COMPUTATIONAL SYSTEMS APPROACHES TO SPATIAL AND MULTI-OMICS ANALYSIS OF TISSUES AND ORGANOID MODELS

CARISSA YAN NAN CHEN

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

This research reported in this thesis was supported by
the award of a Research Training Program scholarship
to the PhD Candidate

Faculty of Medicine and Health
The University of Sydney

28 January 2025

Abstract

The advances of 'omics' technologies have revolutionised our ability to study the global molecular profiles of cells and tissues at an unprecedented scale and precision. This thesis focuses on the application and evaluation of computational methods to analyse data generated from the latest spatial and multi-omics technologies to characterise tissues and organoid models.

In the first chapter of this thesis, I evaluated the performance of current state-of-the-art spatially variable gene (SVG) detection methods. This demonstrated that SVGs are biologically informative in understanding spatial gene expression across various tissue contexts, and provided insights into the current methods and recommendations for future method development. Following this, I explored the application of spatial and multi-omics technologies in the characterisation of *in vitro* 3D neural and retinal organoid development to mimic *in vivo* human development. First, we generated a trans-omic map of early neural organoid development profiling the important kinases which propagate cell identity-specific signalling and downstream gene regulation during neurogenesis. Computational analysis uncovered key signalling cascades and enabled us to modulate kinase signalling toward improving neural organoid generation. Complementary to this work, I characterised retinogenesis under the duress of retinoic acid modulation to perturb retinal cell fate. Using multi-resolution spatial transcriptomics, I investigated the spatiotemporal dynamics of transcriptional activity and its influences on cellular heterogeneity during tissue patterning of retinogenesis. This revealed that retinoic acid signaling affects photoreceptor maturation and retinal interneuron abundance, influencing retinal lamina formation and central retina development, thereby contributing to the increased fidelity of retinal organoids to the human retina.

Together, this thesis showcases the application of computational systems approaches for analysing spatial and multi-omics data in tissues and 3D organoid models. The knowledge discovered from these analyses not only provides guidelines for future method development but also advances the understanding of developmental biology and highlights areas of improvement for the benefit of translational research.

Statement of originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Carissa Yan Nan Chen
Faculty of Medicine and Health

Declaration and authorship attribution

This is to certify that, to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

Chapter 2 of this thesis is based on my co-first author publication in *Genome Biology* (Chen C., Kim, HJ. et al. 2024). I designed the study with Hani J. Kim and my supervisor. I performed the data analysis, interpreted the results and wrote the manuscript with Hani J. Kim.

Chapter 3 of this thesis is based on my first author publication in *Cell Reports* (Chen C. et al. 2024). I designed the study with my supervisor, Anai Gonzalez-Cordero (Stem Cell & Organoid Facility, Stem Cell Medicine Group, CMRI) and Mark E. Graham (Synapse Proteomics, CMRI). I performed the data analysis, interpreted the data and wrote the manuscript. Katherine G. Zyner provided assistance to interpret the results and Lee L. Marshall assisted with the image analysis. Scott Lee, Milan Fernando, Emily Wong, Jesse R. Wark, Mark E. Graham and Anai Gonzalez-Cordero generated the data analysed in the study and performed the wet lab experiments.

Chapter 4 of this thesis features the unpublished collaborative work with Benjamin Y. Lim and Anai Gonzalez-Cordero (Stem Cell & Organoid Facility, Stem Cell Medicine Group, CMRI). I led the design of the computational analysis, analysed and interpreted the results. Data generation was performed by Benjamin Y. Lim and Megan Weatherstone (Single Cell Sequencing Facility, CMRI). The experimental methods for this chapter were provided by Benjamin Y. Lim.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

As a supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Candidate	Supervisor

Publications

Most of the research conducted during my PhD candidature has been published in peer-reviewed journals. Most of the methods, concepts, analyses, and results in this thesis have appeared previously in some of the publications listed below. These materials have contributed significantly to my PhD candidature.

Published in peer-reviewed journals

* publication included in this thesis; ‡: Co-author, equal contribution

(1) Wagle, M., Long, S., **Chen, C.**, Liu, C., Yang, P. (2024) Interpretable deep learning in single-cell omics. *Bioinformatics*, 40(6).

(2) ***Chen, C.**, Lee, S., Zyner, KG., Fernando, M., Nemeruck, V., Wong, E., Marshall, LL., Wark, JW., Arymanesh, N., Tam, PPL., Graham, ME., Gonzalez-Cordero, A., Yang, P. (2024) Trans-omic Profiling Uncovers Molecular Controls of the Early Human Cerebral Organoid Formation. *Cell Reports*, 43(5).

(3) ***Chen, C.**‡, Kim, HJ.‡ and Yang, P. (2024). Evaluating spatially variable gene detection methods for spatial transcriptomics data. *Genome Biology*, 25(18).

(4) Xiao, D., **Chen, C.**, and Yang, P. (2023). Computational systems approach towards phosphoproteomics and their downstream regulation. *Proteomics*, 23, 2200068.

(5) Kim, HJ., Wang, K., **Chen, C.**, Lin, Y., Tam, PPL., Lin, DM., Yang, JYH., Yang, P. (2021) Uncovering cell identity through differential stability with Cepo. *Nature Computational Science*, 1, 784790

Preprints

(1) Toh, H., Xu, L., **Chen, C.**, Yang, P., Sun, AX., Ouyang, J. (2024) BrainSTEM: A multi-resolution fetal brain atlas to assess the fidelity of human midbrain cultures. *Biorxiv*

(2) Salim A., Bhuva, DD., **Chen, C.**, Yang, P., Davis, MJ., Yang, JYH. (2024) SpaNorm: spatially-aware normalisation for spatial transcriptomics data. *Biorxiv*

Presentations

Most of the publications outlined in the previous section have been presented as oral or poster presentations at several international and domestic conferences and various research courses and seminars. The entire list of presentations is outlined below by topic.

Evaluating spatially variable gene detection methods for spatial transcriptomics data

2024 UK Dementia Research Institute Seminar at Imperial College London, United Kingdom, oral presentation

2024 Institute of Human Genetics at French National Centre for Scientific Research, France, oral presentation

2023 Australian Bioinformatics and Computational Biology Society Conference, Australia, oral presentation

2023 Computational Biology Student Symposium, Australia, poster presentation

2023 Children's Medical Research Institute (CMRI) Seminar, oral presentation

2023 Sydney Bioinformatics Research Symposium, Australia, oral presentation

Trans-omic Profiling Uncovers Molecular Controls of the Early Human Cerebral Organoid Formation

2024 Stem Cell Conversations, Australia, virtual oral presentation

2024 International Society for Stem Cell Research Conference, Germany, poster presentation

2024 UK Dementia Research Institute Seminar at Imperial College London, United Kingdom, oral presentation

2024 Institute of Human Genetics at French National Centre for Scientific Research, France, oral presentation

2023 The Australasian Society for Stem Cell Research Conference, poster presentation

2023 Children's Medical Research Institute (CMRI) Seminar, oral presentation

2022 Computational Biology Student Symposium, Australia, oral presentation

2022 Australian Bioinformatics and Computational Biology Society Conference, Australia, poster presentation

2022 Children's Medical Research Institute (CMRI) Seminar, oral presentation

2021 Children's Medical Research Institute (CMRI) Seminar, oral presentation

CLUEY enables knowledge-guided clustering of multi-omic single-cell data

2022 OzSingleCell Conference, Australia, poster presentation

2022 Sydney Bioinformatics Research Symposium, Australia, poster presentation

Acknowledgements

Embarking on a PhD has been a tremendous journey full of growth, scientific learning as well as having the privilege to form a network with like-minded individuals. I would like to also acknowledge that this thesis is additionally supported by a Children's Medical Research Institute PhD Scholarship.

To my supervisor Pengyi, you have taught me how to conduct rigorous, high-quality science that has shaped me for the rest of my research career. I am so grateful for your constant willingness to impart your time, knowledge and expertise. Your work ethic, critical and creative thinking has, and always will continue to inspire me.

To my associate supervisor, Patrick, thank you for your insightful comments and engagement in scientific discourse that reveals the blind spots in my thinking. Your steadfast passion for research and depth of knowledge in stem cell and embryology has been inspiring to witness. To the Embryology Unit - especially Nicole, Pragathi and Megan whom I have spent my lunches with doing New York Times puzzles like it was a daily ritual, it was one of most entertaining times of the day.

To the Sydney Precision Data Science Centre, thank you for creating such a collegiate, supportive and stimulating environment. Shila, for your words of encouragement that always reassures me. To my friends who I have met along the way - Di, Sharon, Chunlei, Sichang, Farhan, Jackson, Raj, Max, Chuhan, Cabiria, Yue, Helen, Lijia, Wenze, Yunwei, Kevin, Xiangnan, Hao, Harry, Peng, it has been fun sharing the same space as you.

Thank you to Hani and Katie, who have always been so open-minded and willing to help in times of need. Your mentorship has been invaluable.

To Sonu, Katie, Anna and Di for proofreading this thesis during crunch time. To my collaborators at CMRI, with special mention to Anai who has been generous with her time and feedback, no matter the hour of the day.

Lastly, to my parents who have supported me throughout, and to my brother, who still does not understand what I do but listens with feigned interest regardless.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Overall aim of the thesis

The dawn of omics technologies and high-throughput sequencing has given us the opportunity to dissect cellular behaviour, decisions and regulation of the cell phenotype at a high resolution. Particularly, spatial transcriptomics allows us to contextualise gene expression in the native spatial microenvironment of a tissue. This offers a deeper understanding of how spatial proximity influences cell function. In lieu, the profiling of multiple omics such as the phosphoproteome, proteome and transcriptome by mass spectrometry and RNA-seq, respectively, we obtain a multi-view perspective on the inner workings of signalling networks that control gene regulation. Namely, the reconstruction of trans-regulatory networks to understand how cell identity is formed. I apply these biological principles to dissect the developmental decisions during human development, using organoid models as a proxy.

Three-dimensional (3D) organoids are *in vitro* systems that mimic the structure and function of human tissues and organs. In contrast to two-dimensional cultures where cells grow on a flat surface, the formation of three-dimensional organoids permits a more complex and accurate representation of *in vivo* organ development. This is crucial since multi-cellularity and tissue organisation naturally occur in 3D environments, which better mimic how cells interact and function *in vivo*. Given this, I demonstrated the capabilities of 3D organoid models to faithfully recapitulate *in vivo* development and their utility to simulate perturbations *in vitro* for a greater understanding of cell identity regulation during human tissue and organ development.

Leveraging cutting-edge high-throughput technologies, this thesis interweaves spatial transcriptomics and multi-omics to study the spatially-resolved global dynamics influencing cellular phenotype, and trans-regulatory networks that drive cell fate decisions and lineage commitment. Specifically, I emphasise the value of the integrative computational analysis

of 3D organoid models to advance our understanding of human development, prefacing the promise of reconstructing spatially-resolved TRNs to improve organoid modelling.

1.2 Background

1.2.1 Spatial transcriptomics to understand spatial organisation of tissues and its cell types

Named method of the year in 2020 by Nature Methods [1], the rise of spatial transcriptomics is motivated by the desire to quantify mRNA expression in their native spatial context, a limitation of single-cell technologies which involve the dissociation of single cells from their tissue micro-environment. There are numerous spatial techniques, of which there is still continuous innovation in the field. They can be categorised into two classes, fluorescence imaging-based techniques and sequencing-based techniques. *In situ* fluorescence-based technologies such as seqFISH [2, 3] and MERFISH [4] which enable the profiling of typically ten thousand genes while compromising on the number of cells detected, typically only up to a thousand cells. On the other end of the spectrum, sequencing-based techniques involve permeabilising tissue onto a library sequencing slide that is populated by beads which contain spatial barcodes to enable mapping of the cells back to their original locations. However, the size of the beads, which will be referred to as 'spots', determines the cellular resolution. For example, the 10x Visium platform contains beads which are 55μm in diameter, which typically is larger than a single cell. Therefore, the gene expression of each spot will be the sum of multiple cells. To decompose the putative cell types present in each spot, many algorithms rely on the assumption that the gene expression of each cell type contributes a percentage to the total gene expression profile of each spot [5, 6] and estimate the cell type proportions by a linear combination of multiple cell types inferred from an appropriate scRNA-seq reference.

More recently, sub-cellular technologies such as BGI Stereo-seq [7] were invented wherein each bead is nanometers in diameter to circumvent the issue of average cell populations per spot. However this introduces new challenges in image analysis to segment single cells based on their organelle composition and morphology, thus advancements in cell segmentation algorithms to accurately segment single cells are required.

Spatial transcriptomics has been crucial in various fields to understand how cellular cross-talk between neighbouring and distant cells inform cell behaviour. In cancer biology, spatial transcriptomics helps reveal how the tumor micro-environment drives clonal growth and metastasis by mapping the spatial spread of tumor cells and their interactions with surrounding tissues [8]. Similarly, in developmental biology, platforms like Stereo-seq have been employed to investigate the spatiotemporal dynamics of regional specification during mouse organogenesis, providing insights into developmental patterning over time [7]. Thus, the field of spatial transcriptomics is becoming increasingly relevant to interrogate complex biological processes that involve interactions with the spatial micro-environment.

1.2.2 Trans-omics analysis reveals the complex interplay of gene regulatory networks across multiple omic layers

All cells share the same genetic code, which provides the foundation for cell- and tissue-specific phenotypes. These phenotypes arise through the activation of specific gene regulatory programs, contributing to the multicellular complexity seen in different organisms. Within each cell, multiple layers of regulation - referred to as 'omic layers' encompass the interactions between molecules such as RNA (transcriptome) and protein (proteome) with the surrounding micro-environment containing metabolites (metabolome) which are influenced by chromatin architecture (epigenome). Each omic layer can be quantified using technologies such as RNA-sequencing [9], mass spectrometry-based proteomics [10], phosphoproteomics [11, 12], metabolomics [13, 14], and chromatin immunoprecipitation sequencing (ChIP-Seq) or assay of transposase accessible chromatin sequencing (ATAC-seq) [15, 16].

Individual omic layers, such as the transcriptome of the human brain [17], proteome of the temporal lobe [18], and phosphoproteome in Alzheimer's disease [19], are frequently studied in isolation. However, global efforts like the Human Proteome Atlas [20] and the Allen Human Brain Atlas [21] are continuously updated to serve as resources for the broader scientific community. However, it is increasingly recognized that these omic layers work synergistically to regulate gene expression and cellular phenotypes. Hence, the concept of 'trans-omics' analysis, coined by Yugi et al. [22], emphasises the integration of these multimodal data measurements, guided by prior biological knowledge [23, 24], to provide deeper insights into cellular dynamics in various biological contexts, including development, health, and disease.

1.2.3 Reconstruction of trans-regulatory networks to understand cell identity formation

Mass spectrometry offers several advantages to study the global proteome and post-translational modifications that affect protein function. It can simultaneously identify and quantify thousands of proteins at a time and the capability to multiplex samples in parallel using techniques such as tandem mass tags (TMT) increases throughput and efficiency [25, 26, 27]. Phosphorylation is the most prevalent form of post-translational modification of a protein in eukaryotes [28]. The addition of a phosphate group by a kinase, or its removal by a phosphatase, induces a conformational change in the protein, influencing protein-protein interactions. This process plays a crucial role in cell behaviour, driving signal transduction and regulating the transcription of genes involved in cell proliferation, differentiation, and metabolism. For example, Akt-driven signalling plays a plethora of roles such as stimulating downstream targets involved in cell cycle regulation and migration and engaging in crosstalk with other pathways such as NFκB and Erk as part of a wider integrated signalling network within a cell [29]. There have also been a number of studies demonstrating the isoform-specific roles of Akt in stimulating context-specific gene regulation despite the ubiquitous expression of all Akt isoforms. Such as during neural development, Akt3 is expressed at a higher level than Akt1/2 [30] and Akt3 deficiency has been implicated in neurodevelopmental disorders [31, 32].

Next-generation sequencing (NGS) has revolutionized the ability to simultaneously sequence millions of DNA fragments, allowing for the comprehensive quantification of the entire transcriptome in a single run, while also enabling sample multiplexing. This offers a more comprehensive view of gene expression within the broader context of global transcriptional activity. Integrated with mass spectrometry-generated proteomics and phosphoproteomics, RNA-sequencing facilitates the reconstruction of trans-regulatory networks (TRNs) that connect multiple regulatory layers. This has greatly enhanced our understanding of the intricate coordination involving signalling networks initiated by kinase-substrate interactions and downstream transcriptional and translational regulation which inform cell fate commitment in a myriad of biological contexts such as pluripotency state transitions [33, 34], gastrulation [35], myogenesis [36] and hepatocyte differentiation [37].

1.2.4 Single-cell transcriptomics to study cellular heterogeneity

Since the earliest demonstration of single-cell transcriptomics of a single mouse blastomere [38], there has been an exponential growth in the scale of single-cell experiments [39]. In contrast to bulk RNA-sequencing experiments wherein whole transcriptomic measurements represent the average levels across the sample population, we now have access to measurements representing each individual single cell. This provides an opportunity to survey rare and major cell type populations within a sample in an unbiased manner. Moreover, allowing the capacity to relate gene transcriptional activity to cell type-specific phenotypes and functions.

The increased complexity in the scale of data generated has also introduced interesting computational challenges to efficiently dissect the heterogeneity. Two approaches are commonly employed to perform cell type annotation of single cells: unsupervised clustering followed by visualisation of key markers to assign clusters to a cell type, and supervised approaches involving an appropriate reference to train a classification model [40] or projection of the reference onto the query dataset in a low dimensional space [41], termed 'transfer learning'. Given the increasing availability of a vast array of single-cell atlases, transfer learning is becoming the preferred approach given its efficiency in handling the increasing size of datasets and its robustness against batch effects or differences in sequencing depth which often plague classification methods [42]. Overall, the suite of cell annotation tools allows the user to dissect the cellular heterogeneity of single-cell data in various contexts.

1.2.5 The era of 3D organoid modelling to understand development and disease

Pluripotent stem cells display a remarkable level of plasticity and a tendency to self-organise and form 3D aggregates. The resulting organoids grown from organ progenitors can mimic the structural architecture or key functions of organs such as the liver, kidney, brain, gut intestine, and retina amongst many others [43]. This has revolutionised our understanding of development [44, 45, 46] and isolated functions of specific organs [47] as it allows us to study changes in real-time. By recapitulating the organic organisation into a 3D structure which encourages more cell diversity, natural progression of maturity and functional fidelity, it is a more accurate model than traditional 2D monolayer cultures.

Organoid models are also useful to model disease pathogenesis, where the flexibility introduces a myriad of possibilities when perturbing the model to study desired effects which have important implications for personalised medicine. For example, cancer organoids [48] are utilised to study patient-specific tumour heterogeneity. Kidney organoids can also be used for nephrotoxicity screening of functional proximal tubules in the nephrons [49] and therefore would prove a useful tool for drug screening for the purposes of personalised medicine.

