Charles University in Prague

Faculty of Science

BACHELOR THESIS

Matouš Laitl

Cell-cell communication analysis in single-cell and spatial transcriptomics

Analýza mezibuněčné komunikace v jednobuněčné a prostorové transkriptomice

Supervisor of the bachelor thesis: Ing. Lukáš Valihrach, Ph.D.

Study programme: Bioinformatics

Prague, 2024

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ....

Podpis autora

I would like to express my sincere thanks to Ing. Lukáš Valihrach, Ph.D. and Mgr. Daniel Žucha for their professional guidance and support in the process of writing this thesis. I also greatly appreciate their great patience with me throughout the writing, which would not be possible without them.

Title: Cell-cell communication analysis in single-cell and spatial transcriptomics

Author: Matouš Laitl

Department: Department of Genetics and Microbiology

Supervisor: Ing. Lukáš Valihrach, Ph.D., Department of Genetics and Microbiology

Abstract: Intercellular communication orchestrates important tissue processes in all multicellular organisms. Recent advantages in single-cell and spatial transcriptomics unveil new possibilities in analysing cell-cell communication. Up-to-date workflows of both revolutionary methods are described and complemented with relevant novel tools. By merging detailed genome-wide single-cell sequencing information with preserved spatial context, it is possible to capture tissue organisation in great detail. Knowledge-graph-based tool SpaTalk is used to reveal signalling pathways in mouse brain after cerebral ischaemia. The results show significant role of adhesion-related and tissue-densification-related signalling events in the intercellular communication at the lesion border. Detailed signalling mechanisms among proliferating microglia, astrocytes, and oligodendrocytes ongoing in the lesion are also revealed.

Keywords: Cell-cell communication single-cell transcriptomics spatial transcriptomics data analysis

Název práce: Analýza mezibuněčné komunikace v jednobuněčné a prostorové transkriptomice

Autor: Matouš Laitl

Katedra: Katedra genetiky a mikrobiologie

Vedoucí bakalářské práce: Ing. Lukáš Valihrach, Ph.D., Katedra genetiky a mikrobiologie

Abstrakt: Mezibuněčná komunikace řídí důležité procesy v tkáních všech mnohobuněčných organismů. Nedávné pokroky v oblastech jednobuněčné a prostorové transkriptomiky otevřely nové možnosti analýzy mezibuněčných komunikací. V práci jsou uvedeny současné postupy obou těchto revolučních metod doplněny o relevantní nástroje. Spojením detailní celogenomové jednobuněčné sekvenace se zachovanou prostorovou informací je možné zachytit detaily v organizaci tkáně. Nástroj SpaTalk je použit pro analýzu signalizačních drah v myším mozku postiženém cerebrální ischemií. Výsledky analýzy ukazují význam mezibuněčné komunikace spojené s adhezí a kompaktací tkáně na hranici léze. Práce také odkrývá signalizační mechanismy mezi buňkami proliferujících mikroglií, astrocytů a oligodendrocytů.

Klíčová slova: Mezibuněčná komunikace jednobuněčná transkriptomika prostorová transkriptomika datová analýza

Contents

Introduction

To understand complex tissue processes in health and disease, one must inevitably comprehend communication events across cellular network. From the plethora of possible ways information is passed in the tissue, the importance of ligand–receptor protein interactions has been obvious for a while now (Bongrand, 1999). Until recently, only low-throughput methods with considerable limitations were widely available (J. Zhu et al., 2023). With recent breakthroughs in transcriptome sequencing, as summarised by T. Hu et al. (2021), in capturing heterogeneity of tissue with spatial information preserved, what Ståhl et al. (2016) invented, and in-depth genome-wide analysis in single cell, as discussed in this article from Kharchenko (2021), new possibilities in studying interactions arises.

One of the novel software tools, SpaTalk from Shao, C. Li, et al. (2022), is used in this thesis to ascertain ongoing communication events in injured mouse brain. The results show significant communication in the affected tissue mostly related to immune response and overall high levels of cell stress in the area. It also localises the venue of most communication events in the lesion and its surroundings. The findings are presented in a biological context.

First, single-cell and spatial transcriptomics are briefly described whilst highlighting important properties to take into account in interpretation of the results. In the second chapter, results are presented and discussed along with data sources and pipeline description.

Chapter 1

Cell-cell communication review

1.1 Cell-cell interactions

Since the beginning of evolution, multicellular organisms have to invent strategies in order to coordinate all of their cells. Cell-cell interactions are the fundamental principles in maintaining the organism in homeostasis and keeping it alive. Studying those interactions is crucial for understanding the processes in a variety of physiological mechanisms, including numerous pathological states. Recent advances in multiple sequencing techniques enable large scale analysis, which helps to gain valuable insights into numerous yet unexplored phenomenons.

1.2 Single-cell transcriptomics

The ability to analyse the whole transcriptome of a single cell in large quantities is the main advantage of single-cell transcriptomics. Usage of next generation sequencing methods (NGS) have shown, how important are sensitive methods for meaningful transcriptome analysis (Tang et al., 2009). Conventional bulk transcriptome sequencing does not provide the information about differences

in expression counts in various cells. It has been observed, that even indistinguishable cells from a homogeneous tissue sample can differ greatly in the gene expression levels (Shalek et al., 2014).

1.2.1 Single-cell transcriptomics workflow

Data preprocessing

Acquiring dissociated cells The first main step is obtaining the single-cell material, from which RNA of interest can be sequenced separately for each cell. The selection of method depends on the accessibility of the undamaged suspended cells. If the tissue of interest does not allow gathering of intact cells, usage of single-nuclei RNA sequencing is also feasible. Those methods dissolve all cellular membranes and directly accessing the RNA (Grindberg et al., 2013). More frequent approach include separation of individual cell by mechanical and enzymatic means (Xin Wang et al., 2003; Muraro et al., 2016).

Cell Isolation To facilitate sequencing of individual cells, they are required to be separated in distinct spaces. Advanced high-throughput methods used for this task are microfluidic-based or plate-based. Microfluidic-based methods allow parallel cell capture using specific devices as presented by Marcus, Anderson, and Quake (2006) where a continuous flow of suspension is separated using valves or systems where each cell is inserted into an emulsion droplet (Guo et al., 2012). Plate-based methods use fluorescence activated cell sorting (FACS) as demonstrated by Bonner et al. (1972) to guide cells into plate wells.

RNA sequencing - Library preparation Next step is to sequence isolated transcripts to obtain raw reads. RNA amplifying and sequencing in each of the isolated volumes is not financially and practically feasible. Usage of cellular bar-

codes and UMIs (unique molecular identifiers) enables pooling and multiplexing of the samples. To distinguish reads from individual cells , transcripts are attached a cellular barcodes (a nucleotide sequence attached to the RNA transcripts, unique for each cell) (Macosko et al., 2015). To distinguish differences in gene expression from amplification bias, transcripts are also attached a UMI (Islam et al., 2014). After reverse transcription and second strand synthesis, to obtain cDNA and amplification of the material to improve the probability sparsely represented genes of being sequenced, raw reads are sequenced with NGS tools (Hwang, J. H. Lee, and Bang, 2018).

Raw data Preprocessing Typical input for the single-cell analysis tools is a matrix of reads. In order to extract meaningful data from such matrices, steps to identify reads from individual cells and to remove bias from cDNA amplification have to be taken (Wilkins et al., 2021). First steps are: alignment of reads, ”demultiplexing” - separating reads appertaining to individual cells from the sequenced pool by their cellular barcodes, and removing amplification noise from PCR reads with same UMI to subsequently obtain original counts in each cell to detect expression level differences (Weber et al., 2021; Parekh et al., 2018). Separated raw reads are mapped to the annotated sequences of studied species using variations of dynamic programming to perform global or local alignments and multiple heuristics (Johansen and Quon, 2019). Available tools differ in alignment approaches and used reference sequences - whether the alignment resource is whole genome or only known transcripts. One advanced tool is STARsolo from Kaminow, Yunusov, and Dobin (2021) capable of incorporating demultiplexing, cell assignment using cellular barcodes and sequence alignment with gene assignment into a single tool.

