SOUHRN

Prostorově rozlišená transkriptomika je inovativní technika, zkoumající genovou expresi ve vzorcích tkání při zachování prostorového kontextu. Tato metoda umožňuje studium distribuce mRNA na úrovni jednotlivých buněk nebo subcelulárně. Díky své schopnosti zachytit genovou aktivitu v prostorových souvislostech se prostorově rozlišená transkriptomika stává cenným nástrojem pro analýzu mezibuněčné komunikace, klíčové pro porozumění buněčným interakcím ve složitých tkáních.

Tato práce představuje srovnávací analýzu výpočetních metod pro inferenci mezibuněčné komunikace. Použité metody – CellChatv2, SpaTalk, NATMI a CellPhoneDB – byly aplikovány na prostorově rozlišená transkriptomická data – VisiumHD, Xenium, CosMx a MERFISH – z myšího mozku, zpracovaná pomocí standardizovaných pipeline v R a Pythonu. Hodnocením technické výkonnosti a biologické relevance metod se studie snaží objasnit silné a slabé stránky jednotlivých metod a přispět k vývoji robustnějších a komplexnějších nástrojů pro studium mezibuněčné komunikace ve složitých tkáních.

SUMMARY

Spatially resolved transcriptomics is an innovative technique that allows for the examination of gene expression within tissue samples while preserving the spatial context. This approach enables the investigation of mRNA distribution at the level of individual cells or even subcellularly. Spatially resolved transcriptomics, through its ability to capture gene activity, has become a valuable tool for analyzing cell-cell communication, which is essential for understanding cellular interactions within complex tissues.

This thesis presents a comparative analysis of computational methods for inferring cell-cell communication. The following methods–CellChatv2, SpaTalk, NATMI, and CellPhoneDB–were applied to spatially resolved transcriptomics datasets–VisiumHD, Xenium, CosMx, and MERFISH–from the mouse brain processed using standardized R and Python pipelines. By assessing the methods’ technical performance and biological relevance, this study aims to elucidate their strengths and limitations, contributing to the development of more robust and comprehensive tools for studying intercellular communication in complex tissues.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisors, Mgr. Michal Kolář, PhD (Head of the Laboratory of Genomics and Bioinformatics at the Institute of Molecular Genetics of the Czech Academy of Sciences) and Lucie Pfeiferová, M.Sc., for their invaluable advice, patience, and guidance throughout the course of this thesis.

Special thanks are due to my beloved parents and sisters for their unwavering support and for always being there to answer my countless phone calls. Finally, many thanks to my close friends for brightening my days through breaks on campus with their company–you know who you are.

TABLE OF CONTENTS

1 INTRODUCTION

Multicellular organisms are very complex, primarily due to the diversity of cell types and the necessity to coordinate their activity. Cell-cell communication (CCC) is a fundamental biological process that synchronizes cellular activity, ensuring the proper regulation of all essential functions of a cell.[1] To guarantee that every cell contributes to the organism's overall goals–survival, growth, and reproduction–unique communication networks are needed to allocate functions among various cell types, tissues, and organs.

CCC involves a range of mechanisms, such as direct contact between neighboring cells and long-distance signaling involving hormones and neurotransmitters. Through these mechanisms, cells are enabled to respond to both internal and external stimuli, regulating essential functions like differentiation, proliferation, and apoptosis. Numerous diseases, such as cancer, are linked to disruptions in these communication pathways, where abnormal signaling can lead to unregulated cell proliferation.

Recent advances in technology, including single-cell RNA sequencing (scRNA-seq) and spatially resolved transcriptomics (SRT), have deepened our ability to study CCC.[2, 3] The use of these methods allows researchers to gather gene expression information at single-cell resolution, with SRT adding spatial context, which provides unprecedented perspectives into how cells communicate within tissues. As data from these technologies grows, so does the need for computational tools that can efficiently dissect CCC. This is particularly important for understanding intricate biological systems, such as tumor microenvironments and organ development.

Despite the progress in studying CCC, our ability to infer and model these interactions still has various shortcomings. The accuracy, scalability, and integration of spatial data vary across known algorithms for the detection of CCC. Capturing the full complexity of multicellular communication, particularly in diverse tissues, remains challenging.