Despite the many challenges such as the reproducibility of organoids, accurate representation of full organ complexity and fidelity of cell identity compared to the primary counterpart [50], the continued research into organoid biology serves as the bridge between theoretical and translational research.

1.2.6 The innovation of organoid models to study human brain development

Human brain development is a sequential, spatiotemporal process that starts with neurulation, where the neural plate folds and fuses to form a neural tube structure. The neural tube then subdivides into lineage-restricted vesicles of neuroepithelial cells that become the four distinct regions of the central nervous system - the forebrain, midbrain and hindbrain and spinal cord [51].

Traditionally, The fruit fly *Drosophila melanogaster* has been one of the main model organisms to study the fundamentals of neural circuitry as well as human neurodegenerative diseases such as Alzheimer's, Parkinson's disease and epilepsy [52]. While genetic manipulation of fruit flies is relatively easy and efficient, we have increasingly turned to other mammalian models such as the mouse as they more closely mimic the structure and function of the human brain. They have been commonly used as the host genetic background to test human-acquired genetic changes, such as identifying the genetic causality of microcephaly-associated genes [53]. However, species-related molecular and anatomical differences still pose a challenge when attempting to predict the full extent of neural complexity or treatment responses in humans. The increased availability of neurotypical and pathological human brain specimens coupled with advancement in high-throughput sequencing and genome-wide association studies have allowed for the multi-omic characterisation of the physiology of the developing human brain. However, these analyses can only provide descriptive evidence and associations with

specific genes. Without further functional studies and genetic perturbations, it is insufficient to conclude the causative effect of candidate genes to the genetic aetiology. As genetic perturbations cannot be performed on human brain tissue for practical reasons, this has motivated the use of human *in vitro* models to perform validation studies [51, 54].

The development of human brain organoid models [55, 56, 46] has been widely applied to study the molecular and cellular processes involved in brain organogenesis [57, 58]. Moreover, numerous efforts have been invested in profiling the single-cell heterogeneity of brain organoids, leading to a unified atlasing effort [59] to generate a 1.77 million-cell atlas covering various human brain organoid protocols. While there has been much focus on examining the transcriptional regulation of human neural organoid development, recent studies have integratively profiled the proteomics [60] or epigenomic landscape [61, 62] to infer the multi-omic coordination of neurogenesis. These studies have allowed a greater understanding of post-transcriptional regulation and chromatin dynamics impacts cell identity reconfiguration during neural maturation. However, the fidelity of cell signalling cascades and their influence on downstream transcriptional regulation during neurogenesis remains to be explored in human brain organoid models.

1.2.7 Developing a faithful model of the human retina with accurate spatial organisation

Human retinal development is a highly orchestrated process, driven by the tight regulation of morphogenetic cues, including Wnt, BMP, Notch, FGF and retinoic acid (RA) [63, 64]. These signalling pathways work in harmony to ensure proper spatial patterning of the retinal laminae and the establishment of the central and peripheral regions of the eye.

Retinogenesis describes the formation of the retinal laminae that comprises of seven major retinal cell types which differentiate sequentially from multipotent retinal progenitor cells (RPCs). The early phase is dictated by the rise of retinal ganglion cells, cone photorecepter cells, horizontal and amacrine cells followed by the later rise of rod photoreceptors, bipolar cells and Müller glia [65]. This process initiates in the central retina, before maturation continues in the later peripheral region of the retina [66]. The delay in developmental timing enables the formation of the macula, a 5mm diameter specialised structure with the highest density of L/M cone photoreceptors responsible for high visual acuity and color perception

[67, 68]. In the chick, formation of a rod-free zone involves the interplay between RA and FGF8 [69]. Although animal models with analogous retinal structures provide insights into potential patterning mechanisms, they do not allow for a comprehensive exploration of the precise timing and signaling events involved in human retinal development. Thus, the introduction of human cell-line derived retinal organoids offer an alternative to study macula development if it can be faithfully recapitulated *in vitro*.

Current retinal organoid protocols successfully mimic key features of retinogenesis, including the distinct retinal layers and functional maturation [70]. Further efforts to achieve finely tuned retinal functions are ongoing, with studies demonstrating that the manipulation of thyroid and RA signaling pathways can promote L/M cone specification [71, 72]. Beyond these studies, further investigation into the role of surrounding retinal interneurons and signaling interactions is needed to understand a holistic picture of how the spatial organisation influences the retinal architecture and function in the developing retina. The generation of a robust and faithful human retinal organoid model will be instrumental in the study and development of therapeutics for retinal degenerative diseases, particularly age-related macular degeneration or other inherited retinopathies.

1.3 Structure of the thesis

This thesis is structured as follows:

- **Chapter 1** provides an overview of the biological background, challenges and contributions of the thesis.
- **Chapter 2** introduces an evaluation framework to assess the performance of spatially variable gene detection methods to select biologically meaningful genes that vary in a spatial context.
- **Chapter 3** presents the multi-modal integrative analysis of early neural organoid development to identify key kinases and characterise transcriptional networks that shape neurogenesis.
- **Chapter 4** investigates the effect of retinoic acid perturbation to manipulate retinogenesis and its influence on the spatial organisation of retinal organoids.
- **Chapter 5** concludes the thesis with a discussion of future directions in the field of multi-omics and spatial transcriptomics to interrogate organoid biology.

1.4 Challenges and thesis contributions

This thesis addresses key analytical challenges in leveraging emerging biotechnologies to study development and cellular complexity through advanced spatial and multi-omics approaches. First, I evaluated computational tools to detect biologically meaningful spatially variable genes in spatial transcriptomics and provided recommendations for future improvements. Next, I demonstrated the value of integrating multi-omics and spatial contextualisation to uncover novel biology surrounding neural and retinal cell fate through 3D organoid models. Altogether, these contributions advocate for the use of organoid models and provide an assessment for future research integrating multi-omic data to holistically understand human development.

1.4.1 Spatially variable gene detection in spatial transcriptomics

The rise of novel biotechnologies to enhance our understanding of cellular complexity naturally leads to the development of novel computational methods to analyse the resulting data to glean biological insights. With the advent of spatial transcriptomics, this allows us

to contextualise gene expression with respect to its location in a tissue to understand how the proximity of cells facilitates signalling and gene regulation involved in cell identity. At the most rudimentary, this has led to the development of numerous spatially variable gene (SVG) detection methods which lay the foundation for understanding spatial gene expression patterns. This has motivated the evaluation of the effectiveness and similarities of the variety of approaches undertaken to translate and integrate spatial location with gene expression to determine spatially-informed gene expression patterns.

In Chapter 2, I evaluated a panel of state-of-art spatial variable gene detection methods on a collection of datasets that span different species, protocols and sequencing depth. I compared the performance of these methodologies under typical analytical scenarios to assess the ability to prioritize biologically meaningful signals, as well as detection accuracy and reveal underlying dependencies on the data. Overall, this work presents a perspective on how current methods use space as an additional layer of information to complement gene expression and highlights weaknesses that could be improved in future method development.

1.4.2 Modelling early neural lineage commitment using cerebral organoids

Transitioning from single-modal transcriptomics, in Chapter 3, I expand the focus to trans-omics data, where multiple omic modalities, including transcriptomics, proteomics, and phosphoproteomics, are analyzed and integrated. This trans-omics approach provides a more comprehensive and holistic view of the underlying biological processes. By combining these data types, I aim to reconstruct the signalling and gene regulatory networks that drive neural lineage commitment in human stem cell-derived brain organoids, offering deeper insights into the molecular mechanisms of early neurogenesis.

Early discovery studies of neural development have been limited to microarray analysis, experimentation of selective genes, mouse models or the single omic layers in isolation. Moreover, due to ethical concerns with acquiring human embryo tissue, it has not been possible to characterise the initial stages of neural lineage commitment. In this chapter, I address the challenges of the field by performing phosphoproteomic, proteomic and (single-cell)-transcriptomic profiling of human cerebral organoid differentiation from the exit of pluripotency to the formation of the neuroectoderm as a model for early neural development.

Through the trans-omic analysis across all three omic layers, I identified crucial kinases that are regulated during this developmental process and link phospho-signalling networks to the downstream transcriptional and translational activity of key molecular factors which is associated with neurogenesis. Additionally, I confirmed the transcriptional fidelity of the cerebral organoid culture with *in vivo* human and orthogonally, mouse embryogenesis. I also demonstrated the utility of the large-scale data to uncover kinases or novel substrates which can be modulated to enhance the efficiency of organoid maturation.

The culmination of this work allowed us to reconstruct the signalling and gene regulatory networks that govern early neurogenesis in stem cell-derived brain organoids and provide valuable insights into neural development.

1.4.3 Characterising the perturbation of retinoic acid signalling during retinogenesis

In Chapter 4, I shift my focus to another organoid model - namely retinal organoids in order to investigate the mechanisms of retinal cell fate determination. This chapter leverages the integrative approaches developed in Chapter 3 to examine changes across a timecourse, but in the context of retinal development. Specifically, I combined spatial transcriptomics and scRNA-seq to explore how retinoic acid signalling influences retinal cell fate and spatial organization.

The role of retinoic acid signalling in organogenesis, particularly in retinal morphogenesis and photoreceptor maturation, is well-established in the literature. However, many of these studies have been conducted in model organisms [73, 74], and there is limited research systematically examining its effects on other retinal cell types during retinogenesis. Additionally, since the human retina is characterized by a laminar structure, spatial organization becomes a critical aspect to consider in these studies which is an emerging direction in the field.

Here, I used a human cell line-derived retinal organoid system to model the effects of altering retinoic acid levels during the course of retinogenesis. We performed spatial transcriptomics and single-cell transcriptomics at key stages of retinogenesis progression to uncover the spatiotemporal dynamics involved in retinal development, expanding the scope of our research to another critical developmental process during human organogenesis.

Finally, I conclude the thesis in Chapter 5 and present future directions in integrating spatial transcriptomics and multi-omics to advance organoid modelling through the reconstruction of spatially-resolved TRNs.

CHAPTER 2

Evaluating spatially variable gene detection methods for spatial transcriptomics data

---

This work is published in the following journal:

Chen, C.[‡], Kim, H.J.[‡] and Yang, P. (2024). Evaluating spatially variable gene detection methods for spatial transcriptomics data. *Genome Biology*, 25, 18. https://doi.org/10.1186/s13059-023-03145-y

My contribution: I co-developed and ran the benchmarking pipeline to reproducibly generate the spatially variable gene statistics for all methods upon which their performance was evaluated on real and simulated datasets. Excluding the accuracy and clustering assessments, I performed the data analysis, visualisation for all figures, and I assisted in the interpretation of the accuracy and clustering assessments.

2.1 Motivation and overview

The fundamental question in spatial transcriptomics revolves around identifying and characterising genes with variable expression patterns in space which inform tissue or cell-type specific regions. From the earliest technologies which were limited in throughput, the field has grown to innovate *in situ* hybridisation and *in situ* sequencing spatial barcoding-based technologies which can now profile thousands of genes and cells at a time. Alongside the technological advancements, computational methods have also evolved to leverage the spatial information and gene expression from spatial transcriptomics. Despite their importance, spatially variable gene detection methods have yet to be thoroughly evaluated. To address this gap in the field, I conducted a systematic benchmark to review and compare a panel of state-of-the-art spatially variable gene detection methods against a range of spatial technologies to examine

the concordance and discrepancies of methods and their performance under a selection of evaluation metrics and relevant analytical scenarios in spatial transcriptomics. Not only does this allow us to determine the competitive methods in the field, we also gain insight into the various methodologies that attempt to incorporate spatial information to identify biologically meaningful genes in space.

2.2 Introduction

Advances in spatial transcriptomics have made it possible to identify genes that vary across spatial domains in tissues and cells. The detection of spatially variable genes (SVGs) is essential for capturing genes that carry biological signals and reducing the high-dimensionality of the spatial transcriptomics data, which is akin to defining highly variable genes (HVGs) [75] in single-cell RNA-sequencing (scRNA-seq) data [76]. These SVGs are therefore useful for various downstream analyses of spatial transcriptomics data. Spatially variable genes are, however, conceptually different from HVGs found in scRNA-seq data as, by definition, SVGs preserve the spatial relationships of tissues and cells in the biological samples whereas HVGs do not necessarily preserve such relationships.

A fast-growing number of methods for SVG detection have been proposed in the recent literature. Some popular examples include SpatialDE [77] based on the Gaussian process; SPARK [78] and SPARK-X [79] based on mixed and non-parametric models, respectively; SOMDE based on the self-organising map (SOM) [80]; Giotto based on statistical enrichment of spatial network in neighbouring cells [81]; nnSVG based on nearest-neighbour Gaussian processes [82]; MERINGUE based on nearest-neighbour spatial auto-correlation [83], and Moran's I as implemented in the Seurat package [84]. While various SVG detection methods have been incorporated into the typical workflows and pipelines for spatial transcriptomics data analysis such as the Giotto and Seurat packages, there is a lack of systematic evaluation and comparison of different methods. Essential questions including the degree of agreement among different methods in terms of the ranking and selection of SVGs, the reproducibility of these methods in terms of SVG detection when the genes included in a given dataset changes, the accuracy and robustness of SVG detection, and the utility of the selected SVGs to perform in downstream data analysis such as spatial domain clustering remain to be addressed. In addition, practical considerations such as running time and memory usage required by each method have not been systematically benchmarked.

To fill this critical gap, I systematically evaluated a panel of eight popular SVG detection methods on a collection of 31 spatial transcriptomics and synthetic spatial datasets. These datasets together capture various sample and tissue types and major spatial biotechnologies with different profiling resolutions, including Visium (10x Genomics), ST [85], Slide-seq [86], Slide-seqV2 [87], MERFISH [88], seqFISH+ [3], Stereo-seq [7], SM-Omics [89], and DBit-seq [90]. The results shed light on the performance of each tested SVG detection method in various aspects and highlight some of the discrepancies among different methods, especially on calling statistically significant SVGs across datasets. Taken together, this work provides useful information for considering and choosing methods for identifying SVGs whilst also serving as a key reference for future development of SVG detection methods from spatial transcriptomics data.

2.3 Materials and methods

2.3.1 SVG detection methods

Datasets were filtered by first removing cells whose top-50 highly expressed genes contributed to 50% of the total counts and then removing genes that were expressed in fewer than 30 cells. Log normalisation of raw counts was performed prior to SVG detection as per the recommended default for each method. The same reproducible seed was set prior to running each method.

Giotto KM and Giotto rank

Giotto requires a spatial Delaunay triangulation network to be built on reduced dimensions to represent the spatial relationships. Then, statistical enrichment using Fisher's exact test of binarized expression in spatial nearest neighbours is performed to determine SVGs. The two methods differ in their binarization method. In Giotto KM, expression values for each gene are binarized using k-means clustering (k=2), otherwise simple thresholding on rank is applied in Giotto rank (default = 30%). Thus, a gene is considered an SVG if it is highly expressed in neighbouring cells. Normalisation was performed using *normalizeGiotto()* under default parameters. SVG detection was thus performed with two different approaches: k-means and rank using *binSpect(bin_method="kmeans")* and *binSpect(bin_method="rank")* respectively,

following the author's tutorial. https://rubd.github.io/Giotto_site/articles/mouse_visium_kidney_200916.html

Moran's I

Moran's I ranks genes by the observed spatial autocorrelation [91, 92] to measure the dependence of a feature on spatial location. Weights are calculated as 1/distance. Raw counts were first normalised using *SCTransform()*. Using Seurat v4.1.1, SVGs were detected using *FindSpatiallyVariableFeatures(selection.method = "moransi")* and statistics for all features were returned. p-value adjustment was manually performed using the BH method.

MERINGUE

MERINGUE identifies SVGs using neighbourhood adjacency relationships and spatial autocorrelation. MERINGUE first represents cells as neighbourhoods using Voronoi tessellation. Then, the resulting Delaunay-derived weighted adjacency matrix and a matrix of normalised gene expression is used to calculate Moran's I. Raw counts were CPM-normalised using *scuttle::normalizeCounts()* and the default filtering distance was used to generate the weighted adjacency matrix. Statistics and p-values for all features were returned. p-value adjustment was manually performed using the BH method.

nnSVG

nnSVG is based on scalable estimation of spatial covariant functions in Gaussian process regression using nearest neighbour Gaussian process (NNGP) models. The BRISC algorithm [93] was used to implement the NNGP model and obtain maximum likelihood parameter estimates for each gene. A likelihood-ratio test is performed to rank genes by estimated LR statistic values. Log normalisation was performed using *scater::LogNormCounts()* prior to running *nnSVG()* with default parameters (k=10). Where default parameters were unsuccessful, the number of nearest neighbours was fine-tuned from k=5 to k=15.

SOMDE

With SOMDE, a SOM neural network is used to adaptively integrate nearest neighbour data into different nodes, achieving a condensed representation of the spatial transcriptome. SVGs are identified on a node-level, using spatial location and gene meta-expression information. A squared exponential Gaussian kernel is applied to generate log-likelihood ratio values wherein

a likelihood-ratio test is performed to rank genes by estimated LLR statistic values. The procedure was performed as per the recommended tutorial at https://github.com/WhirlFirst/somde using Python. k=10 was chosen as the default nearest neighbours when constructing the SOM across all benchmarking datasets to preserve local spatial patterns across both small and large datasets. Where default parameters were unsuccessful, the number of nearest neighbours was fine-tuned from k=5 to k=20.

SPARK-X

SPARK-X is a non-parametric method that relies on a robust covariance test framework, including the Hilber-Schmidt independence criteria test and the distance covariance matrix test. A test statistic is observed by measuring the similarity between two relationship matrices based on gene expression and spatial coordinates respectively. A p-value is computed for each distance covariance matrix constructed and a Cauchy combined p-value is reported. *sparkx()* was run under default parameters.

SpatialDE

SpatialDE fits a linear mixed model for each gene with gaussian kernels and decomposes the gene variation into spatial or non-spatial variation. The non-spatial variation is separately modeled using observed noise, and the spatial variation is explained by an exponential covariance function. For each gaussian kernel, a p-value is calculated from the likelihood test to rank genes by estimated LLR statistics. SpatialDE was run under the Python implementation and the procedure was as follows in the tutorial by the authors as in https://github.com/Teichlab/SpatialDE.

2.3.2 Ranking and identifying significant SVGs

To calculate the pairwise Spearman correlation between each method for each dataset, the corresponding gene statistics were used as outlined in Table 2.1. Where a comparable gene statistic was not reported by a method, the -log10(adjusted p-value) was used to rank the genes.

Significant SVGs were typically defined as genes with an adjusted p-value of < 0.05. Specifically for Moran's I, genes that had a positive spatial autocorrelation coefficient and an adjusted p-value of < 0.05 were selected as significant.

Table 2.1. Description of statistics used to rank SVGs. Statistics used to rank genes and the p-value adjustment methods used by each package.

