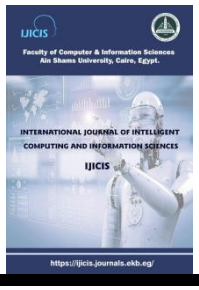
IJICIS, Vol.23, No.3, 1-8
DOI: 10.21608/ijicis.2023.186738.1247

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



CHARTING THE UNCHARTED: COMPUTATIONAL EXPLORATION OF TISSUE ARCHITECTURE USING SPATIALLY RESOLVED TRANSCRIPTOMICS

Mohamad Nossier

Program of Bioinformatics, Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
mohamad.nossier@cis.asu.edu.eg

Sherin M. Moussa*

Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

Laboratoire Interdisciplinaire de l'Université Française d'Egypte (UFEID Lab), Université Francaise d'Egypte, Cairo 11837, Egypt
sherinmoussa@cis.asu.edu.eg, sherin.moussa@ufe.edu.eg

Nagwa Badr

Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
nagwabadr@cis.asu.edu.eg

Received 2023-01-11; Revised 2023-01-11; Accepted 2023-07-14

Abstract: *Exploring and understanding the mechanisms of the complex cellular arrangements' orchestration by gene activity in multicellular organisms has great impact on the advancement of life sciences research. Spatial transcriptomics has been enabled by novel technological breakthroughs in next-generation sequencing- and imaging-based techniques to systematically measure the gene expression levels throughout the tissue, and accordingly, increase our capabilities to draw better biological insights in developmental biology and neuroscience as well as to better understand the cellular composition and landscapes of many complex diseases such as cancer. Such large-scale data made possible population wide genomic sequencing opens the door to answering many unanswered biological questions using exploratory data analysis. In this paper we deliver a review of the different exploratory data analysis aspects of spatial transcriptomic data in order to test different hypotheses using various experimental designs that utilize and compare different genetic or environmental conditions as well as different points in time. Finally, spatial transcriptomics can be integrated with multiple other omics data in order to provide much broader and deeper insights into the cellular composition and organization.*

Keywords: *scRNA-seq, spatial transcriptomics, data analysis, tissue architecture.*

*__Corresponding Author__: Sherin M. Moussa

Information Systems Department, Faculty of Computer and Information Science, Ain Shams University, Cairo 11566, Egypt

Laboratoire Interdisciplinaire de l'Université Française d'Egypte (UFEID Lab), Université Francaise d'Egypte, Cairo 11837, Egypt

Email address: *sherinmoussa@cis.asu.edu.eg, sherin.moussa@ufe.edu.eg*

1. Introduction

The realization that biological functions are implicitly linked to the cellular organization of tissues has enabled various breakthroughs in the life sciences. Spatial relationships between cells represent the basis of multiple main topics in the field of developmental biology like cell-fate decisions and the symmetrical breakage of daughter cells [1]. Since various diseases are characterized by and associated with atypical cellular spatial organization, clinically, histopathology is frequently utilized in the diagnosis of such diseases [2]. Methods in molecular biology [3] provided the chance to map the DNAs, RNAs, and proteins in tissues and thus, the ability to better understand the biological processes and their relationship to the cellular architecture within these tissues. Hence, it is now known that some inflammatory and infectious processes can alter the cellular composition and organization of tissues [5]. However, such methods have limitations regarding the number of genes or proteins studied at a time.

Our ability to understand cells has been profoundly enhanced after the Omics revolution. The newly developed methods are now able to assay full genomes, transcriptomes, and proteomes instead of recognizing just a few genetic markers [6]. That cleared the way to recognize new cell states and types and consequently, helped to provide a much deeper understanding of biological processes [7]. However, such high-throughput technologies result in the loss of cellular spatial information as they could not be performed in situ. Earlier methods tried to overcome such a limitation by the utilization of different techniques such as RNA-seq tomography (tomo-seq) [8] and isolating specific regions for scRNA-seq by microdissection to provide the needed spatial information [9] alongside many other technologies and approaches including PIC-seq [10], ClumpSeq [11], and mapping subsets of genes to deduce the locations of cells in a whole transcriptome [12].