This thesis aims to analyze, experiment with, and compare current algorithms available for CCC detection, specifically focusing on methods using single-cell data and spatial information through SRT. The objective is to discuss the strengths and weaknesses of various existing tools, like CellChatv2[4], SpaTalk[5], NATMI[6], and CellPhoneDB[7]. Furthermore, to identify areas for improvement and suggest future directions for research. Ultimately, this review seeks to contribute to the development of more robust and comprehensive methods for studying CCC.

2 LITERATURE REVIEW

2.1 CCC Overview

CCC is a fundamental process for the functioning of multicellular organisms. While single cells like amoebae can independently perform tasks such as metabolism and movement, multicellular life relies on sophisticated coordination among specialized cells.[8] This communication enables processes like growth, immune responses, tissue repair, and development.[9] Whether through hormones in the bloodstream or neurotransmitters across synapses, these signals transform individual cellular activities into a cohesive system essential for survival.

Recent studies have shown that CCC activity is characteristic to nearly all cells in both vertebrate and invertebrate organisms.[10] This understanding underscores the regulatory roles that communication plays in development and differentiation. The evolution of CCC reflects the need for organisms to adapt and thrive. As signaling pathways became more specialized, distinct mechanisms like paracrine, endocrine, and synaptic communication emerged, enhancing functional specialization.[11]

An important mechanism of CCC involves the concept of cells releasing ligands that diffuse and bind to receptors on neighboring cells, initiating a response dependent on successful ligand-receptor (LR) interactions. Additionally, scRNA-seq and SRT provide insights into these communication networks by allowing researchers to analyze gene expression profiles, enhancing our understanding of cellular connectivity and signaling dynamics.

2.1.1 Fundament of signaling pathways

As previously mentioned, CCC is a crucial process in multicellular organisms, enabling cells to coordinate their activities and maintain homeostasis. This communication often involves signaling pathways, which are complex networks of interactions between various molecules.

The main components of cellular signaling consist of ligands and receptors.[12] Ligands are molecules designed to bind to specific receptors either on the surface or inside target cells. Ligands initiate the signaling process by binding to receptors and through that triggering a cascade of events within the cell. Receptors, on the other hand, are proteins located on the cell surface or within the cell that bind to ligands. They can be classified into several types, such as G-protein coupled receptors (GPCRs), ion channel receptors, and enzyme-linked receptors. Receptors detect and respond to ligands, converting extracellular signals into intracellular actions. In addition, signaling proteins are involved in transmitting signals from receptors to various cellular targets.

Messengers, or second messengers, are molecules relaying signals from receptors to target molecules within the cell. Common examples of these include cyclic AMP, calcium ions, and inositol triphosphate. Their role in CCC is amplifying and propagating the signal within the cell, ensuring that the initial LR interaction leads to a significant cellular response.

Furthermore, transcription factors are proteins that regulate gene expression by binding on specific DNA sequences. They are activated by signaling pathways and control the transcription of target genes. Transcription factors translate extracellular signals into changes in gene expression, affecting cell behavior and function. This is vital for processes like differentiation and apoptosis.

2.1.2 Types of CCC

Cell-cell interaction mechanisms can be divided into categories–autocrine, paracrine, endocrine, and synaptic communication.[13] The differences between these signaling methods are the distance over which signaling occurs and the method of signal delivery.[2] Autocrine signaling occurs when a cell releases a ligand that binds to its own receptors, allowing the cell to regulate its own behavior. On the other hand, paracrine signaling consists of cells that are near one another communicating through the release of chemical messengers. This type of communication is vital in tissue environments in which local signals coordinate functions such as wound healing and immune responses. Last, endocrine signaling is characterized by the release of hormones into the bloodstream, allowing signals to reach distant cells.

2.1.3 Major signaling pathways

CCC relies on several key signaling pathways that regulate various cellular processes. Here are some of the major pathways. The Kinase (RTK) Pathway is involved in the regulation of cell growth, differentiation, and metabolism.[14] When ligands such as growth factors bind to RTKs, they activate intracellular signaling cascades. Second, GPCR Pathway.[15] GPCRs are involved in transmitting signals from a variety of stimuli, including hormones and neurotransmitters. They activate G-proteins, which then trigger various downstream effects. Notch Signaling Pathway is essential for cell differentiation, proliferation, and apoptosis.[16] It involves direct CCC through LR interactions on adjacent cells. Furthermore, the Wnt Signaling Pathway is critical for embryonic development and tissue homeostasis.[17] It regulates gene expression by stabilizing β-catenin, which enters the nucleus and influences transcription. Additionally, the Mitogen-Activated Protein Kinase Pathway is a vital signaling mechanism that transmits extracellular information to the nucleus and induces cellular responses, like proliferation, differentiation, and stress responses.