SVG Method	**Statistic**	**p-value adj.**
Giotto k-means	-log10(adjusted p-value)	BH
Giotto rank	-log10(adjusted p-value)	BH
MERINGUE	observed coefficient	BH
Moran's I	observed coefficient	BH
nnSVG I	LLR statistic	BH
SOMDE I	LLR statistic	q-value
SPARK-X	-log10(adjusted p-value)	BY
SpatialDE	LLR statistic	q-value

2.3.3 Dependency across genes

To assess the dependency across genes in SVG analysis, I randomly down-sampled 50% of genes from all datasets that ran successfully. I next applied each SVG detection method and calculated SVG statistics of remaining genes in the down-sampled dataset as per Table 2.1. The relative rank of these genes was compared with their rank in the original full dataset to assess if there is any change of relative ranking when other genes were included in the dataset. Methods that lead to a different ranking of SVGs in the down-sampled dataset when additional genes are included are considered as calculating spatial variability of a gene depending on the presence and absence of other genes.

2.3.4 Robustness against sparsity

To assess how each method performs against sparse data, I randomly down-sampled 80% of spots from all datasets that ran successfully. After applying each SVG detection method, I evaluated the performance of each method in two aspects. To assess the impact of sparsity

on the relative rankings of the gene statistics, I computed the Spearman correlation of the original dataset and the down-sampled dataset using the statistics reported in Table 2.1. To assess the extent of sparsity on the significantly detected SVGs, I visualised the proportion of uniquely detected SVGs because of the subsampling against the total number of SVGs significantly detected in the original dataset.

2.3.5 Simulation of spatial transcriptomics data

To evaluate the capacity of methods to detect SVGs with high sensitivity and specificity, we simulated a set of spatial transcriptomics data using scDesign3 [94], providing us with ground-truth SVGs. The synthetic data were generated using real spatial transcriptomics datasets from 2.2. Simulation of realistic spatial transcriptomics data was performed following the default settings of scDesign3 [94]. To enable fast computation of the model parameters estimated from the real data, we simulated up to approximately 2000 genes and for each dataset generated 10% of all genes as spatially variable. The synthetic datasets model parameters from nine datasets from seven independent studies that cover different sequencing technologies (Visium and DBiT-seq), tissue histologies (breast cancer, brain, embryo, and cancer), number of spatial spots (369-4895 spatial spots) and sequencing depths (590-1937 genes and 59-194 SVGs).

2.3.6 Benchmarking of simulation studies

To evaluate the performance of the SVG detection methods on the simulated data, we calculated the receiver operating characteristic curve based on the statistics or p-values of the genes, indicating the capacity of methods to rank true SVGs before non-variable ones. We next calculated the true positive rate (TPR) and the false discovery rate (FDR) to evaluate FDR control at six adjusted p-value thresholds (1e-100, 1e-50, 1e-10, 0.01, 0.05, and 0.1) for each simulated dataset. The *cutpointr* package [95] was used to calculate the TPR and FDR performance metrics.

2.3.7 Clustering and concordance quantification

To quantify the utility of SVGs in spatial domain clustering, we used varying number of top significant SVGs (between 100 and 1900 genes) reported from each method to subset the expression matrix, compute principal component analysis, and performed clustering on the

top 20 principal components to cluster the E9.5 mouse embryo spatial transcriptomics data into 13 tissue domains based on the original annotation [7]. We performed 10 repeats by random subsampling of the spatial data to 80% of the total number of spatial spots for each repeat. We performed either spatial clustering using the default settings (unless otherwise stated) of BayesSpace [96] (gamma = 2 and nrep = 1000) and SpaGCN [97] or k-means, hierarchical, Louvain, and Leiden clustering. The total number of clusters was set to the total number of spatial domains observed in the data. In particular, we performed a binary search to tune the resolution parameter as described in SINFONIA [98] to tune the clustering in the two community-based clustering algorithms. To assess the clustering performance of the SVGs defined by various SVG detection methods, we used the adjusted rand index (ARI), the normalized mutual information (NMI), the Fowlkes-Mallows index (FMI), and purity to evaluate the concordance between the clustering labels and the spatial domains. Each metric was calculated as follows:

Adjusted Rand Index

Let T denote the known ground-truth spatial domains of spots, P denote the predicted clustering labels from k-means clustering, N denote the total number of spatial locations, x_i denote the number of spots assigned to the ith cluster of P, y_j denote the number of spots that belong to the jth unique label of T, and n_{ij} denote the number of overlapping spots between the ith cluster and the jth unique label. The Rand index (RI) denotes the probability that the obtained clusters and the spatial domain labels agree on a randomly chosen pair of spots. The adjusted Rand index (ARI) adjusts for the expected agreement by chance.

$$ARI = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{x_i}{2}\sum_j \binom{y_j}{2}\right]/\binom{N}{2}}{\frac{1}{2}\left[\sum_i \binom{x_i}{2} + \sum_j \binom{y_j}{2}\right] - \left[\sum_i \binom{x_i}{2}\sum_j \binom{y_j}{2}\right]/\binom{N}{2}}$$

Normalised Mutual Information Normalised mutual information (NMI) assesses the similarity between the obtained cluster labels and the ground-truth spatial locations, scaled between 0 and 1. The NMI was calculated as follows:

$$NMI = \frac{\mathbf{MI}(P,T)}{\sqrt{H(P)H(T)}}$$

where $H(\cdot)$ is the entropy function.

A comparison of ARI and NMI presented in previous studies [99, 100] suggest ARI is preferred when there are large equal-sized clusters, whilst NMI is preferred in the presence of class imbalance and rare clusters.

Fowlkes-Mallows Index

The Fowlkes-Mallows index (FMI) measures the similarity between two clustering results and is defined as the geometric mean of the precision and recall. The FMI is calculated using the following equation:

$$FMI = \sqrt{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}$$

where TP is the number of true positives, which are pairs of spots that are in the same spatial domain in both the true and predicted labels; FP is the number of false positives, which are pairs of spots that are in the same cluster in the predicted clusters but in different clusters in the ground-truth labels; and FN is the number of false negatives, which are pairs of spots that are in the same cluster in the ground-truth labels but in different clusters in the predicted clusters. The score is adjusted to a range between 0 and 1, where a value of 1 signifies that all the spatial spots are correctly labelled. A higher FMI denotes a greater similarity between the two clustering results.

Purity

Purity is scored in terms of whether the clusters contain only spots of the same spatial domain. Purity equals 1 if all the spots within the same cluster correspond to the same spatial domain. The purity score is computed using the following equation:

$$Purity = 1 - \frac{H(T|P)}{H(T)}$$

where $H(T|P)$ indicates the uncertainty of true labels based on the predicted labels.

2.3.8 Time consumption and memory usage

To measure computational consumption for each method, a standard virtual machine with 16 OCPUs and 256 GB was used. Where methods offered parallelization (Giotto, SPARK-X,

nnSVG, SOMDE, and SpatialDE) and when it was possible to specify, all available cores were utilised to record the running time. For all methods run in R, the elapsed time to run each method was evaluated using the *system.time()* function. The peak memory usage was monitored using *gc()*. For methods run in Python, *perf_counter()* from the time package was used to record the elapsed time. To record the peak memory usage, *get_traced_memory()* was used from the *tracemalloc* package.

2.3.9 Code availability

SVG detection methods were run on R (v4.3) or python (v3.8) and the source code is deposited in Zenodo (https://zenodo.org/doi/10.5281/zenodo.10295502) and is freely available from https://github.com/PYangLab/SVGbench.

2.3.10 Data sources

Summary information of spatial transcriptomics datasets was included in Figure 2.2. Below I provide the accession numbers when available or download links used to obtain each dataset.

- Liu et al., DBiT-seq [101]. Mouse Embryo E12 (GSM4189614_0628cL) and E11 (GSM4364243_E11-2L). Downloaded from GEO accession GSE13798629.
- Xia et al., MERFISH [88]. Human osteosarcoma. Downloaded from the supplementary section of the corresponding paper. https://www.pnas.org/doi/suppl/10.1073/pnas.1912459116/suppl_file/pnas.1912459116.sd12.csv
- Eng et al., SeqFISH+ [3]. Mouse primary visual cortex (VISp). Downloaded from https://github.com/CaiGroup/seqFISH-PLUS. The spatial coordinate of each spot was generated using 'stitchFieldCoordinates' function in Giotto.
- Rodriques et al., SlideseqV1 [102]. Mouse cerebellum. Downloaded the 'Puck_180819_11' sample from https://singlecell.broadinstitute.org/single_cell/study/SCP354/slide-seq-study30.
- Marshall et al., SlideseqV2 [87]. Human kidney cortex. Downloaded the 'HumanKidney_Puck_20011308' sample from https://cellxgene.cziscience.com/datasets.

- Stickels et al., Slide-seqV2 [103]. Mouse hippocampus. Downloaded the 'Puck_200115_08' sample from https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatial-transcriptomics-at-near-cellular-resolution-with-slide-seqv232.
- Vickovic et al., SM-Omics [104] Mouse brain cortex. Downloaded the '10015CN78_C1_stdata_adjusted' and '10015CN89_D2_stdata_adjusted' samples from https://singlecell.broadinstitute.org/single_cell/study/SCP979/sm-omics-an-automated-platform-for-high-throughput-spatial-multi-omics33.
- Ji et al., ST [105]. Human squamous carcinoma. Downloaded from GSM428432234.
- Navarro et al., ST [106]. Mouse hippocampus wild-type replicate 1. Downloaded from https://data.mendeley.com/datasets/6s959w2zyr/136.
- Biancalani et al., Visium [107]. Mouse primary motor cortex. Downloaded from https://storage.googleapis.com/tommaso-brain-data/tangram_demo/Allen-Visium_Allen1_cell_count.h5ad
- Ferreira et al., Visium [108]. Mouse kidney. Downloaded the Sham model and ischemia reperfusion injury model from GSE17140639.
- Hunter et al., Visium [109]. Zebrafish melanoma. Downloaded the 'Visium-A' sample from GSE15970941.
- Janosevic et al., Visium [110]. Mouse kidney. Downloaded from GSE15410743.
- Joglekar et al., Visium [111]. Mouse pre-frontal cortex. Downloaded from GSE15845045.
- Lopez et al., Visium [112]. Mouse lymph node and MCA205 tumour. Downloaded from GSE17377647 and GSE17377348 respectively.
- McCray et al., Visium [113]. Human prostate. Downloaded from GSM483776750.
- Wu et al., Visium [114]. Human breast cancer. https://zenodo.org/record/4739739#.YY6N_pMzaWC52
- E9.5 Mouse Embryo [7]. E9.5 mouse embryo spatial profile. Downloaded from https://db.cngb.org/stomics/mosta/.

2.4 Results

2.4.1 Evaluation framework and data summary

I designed an evaluation framework (Figure 2.1) to gain insight into the performance of different SVG detection methods to call SVGs from a collection of real and simulated spatially resolved transcriptomics datasets.

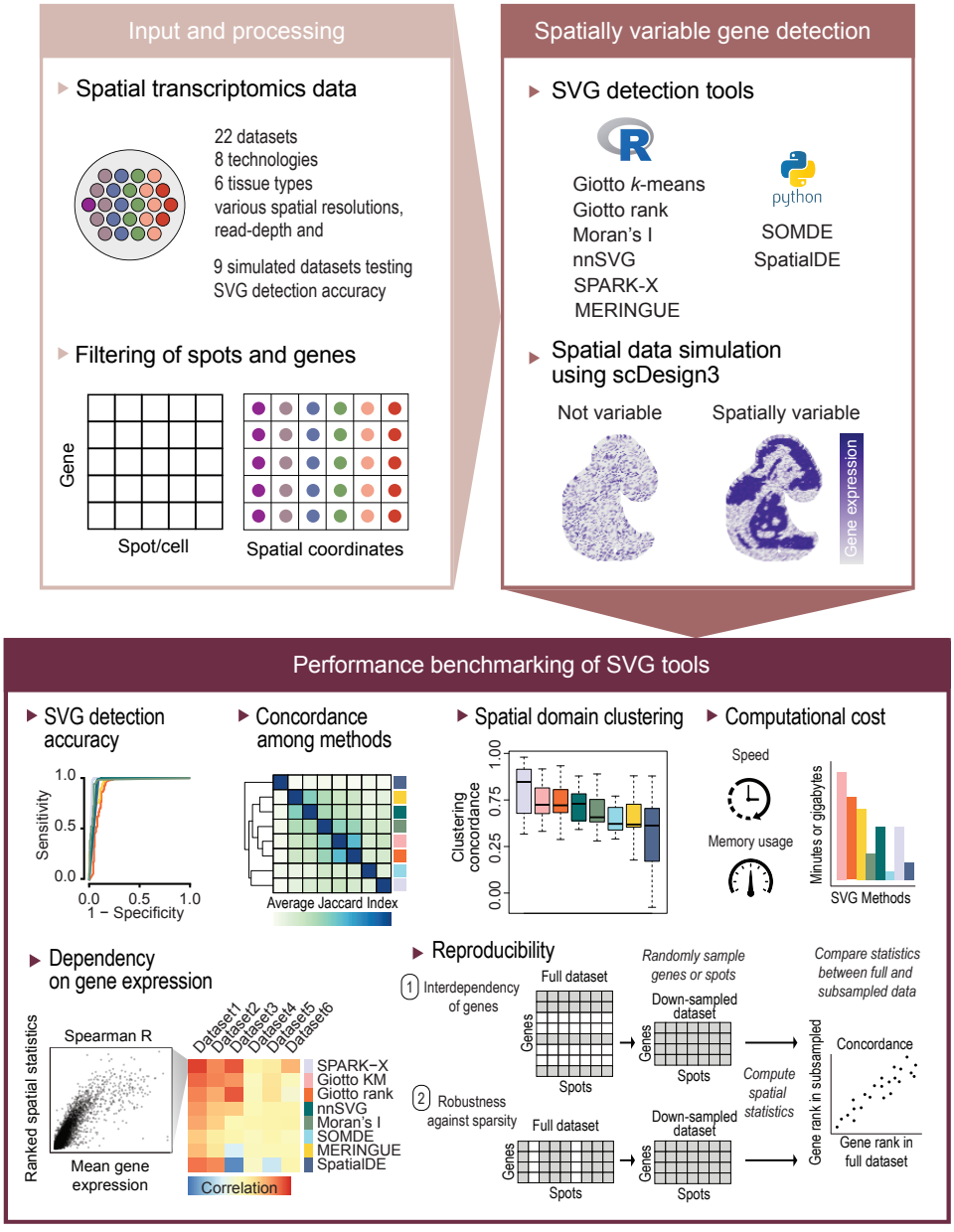

Figure 2.1. Schematic summary of the evaluation framework.

These include spatial transcriptomic data with varying sequencing depths generated from a wide range of spatial profiling platforms, species, tissue types, and spatial resolutions (Figure 2.2, Figure 2.3A-B).

First, I compared the concordance between the overall rankings of the SVGs between SVG tools and evaluated their dependence on mean gene expression to assess the variability among methods and their capacity to account for the bias between gene expression and variance. Next, I investigated the capacity of each SVG method to reproducibly rank SVGs independently of the pool of genes observed in the dataset or with induced sparsity in spots, to call ground-truth SVGs from synthetic spatial data, and to define SVGs required to accurately cluster spatial domains. Finally, using the spatial benchmarking datasets I compared the computational cost in terms of speed and memory required for SVGs to be called by each method.

Reference and platform	Category	Species	Tissue	# Location	# Gene	Source
Liu et al., DBiT-seq, D1	Sequencing	Mouse	Embryo E12	1,615	12,591	GSE137986
Liu et al., DBiT-seq, D2	Sequencing	Mouse	Embryo E11	1,640	12,197	GSE137986
Xia et al., MERFISH	Imaging	Human	Osteosarcoma	645	12,903	Supplementary file
Eng et al., seqFISH+	Imaging	Mouse	Brain cortex	524	10,000	Zenodo#2669683
Rodriques et al., Slide-seqV1	Sequencing	Mouse	Cerebellum	1,574	7,885	Broad Institute
Marshall et al., Slide-seqV2	Sequencing	Human	Kidney cortex	8,794	12,474	CELLxGENE Discover
Stickels et al., Slide-seqV2	Sequencing	Mouse	Hippocampus	3,870	10,658	Broad Institute
Vickovic et al., SM-Omics, D1	Sequencing	Mouse	Brain cortex	830	8,990	Broad Institute
Vickovic et al., SM-Omics, D2	Sequencing	Mouse	Brain cortex	707	6,404	Broad Institute
Ji et al., ST	Sequencing	Human	Squamous carcinoma	1,140	17,823	GSM4284322
Navarro et al., ST	Sequencing	Mouse	Hippocampus	602	18,008	Mendeley#6s959w2zyr.1
Biancalani et al., Visium	Sequencing	Mouse	Primary motor cortex	2,177	31,053	Google Cloud
Ferreira et al., Visium, D1	Sequencing	Mouse	Kidney sham	1,833	13,586	GSE171406
Ferreira et al., Visium, D2	Sequencing	Mouse	Kidney ischemia	2,050	13,734	GSE171406
Hunter et al., Visium	Sequencing	Zebrafish	Melanoma	1,548	32,268	GSE159709
Janosevic et al., Visium	Sequencing	Mouse	Kidney sepsis	1,888	13,710	GSE154107
Joglekar et al., Visium	Sequencing	Mouse	Prefrontal cortex	3,017	15,440	GSE158450
Lopez et al., Visium, D1	Sequencing	Mouse	Lymph node	369	13,948	GSE173776
Lopez et al., Visium, D2	Sequencing	Mouse	MCA205 tumour	2,125	15,976	GSE173773
McCray et al., Visium	Sequencing	Human	Prostate	722	36,601	GSM4837767
Wu et al., Visium, D1	Sequencing	Human	Breast cancer	4,781	15,325	Zenodo#4739739
Wu et al., Visium, D2	Sequencing	Human	Breast cancer	4,821	15,901	Zenodo#4739739

Figure 2.2. Summary information of the spatial transcriptomics datasets used for evaluating concordance, statistical significance, and reproducibility of SVG detection methods. Dataset reference table, summarising basic information and source of each dataset.

