considered (Figure 1). 1) Sample types: Freshly frozen samples are generally preferred for most ST studies as they provide relatively comprehensive gene expression profiles. However, for projects aiming to investigate the chronological changes in rare diseases, FFPE samples are more suitable. Moreover, FFPE samples offer a higher level of biosafety when working with pathogen-infected tissues. 2) Biological replicates: Although ST data from one sample already contain thousands or millions of cells, biological replicates are crucial to ensure sufficient statistical significance. 3) Cohorts and multi-omics: Despite the high cost of ST, a well-thought-out experimental design that incorporates appropriate cohorts/treatments (including control, pre-treatment, and post-treatment) concordant with expert clinical advice, supplement technologies (scRNA-seq, genome sequencing, epigenetic assessment, etc.), and necessary clinical records is essential to enlightening and meaningful discoveries.

CHALLENGES IN DATA STORAGE AND BIOINFORMATICS ANALYSIS

The data associated with ST include high-resolution images and large volumes of sequencing data. The incorporation of multidimensional data from different cohorts in cancer studies further increases the requirements for sequencing capacity, storage space, processing speed, and analytical

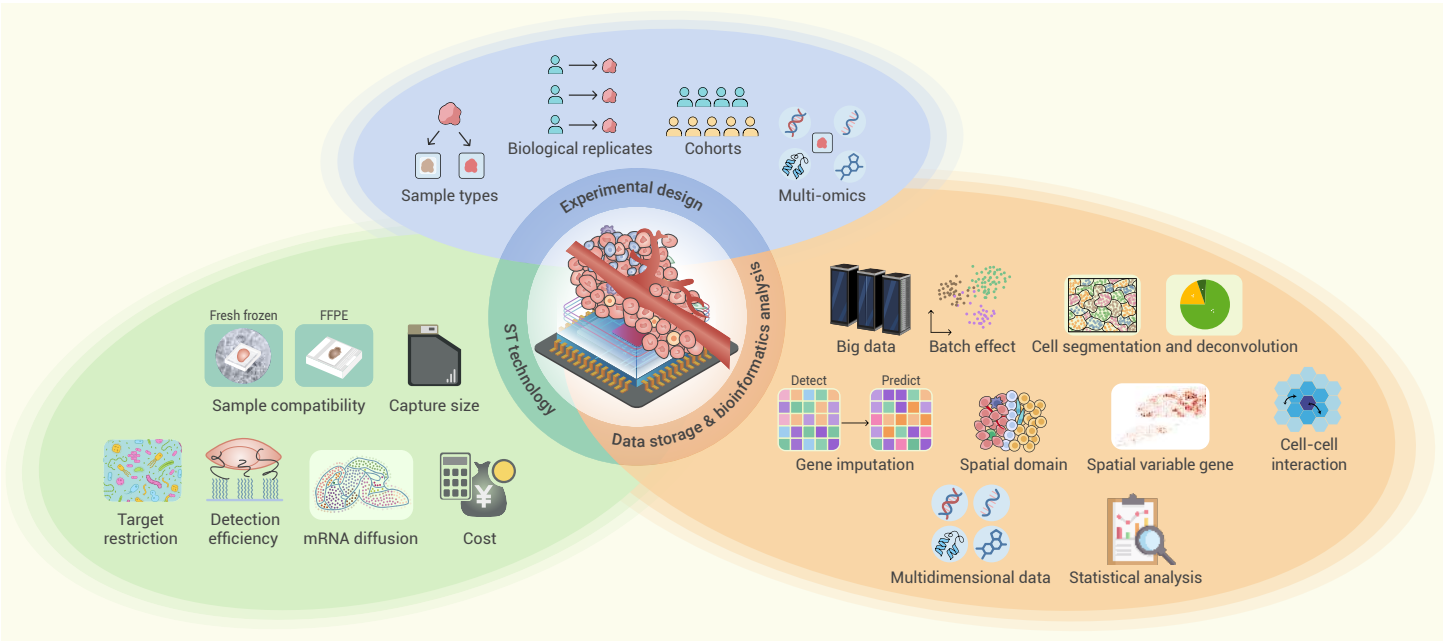

Figure 1. Challenges of ST application in cancer research Advancements in ST technology offer significant opportunities for cancer research. However, multiple challenges must be addressed, including ST technology, experimental design, data storage, and bioinformatics analysis.