Quality Control This step aims to filter out abnormal cell reads which correspond to defective or damaged cells or an error in workflow. Unwanted situations include cells with torn membrane, dying cells, doublets (a situation where two (or more) cells are incorrectly captured in a single, discrete space instead of one cell (Germain et al., 2021)) and unsuccessful captures (Lun, McCarthy, and Marioni, 2016). Popular and well established doublet detection tool is Scrublet by Wolock, Lopez, and Klein (2019) or a novel tool SoCube by H. Zhang et al. (2023) which utilises neural networks for doublet detection.

Normalisation Matrices of scRNA-seq data are sparse. Zeros are predominant due to high variability in gene expression among individual cells and technical hindrances in processing low quantities of mRNA (sampling effect) (L. Lun, Bach, and Marioni, 2016). Compared to bulk sequencing, single-cell resolution highlights the expression differences greatly. Normalisation step addresses this issue and tries to make expression profiles comparable. It is also important to account for ”batch effects” which arises from noise, artifacts and dissimilarities in isolated experiments (Leek, 2014). Simple and commonly used but non-optimal method is CPM (counts per million) which does not assumes differences in count depth (Luecken and Theis, 2019). Selection of the normalisation method always affects downstream analysis and has to be tailored for individual applications (Vallejos et al., 2017). Recently Lause, Berens, and Kobak (2021) showed, that method utilising analytic approximation of Pearson residuals presented by Hafemeister and Satija (2019) outperforms other methods and is capable of capturing more relevant biological variation. It utilises the Pearson residuals to estimate the technical artifacts of the experiment and thus enabling to remove them.

Feature selection Filtering the large amount of genes is necessary to capture biological relevance. Feature selection isolates only informative genes for down-

stream analysis. Classical approach from Brennecke et al. (2013) is to identify highly variable genes. This is unfortunately susceptible to normalisation methods. Therefore, Townes et al. (2019) suggested using deviance for feature selection, which describes how well each gene matches constant, null expression model. Informative genes are those with dissimilar expression among cells, such genes fit purely the null model.

Dimensionality reduction With a similar purpose as feature selection dimensionality reduction extracts informative data to reduced space for further downstream algorithms to work with. Such step is possible due to the fact that scRNA-seq data have intrinsically low dimensionality (Moon et al., 2018). While commonly used PCA (principal component analysis) suffers from normalisation techniques as Townes et al. (2019) showed, generalised PCA is more suitable. However, both are not the best suitable for 2D visualisation, unlike superior graph based UMAP. Besides that, authors McInnes, Healy, and Melville (2018) state its weakness in interpretability in comparison with PCA.

Cellular structure and cell populations (Intermediate analysis)

Clustering For scRNA-seq data, without spatial information, clustering provides way of grouping cells together based on similarity of their expression profiles. Community detection algorithms are one way to divide distinct groups. Recently developed Leiden algorithm by Traag, Waltman, and Van Eck (2019) is a successor to widely used Louvain algorithm which may not recognise all groups sufficiently. Both utilises KNN (k-nearest neighbours graph) as a base structure. Clustering can also be performed using machine learning methods. Such approaches iteratively recalculate centroids of clusters and form groups from closest cells (Petegrosso, Z. Li, and Kuang, 2020).

Annotation Created clusters distinguish cells with similar expression profiles, which corresponds to similar cell types. Annotation of these clusters assign each cell a particular cell type. This is ambiguous because of the nature of scRNA-seq data, which captures a single snapshot in life of a cell and therefore different cell states or subtypes modify the observation (Zeng, 2022). Depending on application, different level of distinction is desired to be captured. Annotation always work on previously annotated atlases of reference data or marker genes databases. Annotating whole clusters is faster and may benefit from consensus expression profile among cells in cluster, although annotating each cell separately will not hide subtle differences, which may be crucial (Clarke et al., 2021). One example is an automated annotation tool scAnvi by C. Xu et al. (2021). It merges multiple datasets to gain less dataset-biased information.

Trajectory inference Differentiation and developmental transitions are important in understanding underlying processes in health and disease (Bendall et al., 2014). Current trajectory inference methods differ in types of topologies they are capable detecting and whether the inferred topology is fixed or changeable in progress (Saelens et al., 2019). Slingshot algorithm by Street et al. (2018) uses a minimum spanning tree to identify branches and fit the structure with examined data afterwards. In reverse order works novel scTEP from authors Y. Zhang et al. (2023), who first determine the pseudotime (path of expression profiles similarity of cells) from which the trajectory is modelled. It is always important to assess the proposed trajectories with awaited results as they don't necessary represent real biological processes (Sundell et al., 2023).

Gene regulatory networks Another useful information to obtain from scRNA-seq data is to determine transcription factors pathways. It is possible to identify groups of co-regulated genes and the regulatory structure called regulon (Keene,

2007). Although Efremova and Teichmann (2020) stated, that inferring regulons only from scRNA-seq is not sufficient due to histons, accessibility of the chromatin and other DNA modifications, several tools exists. GRNBoost2 uses machine learning method gradient boosting, which is reducing bias (Moerman et al., 2019). Very recent SCENIC+ combines single-cell count matrix with chromatin accessibility to infer regulons (Bravo González-Blas et al., 2023).

Cell-cell signalling The core of inferring cell-cell communication and signalling pathways are protein-protein (mostly ligand-receptor) interactions (De Las Rivas and Fontanillo, 2010). By utilising prior knowledge of interacting protein pairs, it can be predicted that two cells expressing each protein of the pair can interact. However, as Armingol, Officer, et al. (2021) summarised in their review, multiple factors need to be considered. Those include: sequenced mRNA counts might not reflect true protein quantities in a cell, the pure fact that protein is synthesised in a cell does not have to guarantee communication events to happen, and proteins composed of multiple sub-units. Besides these limitations, spatial information is due to the tissue dissociation step inherently lost. This crucial disadvantage in resolved with incorporation of spatial transcriptomics.

1.3 Spatial transcriptomics

In 2016, the first method allowing collection of transcriptomic data while preserving spatial information was published by Ståhl et al. (2016). The method uses oligonucleotides on a glass slide, which are barcoded with unique sequence for distinct positions. On those probes, mRNAs from tissue placed above are caught and sequenced with their spatial position in the sample preserved. Since then different technologies were used to profile gene expression while preserving spatial information.

Emerging spatial transcriptomics methods allows single-cell or even sub-cellular resolution of space (Savulescu et al., 2020). However, such methods are usually costly, have insufficient sensitivity or are not transcriptome wide (Yue et al., 2023). When predicting cell-cell communication, integrating spatial data with single-cell data allow employing spatial methods with multi-cell resolution without losing the accuracy.

1.3.1 Spatial transcriptomics workflow

The wide range of available spatial transcriptomics methods can be divided in two main groups: sequencing-based and imaging-based. They differ in the way they capture the spatial and mRNA abundance information. While imaging-based insert (usually fluorescent) visible markers into the sample, either directly, as Ke et al. (2013) first demonstrated, or by hybridisation, sequencing-based methods rely on NGS after adding position-specific barcodes (Ye Wang et al., 2023).

Preprocessing For sequencing-based methods initial quality control steps as well as normalisation and selection of highly (spatially) variable genes are similar to single-cell analysis procedure (Walker et al., 2022). State-of-the-art, recently updated toolkit Seurat by Hao et al. (2023) offers method SCTransform for statistical normalisation among many other analytical tools for spatial transcriptomics data pipeline.