2.1.4 CCC in Development, Homeostasis, and Disease

CCC is vital for the proper functioning of multicellular organisms. It plays a significant role in development, maintaining homeostasis, and the progression of diseases.

During cellular development, CCC coordinates the activities of different cell types throughout embryogenesis, ensuring proper tissue and organ formation.[18] Signaling pathways like Notch and Wnt are crucial for guiding cell differentiation and morphogenesis. In adult organisms, CCC maintains homeostasis–a stable internal environment of an organism–by regulating tissue repair, immune responses, and metabolic balance. For example, the RTK pathway helps regulate cell growth and survival, ensuring tissue integrity. On the other hand, disruptions in CCC can lead to various diseases, including cancer, autoimmune disorders, and neurodegenerative diseases.[18] Abnormal signaling can result in uncontrolled cell proliferation, immune system malfunctions, and impaired cell communication.

Therefore, understanding CCC is crucial for developing targeted therapies.[18, 19] By identifying and modulating specific signaling pathways, researchers can design treatments for diseases caused by communication breakdowns. Studying CCC not only provides insights into fundamental biological processes but also opens avenues for new therapeutic strategies, making it a critical area of research.

2.2 Technological Advancements

2.2.1 scRNA-seq

With the introduction of scRNA-seq, researchers gained the ability to investigate gene expressions at the individual cell level, providing valuable insights into communication patterns based on the presence of ligands and receptors across different cells.[2] This technology is essential for studying cellular interactions, particularly in heterogeneous environments like tumors, where it helps to elucidate how tumor cells communicate with immune cells, stromal cells, and other surrounding cells, thus facilitating disease diagnosis and progression understanding.

The scRNA-seq process begins with the isolation of individual cells from tissue samples, which are then converted into a sequenceable form. This generates an RNA profile for each cell, detailing which genes are active, ultimately allowing researchers to identify which cells express specific ligands and their corresponding receptors. This data is crucial for computational methods such as CellChatv2 and NicheNet, which are specifically designed to analyze cell-to-cell communication networks using scRNA-seq data.[20]

One of the most widely used techniques for scRNA-seq preparation is the 10× Genomics workflow.[21] In simple terms, the 10× scRNA-seq process begins with tissue dissociation and encapsulation of individual cells.[22] The microfluidic device partitions individual cells into tiny droplets containing ideally just one cell. The cells are then lysed within the droplets, and mRNA is equipped with barcoded primers that contain unique identifiers. This mRNA is subsequently converted to cDNA through the reverse transcription process. The cDNA from all cells is pooled and amplified via polymerase chain reaction, preserving the cell identity due to barcoding. Finally, the resulting library is sequenced.

Recent advances in computational algorithms tailored for analyzing CCC through scRNA-seq data have opened new avenues for research. Studies have highlighted the importance of single-cell analysis in revealing cellular heterogeneity and tracking gene expression dynamically.[23] Furthermore, improvements in sensitivity and throughput have been achieved, though ongoing challenges remain in analysis and integration with other technologies.[24]

Nevertheless, scRNA-seq has inherent limitations, particularly its lack of spatial context. While it provides comprehensive gene expression profiles, it does not capture the spatial organization of cells within tissues, which is critical for understanding the dynamics of cell communication. Inferring communication networks accurately requires integrating this extensive data while considering spatial information, which remains a significant challenge in the field.[25]

2.2.2 Spatially Resolved Transcriptomics

SRT is a novel technique enabling the study of gene expression in relation to tissue architecture. The gene expression information is obtained from intact tissue sections in the original physiological context at a spatial resolution.[3, 26] This approach allows researchers to elucidate CCC and understand the functional roles of specific cell types in their native environments.