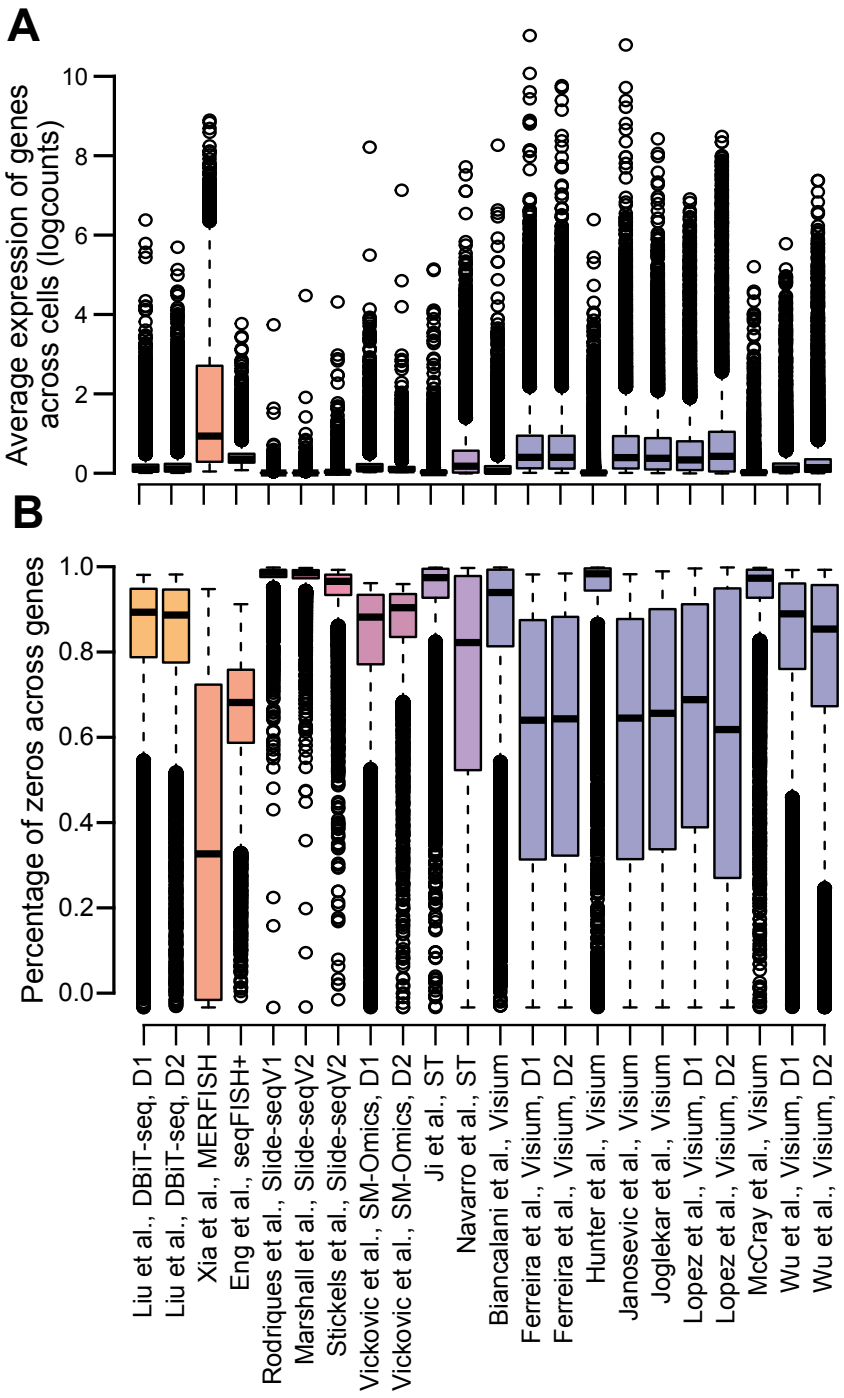

Figure 2.3. Quality of the spatial transcriptomics datasets used in the study. (A, B) Boxplot of mean gene expression and proportion of zero of gene across cells in each dataset.

2.4.2 Concordance among SVG detection methods

To quantify the degree of agreement amongst the different SVG detection methods, I first obtained the ranking of genes in each dataset ordered from the most to least spatially variable based on the statistics reported by each method (Table 2.1) and correlated the SVG rankings from each pair of methods. These correlation results were summarised for each SVG detection method with respect to other methods across the spatial datasets in Figure 2.4 and visualised individually in Figure 2.5.

The overall concordance results showed two groups of methods that highlighted an average similarity (measured as the Spearman correlation of SVG statistics) of > 0.8 across the spatial datasets Figure 2.4A). The most correlated pair of methods were Giotto K-means and Giotto rank, as expected, because of a large overlap in their framework to perform spatial network enrichment. The next group of correlated methods were MERINGUE, Moran's I, and nnSVG. SOMDE, SPARK-X, and SpatialDE showed the least concordance with the other methods, suggesting the prioritisation of SVG statistics by these methods, in particular SpatialDE, are different to other methods.

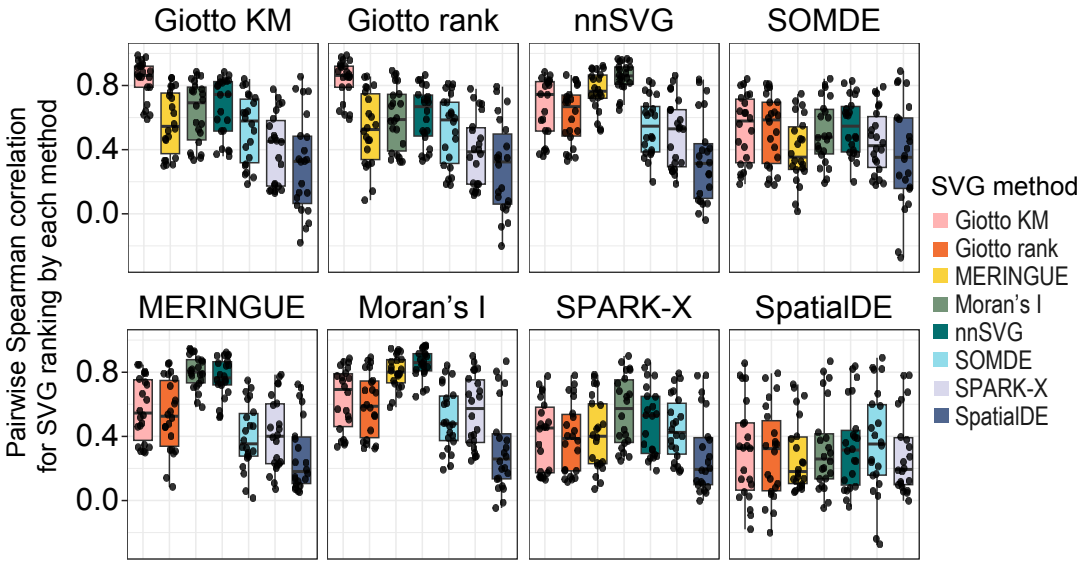

Figure 2.4. Concordance of SVGs detected by different methods. Concordance of SVG rankings reported from each SVG detection method. Each panel uses one SVG detection as an anchor and the y-axes are pairwise Spearman correlation coefficient for quantifying concordance in ranking of each pair of SVG detection methods. Points in each boxplot represent the result from a dataset. Boxplot centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 times the interquartile range.

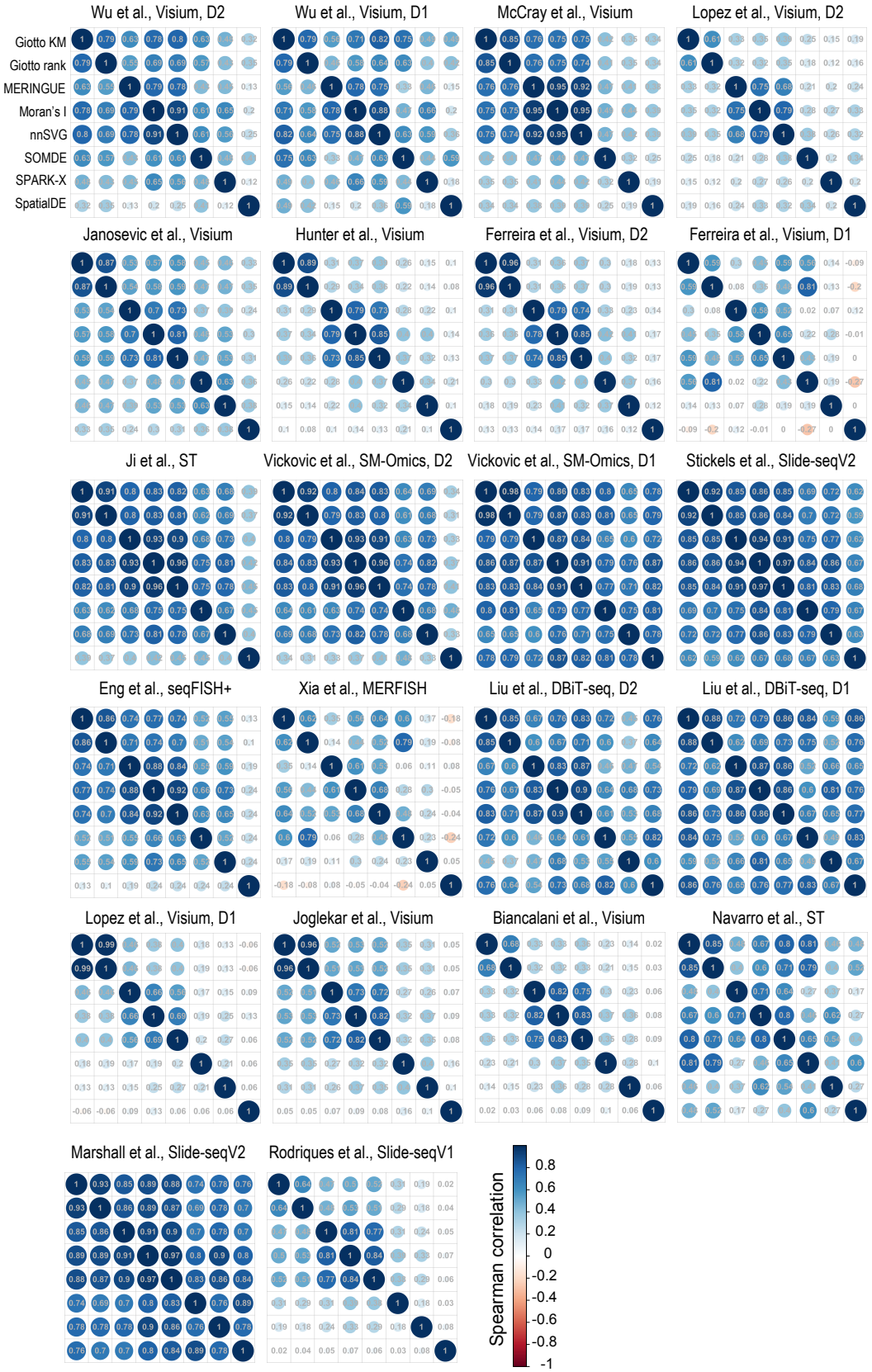

Figure 2.5. Pairwise correlation of SVG rankings reported by each method for individual spatial transcriptomics datasets.

The choice to rank the genes based on transformed raw p-values, Benjamini Hochberg-adjusted p-values, or test statistics had negligible impact on most methods as there was an observed linear relationship between ranked p-values and ranked test statistics Figure 2.6A, B). However, I found it was necessary to rank Moran's I based on the observed coefficient as genes with positive spatial autocorrelation would be highly ranked, whereas the adjusted p-values exhibit a symmetric relationship at the two extremities Figure 2.6C).

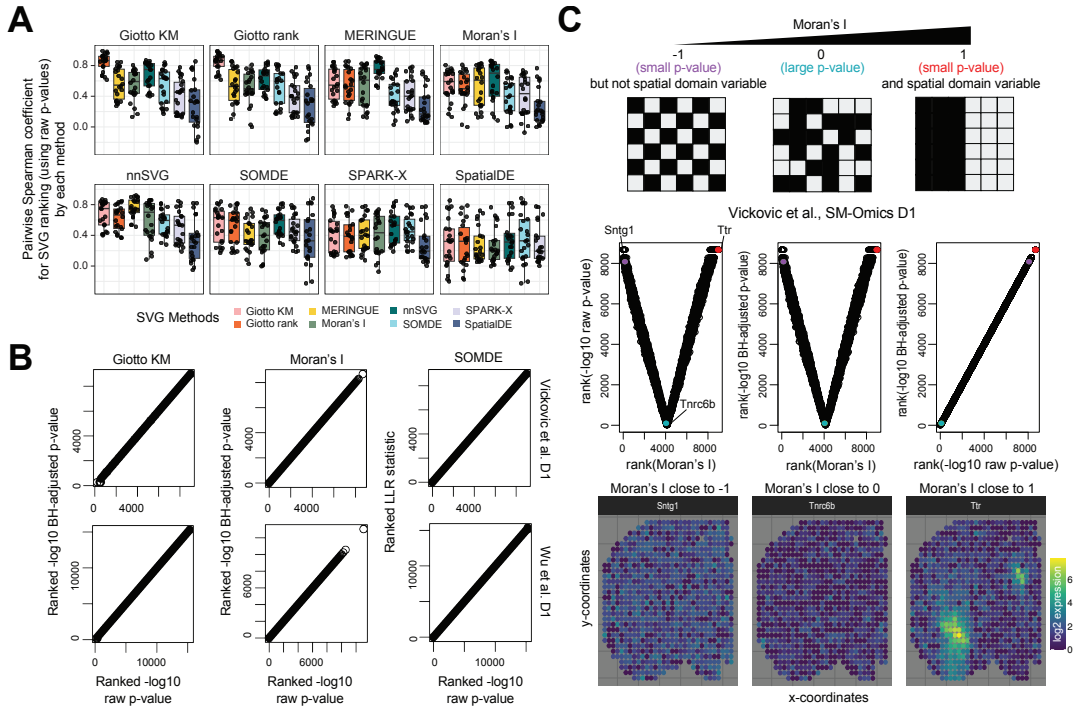

Figure 2.6. Comparison of different gene statistics to rank SVGs. (A) Concordance of SVG rankings using raw-pvalues reported from each SVG detection method. (B) SVG statistics from Giotto KM, Moran's I and SOMDE from two representative datasets compared against their transformed raw p-values. (C) Top, Schematic of Moran's I spatial autocorrelation. Middle, Scatter plots of ranked Moran's I statistic versus transformed raw p-values and BH-adjusted p-values with three genes representing distinct spatial relationships highlighted an example dataset. Bottom, Visualisation of the gene expression pattern of the three representative genes.

Amongst the methods, I observed that SpatialDE demonstrated the highest variability across datasets. Colouring the data points in Figure 2.4A by the total number of spatial spots and technology (Figure 2.7A, B) revealed an interesting trend, which was most striking in SpatialDE, where despite an overall low correlation in spatial statistics with all other methods a high correlation was observed in specific datasets derived from the 10x Visium platform. These results demonstrate that whilst I observed moderate-to-high correlation between SVG

detection tools in terms of SVG ranking, I found considerable variability of reported SVG statistics across the computational methods, platforms, and datasets.

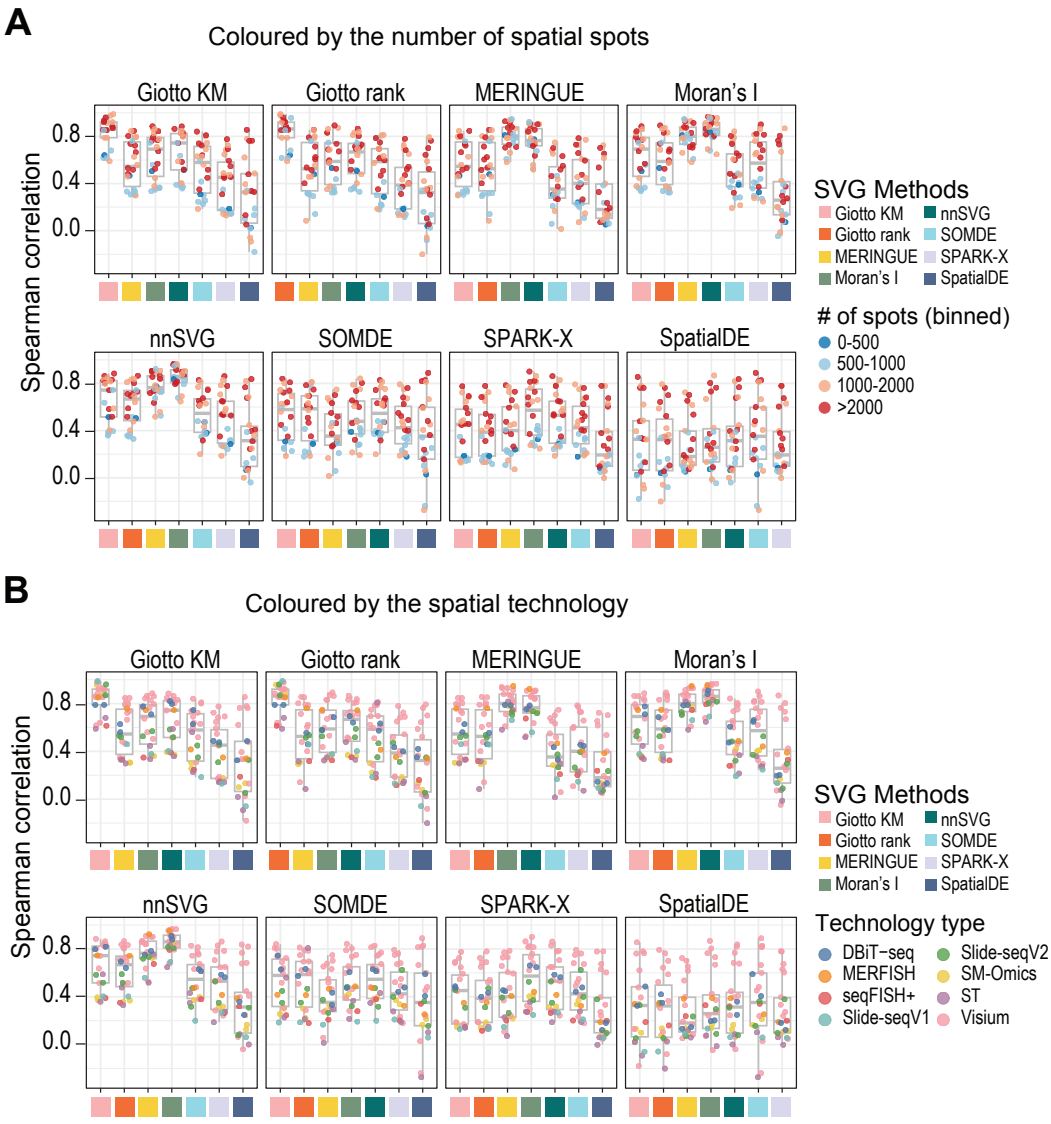

Figure 2.7. Boxplot of correlations of SVG rankings reported by each method against all other methods. Each dot denotes the results for a spatial transcriptomics dataset. The dots are coloured (A) by the total number of spatial spots in the dataset or (B) by the spatial technology platform.

While the ranking of SVGs is useful for selecting the top candidates for subsequent analysis, in practice, statistical significance such as p-values is frequently used for selecting SVGs. To this end, I first partitioned the SVGs into three categories (i.e., $p = 0$; $0 < p \leq 0.05$; $p > 0.05$) based on the adjusted p-value reported from each computational method (Figure 2.8A).