Image-based methods require processing of acquired images using computer vision techniques. Noise reduction algorithms are used to remove technical artifacts from imaging and backgrounds are removed (Williams et al., 2022). Each hybridisation or sequencing round, an image is taken, again due to limitations of microscopic imaging techniques, all images must be aligned so that a single observed point is matched throughout the imaging (Kleino et al., 2022). Once

aligned, sequence of intensities is measured and matched to known transcripts fluorescence signals sequence, if possible (Cisar et al., 2023). Segmentation is a preprocessing method, which can be used to detect cell boundaries. One of the latest tools for segmentation is Baysor by Petukhov et al. (2022).

Spatial clustering Common approach to clustering in spatial data is adapting single-cell methods and algorithms for spatial usage (J. Li, Shyr, and Q. Liu, 2023). Many of these methods use k-nearest neighbours graph. Quality of results based on kNN graphs strongly depends on the selection of k, which leads to inaccuracies in clustering rare or abundant cell types (Dries et al., 2021). Novel DR-SC from authors W. Liu et al. (2022) unifies dimensionality reduction with clustering to enhance the quality of its overall output.

Deconvolution Most broadly available genome-wide methods for spatial transcriptomics with sufficient sequencing depth are still in multi-cell resolution. This limitation is tackled with algorithms capable of inferring specific cell type proportions on each spot (Cable et al., 2022). Some methods can decompose cell types without single-cell reference data, one of them is STdeconvolve from Miller et al. (2022) which uses generative probabilistic model. Different approach to decomposition is coalescing spatial with single-cell data as a reference for cell type inference. Gold standard is a tool from Cable et al. (2022) called RCTD. It calculates mean expression of marker genes in reference scRNA-seq dataset and uses it to model probabilities how would a particular single cell contribute to spot expression levels. Redeconve is a recent tool by Z. Zhou et al. (2023) who claims it outperforms other deconvolution tools by using linear regression model. Performance of these tools depends greatly on the strength of match between spatial and single-cell data. Authors Lopez et al. (2022) created tool named DestVI which might benefit from also considering sub-cell-type variations.

Gene imputation When single-cell data are used as a supporting evidence, it is possible to fill in the missing gene expression information (Biancalani et al., 2021). Gene imputation tools aim to learn translation function from overlapping genes (in spatial and single-cell datasets) and utilise that knowledge to fill the gaps in spatial expression profiles with profiles, present in single-cell dataset (Qiao and Huang, 2023).

Data integration The most widespread spatial transcriptomics methods come with limited sample sizes for a single run. This might lead to necessity of multiple runs. Other possibility is that the nature of experiment requires runs separated in time. Both of these eventualities result in data which cannot be simply merged owing to batch effects and other non-identical conditions (R. Li, X. Chen, and X. Yang, 2024). The issue is solved by the methods of integrative analysis. The objective is to capture similar spots between each run, anchor those and reconstruct 3D space from gathered connections (X. Zhou, K. Dong, and S. Zhang, 2023).

1.4 New generation of tools for inferring cell-cell communication in spatial data

From the numerous ways cells interact in organisms, usage of transcriptomic methods limit the spectrum of detectable interactions to those where both ends of the communication (sender and receiver) are proteins or other sequenceable molecules (Dimitrov, Schäfer, et al., 2023). Prevalent target of studies are ligand-receptor pairs, because of the simplicity of their acquirement and extensive literature-curated databases (Armingol, Officer, et al., 2021).

The basic premise is that it is possible to identify communicating cells by analysing expression levels of ligand and receptor in the corresponding cells and

that those levels can be used to measure communication strength (Armingol, Ghaddar, et al., 2022). Directly measuring if a communication event occurs in living tissue would provide better image, but this is currently not possible. For algorithm to successfully reveal communication it is required that both sender and receiver cell types are sufficiently abundant, expressions of both interacting proteins are sufficient for capturing sensitivity in sequencing algorithms, and previous knowledge of interactions between these specific proteins is described and ideally confirmed (Almet et al., 2021).

1.4.1 Selected tools

According to very recent study by Armingol, H. M. Baghdassarian, and Lewis (2024) there are over one hundred tools for communication inference with different localisation, scalability, accuracy, sensitivity and performance. SpaTalk from Shao, C. Li, et al. (2022) is the one chosen for the analysis in the second chapter. Here is presented a brief theoretical summary and comparison of tools comparable with SpaTalk.

CellChat and CellChat v2 First CellChat tool presented in article by Jin, Guerrero-Juarez, et al. (2021) was able to calculate probability of intercellular interactions from scRNA-seq data. Although it has been highly scored as a tool with stable consistency in this study from Z. Liu, Sun, and C. Wang (2022) comparing CCC (cell-cell communication) inference tools, hardly any other studies benchmarking CCC tools exists. CellChat package includes its own custom ligand-receptor database CellChatDB and authors claim, that its advantage is inclusion of multimeric protein complexes and cofactors. Although study comparing CCC tools and resources by Dimitrov, Türei, et al. (2022) clearly states, that most available resources share a common origin and therefore lacks overall uniqueness.

That perhaps opens possibilities for a new, frequently updated universal consensus database.

The second, updated version of this tool is able to incorporate spatial dataset as a guidance for proximal interactions (Jin, Plikus, and Nie, 2023). Major disadvantage of both versions is the conception of one "spot". Cell groups, as they call them, consist of single most abundant cell type, this wastes measured heterogeneity and limits the potential of full performance.

stLearn Python based software stLearn offers spatially-centred approach with great efforts to minimise false positives (Pham et al., 2023). For its optimal performance it also accepts, aside of gene expression and spatial guidance, tissue morphology images. Those images facilitates adjustments of the sequencing data. In the same comparison study from Z. Liu, Sun, and C. Wang (2022) where CellChat scored high, stLearn succeeds in human dataset, but on the whole was below-average. It should be noted that even authors ascribe that to properties of their specific metric.

SpaTalk This recent R-compatible graph-based tool for CCC inference created by Shao, C. Li, et al. (2022) directly combines scRNA-seq and spatial data. No relevant let alone independent studies compared SpaTalk with other CCC inference tools as of time of writing this thesis. Nevertheless, authors J. Zhu et al. (2023) of this article declare, that far too much predicted interactions are shared among dissimilar cell types. On a theoretical level, SpaTalk constructs KNN graph from euclidean distance matrix and obtains significantly co-expressed ligand-receptor pairs from sufficiently neighbouring cells. It than scores the predicted interaction pairs by their influence on the corresponding receptor cells. The greater the number of activated transcription factors and terminal downstream target genes, the higher score the interaction gets. Important benefit of SpaTalk is the

possibility to input data after deconvolution, which gives users the opportunity to incorporate different deconvolution tools. Unlike CellChat, SpaTalk considers cell types proportions in each spatial spot.

1.4.2 Specific approaches

Among all tools based on coalescing spatial with single-cell data, some methods approach CCC inference differently. One example comes from Boisset et al. (2018) whose method guarantees spatial closeness by intentionally dissecting doublets or triplets for sequencing.

Chapter 2

Implementation of SpaTalk on cells in the mouse brain affected by ischaemic brain injury

In this section is described, how the previously mentioned SpaTalk tool from Shao, C. Li, et al. (2022) was used to infer communication between mouse brain cells after cerebral ischaemia. As an input, outputs from recent spatial transcriptomics and single-cell sequencing methods were used in tandem. The package version 1.0 was running in R version 4.3.3 environment.