While such approaches allowed tissue organization reconstruction, they have also demonstrated the importance of developing spatial resolution methods that scans across the whole transcriptome. That need has been met over the past decade through the immersion of a set of technologies that combine the retention of spatial information and the ability to operate on whole transcriptomes. The emergence of this novel ‘spatially resolved transcriptomics’ approach has led to breakthrough discoveries in many research fields such as neuroscience, and cancer research. In this paper, we briefly review common spatial transcriptomics technology while focusing on discussing the main principles of spatially resolved transcriptomics data exploration and highlighting the promise of such technology for providing new valuable biological insights.

2. Spatially resolved transcriptomic technologies

The computational approaches addressed here concentrate on spatial transcriptomic technologies that provide tissue-level transcriptome wide information since such technologies usually broadly differ with regard to the examined tissue size and the total number of probed genes. These technologies are commonly classified into two main groups [13]; the first group is based on incorporating spatial information in the studied transcripts prior to next-generation sequencing, while the second one involves imaging-based methodologies such as in situ sequencing (ISS) and hybridization (ISH) based techniques [14]. Further, it is worth mentioning that such classification may not always be binary, and that some computational approaches may require the utilization of information from both groups.

2.1. Next-generation sequencing-based techniques

Next generation sequencing based techniques rely on adding spatial markers before performing single-cell RNA sequencing; This technique was initially used in 2016 and enabled the capturing of tissue-level spatial data across whole transcriptomes [15]. That was done by ensuring the ability to map each transcript to its original microarray spot by utilizing positional barcodes prior to reverse transcription to detect poly-adenylated RNA. Large areas of tissues can be thoroughly examined without choosing to focus on a specific area or being limited to a group of gene markers, that is due to the high resolution of the used slides [16]. This primary concept has been previously applied by numerous research groups [15]. Thereafter, 10X Genomics further enhanced both the sensitivity and resolution of the technology which was later used in several domains such as developmental and cancer biology [17].

2.2. Imaging-based techniques

Both the in-situ sequencing (ISS) and hybridization (ISH) based techniques come at the center of Imaging-based techniques. Target genes were investigated in studies of brain development [18] and cancer [19] by applying Sequencing-by-ligation (SBL) after reverse transcription probes targeting. By further enhancing this technique, thousands of genes were successfully profiled in the cerebral cortex of mice [20]. Moreover, other techniques utilized sequencing by synthesis [21] or hybridization have yielded longer reads, thus leading to enhanced barcoding and increased throughput [22]. Further, in some studies, both imaging- and NGS-based techniques were combined in pursuit of enhanced performance and deeper insights [23].

3. Exploratory data analysis

The gene-expression matrix produced by spatial transcriptomic technologies can be analyzed to verify current hypotheses alongside making novel observations through performing exploratory spatial transcriptomic data analysis. Novel insights can be figured out by openly exploring the vastly complex high dimensional data looking for unexpected relationships. Typically, the outcome of a bench experiment acts as a guide to the next one; Similarly, the result of one exploratory data analysis leads to the choice of the following one [24]. That does not mean that existing hypotheses and knowledge are not taken into account, rather that they are utilized to better interpret the results of the analysis and direct it. Hence, there is no typical exploratory data analysis protocol or pipeline for how to investigate a spatially resolved transcriptomic dataset. Instead, there are main logical guidelines and principles for how the data may be studied alongside an expectation of the possible outcomes for each analysis [25].

The analysis of spatial transcriptomics data often starts with quality control steps and some preliminary transformations on the gene-expression matrix to reduce noise in the data and, accordingly, amplify the signal-to-noise ratio; that can be done by utilizing data analysis packages like Seurat [26] and Giotto [27]. A preliminary insight of the technical and biological characteristics of the data can be provided by the number of detected transcripts in a given location. A lower average number of transcripts may signify a technical issue; namely, cell-density differences or insufficient permeabilization in some regions. Nevertheless, there may be biological sources to such variations; including transcriptional activity variations across different cell types or death of cells, this may result in a state of confusion to the downstream analyses. In order to increase sensitivity and get rid of both undesirable technical and biological variation sources, smoothing algorithms may be used. Also, using a moving window across the spatial coordinates while calculating the average gene expression within it can help in reducing noise since neighboring spots usually have common or shared information [28]. Normalization of transcriptomes is usually done using regularized negative binomial regression or by simply dividing by

the expression values the sum of transcripts (transcripts per million) [29] allowing the comparison between gene expression among spots. In a similar fashion, such comparisons are often supported by z-score data scaling, that is to unify both mean and variance values across all the spots.