Methodological Approaches

SRT methods can be categorized into two primary groups: sequencing-based and imaging-based techniques. Sequencing-based methods utilize next-generation sequencing after integrating position-specific barcodes to capture both spatial and mRNA abundance data.[27] In contrast, imaging-based methods involve the application of fluorescent markers either through direct insertion or hybridization.[28]

SRT aims to count the number of transcripts of a gene at distinct spatial locations in a tissue. Different techniques have different technical parameters.[29] The tissue size can vary from a small (<1mm^2) section to whole organ sections from model organisms. The number of genes counted varies from tens to thousands or even the whole genome. A spatial location may range from a whole tissue domain to a large 500 μm × 500 μm region of interest, down to a single cell or even subcellularly. With current technologies, there is often a trade-off between the number of genes profiled and the efficiency of the technique—the proportion of transcripts of interest that are successfully counted, ranging from nearly 100% to as low as 1%.

10× Genomics Visium

10× Genomics Visium is one of the most widely used spot-based, sequencing-based SRT methods.[29-31] It builds upon the original SRT assay, offering improved sensitivity and resolution.[3] The method relies on spatially barcoded oligonucleotide probes printed onto the surface of glass slides, allowing the capture of mRNA from tissue sections while preserving spatial information.

Each Visium Spatial Gene Expression slide contains four capture areas, with approximately 5,000 barcoded spots per area. Each spot has a diameter of 55 μm, with a center-to-center distance of 100 μm, and is designed in a staggered pattern to minimize spacing. Each spot captures mRNA from approximately 1 to 10 cells, providing near single-cell resolution while maintaining high transcript coverage. This technology has been widely adopted in various fields, including cancer research, neuroscience, and developmental biology.

Visium HD is an advanced version of the classic Visium platform, designed to achieve significantly higher spatial resolution. Unlike the original Visium assay, which captures mRNA from multiple cells per spot, Visium HD features much smaller spots**,** enabling true single-cell resolution or even subcellular resolution in some cases. The increased spot density allows for a more detailed spatial transcriptomic map, making it particularly useful for studying complex tissue architectures or heterogeneous microenvironments. With Visium HD, the number of spots per capture area is dramatically increased, reducing the distance between spots and allowing for finer-grained spatial mapping of gene expression.

10× Genomics Xenium

Xenium is a high-resolution imaging-based in situ spatial profiling technology from 10× Genomics that allows for simultaneous expression analysis of RNA targets (currently in the range of thousands) within the same tissue section.[31, 32] Xenium starts with Formalin-Fixed Paraffin-Embedded (FFPE) tissue or fresh frozen sections on specific Xenium slides. The sectioned tissue is processed, and selected probes are hybridized to the RNA targets within the tissue. Ligated bound probes create a circular template that facilitates rolling circle amplification of the probe and a distinct barcode. On the Xenium platform, the tissue with amplified probe products is decoded through a series of cycles and onboard image data processing that allows identification and location of each transcript target. The resulting data is immediately viewable and is compatible with an easy-to-use data viewer and with many third-party downstream analysis tools, which continue to develop rapidly.

Vizgen MERSCOPE

The MERSCOPE platform is an in situ spatial genomics technology based on MERFISH (Multiplexed Error-Robust Fluorescence In Situ Hybridization), designed to integrate single-cell and spatial gene expression profiling.[33] MERFISH facilitates the detection and quantification of hundreds of gene targets within intact tissue sections, leading to high multiplexing capacity with subcellular resolution. This instrument uses error-robust barcoding mechanisms enabling the detection and correction of certain imaging errors, increasing the reliability and specificity of transcript identification even in noisy tissue regions.

The MERSCOPE workflow is as follows. Tissue samples are sectioned and positioned on the appropriate MERSCOPE slide, which resembles a cover slip-like slide and is designed to provide optimum optical clarity. The order of workflow operations varies slightly depending on the sample type, but major steps are synonymous: sectioning, hybridization, gel embedding, and clearing, followed by imaging. After sample preparation, the slide is placed into the MERSCOPE instrument's flow chamber, where the system automates imaging and analysis, including cyclic hybridization with reporter probes. A low-magnification (10×) mosaic of the tissue section is displayed, enabling users to choose specific areas or the entire tissue for imaging. The system can capture up to one square centimeter in total, either as a single continuous area or multiple regions. Following imaging, the data undergoes automatic cell segmentation and transcript analysis. The entire sample preparation process typically takes three to four days, though most of this time is spent on incubation, with only three to four hours of hands-on work. Imaging lasts approximately 24 to 30 hours, depending on panel size, followed by another 24 to 30 hours of analysis. Throughput is optimized by loading the next sample for imaging while the previous dataset is being analyzed.