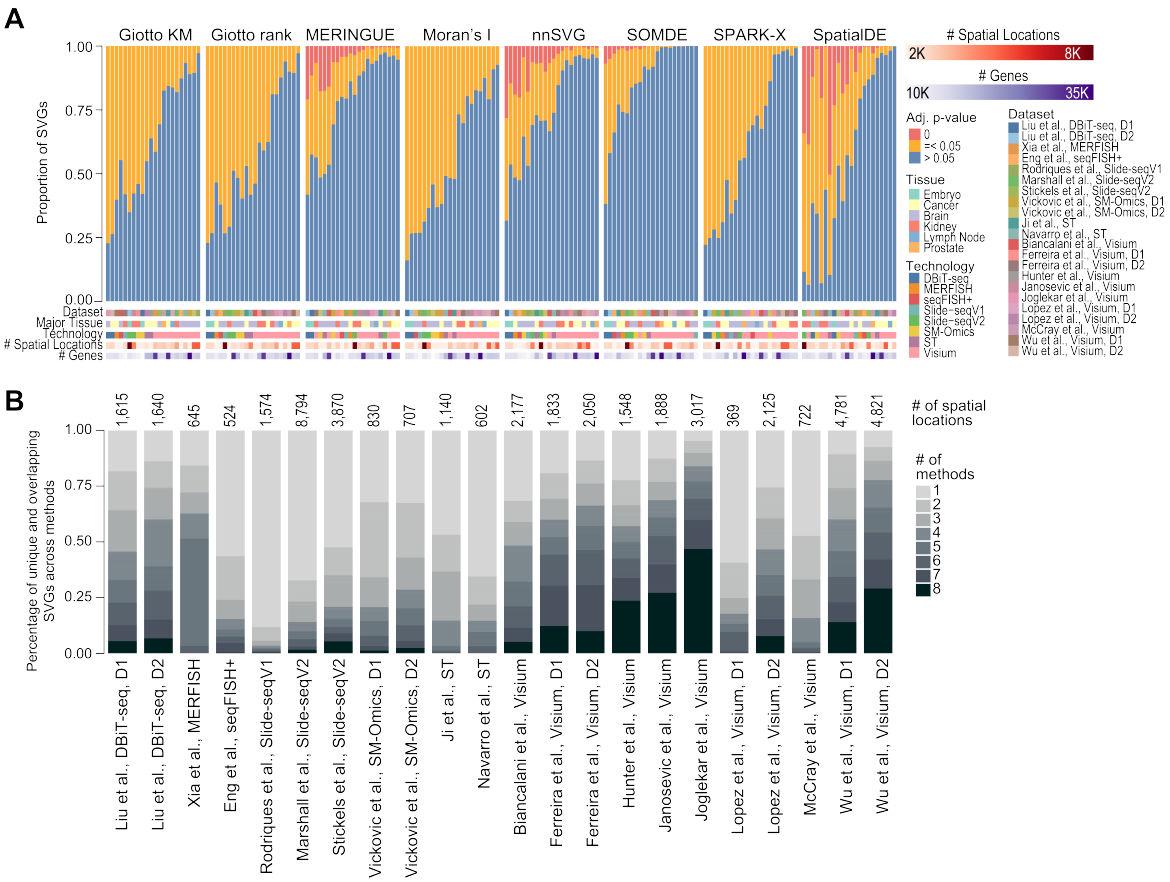

Figure 2.8. Statistical significance, and overlap of SVGs detected by different methods. (A) Statistical significance of SVGs detected by each method. SVGs are partitioned into three categories based on the adjusted p-values reported by each method (i.e., $p = 0$; $0 < p \leq 0.05$; $p > 0.05$) and presented as a percentage (y-axis). The datasets are ordered in terms of the decreasing proportion of genes observed in the orange category. The colour bars denote various characteristics of the spatial dataset including the tissue type, spatial technology, number of spatial locations, and total number of genes expressed. (B) A proportional bar plot showing the percentage of unique (# of method = 1) and overlapping (# of method > 1) significant SVGs (adjusted p-value ≤ 0.05) reported by the SVG detection methods for each spatial transcriptomics dataset.

I found that most methods report a large proportion of SVGs at an adjusted p-value threshold of 0.05 on many datasets. Among the eight methods, nnSVG, MERINGUE, and SpatialDE, and to a lesser degree SOMDE, reported a sizable proportion of SVGs with an adjusted p-value of 0. Interestingly, SOMDE reported on average the fewest number of significant SVGs with some datasets having almost no significant SVGs (Figure 2.8A and Figure 2.9). Intriguingly, I observed that despite the high correlation in SVG statistics (Figure 2.4A), different methods predicted a vastly differing number of SVGs as significant using a p-value threshold of 0.05.

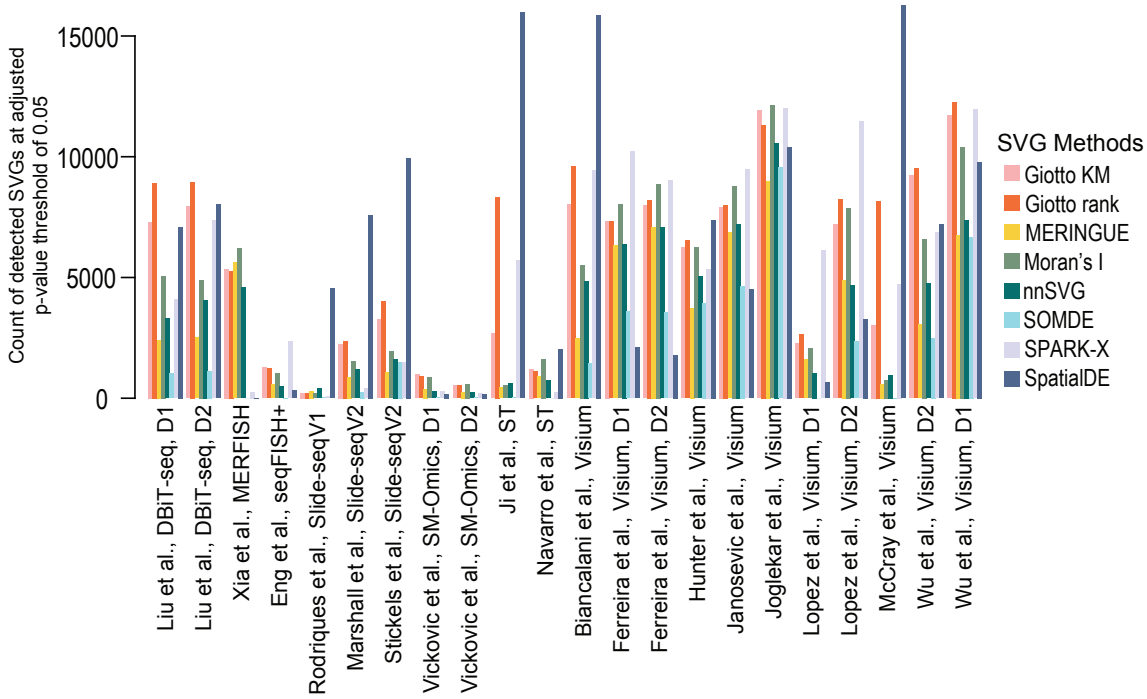

Figure 2.9. Number of statistically significant SVGs reported for each spatial transcriptomics dataset. Bar plot quantifying the number of statistically significant SVGs at an adjusted p-value threshold of 0.05 reported by each method for each dataset.

However, I note that the overall pattern between methods is still similar when I computed the average concordance in gene sets of the top 200, 500, 1000, and all significant SVGs across all the datasets between methods (Figure 2.10). As before, SpatialDE demonstrated the least similarity against all other methods, followed by SPARK-X and SOMDE (Figure 2.10). Giotto KM and Giotto ranks demonstrated a high similarity, but this time Moran's I's gene sets tended to show a higher concordance with the Giotto methods rather than MERINGUE and nnSVG, suggesting that whilst the overall ranking in gene statistics may be similar between Moran's I and MERINGUE and nnSVG, the top most significant SVGs identified by Moran's I appear to be more similar to those of the Giotto methods (Figure 2.10). Importantly, despite the relatively high correlation in SVG statistics observed between methods, the number of SVGs found by all methods are strikingly low with many datasets having close to no overlapping SVGs across all eight computational methods (Figure 2.4C). In addition, many unique genes were found by various individual methods in most datasets (Figure 2.4C). Together, these findings highlight the discrepancy among methods when an adjusted p-value threshold of 0.05 was used for calling statistically significant SVGs.

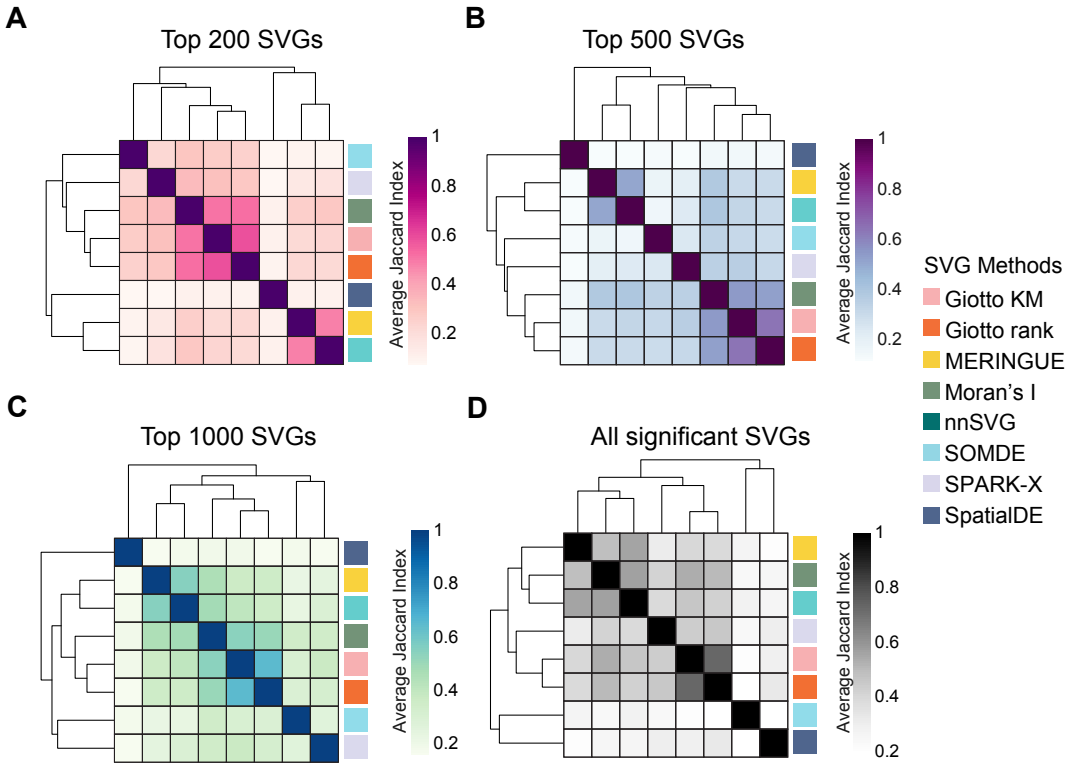

Figure 2.10. Heatmaps of the overlap of SVGs reported by each method for each spatial transcriptomics dataset. The overlap calculated as the Jaccard Index was computed for the (A) top 200 SVGs, (B) top 500 SVGs, (C) top 1000 SVGs, and (D) all the significant SVGs (adjusted p-value <= 0.05) reported by each method for each spatial transcriptomics dataset. The average Jaccard Index across dataset and then plotted as a clustered hierarchical heatmap, where a higher Jaccard Index (darker colour) denotes a greater overlap of SVGs between methods and a lower Jaccard Index (lighter colour) denotes a lower overlap. The discrete colour bars denote the different SVG methods.

2.4.3 Dependency of SVG statistics on gene expression levels

In scRNA-seq data, it is known that variance in gene expression is positively correlated with gene expression level; therefore, most HVG detection methods implement procedures to account for this bias [75]. To test whether methods designed for SVG detection have the tendency to select genes with higher expression levels, I investigated the correlation between mean gene expression and the SVG statistics for each method and dataset pair. I found that indeed the rankings of SVGs from most methods correlated positively with the mean gene expression (Figure 2.11A, B). In particular, SPARK-X showed average correlations of around 0.8 across the datasets (Figure 2.11C), and the Giotto methods and nnSVG showed correlations of around 0.5 across the datasets, suggesting a high dependency of SVG ranking on gene expression for these methods.

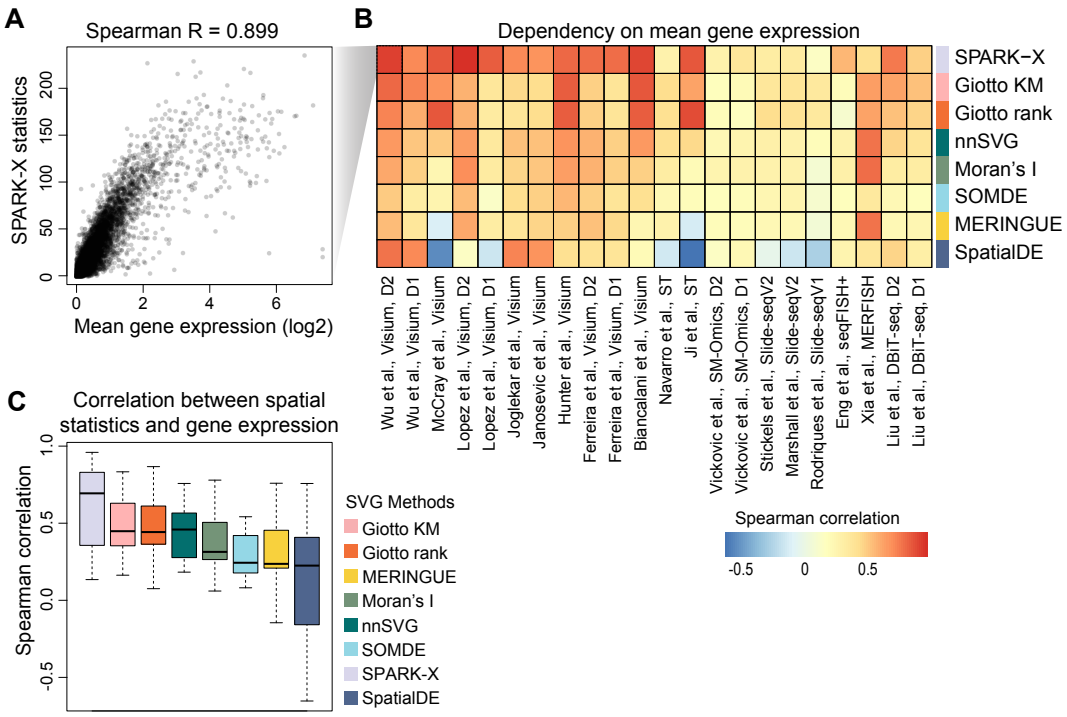

Figure 2.11. Dependency of SVG statistics on gene expression level. (A) An example showing the positive correlation between SVG statistics reported from SPARK-X and mean gene expression across cells in the 'Wu et al., Visium D2' dataset [114]. (B) Heatmap summarising the Spearman correlation of SVG statistics reported by each method and the mean gene expression in each dataset. The rows are ordered from the highest to lowest average dependency. (C) Boxplot of Spearman correlation of SVG statistics reported by each method and the mean gene expression across the spatial transcriptomics datasets.

I also correlated the proportion of zeros in gene expression across cells against SVG ranking for each method (Figure 2.12A, B). Since the proportion of zeros is known to be negatively correlated with their expression levels, the negative correlation observed among each method and dataset pair further confirms the dependency that was found between SVG ranking and gene expression among current SVG detection methods.

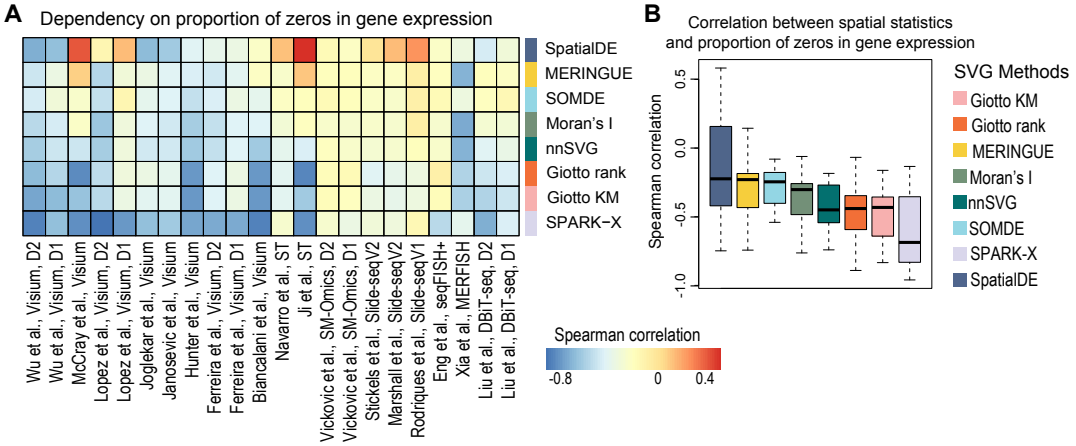

Figure 2.12. Relationship between SVG statistics and proportion of zero of genes. (A) Heatmap summarising the Spearman correlation of SVG statistics reported by each method and the proportion of zero of genes across cells in each dataset. (B) Boxplot of Spearman correlation of SVG statistics reported by each method and the proportion of zero of genes across the spatial transcriptomics datasets.

2.4.4 Dependency of SVG statistics across genes and spatial spots

Next, I assessed the reproducibility of gene ranks based on the SVG statistics reported from each method when either the number of genes or the total number of spatial spots included in a dataset changes. To this end, I randomly down-sampled the genes in all benchmarking datasets (Figure 2.13A) to 50% and re-calculated the ranks of genes from the reported SVG statistics of each method on the down-sampled datasets. Most methods except for SpatialDE, and to a lesser extent nnSVG and Giotto KM, demonstrate a high fidelity in gene ranks across all datasets. Therefore, the methods that have a lower correlation when the genes included in a dataset change, do not independently calculate the SVG statistics for each gene (Figure 2.13B). Although there is some variability in MERINGUE, SPARK-X, Moran's I, Giotto rank and SOMDE, this variability may not have a significant impact on downstream analysis. These analyses reveal that the decisions made on gene filtering, a common step in data pre-processing, may result in a change in SVG statistics and their ranking for some of the SVG detection methods.