2.1 Data analysis using SpaTalk

2.1.1 Used datasets

The spatial dataset used for the analysis from Zucha et al. (2023) is a section of mouse forebrain after middle cerebral artery occlusion. The data processing was performed as described in Zucha et al. (2023). In short, the spatial dataset

was processed using 10X Visium platform and the data processed using Seurat pipeline (Hao et al., 2023). For deconvolution of the spatial dataset, RCTD deconvolution algorithm from Cable et al. (2022) was used with reference single-cell dataset from Milich et al. (2021). The same spinal cord injured mouse dataset by Milich et al. (2021) was used as a single-cell input for the SpaTalk communication inference module. The merged data contains 17 distinct cell types of two categories: CNS cells and immune system cells. The default SpaTalk databases served as knowledge base for the analysis, namely: ligand-receptor database CellTalkDB from Shao, Liao, et al. (2021), database of transcription factors AnimalTFDB curated by H. Hu et al., 2019, and downstream pathways databases Reactome from Milacic et al. (2024) and KEGG (Kanehisa et al., 2023).

2.1.2 SpaTalk workflow description

First of all, SpaTalk object was created with `createSpaTalk()` function with preprocessed spatial dataset and parameters: `species = "Mouse"`, `if_st_is_sc = FALSE` and `spot_max_cell = 25`. Initial object creation was followed by incorporation of single-cell data. The `dec_celltype()` function was called with single-cell expression matrix and already deconvoluted spatial data as `dec_result` argument. Next, by calling `find_lr_path()` function, communication pathways were found using the default ligand-receptor database. At last with the `dec_cci_all()` function, all proximal cells interactions were found. All information for the following plots was contained in the SpaTalk object at this point.

2.2 Results of the performed data analysis

As already mentioned in the section 2.1.1 Used datasets, 17 distinct cell types were analysed with SpaTalk. Out of 272 result pairs produced by the algorithm (all

pairs minus signalling between cells of the same type), 4 selected are presented here. Namely proliferating microglia sending to astrocytes, the reverse direction, where astrocytes are sending, microglia sending to oligodendrocytes, and its reverse. Those were selected mainly for their high abundance in the datasets and their crucial role in response to the injury.

2.2.1 Proliferating microglia to astrocytes signalling

The analysis shown considerable changes occurring in the lesion area in the sample from day 3 post injury. Astrocytes are equally distributed in healthy areas as expected, whereas proliferating microglia are gathered in lesion, especially at the border of healthy and damaged tissue (figure 1A). Neither astrocytes nor microglia are plentifully present in already destroyed tissue (figure 1A and figure 1B). The maximum of inferred communication events occurs at the lesion boundary (figure 1B). Lipid associated Apoe ligand in proliferating microglia supports the inflammatory state of those cells and signals to the astrocytes through the Sdc4 receptor, significantly activating three downstream pathways (figure 1C and figure 1D). Great levels of transmission are realised via Itgb1 and Itgav receptors activation (figure 1C). This integrin-related signalling conforms with the cell adhesion in tissue undergoing immune response. The complex branching (figure 1E) shows the extensive downstream activation power of Sdc4 receptor.

2.2.2 Astrocytes to proliferating microglia signalling

Although the overall number of communication events measured in the direction from astrocytes to proliferating microglia is greater than the opposite (figure 1A and figure 2A), when monitoring ligand Vcan and receptor Tlr2, the plot is much more sparse (figure 2B). Much less astrocytes express Vcan compared with Sdc4 across the whole tissue sample, not clustering on the lesion area border (figure 2B).

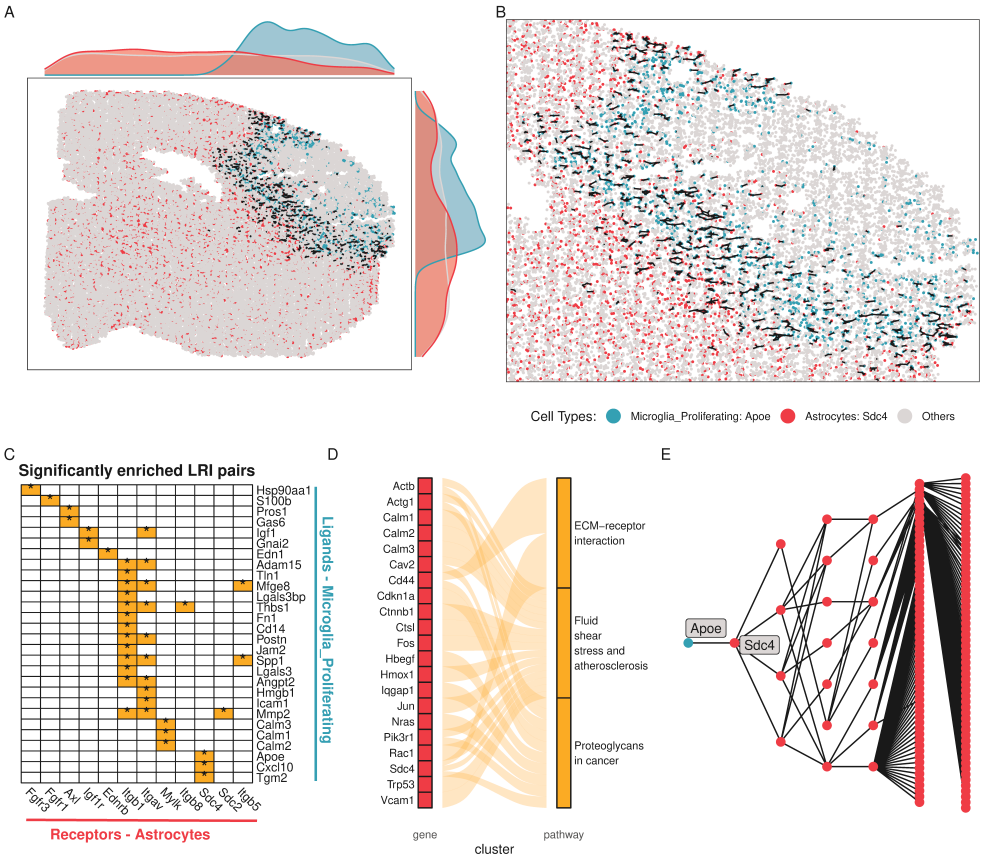

Figure 1 Visualisation of proliferating microglia signalling to astrocytes

A) Point plot with spatial distribution of proliferating microglia and astrocytes with highlighted significantly communicating cells. B) Inset point plot of lesion area. Straight lines represent ligand-receptor pairs of Apoe and Sdc4 genes in interacting proliferating microglia with astrocytes. C) Table containing ligand-receptor pairs significantly enriched in proliferating microglia and astrocytes. Ligands produced by proliferating microglia are on the y-axis, while on thy x-axis are corresponding astrocytes receptors. Fields marked with an * (asterisk) are the ones most enriched. D) Alluvial plot of significantly activated pathways downstream to Sdc4. On the gene-axis, strata are related downstream genes of astrocytes and strata on pathway-axis are significantly used pathways. E) Network plot mapping the extent of downstream gene pathways in receipting cells.