NanoString CosMx

NanoString's newer system, CosMx spatial molecular imager (CosMx SMI), is an imaging-based SRT and proteomics platform. Currently, CosMx SMI supports simultaneous imaging and quantification of more than 1000 RNA and 64 protein targets at subcellular resolution.[34]

This technology uses a hybridization strategy involving five gene-specific probes per target, each with a unique target-binding domain and a shared readout domain composed of 16 sub-domains, and their signal is imaged before being cleaved by UV light.[35]

This cycle is repeated 16 times, with the specific combination of fluorescent signals uniquely identifying each transcript. The design is optimized for FFPE samples, enhancing sensitivity in degraded RNA contexts. Although CosMx provides excellent spatial resolution and detection sensitivity, it is limited to pre-defined gene panels and constrained by challenges such as optical crowding, long imaging times, and small imaging areas. Nonetheless, CosMx SMI is among the most advanced platforms for targeted spatial profiling in intact tissue sections.

Spatial Clustering and Deconvolution

Table 2.1: Differences in Sequencing-Based and Imaging-Based SRT.

Feature	**Sequencing-Based Low-Definition (Visium)**	**Sequencing-Based High-Definition (VisiumHD)**	**Imaging-Based (Xenium, MERSCOPE, CosMx)**
Resolution	Spot-level (~1–10 cells per spot)	Single-cellular or subcellular	Single-cell or subcellular

Table 2.1 - continued

Feature	Sequencing-Based Low-Definition (Visium)	Sequencing-Based High-Definition (VisiumHD)	Imaging-Based (Xenium, MERSCOPE, CosMx)
Data Type	RNA sequencing reads		Fluorescence microscopy images
Preprocessing	Read alignment, barcode assignment		Background correction, transcript detection
Gene Coverage	Transcriptome-wide (unbiased)		Targeted panel
Cell Segmentation	Not needed (multi-cell spots)	Required (each spot may correspond to a single cell or part of one)	Required to define cell boundaries
Deconvolution	Needed to infer cell types per spot	Less needed or not needed (near single-cell resolution)	Not needed (single-cell resolution)

2.3 Computational Methods for CCC

According to recent studies, there are over one hundred tools for communication inference with different localization, scalability, accuracy, sensitivity, and performance.[36] This research will only focus on selected tools CellChatv2[4], NicheNet[37], SpaTalk[5], LIANA+[38], NATMI[6], and CellPhoneDB[7].

Computational tools have been developed for CCC inference from scRNA-seq at both the individual cell and cell cluster levels.[39, 40] Due to the simplicity of their acquirement and extensive literature-curated databases, the main concept of algorithms identifying communication between cells is the analysis of expression levels of a ligand and receptor in the corresponding cells. These levels are used to measure communication strength.

CellChatv2

CellChatv2 is a computational method that allows to infer, quantify, and analyze cell-to-cell interaction networks using scRNA-seq data.[4] The tool, implemented in R, operates by leveraging a comprehensive LR interaction database, CellChatDB, which integrates known molecular interactions, including multimeric LR complexes and cofactors such as soluble agonists, antagonists, and co-stimulatory or co-inhibitory receptors. This database serves as the foundation of CellChatv2, encompassing over 2,000 validated interactions, with a significant proportion involving secreted interactions and heteromeric molecular complexes, ensuring an accurate representation of intercellular signaling.

The CellChatv2 analysis pipeline begins with the construction of interaction networks, where statistical methods such as permutation tests identify significant cell-type interactions.[4, 41] It then models these interactions by creating a network in which nodes represent cell types and weighted edges indicate the probability of communication. By measuring the expression levels of interacting molecules, CellChatv2 quantitatively assesses each cell's signaling input and output, providing a comprehensive view of intercellular communication dynamics.