Preliminary insights at the level of single genes are then based on the normalized gene-expression matrix. Further analysis of the matrix is often required as a means to reveal underlying relationships in the data such as functional modules of genes or characteristics of cell types. While many other computational processes may be utilized, here, we briefly mention five main categories that have been commonly applied in an effort to investigate and understand spatially resolved transcriptomic data. Immediate insights may not be derived by implementing any single process; still, applying multiple processes in a serial fashion and interpreting the results of each stage may eventually lead to valuable insights.

The first of these categories is the clustering of data in order to uncover intricate structures. There are two basic options for clustering spatial transcriptomics data. The first option is to cluster spots with regards to transcriptomic similarity among them, the resulting clusters usually represent unique cell types or regions in the tissues of interest [30]. The second option is to cluster genes in order to identify sets of co-expressed genes that correspond to different cell types or states [31]. Various dimensionality reduction methods [32] are often used before the choice of suitable clustering method [33].

The second category is the choice of specific region to study in order to increase the interpretability of the analysis by limiting the investigation to a specific region, for instance, a distinct brain layer [34]. The third one is the summarization of the data to help in the recognition of spatially related functional characteristics within tissues and organs. This can be done by averaging the counts within each cluster or by utilizing a null model in order to relatively grade expression values [35]. The fourth is the annotation of the clustered data whether by integrating other datasets, using marker genes or mapping to already annotated gene sets (e.g., multimodal intersection analysis, gene-set enrichment analysis) [36]. While the fifth is the comparison of modules of genes and regions of tissues in order to point out, study, and understand the similarities, variations, and relationships between them. These comparisons can be done considering multiple factors such as RNA velocity [37] and modes of cellular interaction [38].

4. Analytical challenges and opportunities

While there are various challenges to the analysis of spatial transcriptomics data, it also offers numerous opportunities. To attain single-cell resolution transcriptomic profiling, identified RNA molecules should be aggregated individually into the cells in the case of in situ imaging-based approaches. Thus, as a means to correctly analyze the morphological properties and spatial heterogeneity among single cells, proper cell segmentation is required. Multiple current cell segmentation workflows perform well with fluorescent labeled and culture cells [39]. Segmentation performance can be further improved by integrating extra information; for example, cellular composition of the transcriptome and previously known gene expression per cell type [40]. However, new computational approaches for proper cell segmentation are still required in the case of cells with more complicated morphological features (e.g. neurons). Proper cell segmentation, alongside enhancing gene counts estimation in single cells, paves the road for the integration of spatial information by more computational approaches down the analysis stream. One instance is the prediction of transcriptional states of cells by taking the transcriptomic subcellular location into account; that is done by analyzing the temporal and spatial tissue-level heterogeneity and their functional consequences [41].

Similarly, varying distinct challenges arise regarding the analysis of spatial transcriptomic data depending on the technological source whether it is NGS-based or Imaging-based. Especially if transcripts across several cells were captured in each pixel due to larger pixels used in some technologies. This may significantly Impede the characterization of architectural patterns of single cells. Various computational approaches have been implemented in order to help overcome this obstacle by building and utilizing generative models [42] or by combining transcriptomic profiles for single cells [43] to predict the relative degree to which each cell type is represented in pixels with multi-cellular transcripts. For this reason, in order to examine the pixel- and sub-pixel-level spatial architectural patterns of distinct cell types, new computational approaches are required.

Nevertheless, to enable statistical assessment and methodical identification of the mechanisms underlying the relationship between cellular architectural patterns and phenotypical features, additional computational approaches for studying temporal and spatial transcriptomic patterns are required to overcome the limitations of the current computational approaches such as their inability to analyze multiple tissue sections over multiple time points [44].