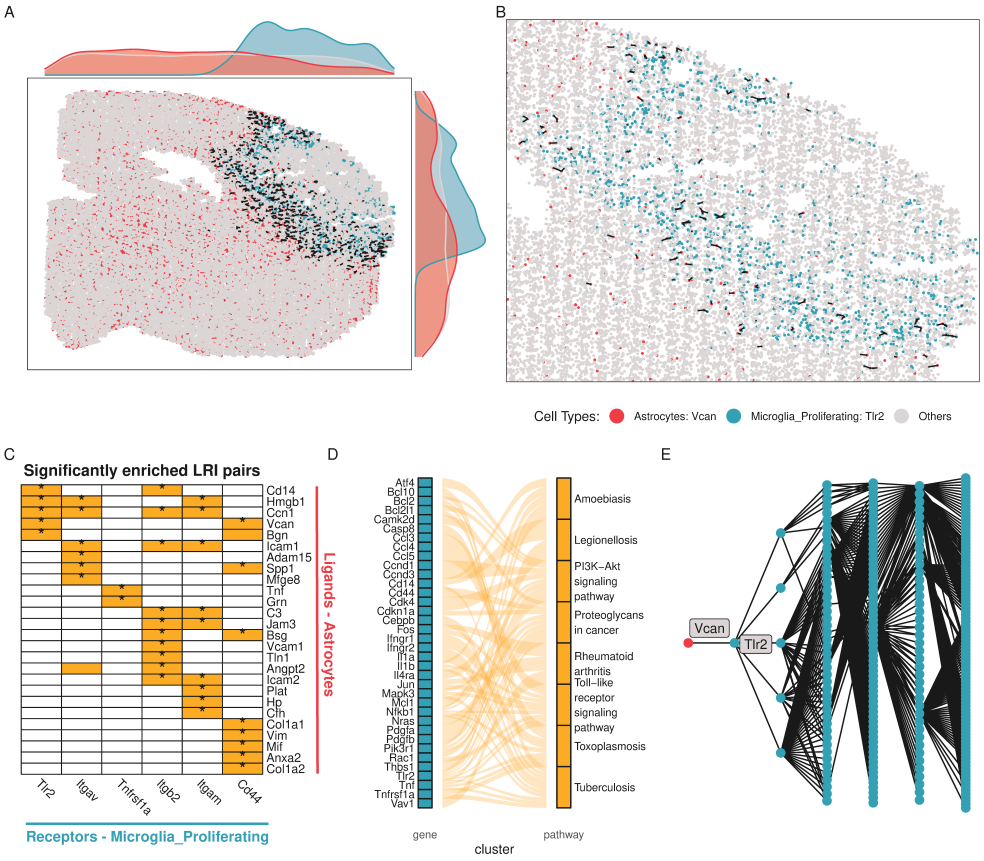

Figure 2 Visualisation of astrocytes signalling to proliferating microglia

A) Point plot with spatial distribution of astrocytes and proliferating microglia with highlighted significantly communicating cells. B) Inset point plot of lesion area. Straight lines represent ligand-receptor pairs of Vcan and Tlr2 genes in interacting astrocytes with proliferating microglia. C) Table containing ligand-receptor pairs significantly enriched in astrocytes and proliferating microglia. Ligands produced by astrocytes are on the y-axis, while on thy x-axis are corresponding proliferating microglia receptors. Fields marked with an * (asterisk) are the ones most enriched. D) Alluvial plot of significantly activated pathways downstream to Tlr2. On the gene-axis, strata are related downstream genes of proliferating microglia and strata on pathway-axis are significantly used pathways. E) Network plot mapping the extent of downstream gene pathways in receipting cells.

By contrast, proliferating microglia with abundantly expressed Tlr2 are almost solely present in the lesion boundary (figure 2B). As a result, number of detected communication events through the mediation of Vcan-Tlr2 pair, is low (figure 2B). However apart from Vcan, other astrocytes ligands (Cd14, Hmgb1, Ccn1, and Bgn) are also received by Tlr2 (figure 2C). Similarly to communication in section 2.2.1, different integrin subunit gene Itgb2 is also redundantly activated by multiple ligands (Icam1, C3, Tln1, Cd14, Vcam1, Jam3, and others), again showing high levels of cell adhesion (figure 2C). Another significantly enriched receptor receiving in microglia is the Cd44 (figure 2C) involved in lymphocyte homing and again in cell adhesion. The network of activated downstream pathways in microglia is extensive and highly interconnected with Tlr2 as a key gene in most of the pathways (figure 2D and figure 2E). Tlr2 being a member of Toll-like receptor family, plays significant role in pathogen recognition and immunity triggering. This fact is also supported in figure 2D with the number of cell defence-related activated pathways. In summary, astrocytes response to proliferating microglia is triggering multiple mostly stress-related pathways.

2.2.3 Proliferating microglia to oligodendrocytes signalling

The topological structure of Tnfrsf1a downstream pathways show the richest branching in the last level (figure 3A). While the Tnf-Tnfrsf1a cascade show few major controlling genes, triggering large quantity of downstream targets in the last level, Vcan-Tlr2 cascade is much more interconnected (figure 2E and figure 3A). Areas of inferred communication between proliferating microglia and oligodendrocytes are especially in places where nerve fibres, myelinated by oligodendrocytes, meet damaged tissue. In the mentioned area in figure 3B, plotted ligand-receptor pair Tnf and Tnfrsf1a interacts in similar numbers as Vcan and Tlr2. Similarly to both previous communicating pairs, integrin subunits genes are also

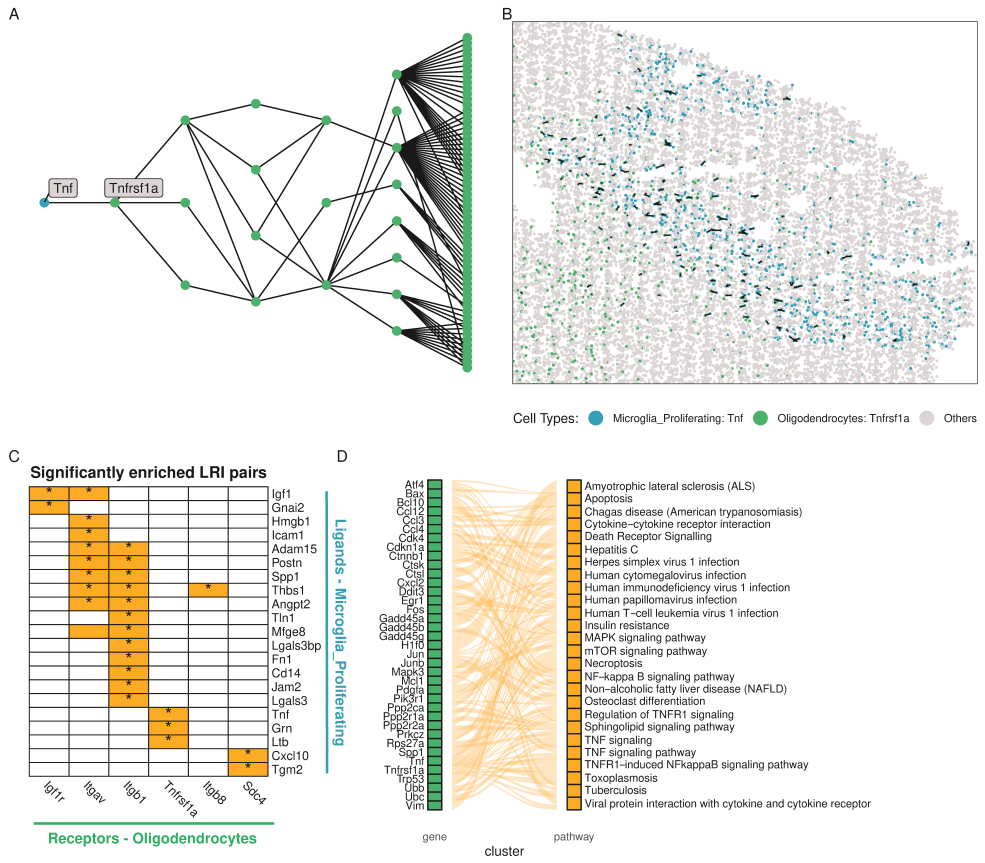

Figure 3 Visualisation of proliferating microglia signalling to oligodendrocytes

A) Network plot mapping the extent of downstream gene pathways in receipting cells. B) Inset point plot of lesion area. Straight lines represent ligand-receptor pairs of Tnf and Tnfrsf1a genes in interacting proliferating microglia with oligodendrocytes. C) Table containing ligand-receptor pairs significantly enriched in proliferating microglia and oligodendrocytes. Ligands produced by proliferating microglia are on the y-axis, while on thy x-axis are corresponding oligodendrocytes receptors. Fields marked with an * (asterisk) are the ones most enriched. D) Alluvial plot of significantly activated pathways downstream to Tnfrsf1a. On the gene-axis, strata are related downstream genes of oligodendrocytes and strata on pathway-axis are significantly used pathways.