To gain deeper insights into signaling organization, CellChatv2 applies graph theory, pattern recognition, and manifold learning techniques, which help distinguish between two categories of signaling pathways - conserved and context-specific categories.[4] Conserved pathways are common across biological contexts, while context-specific pathways highlight interactions unique to particular developmental stages or disease conditions. These classifications facilitate a more nuanced understanding of how signaling mechanisms adapt across different biological environments.

NicheNet

NicheNet, also R-based, provides a mechanistic approach to CCC, which, unlike many other tools, only identifies co-expressed LR pairs.[37] NicheNet goes further by estimating the downstream regulatory effects of LR interactions. At its core, NicheNet relies on a weighted prior knowledge model integrating LR interactions, intracellular signaling pathways, and transcription factor target regulation. This enables the tool to go beyond the cell surface, tracing signaling cascades from ligand binding to gene expression changes. NicheNet applies network propagation to compute a ligand activity score and a regulatory potential score, prioritizing ligands based on their inferred ability to explain observed differential gene expression in the receiver cell type.

This approach allows NicheNet to generate hypotheses about which ligands are likely to drive transcriptional changes, what genes are targeted, and what signaling intermediates are involved.[37] This makes NicheNet particularly valuable for studies aiming to uncover functional regulatory mechanisms of CCC, offering mechanistic insights that go beyond correlation.

SpaTalk

SpaTalk is a recent graph-based tool designed for cell-cell communication inference developed in 2022.[5] It uniquely integrates scRNA-seq data with SRT data, filling a niche in CCC analysis that allows direct spatial resolution of cellular interactions.

SpaTalk's core methodology relies on K-Nearest Neighbors graph construction, mapping cells based on their Euclidean distances. This mapping identifies LR pairs between spatially close cells, and subsequently, SpaTalk scores these pairs based on their influence on receptor-expressing cells. The scoring increases with the number of transcription factors and terminal target genes activated in the receiving cells, thus emphasizing interactions that trigger downstream gene expression.

SpaTalk's workflow involves two main steps. First, it identifies cell-type composition in spatial transcriptomic data using a non-negative linear model to deconvolve spatial transcriptomic matrices, taking scRNA-seq data as a reference. For spot-based data, cell types are projected spatially to reconstruct single-cell resolution within each spatial spot, through sampling and iteration.

The second step focuses on CCC inference, beginning with constructing a cell graph network via K-Nearest Neighbors. Here, ligand and receptor pairs are identified within the graph, with a permutation test applied to assess the statistical significance of each LR interaction. The process continues by modeling intracellular signaling from receptors to transcriptomic factors and downstream target genes using a knowledge graph composed of pathways and transcription factors. A random walk algorithm traverses the knowledge graph, producing intracellular scores derived from co-expressed transcriptomic factors and targets. Ultimately, SpaTalk uses both inter- and intracellular scores to highlight spatially relevant interactions.

LIANA+

LIANA+ is a Python-based framework that unifies and extends various existing CCC inference tools.[38] It supports multi-omics and spatial data inputs in AnnData or MuData formats. LIANA+ offers both hypothesis-driven and hypothesis-free methods for CCC inference and includes options such as Tensor-cell2cell and PyDESeq2. A key strength of LIANA+ is its ability to integrate information from multiple knowledge bases, including OmniPath and BioCypher, as well as a wide range of database resources like ConnectomeDB, CellTalkDB, CellPhoneDB, CellChatDB, and others. Users can leverage a variety of CCC inference methods within the framework, such as CellPhoneDB, Connectome, CellChatv2, NATMI, logFC, SingleCellSignalR, Crosstalk, and Scores.

Beyond its comprehensive method integration, LIANA+ offers features like causal subnetwork analysis and multi-view learning, enabling researchers to decode coordinated inter- and intracellular signaling events across different biological contexts. It is a scalable and flexible framework, providing synergistic tools to study CCC through diverse molecular mediators–including those captured in multi-omics and spatially resolved data–making it a powerful resource for exploring complex cellular interactions.