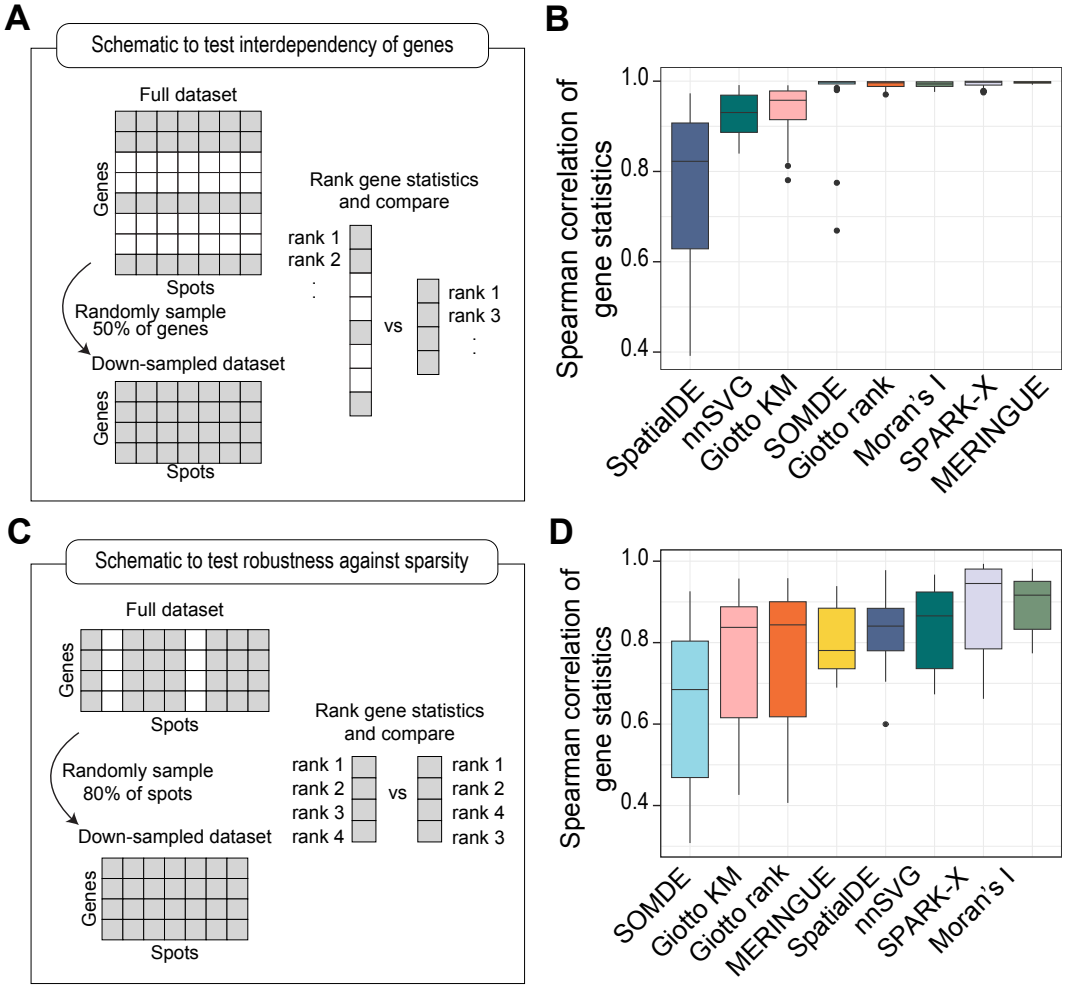

Figure 2.13. Reproducibility of SVG detection tools with down-sampling of the data. (A) Schematic of the experimental approach to determine the interdependency of genes in the calculation of SVG statistics. For the down-sampling, 50% of the genes were randomly down-sampled whilst keeping the number of spots equal. (B) Boxplots of the Spearman correlation results performed on all datasets coloured by SVG method and ordered by increasing mean correlation coefficient. (C, D) as in (A, B) but for the investigation of the reproducibility of SVG methods with increased sparsity in spatial spots. The datasets were randomly down sampled to 80% of the total number of spots.

Each spatial technology has a different capacity to capture spatial locations (Figure 2.2) which may be due to the relatively low-throughput nature of some spatial technologies or inefficiencies in sample preparation. To test the robustness of each method against the sparsity of spatial locations, I down-sampled all datasets to 80% of the total number of spatial spots and repeated the SVG detection (Figure 2.13C). Across all methods, there is some degree of variability in the Spearman correlation amongst datasets due to the induced sparsity

(Figure 2.13D). In particular, I found that the variability amongst datasets and the degree of sensitivity to spot sparsity tend to be greater for methods that rely on neighbourhood adjacency relationships like nnSVG (uses spatial covariance functions in Gaussian Processes using a nearest-neighbour Gaussian process model), SOMDE (uses self-organising map to cluster neighbouring cells into nodes), MERINGUE (uses neighbourhood relationships encoded by a Voronoi Tessellation and Delaunay-derived weighted adjacency matrix), and the Giotto methods (uses a Delaunay triangulation network based on cell centroid physical distances). Conversely, methods that were less sensitive were SPARK-X, SpatialDE, and Moran's I. The reliance on such nearest neighbourhood maps or distance-based networks in the former group of methods may explain the sensitivity to sparsity as it affects the detection of SVGs based on their expression between neighbours in a spatial network.

To investigate the capacity of the methods to correctly identify SVGs and avoid the detection of false positive SVGs with induced down-sampling of the spatial spots, next I quantified the proportion of SVGs that were uniquely identified in the down-sampled data (Figure 2.14A).

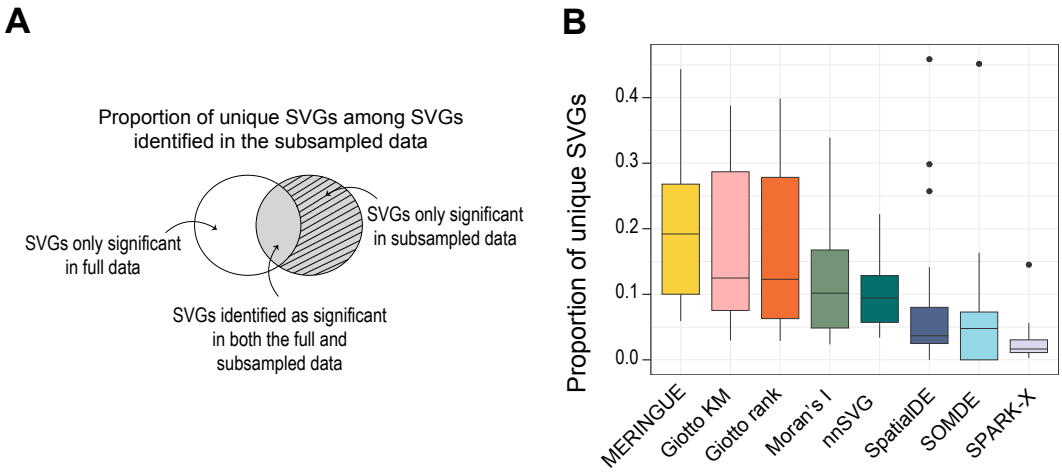

Figure 2.14. False detection of SVGs with down-sampling of the data. (A) Venn diagram illustrates the proportions of uniquely identified SVGs as described in (B). Boxplots of the proportion of false positive SVGs calculated as the proportion of SVGs uniquely identified in the down-sampled data divided by all the significant SVGs identified in the down-sampled data.

It was considered that the original full dataset has the most power to detect SVGs and any significant SVGs that were detected in the down-sampled data but not in the original data were false positives. I visualised the proportions of all significant SVGs identified in the down-sampled data that are either identified as significant in the full data or unique to the down-sampled data (Figure 2.14B). The findings show that SPARK-X, SOMDE and

SpatialDE performed the best in terms of identifying the lowest proportion of false positive SVGs with down-sampling of the data. Although the performance of SOMDE suffered under induced sparsity, the low proportion of false positive SVGs may be explained by the fact that SOMDE tends to select fewer SVGs overall compared to other methods (Figure 2.8A). Again, for most methods there was high variability amongst datasets, which suggests that a method's performance may be dataset-dependent under sparse conditions.

Overall, the down-sampling experiments of genes and spots show that the performance of most methods in detecting significant SVGs may be affected by changes in the gene number and sparsity of spatial spots. This has important implications when considering the most suitable method that is insensitive to gene filtering and dataset quality.

2.4.5 Accuracy of SVG methods in detecting SVGs using synthetic spatial transcriptomics data

To test the accuracy of the SVG detection methods, we next simulated spatial transcriptomics datasets with ground-truth SVGs and spatially invariant genes using scDesign3 [94] (Figure 2.15, Figure 2.16). To enable representation of the diverse sequencing technologies and tissue histologies in real spatial data, we simulated *in silico* data from nine data sources covering nine distinct spatial masks, five tissue histology types, two spatial platform technologies, and diverse sequencing depths (590-1937 genes, 59-194 spatially variable genes, and 369-4895 spatial spots). Then I performed SVG detection on the simulated datasets using the eight methods and evaluated their performance by calculating the true positive rate (TPR) and the false discovery rate (FDR) across three adjusted p-value thresholds (0.01, 0.05, and 0.1).

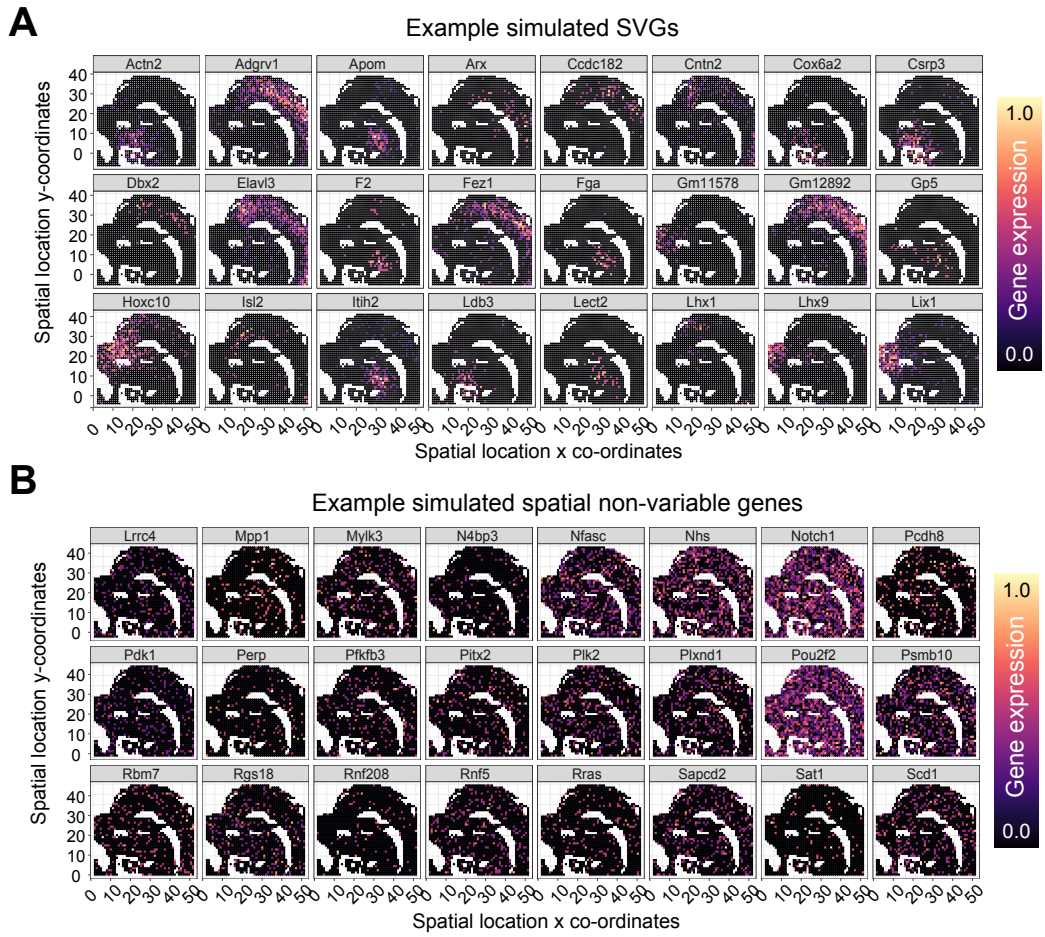

Figure 2.15. Spatial patterns of spatially variant and non-variant genes. spatial expression patterns of example (A) spatially variable genes and (B) spatially non-variable genes of an example simulated spatial transcriptomics data (derived from Liu et al., DBiT-seq, Embryo [90]).

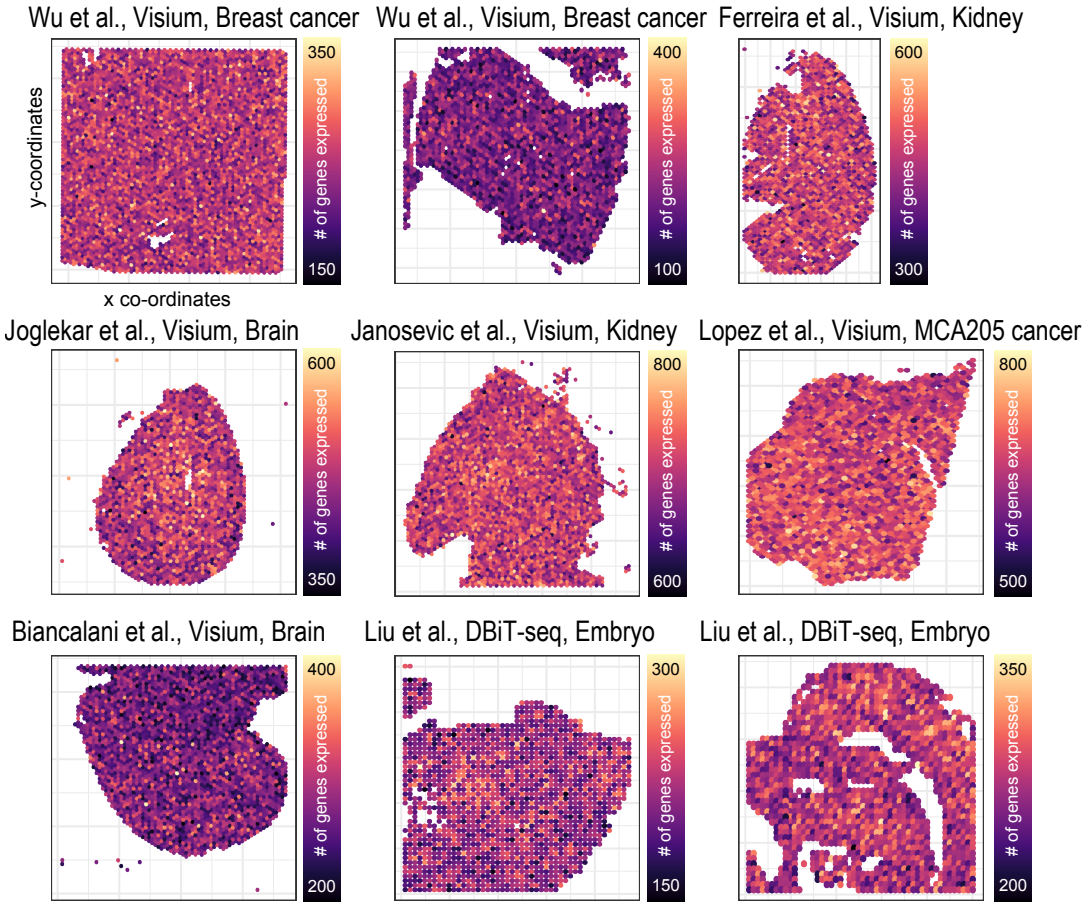

Figure 2.16. Simulation of spatial transcriptomics data. Spatial locations and the total number of genes expressed per spatial spot in nine simulated spatial transcriptomics datasets generated from different sources of data.

At the adjusted p-value thresholds of 0.01 and 0.05, we found that SPARK-X, SOMDE, nnSVG, and SpatialDE performed well with a high TPR and a low FDR (Figure 2.17A and Figure 2.18). Under adjusted p-value thresholds of 0.01, 0.05, and 0.1, Giotto rank, Moran's I, and nnSVG all demonstrated a high TPR but suffered from a high level of false positive identification. Compared to the other methods, the Giotto methods and Moran's I performed relatively poorly in the simulation, displaying the highest FDRs in most datasets (Figure 2.17A). This is alternatively reflected in Figure 2.17B as there was a marked reduction in the proportion of ground truth SVGs detected within the set of statistically significant SVGs at an adjusted p-value threshold of 0.05. However, under a more stringent threshold of 0.01, the Giotto methods and Moran's I maintained a high proportion of correctly detected ground truth SVGs. These methods also tended to identify a greater proportion and number of significant SVGs (Figure 2.19A, B) which contributed to the higher proportion of false detection.

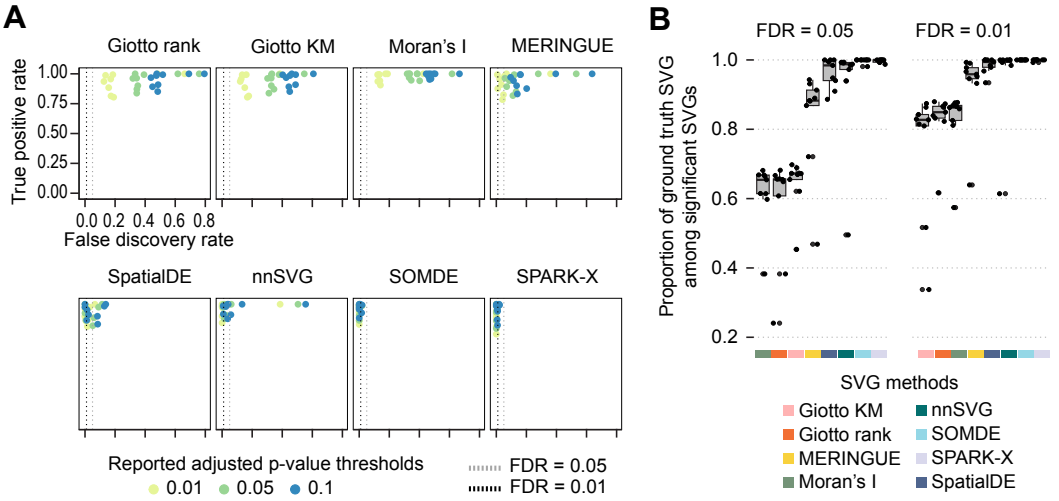

Figure 2.17. Spatially variable gene detection performance across 9 simulated datasets. (a) Scatter plot of observed true positive rate (y-axis) and false discovery rate (x-axis) of spatially variable gene detection by the benchmarked tools at three adjusted p-value cut-offs at 0.1, 0.01, and 0.05. Each dot is colour-coded by the cut-off used. The two horizontal lines represent the true FDRs of 0.01 and 0.05. (b) The proportion of ground truth SVGs among significant SVGs determined by each method at an FDR-adjusted p-value of 0.01.