5. Future Perspective

The domain of spatial transcriptomics is developing at a great rate due to the emergence of new related technologies and datasets on a daily basis. Quickly overcoming various challenges to the existing spatially resolved transcriptomics methods such as sensitivity, resolution, and accessibility limitations. Lately, retrospective analyses of previously collected samples have been enabled by applying spatially resolved transcriptomics methods to paraffin-embedded tissues [45]. Looking forward, the ability to analyze larger tissue areas to assemble 3-dimensional cellular atlases for organs and to visualize temporal gene expression variations across full transcriptomes will be enabled by future innovations. The development of novel computational approaches and creatively analyzing their results will come beside tackling these technological limitations in the future. These will jointly help with the exploration, identification, and understanding of spatial patterns and their related biological factors. While the functionalities of various genomic regions are still being studied, future spatially resolved transcriptomics experiments may exploit from human reference genome (that is considered a reference for investigating the origins and effects of genetic variations and was firstly drafted and published by the Human Genome Project in 2001 [46]) to study different conditions. Nonetheless, charting the levels of gene expressions will only be an opening move to shed light on how tissues are organized. The pairing of such advanced cellular atlases with creative analytical thinking is what will reveal the consequences of tissue composition and organization in physiology and disease.

One of the main questions in the field is how we can create a model that utilizes the characteristics of single cells in order to predict multicellular spatial patterns. Intricate spatial patterns were recovered by executing on a basic hypothesis that neighboring cells mostly have similar levels of gene expressions [47]. Hence, by exploring and analyzing spatial transcriptomic data, it is expected that other elementary principles will be discovered. Such principles and insights may steer spatial architecture modeling and assist mechanistic studies of spatial patterns and their impacts. This will help deepen our understanding of complex tissue architectural patterns in healthy and diseased tissues, providing us with the opportunity to chart the uncharted biological territory.

6. Conclusion

Eventually, spatially resolved transcriptomics provides an interesting novel approach to uncover the complex spatial systematic mechanisms within the tissues. Datasets carrying spatial information generated by such technologies alongside novel computational tools especially developed to analyze such data enable the identification of the sophisticated architecture of tissues. In order to confirm the results and insights derived from employing these computational tools, their generalizability is still to be tested, the experiments are to be validated, and targeted perturbation is to be performed. One example is the need for additional validation in order to figure out whether newly identified cell types through the computational integration of spatial and morphological information illustrate real biological insights into the functional heterogeneity among these cells. Further, observations about other elements that affect cell phenotype can be made by studying the degree to which gene expression is correlated to the spatial and morphological features of cells. Similarly, the impact of spatial heterogeneity among the cells on their functions remains to be identified. Ultimately, charting the cellular landscape and studying its temporal and spatial heterogeneity using specialized computational approaches can provide vital biological insights regarding tissue composition and organization in physiology and disease.

References

[1] Barresi, et al. "Developmental biology." Sinauer Associates 12 (2019).

[2] Damjanov, et al. Histopathology: A color atlas and textbook. Lippincott Williams & Wilkins, 1996.

[3] Lehmann, et al. "Methods in cell biology, Vol 44." (1994): 575-598.

[4] Swanson, et al. "Foundations of immunohistochemistry." American journal of clinical pathology 90.3 (1988): 333-339.

[5] Safai, Bijan, and Robert A. Good. Immunodermatology. Vol. 7. Springer Science & Business Media, 2013.

[6] Mincarelli, et al. "Defining cell identity with single-cell omics." Proteomics 18.18 (2018): 1700312.

[7] Stubbington, et al. "Single-cell transcriptomics to explore the immune system in health and disease." Science 358.6359 (2017): 58-63.

[8] Combs, et al. "Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression." PloS one 8.8 (2013): e71820.

[9] Nichterwitz, et al. "Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling." Nature communications 7.1 (2016): 1-11.

[10] Giladi, et al. "Dissecting cellular crosstalk by sequencing physically interacting cells." Nature Biotechnology 38.5 (2020): 629-637.

[11] Manco, et al. "Clump sequencing exposes the spatial expression programs of intestinal secretory cells." Nature communications 12.1 (2021): 1-13.

[12] Satija, et al. "Spatial reconstruction of single-cell gene expression data." Nature biotechnology 33.5 (2015): 495-502.

[13] Zhuang, et al. "Spatially resolved single-cell genomics and transcriptomics by imaging." Nature methods 18.1 (2021): 18-22.

[14] Crosetto, et al. "Spatially resolved transcriptomics and beyond." Nature Reviews Genetics 16.1 (2015): 57-66.