NATMI

In 2020, NATMI (Network Analysis Toolkit for Multicellular Interactions) was introduced, a tool implemented in Python that is user-friendly and primarily designed to analyze single-cell gene expression data but can also process bulk transcriptomics and proteomics data.[6]

NATMI works with LR interactions by creating a network in which nodes represent cell types and directed, weighted edges represent specific LR pairs connecting these cell types. NATMI assesses the expression levels of ligands and receptors to determine each edge weight and infers the likelihood and strength of intercellular interactions.[5, 42] ConnectomeDB2020 serves as a database for NATMI, which includes 2,293 manually curated LR pairs. To enhance specificity, this method only considers a ligand or receptor as expressed if it is present in at least 20% of the cells within a given cell type. This threshold helps ensure that detected interactions are biologically meaningful. Based on edge weights, NATMI constructs a cell connectivity summary network, highlighting cellular communication interactions for further analysis. The network can be customized by setting user-defined thresholds for interaction strength, allowing researchers to focus on the most significant interactions.

For further analysis, the NATMI implementation available within the LIANA+ framework was used.

CellPhoneDB

CellPhoneDB, another Python tool, is one of the first methods to account for the multimeric nature of LR complexes.[7, 43] It uses a curated database with detailed annotations and performs permutation-based testing to identify significant interactions. Its default LR database is based on human interactions and curated human gene annotations. To use CellPhoneDB for other organisms, it is possible to either use a custom database or map gene symbols to those used in the database. Significant LR interaction pairs are then defined based on the likelihood estimation of their respective cell-type enrichment. The number of significant LR interaction pairs is used to prioritize the cell-cell interactions (CCIs) specific to two given cell types. The networks based on CCIs can be constructed to assess cellular crosstalk between varopis cell types. The cell subsampling technique is used to minimize memory usage and runtime. Outputs include mean expression values and p-values for each LR interaction between cell-type pairs.

Comparative Analysis

While the initial comparative analysis included a broader range of CCC detection methods, the focus was subsequently refined to encompass only those tools that are directly applicable to the practical part of this study. NicheNet was excluded from the final comparison, as its methodological framework was not comparable within the context of the experimental analyses conducted. Consequently, **Tab 2.2** presents a comparative overview of CellChatv2, SpaTalk, NATMI, and CellPhoneDB, which represent the set of tools selected for the experimental part of this thesis. The listed strengths and limitations in **Tab 2.2** are based on information from literature reviews that compare different CCC tools.

Table 2.2: Comparison of CCC Detection Methods: CellChatv2, SpaTalk, NATMI, and CellPhoneDB.

Feature	**CellChatv2**	**SpaTalk**	**NATMI**	**CellPhoneDB**
Programming Language	R	R	Python	Python
Used Version	CellChat 2.1.0	SpaTalk v1.0	via LIANA+ v1.5.1	CellPhoneDB v2.1.17
Incorporation of Spatial Data	Yes (spatial integration in v2)	Yes	No	No
Analysis level	Cluster-level	Single-cell level (with spatial proximity consideration)	Cell-type level (depends on input data)	Cluster-level

Table 2.2 - continued

Feature	**CellChatv2**	**SpaTalk**	**NATMI**	**CellPhoneDB**
Default database	CellChatDB (human and mouse)	Default LR list + downstream signaling activity	ConnectomeDB2020 (human) or HomoloGene Database (21 species)	Predefined LR database (human only)
Database customization	Supports custom databases	Supports custom databases	Limited user customization	Supports user customization
Customizability	Modular, but clustering-based resolution limits flexibility	High flexibility with downstream pathway modeling	High flexibility (thresholds for interaction weights)	Limited (fixed LR database and pipeline)
Performance in Spatial Data	Strong (with spatially organized tissues)	High (particularly across various spatial transcriptomics methods)	Not applicable	Not applicable
Strengths[20, 44]	Good for broad CCC analysis across large datasets	Strong spatial applicability, supports visualization of CCC	Easy to use, flexible, customizable	Rigorous statistical testing of LR pairs
Limitations[20, 44]	Clustering may obscure cellular heterogeneity	Performance varies with gene coverage	Limited adaptability (no spatial data, pre-defined database)	No modeling of downstream effects; limited customization

3 DATA AND METHODS

3.1 Methodological approach

In this study, we analyze and compare four computational methods designed to detect cell-cell interactions from single-cell and SRT data: CellChatv2[4], SpaTalk[5], NATMI[6] (via LIANA+[38] framework), and CellPhoneDB[7]. The objective is to assess their performance and suitability for CCC analysis across different datasets and provide insights into their practical and biological outputs.