significantly enriched in oligodendrocytes. This is done through multiple ligands (Adam15, Spp1, Thbs1, Postn and others) again indicating strong adhesion and tissue densification (figure 3C). Tnfrsf1a triggers a plethora of activated pathways, most of which are related with cell stress, survival, apoptosis, and inflammation (figure 3D). The large number of significantly activated pathways show how massive is the response to lesion-related events in participating oligodendrocytes. It is also worth to notice, how some genes (Tnfrsf1a, Pik3r1, Jun, Fos, and Mapk3) are involved in multiple activated pathways and others (Prkcz, Vim, Pdgfa and, others) only in one or few pathways (figure 3D).

2.2.4 Oligodendrocytes to proliferating microglia signalling

In comparison with astrocytes distribution across the sample tissue as discussed in section 2.2.2, oligodendrocytes are located more densely in the inner regions of the hemisphere and the corpus callosum (figure 4A). Semaphorine gene Sema6d abundantly expressed in some oligodendrocytes being received in microglia indicates chronic inflammation of some of the cells at the lesion border (figure 4B). Chemokine signalling is also enriched in signalling between oligodendrocytes and proliferating microglia through ligands Ccl3, Ccl4, Ccl5, etc. and receptors Ccr1 and Ccr5. This suggests signal transduction and effector immune cells homing (figure 4C). The figure 4D shows very high abundance of the proteasome subunit genes (Psma4, Psmb5, Psmc1, Psmd12, etc.) suggesting that protein metabolism in proliferating microglia is at full speed. Likewise speaks the activated pathways, ongoing immune response utilise all pathways including tyrosine-based DAP12 immunoreceptor pathway (figure 4D). The topology of pathways downstream to Trem2 is noticeably different from Tnfrsf1a downstream topology (figure 4E). While pathways downstream to Tnfrsf1a branches mostly in the last level and also most of the genes in the pathways tree are leaves, Trem2 as well as Tlr2

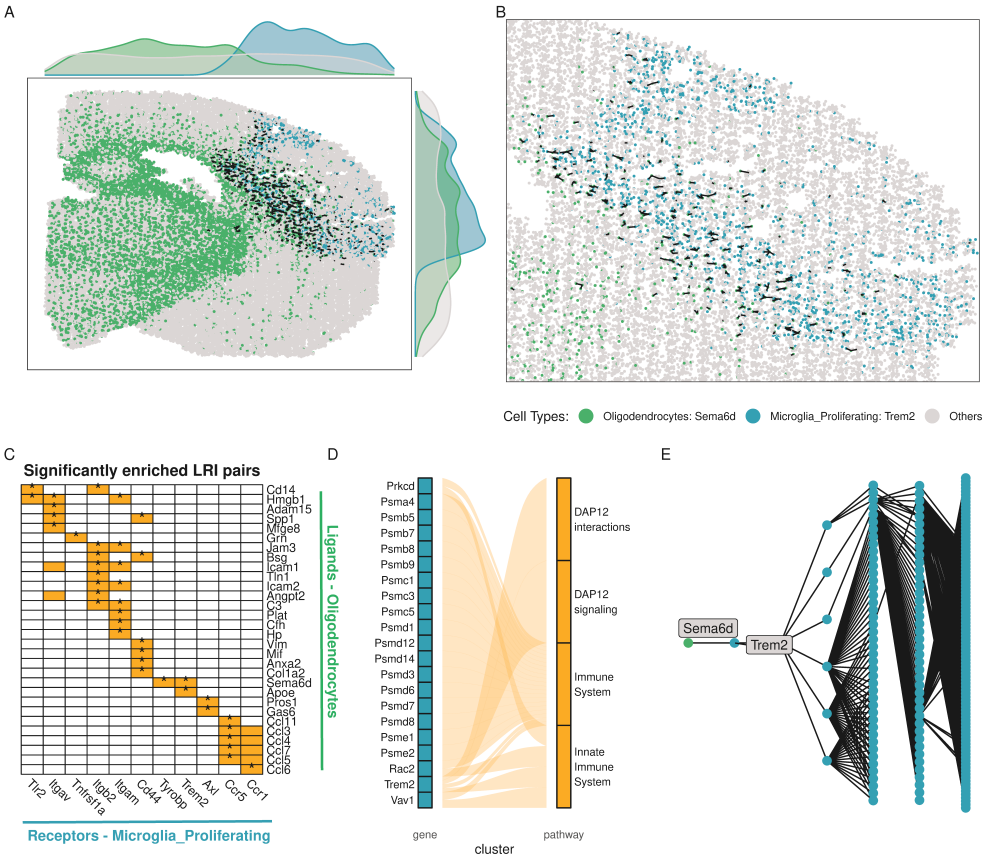

Figure 4 Visualisation of oligodendrocytes signalling to proliferating microglia

A) Point plot with spatial distribution of oligodendrocytes and proliferating microglia with highlighted significantly communicating cells. B) Inset point plot of lesion area. Straight lines represent ligand-receptor pairs of Sema6d and Trem2 genes in interacting oligodendrocytes with proliferating microglia. C) Table containing ligand-receptor pairs significantly enriched in oligodendrocytes and proliferating microglia. Ligands produced by oligodendrocytes are on the y-axis, while on thy x-axis are corresponding proliferating microglia receptors. Fields marked with an * (asterisk) are the ones most enriched. D) Alluvial plot of significantly activated pathways downstream to Trem2. On the gene-axis, strata are related downstream genes of proliferating microglia and strata on pathway-axis are significantly used pathways. E) Network plot mapping the extent of downstream gene pathways in receipting cells.

downstream pathways branches more, even in the first levels, and the numbers of genes in each level are more similar across levels (figure 4E).

2.2.5 Communication summary

All of the tracked communication pairs, show strong signalling related to cell adhesion and tissue densification through multiple subtypes of integrines (figures 1 to 4C). Proliferating microglia at both receiving ends expressed abundantly receptors Tlr2, Cd44, Itgam, Itgb2, Itgav, Tnfrsf1a. It is also noticeable, how proliferating microglia are gathered at the borderline between healthy and damaged tissue (figures 1, 2 and 4A). This is with agreement with the expectations, as microglia being the pivotal cell type in the immune response in brain tissue. It is also in accordance with the expected results, that an intensive communication would be observed between proliferating microglia and two other neuroglial cell types in the affected area at the global scale.

Conclusion

Recent tools allowing cell-cell communication inference emerged recently. Their functionality depends on advances in single-cell sequencing and spatial transcriptomics techniques. This thesis summarises latest workflow methodologies of both of these methods and highlights their significant aspects. It also states possible hindrances which may arise from the workflow and provides guidance through the complex pipeline. Selected tools with some significant properties were also stated.

In this rapidly developing field, the latest cell-cell communication inference tools lack unbiased and independent benchmarks forbidding other than theoretical comparison. Despite that, very recent tools were stated and their strengths and limitations were articulated.

In the second chapter, one of the most recent tools, SpaTalk, was used on dataset from mouse cells after ischaemic brain injury. The analysis have revealed extensive communication occurring in the lesion area. All of the examined cell types shown high levels of signalling related with cell adhesion and tissue densification. Immune response triggering and other stress-related communication was also strongly present in all cell types. The analysis also highlighted the localisation of most of the communication events is in the injured area or at its borders. This results might be used to better understand ongoing processes in brain tissue after injury and the nature of its functioning generally.