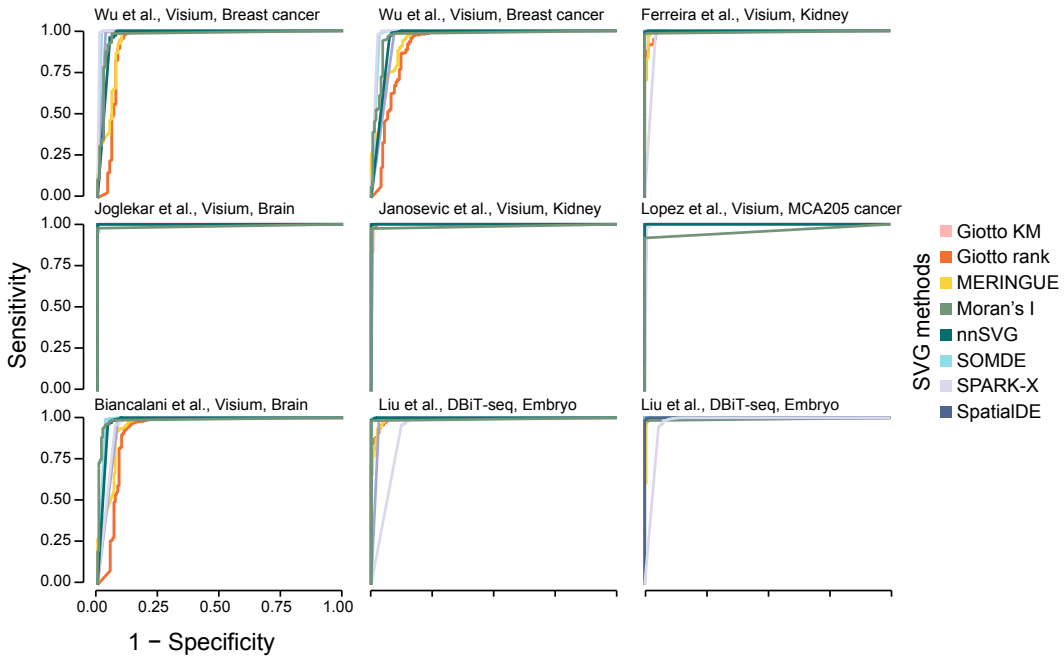

Figure 2.18. ROC curves of spatially variable gene detection. ROC curves were used to assess the capacity of each method to detect spatially variable genes between two spatial domains. The curves are colour-coded by methods. The 1-specificity and sensitivity of spatially variable gene detection are plotted on the x- and y-axes, respectively.

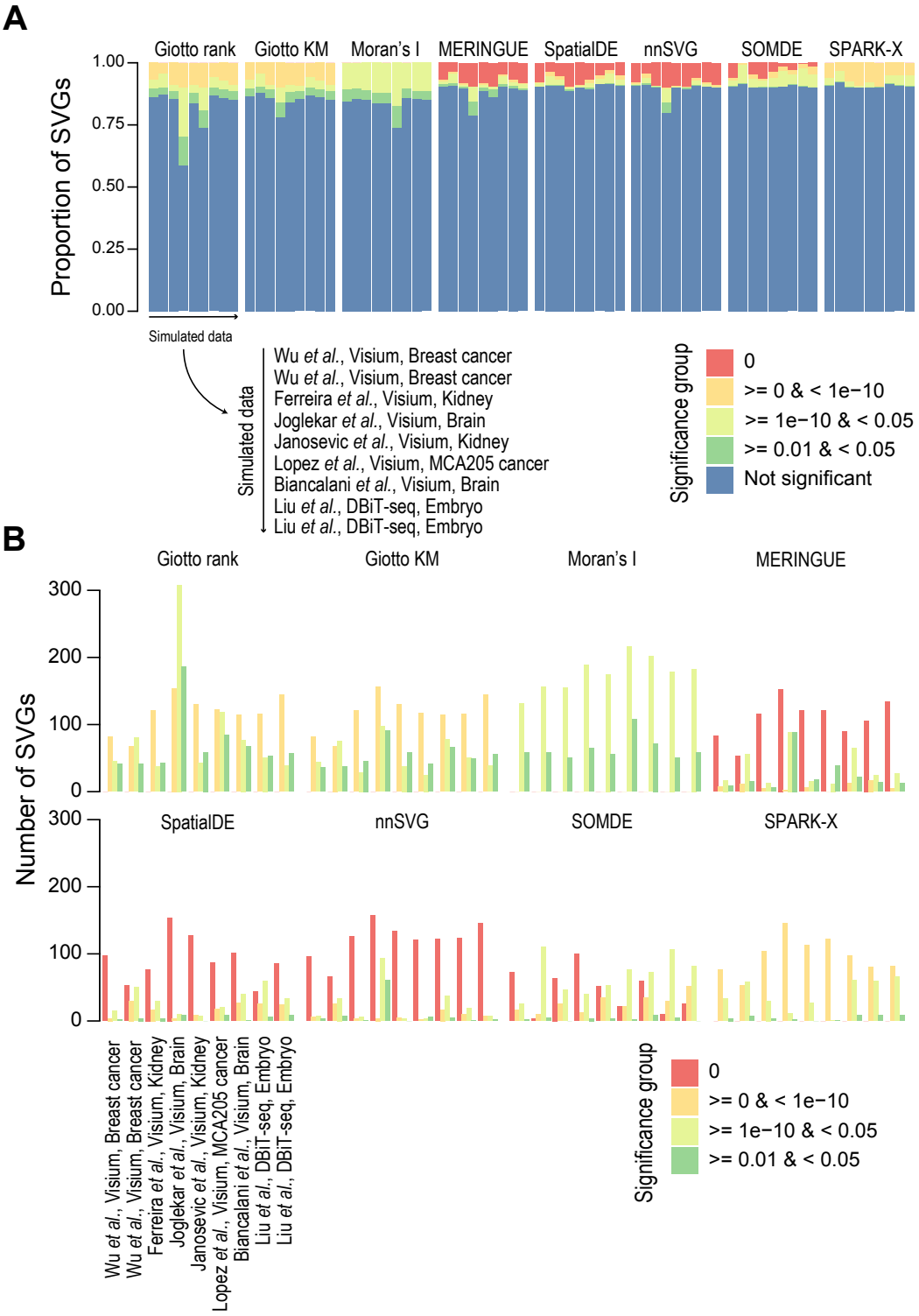

Figure 2.19. Proportion and total number of simulated SVGs detected. (A) Proportion bar plots of SVGs of different statistical significance levels identified by each method. SVGs are partitioned into five categories based on the adjusted p-values reported by each method (i.e., $p = 0$; $0 < p <= 1e\text{-}10$; $1e\text{-}10 < p <= 0.01$; $0.01 < p <= 0.05$; and $p > 0.05$) and presented as a percentage (y-axis). (B) Bar plots of the total number of SVGs reported by each method, for each simulated data, and for each statistical significance level as in (A).

Similar to the trend observed in Figure 2.8A, I observed that MERINGUE, SpatialDE, nnSVG and SOMDE had a higher proportion of significant SVGs reported with an adjusted p-value of 0. These findings reveal that for the methods, except SPARK-X and SOMDE, the estimated FDRs (i.e., adjusted p-value thresholds) do not accurately represent the true FDRs for SVG detection in these simulated datasets. Moreover, some methods lack sensitivity to distinguish between extremely small p-values.

2.4.6 Performance on clustering spatial domains

A key task in spatial transcriptomics data analysis is to identify spatial domains that mark distinctive cell and tissue types in a biological sample. One approach to achieve this is to cluster profiled locations into spatial domains using SVGs. To compare the capacities of SVGs identified by each method in clustering the spatial domains, we took advantage of the spatial transcriptomics data of an E9.5 mouse embryo given the availability of tissue annotations in these samples (Figure 2.20A).

First, I performed SVG calling using each SVG detection method. Then taking a varying number of top SVGs, we computed the top 20 principal components (PCs) using the feature-selected spatial transcriptomics data. Using either spatially-aware clustering tools (BayesSpace [92] and SpaGCN [97]) (Figure 2.20B) or canonical clustering approaches (k-means, hierarchical, Louvain, and Leiden clustering using the SINFONIA framework [98]) (2.20C), we performed clustering on the top 20 PCs and calculated the concordance between the clustering results and the pre-defined spatial domains to measure the performance of the SVGs to delineate the anatomical locations.

By taking a large range in the number of features used (between 100 and 1900 features), we were able to observe an overall increasing trend in performance with an increasing number of SVGs used for all SVG methods, with the accuracy in classification peaking at around 900-1100 SVGs (Figure 2.20B). Whilst this observation was broadly consistent, the pattern differed for some clustering and SVG method combinations. For example, hierarchical clustering demonstrated a decreasing trend in accuracy with increasing number of SVGs used, unlike most clustering methods. The overall pattern was consistent between different concordance measures, including Fowlkes-Mallows index (FMI), normalised mutual information (NMI), and purity score (Figure 2.21, Figure 2.22, Figure 2.23).

Of note, under the default parameters, the spatially-aware methods estimate the expected cell type clusters and in contrast, canonical clustering methods were set or tuned to the number of clusters to match the number of cell types. When I observe the range of concordance across the ARI, NMI, FMI and Purity, I did not find that the spatially-aware methods led to higher concordance scores compared to the canonical methods. This may indicative that the spatial domains in the mouse embryo are most likely clearly discretised based on transcriptomic expression, thus the spatial information may not be contributing strong additive signal.

Overall, these results suggest that whilst the selection of the number of top SVGs used in clustering will depend on the data, using approximately 900-1300 genes for the dataset tested led to the highest accuracy in clustering of spatial domains across most conditions.

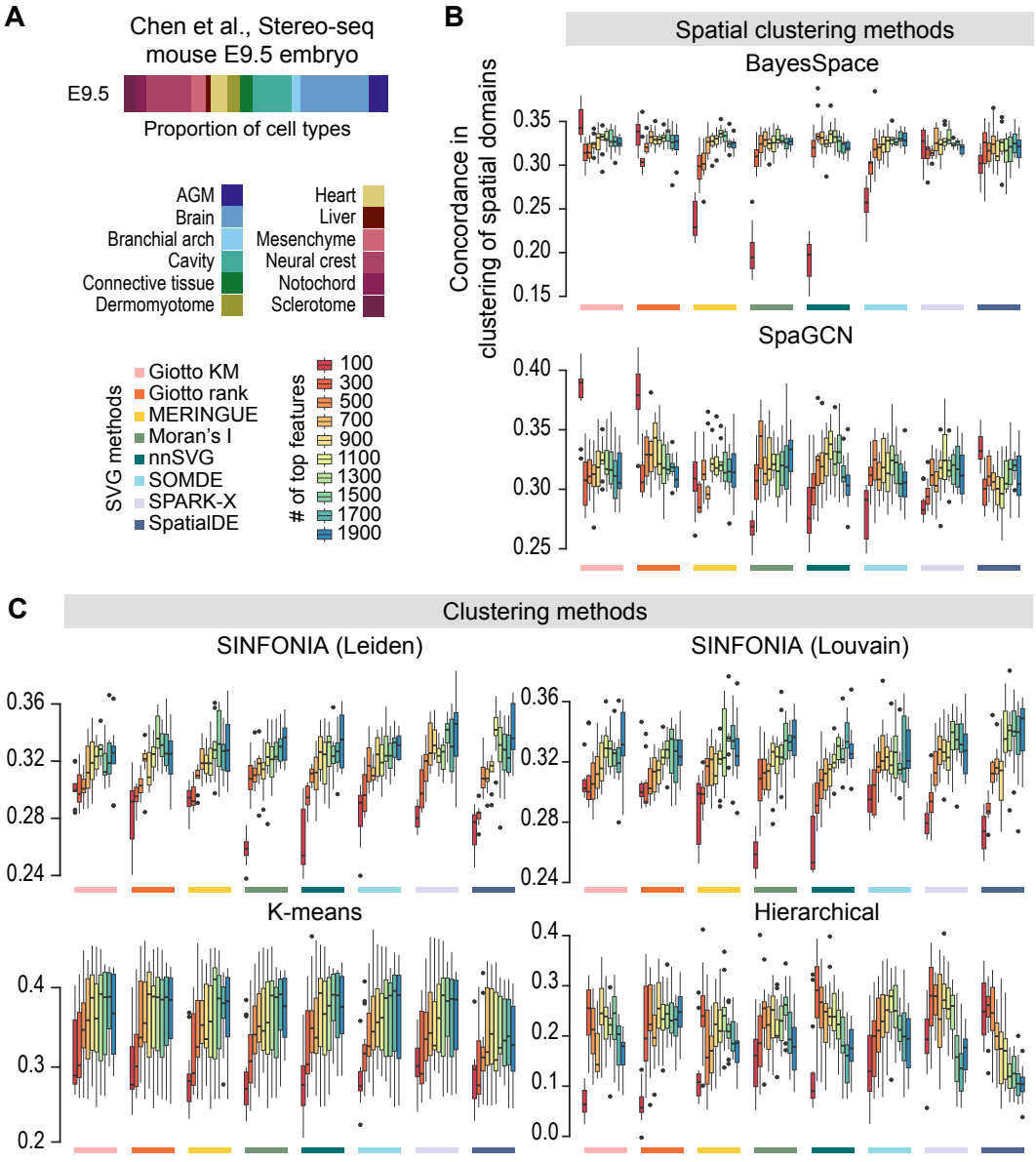

Figure 2.20. Performance of SVGs selected by each method for clustering spatial domains in mouse embryos. (A) Proportion of spatial locations annotated to one of sixteen tissue domains in the E9.5 mouse embryo. Concordance in the clustering outputs and the pre-defined spatial domains in the mouse embryo was computed across a range of top SVGs (between 100 and 1900 genes) selected by each method. Clustering is performed using (B) two spatial clustering methods (BayesSpace and SpaGCN) and (C) four non-spatial clustering methods (SINFONIA's Louvain and Leiden, k-means, and hierarchical clustering). Concordance between the clustering outputs and the pre-defined spatial domains is quantified in terms of the Adjusted Rand Index.

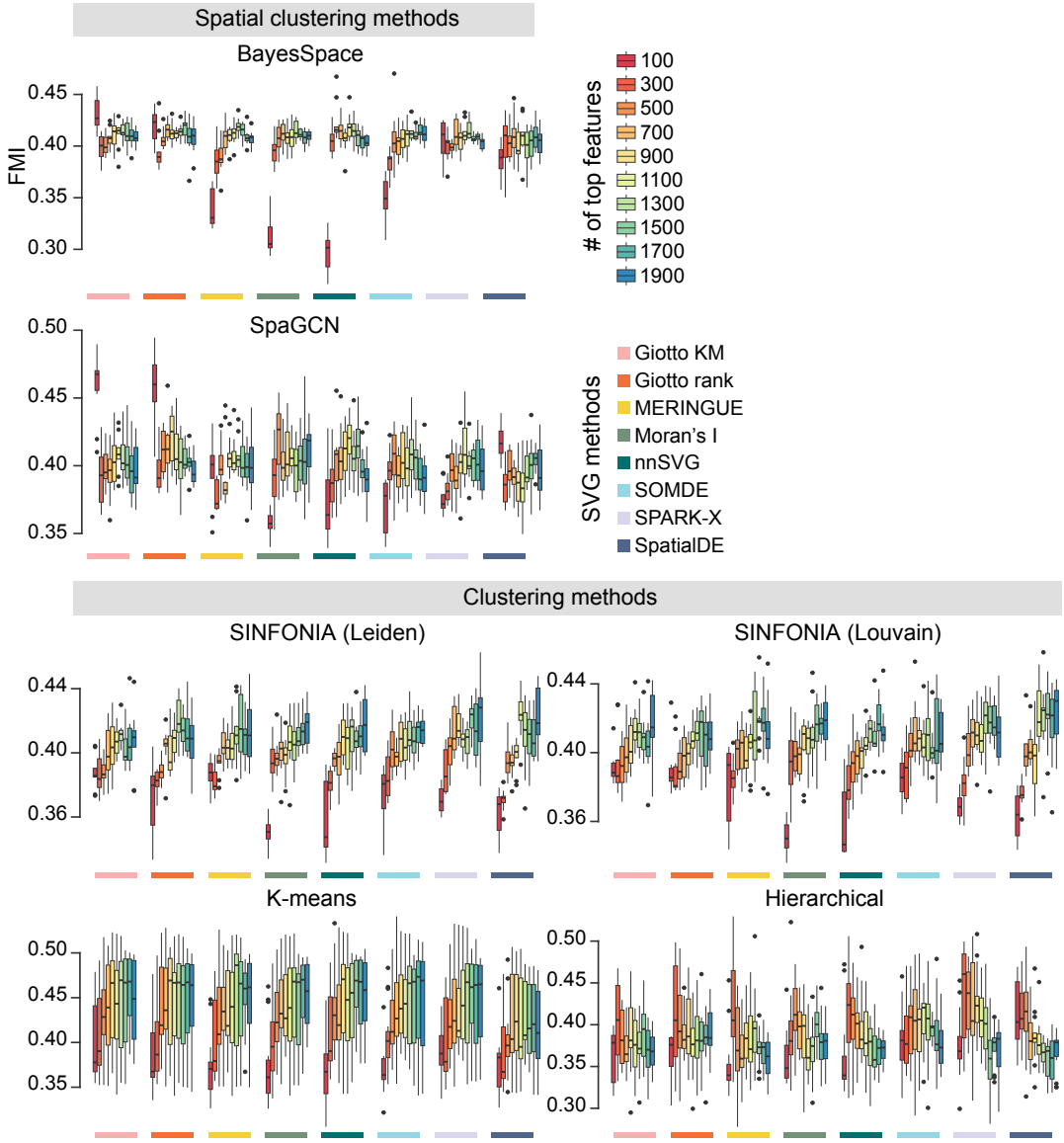

Figure 2.21. FMI concordance SVGs selected by each SVG method for clustering spatial domains in the mouse embryo. Concordance measured by Fowlkes-Mallows index (FMI) in the clustering outputs from spatial clustering methods (BayesSpace and SpaGCN) and non-spatial clustering methods (SINFONIA's Louvain and Leiden, k-means, and hierarchical clustering) and the pre-defined spatial domains in the mouse embryo was computed using across a range of top SVGs (between 100 and 1900 genes).