[15] Ståhl, et al. "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics." Science 353.6294 (2016): 78-82.

[16] Jemt, et al. "An automated approach to prepare tissue-derived spatially barcoded RNA-sequencing libraries." Scientific reports 6.1 (2016): 1-9.

[17] Grauel, et al. "TGFβ-blockade uncovers stromal plasticity in tumors by revealing the existence of a subset of interferon-licensed fibroblasts." Nature communications 11.1 (2020): 1-17.

[18] Tiklová, et al. "Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development." Nature communications 10.1 (2019): 1-12.

[19] Ke, et al. "In situ sequencing for RNA analysis in preserved tissue and cells." Nature methods 10.9 (2013): 857-860.

[20] Wang, et al. "Three-dimensional intact-tissue sequencing of single-cell transcriptional states." Science 361.6400 (2018): eaat5691.

[21] Chen, et al. "Efficient in situ barcode sequencing using padlock probe-based BaristaSeq." Nucleic acids research 46.4 (2018): e22-e22.

[22] Gyllborg, et al."Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue." Nucleic acids research 48.19 (2020): e112-e112.

[23] Fürth, et al. "In situ transcriptome accessibility sequencing (INSTA-seq)." BioRxiv (2019): 722819.

[24] Tukey, John Wilder. Exploratory data analysis: Limited preliminary Ed. Ann Arbor, MI, USA: Addison-Wesley Publishing Company, 1970.

[25] Yanai, et al. "What is the question?." Genome Biology 20.1 (2019): 1-5.

[26] Stuart, et al. "Comprehensive integration of single-cell data." Cell 177.7 (2019): 1888-1902.

[27] Dries, et al. "Giotto: a toolbox for integrative analysis and visualization of spatial expression data." Genome biology 22.1 (2021): 1-31.

[28] Ji, et al. "Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma." Cell 182.2 (2020): 497-514.

[29] Hafemeister, et al. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression." Genome biology 20.1 (2019): 1-15.

[30] Lundmark, et al. "Gene expression profiling of periodontitis-affected gingival tissue by spatial transcriptomics." Scientific reports 8.1 (2018): 1-9.

[31] Dries, et al. "Giotto: a toolbox for integrative analysis and visualization of spatial expression data." Genome biology 22.1 (2021): 1-31.

[32] Xiang, et al. "A comparison for dimensionality reduction methods of single-cell RNA-seq data." Frontiers in genetics 12 (2021): 646936.

[33] Kiselev, et al. "Challenges in unsupervised clustering of single-cell RNA-seq data." Nature Reviews Genetics 20.5 (2019): 273-282.

[34] Rodriques, et al. "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution." Science 363.6434 (2019): 1463-1467.

[35] Butler, et al. "Integrating single-cell transcriptomic data across different conditions, technologies, and species." Nature biotechnology 36.5 (2018): 411-420.

[36] Moncada, Reuben, et al. "Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas." Nature biotechnology 38.3 (2020): 333-342.

[37] La Manno, et al. "RNA velocity of single cells." Nature 560.7719 (2018): 494-498.
[38] Armingol, et al. "Deciphering cell–cell interactions and communication from gene expression." Nature Reviews Genetics 22.2 (2021): 71-88.
[39] Stringer, et al. "Cellpose: a generalist algorithm for cellular segmentation." Nature methods 18.1 (2021): 100-106.
[40] Petukhov, et al. "Bayesian segmentation of spatially resolved transcriptomics data." bioRxiv (2020).
[41] Cajigas, et al. "The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging." Neuron 74.3 (2012): 453-466.
[42] Miller, et al. "Reference-free cell-type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data." bioRxiv (2021).
[43] Andersson, et al. "Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography." Communications biology 3.1 (2020): 1-8.
[44] Rood, et al. "Toward a common coordinate framework for the human body." Cell 179.7 (2019): 1455-1467.
[45] "Spatial Transcriptomics." 10x Genomics, 2021, http://www.10xgenomics.com/spatial-transcriptomics/.
[46] Lander, et al. "Initial sequencing and analysis of the human genome." Nature 409.6822 (2001): 860-921.
[47] Nitzan, et al. "Gene expression cartography." Nature 576.7785 (2019): 132-137.