The results demonstrate the extent of information obtainable from the gathered spatial and single-cell sequencing data. This can be done with data from other single-cell and spatial transcriptomics experiments beyond their initial objective to deepen the knowledge.

It is possible to extend the analysis to all other cell types and through days 1 and 7 post injury in this exact dataset. Further applications of this exact or similar computational tools opens up great new opportunities to study cellular processes and identify novel targets for therapy. It can also be applied in many other biological fields. As long as the required background knowledge grows and improves in quality, those tools and their future successors will continue to provide valuable insight into the functioning of living organisms.

Code and data availability

All used datasets are available from the corresponding articles as specified in section 2.1.1. Source codes are provided in the supplementary files.

Bibliography

Almet, Axel A et al. (2021). “The landscape of cell–cell communication through single-cell transcriptomics”. In: *Current opinion in systems biology* 26, pp. 12–23.

Armingol, Erick, Hratch M Baghdassarian, and Nathan E Lewis (2024). “The diversification of methods for studying cell–cell interactions and communication”. In: *Nature Reviews Genetics*, pp. 1–20.

Armingol, Erick, Abbas Ghaddar, et al. (2022). “Inferring a spatial code of cell-cell interactions across a whole animal body”. In: *PLoS computational biology* 18.11, e1010715.

Armingol, Erick, Adam Officer, et al. (2021). “Deciphering cell–cell interactions and communication from gene expression”. In: *Nature Reviews Genetics* 22.2, pp. 71–88.

Bendall, Sean C et al. (2014). “Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development”. In: *Cell* 157.3, pp. 714–725.

Biancalani, Tommaso et al. (2021). “Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram”. In: *Nature methods* 18.11, pp. 1352–1362.

Boisset, Jean-Charles et al. (2018). “Mapping the physical network of cellular interactions”. In: *Nature methods* 15.7, pp. 547–553.

Bongrand, Pierre (1999). “Ligand-receptor interactions”. In: *Reports on progress in Physics* 62.6, p. 921.

Bonner, WA et al. (1972). “Fluorescence activated cell sorting”. In: *Review of Scientific Instruments* 43.3, pp. 404–409.

Bravo González-Blas, Carmen et al. (2023). “SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks”. In: *Nature methods* 20.9, pp. 1355–1367.

Brennecke, Philip et al. (2013). “Accounting for technical noise in single-cell RNA-seq experiments”. In: *Nature methods* 10.11, pp. 1093–1095.

Cable, Dylan M et al. (2022). “Robust decomposition of cell type mixtures in spatial transcriptomics”. In: *Nature biotechnology* 40.4, pp. 517–526.

Cisar, Cecilia et al. (2023). “A unified pipeline for FISH spatial transcriptomics”. In: *Cell Genomics* 3.9.

Clarke, Zoe A et al. (2021). “Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods”. In: *Nature protocols* 16.6, pp. 2749–2764.

De Las Rivas, Javier and Celia Fontanillo (2010). “Protein–protein interactions essentials: key concepts to building and analyzing interactome networks”. In: *PLoS computational biology* 6.6, e1000807.

Dimitrov, Daniel, Philipp Sven Lars Schäfer, et al. (2023). “LIANA+: an all-in-one cell-cell communication framework”. In: *BioRxiv*, pp. 2023–08.

Dimitrov, Daniel, Dénes Türei, et al. (2022). “Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data”. In: *Nature communications* 13.1, p. 3224.

Dries, Ruben et al. (2021). “Giotto: a toolbox for integrative analysis and visualization of spatial expression data”. In: *Genome biology* 22, pp. 1–31.

Efremova, Mirjana and Sarah A Teichmann (2020). “Computational methods for single-cell omics across modalities”. In: *Nature methods* 17.1, pp. 14–17.

Germain, Pierre-Luc et al. (2021). “Doublet identification in single-cell sequencing data using scDblFinder”. In: *F1000Research* 10.

Grindberg, Rashel V et al. (2013). “RNA-sequencing from single nuclei”. In: *Proceedings of the National Academy of Sciences* 110.49, pp. 19802–19807.

Guo, Mira T et al. (2012). “Droplet microfluidics for high-throughput biological assays”. In: *Lab on a Chip* 12.12, pp. 2146–2155.

Hafemeister, Christoph and Rahul Satija (2019). “Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression”. In: *Genome biology* 20.1, p. 296.

Hao, Yuhan et al. (2023). “Dictionary learning for integrative, multimodal and scalable single-cell analysis”. In: *Nature biotechnology*, pp. 1–12.

Hu, Hui et al. (2019). “AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors”. In: *Nucleic acids research* 47.D1, pp. D33–D38.

Hu, Taishan et al. (2021). “Next-generation sequencing technologies: An overview”. In: *Human Immunology* 82.11, pp. 801–811.

Hwang, Byungjin, Ji Hyun Lee, and Duhee Bang (2018). "Single-cell RNA sequencing technologies and bioinformatics pipelines". In: *Experimental & molecular medicine* 50.8, pp. 1–14.

Islam, Saiful et al. (2014). "Quantitative single-cell RNA-seq with unique molecular identifiers". In: *Nature methods* 11.2, pp. 163–166.

Jin, Suoqin, Christian F Guerrero-Juarez, et al. (2021). "Inference and analysis of cell-cell communication using CellChat". In: *Nature communications* 12.1, p. 1088.

Jin, Suoqin, Maksim V Plikus, and Qing Nie (2023). "CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics". In: *BioRxiv*, pp. 2023–11.

Johansen, Nelson and Gerald Quon (2019). "scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data". In: *Genome biology* 20.1, p. 166.

Kaminow, Benjamin, Dinar Yunusov, and Alexander Dobin (2021). "STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data". In: *Biorxiv*, pp. 2021–05.

Kanehisa, Minoru et al. (2023). "KEGG for taxonomy-based analysis of pathways and genomes". In: *Nucleic acids research* 51.D1, pp. D587–D592.

Ke, Rongqin et al. (2013). "In situ sequencing for RNA analysis in preserved tissue and cells". In: *Nature methods* 10.9, pp. 857–860.

Keene, Jack D (2007). "RNA regulons: coordination of post-transcriptional events". In: *Nature Reviews Genetics* 8.7, pp. 533–543.

Kharchenko, Peter V (2021). “The triumphs and limitations of computational methods for scRNA-seq”. In: *Nature methods* 18.7, pp. 723–732.

Kleino, Iivari et al. (2022). “Computational solutions for spatial transcriptomics”. In: *Computational and structural biotechnology journal* 20, pp. 4870–4884.

L. Lun, Aaron T, Karsten Bach, and John C Marioni (2016). “Pooling across cells to normalize single-cell RNA sequencing data with many zero counts”. In: *Genome biology* 17, pp. 1–14.

Lause, Jan, Philipp Berens, and Dmitry Kobak (2021). “Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data”. In: *Genome biology* 22, pp. 1–20.

Leek, Jeffrey T (2014). “Svaseq: removing batch effects and other unwanted noise from sequencing data”. In: *Nucleic acids research* 42.21, e161–e161.

Li, Jia, Yu Shyr, and Qi Liu (2023). “Single-cell and Spatial Transcriptomics Clustering with an Optimized Adaptive K-Nearest Neighbor Graph”. In: *bioRxiv*.

Li, Runze, Xu Chen, and Xuerui Yang (2024). “Navigating the landscapes of spatial transcriptomics: How computational methods guide the way”. In: *Wiley Interdisciplinary Reviews: RNA* 15.2, e1839.