algorithms. Among the emerging challenges in bioinformatics analysis of ST data, the most pressing include (Figure 1): 1) Batch effect correction: Currently, there are no batch correction tools specially designed for ST data. Researchers working with ST data may have to use scRNA-seq-based programs such as Seurat, Harmony, and LIGER sacrificing spatial information.[3] 2) Cell segmentation and deconvolution: It is crucial to adequately segment cell boundaries to annotate cell identity and crosstalk with their immediate neighbors. Careful consideration should be given to the necessity of cell segmentation based on the data quality and research purpose. Robust deconvolution algorithms, specific cellular markers, and high-quality scRNA-seq data are required to decipher cell composition in ST spots. Methods such as Cell2location, RCTD, SpatialDWLS, and Tangram perform well for cell type deconvolution.[4] 3) Gene imputation: Due to the low detection efficiency of certain ST technologies and the low RNA abundance of non-proliferating cells, expression signals of some genes may be missing in ST data. Programs such as Tangram, gimVI, SpaGE, Seurat, and ENVI can predict the possible spatial distribution of missing genes based on associated scRNA-seq datasets.[4] 4) Spatial domain determination: Identifying interaction hotspots between different cell types, such as the invasive front of tumors, provides critical insights into disease progression. Analytical methods for the identification of pathological regions of interest, including STAGATE, BayesSpace, DeepST, are encouraged.[3] 5) Spatial variable gene identification: Tools like SpatialDE, SPARK, and HRG consider spatial patterns to enhance the reliability of differentially expressed genes,[3] facilitating the identification of genes associated with cancer characteristics such as tumor cell proliferation, angiogenesis, necrosis, etc. 6) Cell-cell interaction: Identifying the ligand-receptor network between neighboring cells in the TME is crucial for the understanding of the driving factors in tumorigenesis. Programs such as CellChat, Giotto, NicheNet, NCEM, etc. can be used[3], although the interactions inferred may vary among different algorithms and require experimental validation. 7) Multidimensional data: Advanced processing and visualization tools are in high demand for the joint analysis of ST data with other dimensional data such as scRNA-seq data, *in situ* genome sequencing data, protein atlas, and metabolic profiles. Tools such as Squidpy and MUON are gaining popularity.[5] 8) Statistical analysis: The large number of spots associated with spatial information in ST datasets poses new challenges for statistical analysis. Robust statistical methods embedded within bioinformatics pipelines, along with usage recommendations are required.

PERSPECTIVES

Recently, the integration of multi-omics technologies with ST, including single-cell transcriptomics, *in situ* genome sequencing, proteomics, metabolomics, and epigenomics, has gained popularity in cancer research. Paired samples from the same individual are essential to combine multi-omics data. Additionally, multiplexed spatial technologies which collect multi-dimensional data, such as H&E (10x Visium), TCR/BCR (Slide-TCR-seq), protein expression (GeoMx DSP, DBiT-seq, STARmap PLUS, spatial-CITE-seq, etc.), and functional modification (Perturb-map), from the same tissue section are also under active development.[1] The application of multi-omics further increases the demands for project funding, group coordination, experimental design, and analytical pipelines. Therefore, inter-institutional and interdisciplinary collaborations are encouraged to accelerate the advancement of ST and associated omics technologies. As the application of ST reveals more and more molecular and cellular factors associated with cancer progression and evolution, more targets would be identified for the diagnosis, prognosis, and treatment of cancers, further facilitating the development of detection assays, drugs, and therapies.

REFERENCES

1. Cheng, M., Jiang, Y., Xu, J., et al. (2023). Spatially resolved transcriptomics: a comprehensive review of their technological advances, applications, and challenges. Journal of Genetics and Genomics, S1673852723000759.
2. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery ***12***, 31–46.
3. Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. Nat Methods ***19***, 534–546.
4. Li, B., Zhang, W., Guo, C., et al. (2022). Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. Nat Methods ***19***, 662–670.
5. Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet.

ACKNOWLEDGMENTS

The authors are grateful to Drs. Yinqi Bai and Ying He for their constructive comments on this manuscript. The authors also thank Yi Liu and Qian Zhang for their assistance in preparing graphics.

DECLARATION OF INTERESTS

The authors declare no competing interests.