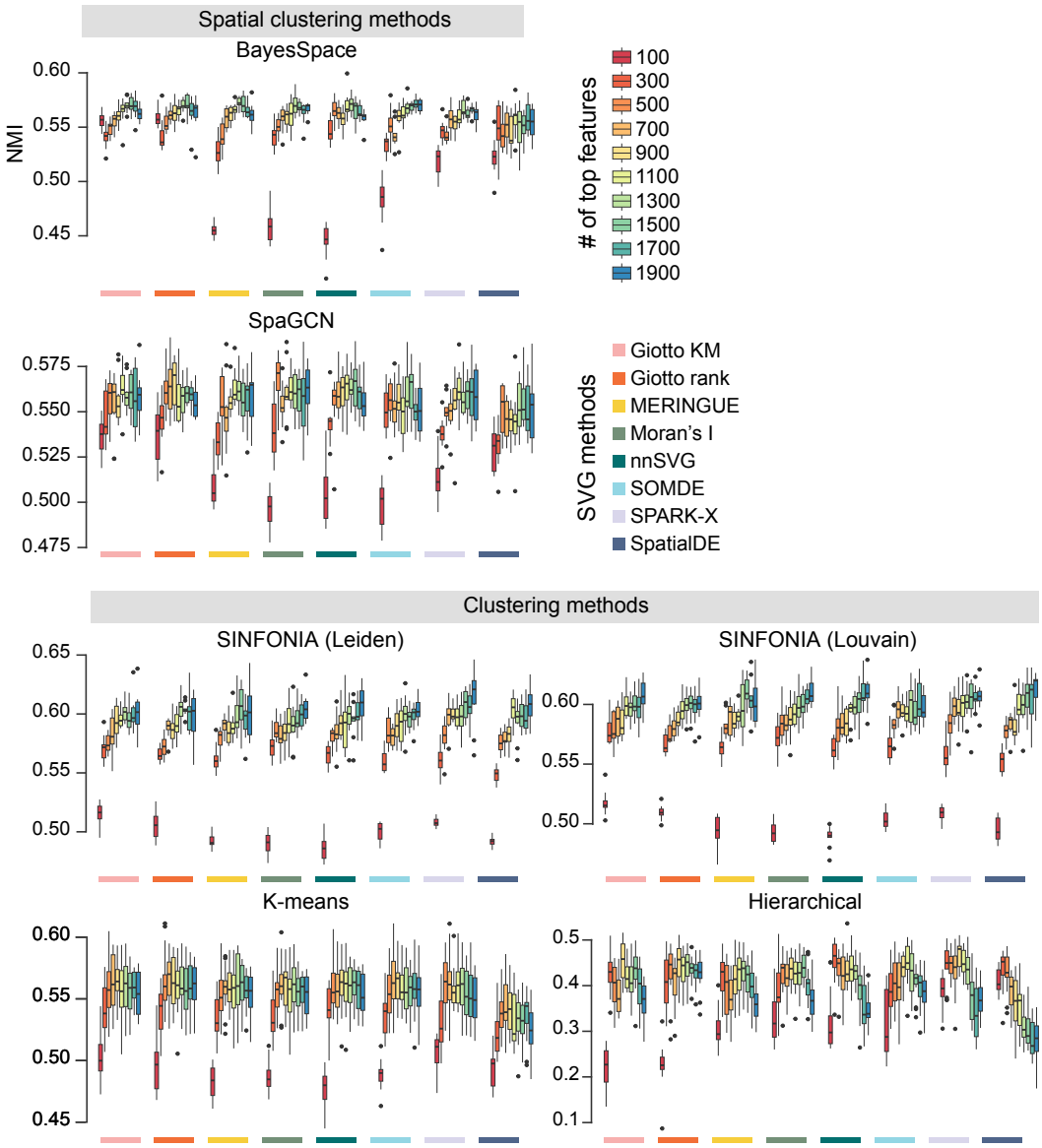

Figure 2.22. NMI of SVGs selected by each SVG method for clustering spatial domains in the mouse embryo. Concordance measured by normalized mutual information (NMI) in the clustering outputs from spatial clustering methods (BayesSpace and SpaGCN) and non-spatial clustering methods (SINFONIA's Louvain and Leiden, k-means, and hierarchical clustering) and the pre-defined spatial domains in the mouse embryo was computed using across a range of top SVGs (between 100 and 1900 genes).

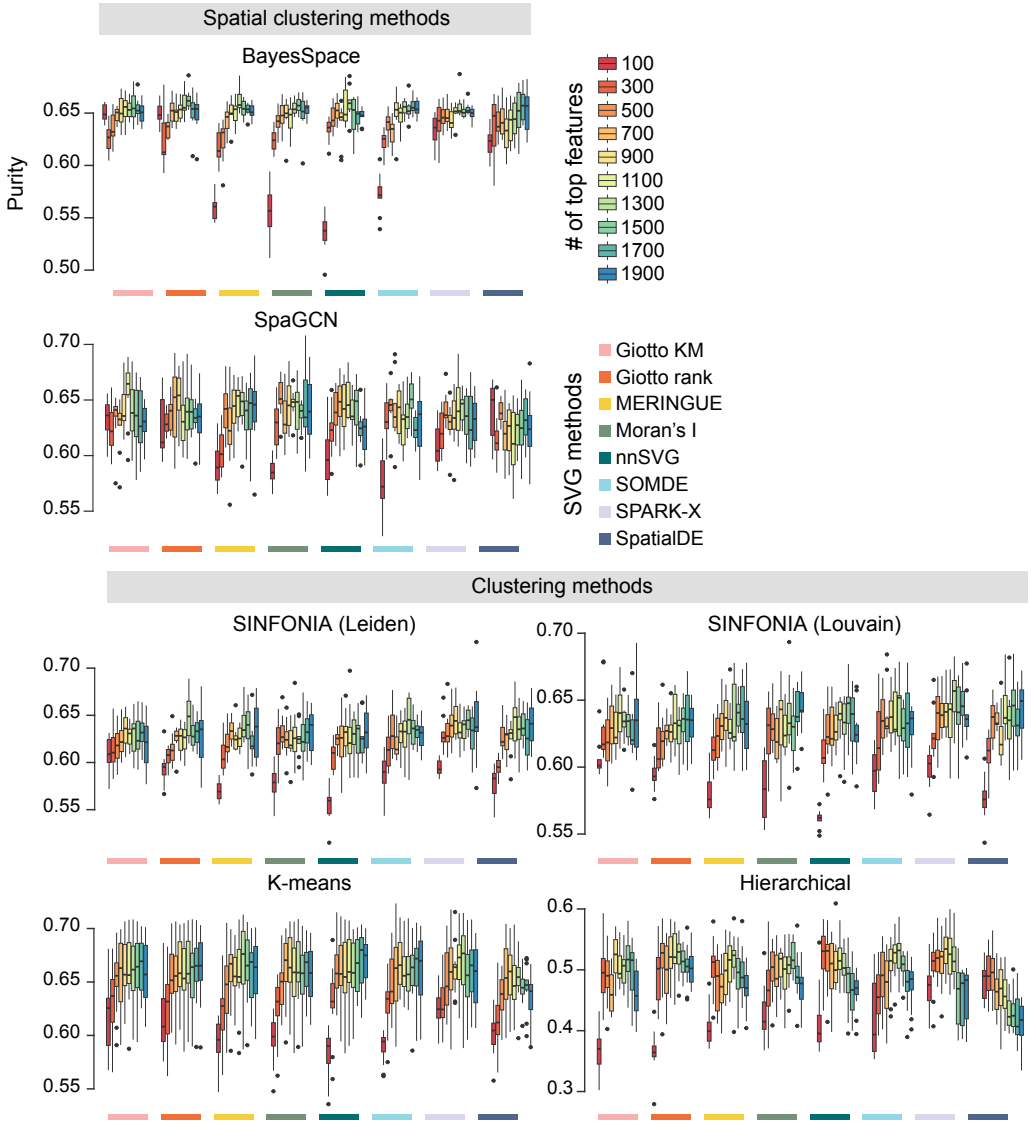

Figure 2.23. Purity concordance of SVGs selected by each SVG method for clustering spatial domains in the mouse embryo. Concordance measured by Purity in the clustering outputs from spatial clustering methods (BayesSpace and SpaGCN) and non-spatial clustering methods (SINFONIA's Louvain and Leiden, k-means, and hierarchical clustering) and the pre-defined spatial domains in the mouse embryo was computed using across a range of top SVGs (between 100 and 1900 genes).

2.4.7 Computational time and memory usage

Computational time and memory usage are key considerations in practical applications, especially for large spatial transcriptomics data analyses. In our evaluation, I configured a standard virtual machine, with 16 OCPUs and 256 GB of memory and recorded the runtime and the peak memory usage for each SVG detection method on each dataset (Figure 2.24).

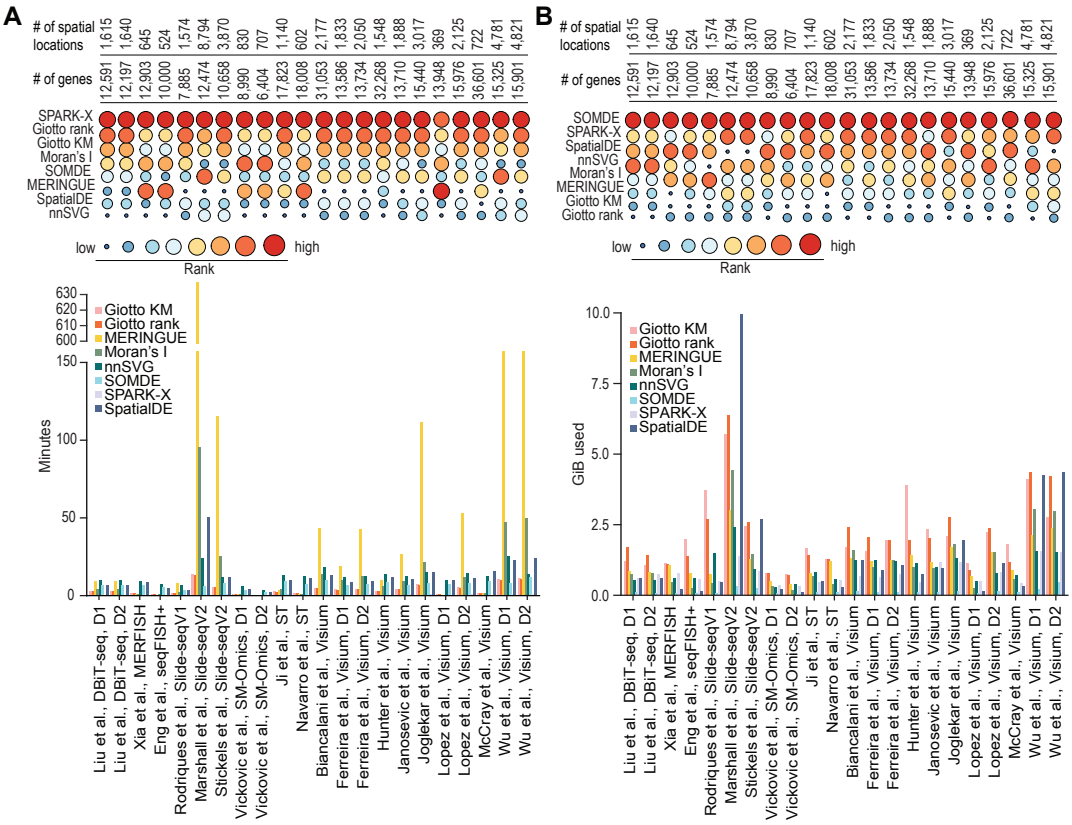

Figure 2.24. Evaluation of computational speed and peak memory usage of SVG detection methods. (A) Computation time in units of minutes and (B) peak memory usage in units of GiB across datasets.

I found the computational time and the peak memory usage were both positively correlated with the number of spatial locations in the datasets. In terms of computational time, comparison across methods revealed that SPARK-X was the fastest method and scaled extremely well with the number of spatial locations. While SOMDE was the second best in most cases, it was significantly slower compared to SPARK-X. In contrast, SpatialDE performed poorer especially on datasets with large numbers of spatial locations. Giotto KM performed poorly in most of the datasets but scaled better than SpatialDE with the number of spatial locations in datasets. Similarly, nnSVG scaled better with the number of spatial locations than SpatialDE

but was slower on datasets with many genes. In terms of peak memory usage, I found that SOMDE used the least peak memory across all datasets and SPARK-X ranked the second in most cases, although it has a significantly higher peak memory usage. In comparison, the two methods implemented in Giotto and SpatialDE showed high peak memory usage especially in datasets with many spatial locations. While there is a trade-off between speed and memory usage, taken together, these results suggest that SPARK-X and SOMDE are the most efficient methods in terms of speed and memory usage for SVG detection.

2.5 Discussion

For most methods, a significant proportion of genes were detected as SVGs under the adjusted p-value of 0.05 in most of the tested datasets (Figure 2.4B and Figure 2.7C). However, the overlaps across the eight methods were relatively small considering the large numbers of SVGs identified from each SVG detection method (Figure 2.8B), suggesting large discrepancies among SVG detection methods when a significance cut-off is used to filter for SVGs. This highlights that it may be difficult to identify a concordant set of statistically significant SVGs across multiple packages. Thus, an alternative approach may involve using gene statistics to find a concordant set of top ranking SVGs. Consistent with this, in our simulation study where ground truth SVGs were introduced into simulated spatial transcriptomics data, I found that for some methods, in particular the Giotto methods and Moran's I, the estimated FDRs did not accurately represent the true FDRs in most of the synthetic datasets (Figure 2.17). These results highlight that the estimation of statistical significance is difficult and there is much room for improvement. It also cautions the reliance on statistical significance from current SVG detection tools when drawing biological conclusions from the data.

I also discovered that SVGs identified by most methods show a strong positive correlation with their expression levels (Figure 2.11). It is of note that a similar relationship was found between gene variability and expression level in scRNA-seq data and most computational methods designed for HVG detection actively correct for such 'bias' [75]. While it cannot be ruled out that genes which vary spatially are also highly expressed, future work should be performed to investigate the biological basis and plausibility for such a correlation. During practical application, it is important to be aware of the tendency of current SVG detection tools to select genes with high expression levels. Future method development will be required to account for this effect such as to retain relatively lowly expressed genes such as transcription factors in

downstream analysis. In addition, I found that for most methods the relative rankings of SVGs changed when different pools of genes and spots were included in the datasets (Figure 2.13C, F). While considering the interdependency among genes may provide useful information for identifying SVGs, it is important to be aware that different SVG detection results may be obtained when different pre-processing criteria are used to filter genes prior to SVG analysis.

Considering the various metrics used to evaluate these SVG methods, dataset dependency should be considered when determining the most suitable SVG method to apply. With sparse spatial technologies, it is recommended to use an SVG method that does not incorporate a network topology such as SPARK-X or Moran's I as other methods will likely be susceptible to inaccurate SVG detection due to incomplete information of spatial neighbourhoods due to sparsity (Figure 2.13D). Based on SVG detection accuracy, SPARK-X, SOMDE, nnSVG and SpatialDE are the top-performing methods. Although high accuracy is desirable, another important consideration is computational resource usage given the ever-increasing scale of spatial transcriptomics data. The computational time of SpatialDE has been reported to scale cubically according to the number of spots in the data [77], and nnSVG has been found in our evaluation to scale according to the number of genes. Therefore, it may not be computationally feasible to apply these methods on spatial technologies which have the capacity to sequence thousands of spots or profile the whole transcriptome, respectively.

Evaluating the performance of SVG methods presented certain caveats in the interpretation of concordance and clustering outcomes. When assessing the concordance of SVG rankings across different methods (Figure 2.4), a large spread in Spearman's correlation values was observed, and the segregation was found to be based on the spatial technology used (Figure 2.7). Consistent with this observation, analysis of the proportion of detected SVGs relative to all genes revealed that the 10x Visium datasets, on average, exhibited higher proportions of detected SVGs (Figure 2.8). This finding may explain the observed variability in correlation values; when a substantial number of SVGs are detected, the ranking of gene statistics is meaningful. However, in cases where few or no SVGs are detected, the rankings become less informative. As the associated statistics or p-values may follow a uniform distribution, leading to random rankings and thus impacting the variability of concordance between SVG methods. In a separate evaluation, the comparison of spatial clustering to canonical clustering methods demonstrated that similar clustering concordance was achieved in spatial domain detection within the mouse embryo. In cases where biological signals are well-defined across spatial

domains, the addition of spatial information may not contribute greatly to the improvement of the clustering performance. Alternatively, to comprehensively assess the potential advantages of spatial clustering methods in leveraging information from neighboring spots, it would be beneficial to test scenarios involving continuous cell states or weak gene expression signals, where such methods could outperform canonical approaches.

SVG detection can be viewed as a feature selection step in spatial transcriptomics data analysis, where useful features (i.e. SVGs) are selected and/or uninformative ones are removed. In particular, the current SVG detection methods can be considered as unsupervised approaches where no information such as cell types, cell states, or spatial domains are required. A great amount of work has been done in feature selection in single-cell data analysis [115], including unsupervised methods and more advanced methods that perform combinatorial feature selection using supervised learning, such as embedded feature selection using random forest and wrapper feature selection using genetic algorithms. I anticipate that future development of SVG detection methods will explore the utility of information such as cell types and states to identify SVGs that not only independently mark the spatial variability but also those that cooperate across multiple genes and together define spatial variability, which would be highly applicable to study contexts such as developmental patterning.

Based on our evaluation, the most common method for prioritising variable gene expression between cells involves first summarizing the spatial relationships between cells. This is achieved either by creating a weighted matrix, where nearby cells are assigned higher weights, or by constructing a nearest neighbor network that connects only the closest cells. Building on this framework, we can apply a parametric model incorporating cell type as a covariate to calculate the variance in gene expression within specific cell type regions relative to the rest of the tissue. An important consideration is the distribution of cell types across different areas of the tissue. Since weighted matrices or nearest-neighbor approaches prioritise nearby cells, an additional parameter may be needed to account for multiple cell type regions that are spatially distant from each other.

The notion of identifying the cooperative activity of more than one gene to uncover cell function or cell fate in spatial transcriptomics has been previously explored, however it has so far been limited to examining gene pairs. For example, Ghazanfar et al. [116] developed scHOT, a method to identify gene pairs with differential correlation across space. Moreover, the fundamental principle of cell-cell communication analyses revolves around

identifying the co-expression of genes (that is, ligand and receptor pairs) between cell types [117, 118]. Similarly, Dries et al. [81] explores an ad-hoc approach to identify the differential co-expression of genes dependent on the co-localisation of pairs of cell types. To move beyond the restriction of examining gene pairs, we may consider implementing a framework like WGCNA [119] to group similar genes into gene modules and measure the expression variability of gene modules across cell type neighbourhoods.

Overall, while these implementations will be computationally challenging but will undoubtedly lead to new biological insight in spatial transcriptomics analysis.

2.6 Summary

Chapter 2 presents an evaluation of current state-of-the-art computational methods for detecting SVGs across a wide array of imaging-based and sequencing-based spatial transcriptomics data. Since the popularisation of spatial transcriptomic platforms, numerous computational methods have emerged, each attempting to incorporate spatial coordinate information to identify biologically meaningful genes across tissues. I evaluated the performance of these methodologies using several metrics, including similarity in prioritizing SVGs, the impact of sparsity and gene filtering on detection, accuracy, and how informative the genes are in delivering high clustering concordance. It should be noted that the majority of datasets analyzed in this study were generated from the 10x Visium platform, which, being the most commercially available, introduces a bias during both method development and evaluation. Although I included data from eight unique technologies, more replicate datasets from each technology would have been beneficial. Despite this limitation, the benchmark and overview of spatially variable gene detection methods provide a foundational understanding of how spatial information can be integrated with gene expression to uncover biological insights. This lays the groundwork for future improvements in methodologies for spatial transcriptomics.