Liu, Wei et al. (2022). “Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data”. In: *Nucleic acids research* 50.12, e72–e72.

Liu, Zhaoyang, Dongqing Sun, and Chenfei Wang (2022). “Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information”. In: *Genome Biology* 23.1, p. 218.

Lopez, Romain et al. (2022). “DestVI identifies continuums of cell types in spatial transcriptomics data”. In: *Nature biotechnology* 40.9, pp. 1360–1369.

Luecken, Malte D and Fabian J Theis (2019). “Current best practices in single-cell RNA-seq analysis: a tutorial”. In: *Molecular systems biology* 15.6, e8746.

Lun, Aaron TL, Davis J McCarthy, and John C Marioni (2016). “A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor”. In: *F1000Research* 5.

Macosko, Evan Z et al. (2015). “Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets”. In: *Cell* 161.5, pp. 1202–1214.

Marcus, Joshua S, W French Anderson, and Stephen R Quake (2006). “Microfluidic single-cell mRNA isolation and analysis”. In: *Analytical chemistry* 78.9, pp. 3084–3089.

McInnes, Leland, John Healy, and James Melville (2018). “Umap: Uniform manifold approximation and projection for dimension reduction”. In: *arXiv preprint arXiv:1802.03426*.

Milacic, Marija et al. (2024). “The reactome pathway knowledgebase 2024”. In: *Nucleic acids research* 52.D1, pp. D672–D678.

Milich, Lindsay M et al. (2021). “Single-cell analysis of the cellular heterogeneity and interactions in the injured mouse spinal cord”. In: *Journal of Experimental Medicine* 218.8, e20210040.

Miller, Brendan F et al. (2022). “Reference-free cell type deconvolution of multicellular pixel-resolution spatially resolved transcriptomics data”. In: *Nature communications* 13.1, p. 2339.

Moerman, Thomas et al. (2019). “GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks”. In: *Bioinformatics* 35.12, pp. 2159–2161.

Moon, Kevin R et al. (2018). “Manifold learning-based methods for analyzing single-cell RNA-sequencing data”. In: *Current Opinion in Systems Biology* 7, pp. 36–46.

Muraro, Mauro J et al. (2016). “A single-cell transcriptome atlas of the human pancreas”. In: *Cell systems* 3.4, pp. 385–394.

Parekh, Swati et al. (2018). “zUMIs-A fast and flexible pipeline to process RNA sequencing data with UMIs”. In: *Gigascience* 7.6, giy059.

Petegrosso, Raphael, Zhuliu Li, and Rui Kuang (2020). “Machine learning and statistical methods for clustering single-cell RNA-sequencing data”. In: *Briefings in bioinformatics* 21.4, pp. 1209–1223.

Petukhov, Viktor et al. (2022). “Cell segmentation in imaging-based spatial transcriptomics”. In: *Nature biotechnology* 40.3, pp. 345–354.

Pham, Duy et al. (2023). “Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues”. In: *Nature communications* 14.1, p. 7739.

Qiao, Chen and Yuanhua Huang (2023). “Reliable imputation of spatial transcriptome with uncertainty estimation and spatial regularization”. In: *bioRxiv*, pp. 2023–01.

Saelens, Wouter et al. (2019). “A comparison of single-cell trajectory inference methods”. In: *Nature biotechnology* 37.5, pp. 547–554.

Savulescu, Anca F et al. (2020). “Pinpointing cell identity in time and space”. In: *Frontiers in Molecular Biosciences* 7, p. 209.

Shalek, Alex K et al. (2014). “Single-cell RNA-seq reveals dynamic paracrine control of cellular variation”. In: *Nature* 510.7505, pp. 363–369.

Shao, Xin, Chengyu Li, et al. (2022). “Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk”. In: *Nature Communications* 13.1, p. 4429.

Shao, Xin, Jie Liao, et al. (2021). “CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice”. In: *Briefings in bioinformatics* 22.4, bbaa269.

Ståhl, Patrik L et al. (2016). “Visualization and analysis of gene expression in tissue sections by spatial transcriptomics”. In: *Science* 353.6294, pp. 78–82.

Street, Kelly et al. (2018). “Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics”. In: *BMC genomics* 19, pp. 1–16.

Sundell, Timothy et al. (2023). “Single-cell RNA sequencing analyses: interference by the genes that encode the B-cell and T-cell receptors”. In: *Briefings in Functional Genomics* 22.3, pp. 263–273.

Tang, Fuchou et al. (2009). “mRNA-Seq whole-transcriptome analysis of a single cell”. In: *Nature methods* 6.5, pp. 377–382.

Townes, F William et al. (2019). “Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model”. In: *Genome biology* 20, pp. 1–16.

Traag, Vincent A, Ludo Waltman, and Nees Jan Van Eck (2019). “From Louvain to Leiden: guaranteeing well-connected communities”. In: *Scientific reports* 9.1, p. 5233.

Vallejos, Catalina A et al. (2017). “Normalizing single-cell RNA sequencing data: challenges and opportunities”. In: *Nature methods* 14.6, pp. 565–571.

Walker, Benjamin L et al. (2022). “Deciphering tissue structure and function using spatial transcriptomics”. In: *Communications biology* 5.1, p. 220.

Wang, Xin et al. (2003). “The origin and liver repopulating capacity of murine oval cells”. In: *Proceedings of the National Academy of Sciences* 100.suppl_1, pp. 11881–11888.

Wang, Ye et al. (2023). “Spatial transcriptomics: Technologies, applications and experimental considerations”. In: *Genomics*, p. 110671.

Weber, Lukas M et al. (2021). “Genetic demultiplexing of pooled single-cell RNA-sequencing samples in cancer facilitates effective experimental design”. In: *GigaScience* 10.9, giab062.

Wilkins, Oscar G et al. (2021). “Ultraplex: A rapid, flexible, all-in-one fastq demultiplexer”. In: *Wellcome Open Research* 6.

Williams, Cameron G et al. (2022). “An introduction to spatial transcriptomics for biomedical research”. In: *Genome Medicine* 14.1, p. 68.

Wolock, Samuel L, Romain Lopez, and Allon M Klein (2019). “Scrublet: computational identification of cell doublets in single-cell transcriptomic data”. In: *Cell systems* 8.4, pp. 281–291.

Xu, Chenling et al. (2021). “Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models”. In: *Molecular systems biology* 17.1, e9620.

Yue, Liangchen et al. (2023). “A guidebook of spatial transcriptomic technologies, data resources and analysis approaches”. In: *Computational and Structural Biotechnology Journal* 21, pp. 940–955.

Zeng, Hongkui (2022). “What is a cell type and how to define it?” In: *Cell* 185.15, pp. 2739–2755.

Zhang, Hongning et al. (2023). “SoCube: an innovative end-to-end doublet detection algorithm for analyzing scRNA-seq data”. In: *Briefings in Bioinformatics* 24.3, bbad104.

Zhang, Yifan et al. (2023). “A robust and accurate single-cell data trajectory inference method using ensemble pseudotime”. In: *BMC bioinformatics* 24.1, p. 55.

Zhou, Xiang, Kangning Dong, and Shihua Zhang (2023). “Integrating spatial transcriptomics data across different conditions, technologies and developmental stages”. In: *Nature Computational Science* 3.10, pp. 894–906.

Zhou, Zixiang et al. (2023). “Spatial transcriptomics deconvolution at single-cell resolution using Redeconve”. In: *Nature Communications* 14.1, p. 7930.

Zhu, James et al. (2023). “Mapping Cellular Interactions from Spatially Resolved Transcriptomics Data”. In: *bioRxiv*, pp. 2023–09.

Zucha, Daniel et al. (2023). “Spatiotemporal transcriptomic map of ischemic brain injury”. In: *bioRxiv*, pp. 2023–03.