The comparison is based on two main aspects: technical performance and biological relevance. Technical features include practical considerations, such as installation and setup requirements, computational efficiency–runtime, and memory usage. Biological relevance is evaluated through metrics such as the number of detected inferred interactions, ligands, and receptors. Moreover, the detected interactions are cross-validated across methods applied to the same datasets to assess consistency.

These methods are applied to four distinct datasets with the aim of generalizing and highlighting the respective strengths and limitations of each method in different experimental contexts. The goal of this comparison is to provide practical recommendations for researchers selecting CCC inference tools, as well as to contribute to a broader understanding of how methodological choices influence downstream biological interpretation.

3.2 Datasets

Four spatial transcriptomics datasets derived from the mouse brain were used for the analysis. The datasets used in this study were obtained from Lucie Pfeiferová, Institute of Molecular Genetics of the Czech Academy of Sciences. The datasets were preprocessed both in R and Python environment, with details below.

3.2.1 Processing in R

All datasets were processed in R (v4.4.3) package Seurat (v5.1.0), based on official Seurat tutorials. For individual cell annotations, R Bioconductor (3.20) SingleR (2.20) package was selected. As a reference scRNA-seq was used mouse cortex from Allen Brain Institute (https://www.dropbox.com/s/cuowvm4vrf65pvq/allen_cortex.rds?dl=1).

Both Visium HD and Xenium datasets were obtained from official 10× Genomics datasets. Due to its high resolution and large size, the VisiumHD dataset posed particular computational challenges. For this reason, only 8μm layer was used.

The CosMx dataset, downloaded from the official CosMx website, represents a single tissue section from a larger CosMx experiment. The dataset was subsetted for a specific slide (slide_ID_numeric == 2).

MERFISH experiment was obtained from the CZ CellXGene website as Seurat object, as a part of analysis from Allen Brain Institue. A specific brain section (C57BL6J-2.030) was selected for further analysis. This dataset was already annotated by the Allan Brain Consortium.

3.2.1 Processing in Python

All the datasets mentioned above were independently preprocessed using Python 3.10, following a standard pipeline based on publicly available tutorials. The preprocessing was performed with the help of the SpatialData (0.4.0) and scanpy (1.10.0)/squidpy (1.6.2) libraries. With the exception of the MERFISH experiment, which was obtained as annotated anndata object from CZ CellXGene (from the same place as the Seurat object, using the same section), all the objects were annotated using R SingleR package (mentioned above).

3.3 Methods

This comparative analysis includes four computational CCC tools–CellChatv2 (v2.1.0), SpaTalk (v1.0), NATMI (via LIANA+ v1.5.1), and CellPhoneDB (v2.1.17). To ensure standardized conditions and reproducibility, all methods used default parameters, unless explicitly stated otherwise. To maintain practical feasibility and timely progress during the analysis, we imposed a computational time limit of 16 hours per method. This decision was made to avoid excessive waiting times and to maintain a consistent and reasonable evaluation period for each CCC analysis tool.

CellChatv2

The CellChatv2 analysis was performed in R (v4.2.2) using RStudio. Required packages included Seurat (v5.1.0), devtools (v2.4.5), and CellChatv2 (v2.1.2), along with Bioconductor dependencies (BiocManager (v1.30.25), BiocGenerics (v1.52.0), BiocNeighbors (v2.0.1), and Biobase (v2.66.0)). Installation was completed iteratively to resolve dependencies.

For each dataset, a Seurat object was loaded (RDS file), obtaining a table of normalized values, cell types, and spatial coordinates. Next, two spatial parameters were defined: the *conversion.factor* was set to 1 ensuring that the spatial data was on the correct scale for analysis. The parameter *spot.size* defining the size of an average cell in the spatial transcriptomics data is set to 10. Furthermore, we created a metadata dataframe with labels, group information, sample names, data type, coordinates, and spatial factors.

The CellChatv2 workflow consists of creating a CellChatv2 object using the prepared data and metadata. Afterward, the database was subsetted to include only relevant LR interactions. Finally, a function identifies over-expressed genes and interactions within the dataset. The *computeCommunProb* function is used to compute the communication probability, which is a crucial step in identifying significant cell-cell interactions. Key parameters are summarized in **Tab 3.1**. The only parameter that varied in different dataset analyses is the *scale.distance* which scales the spatial distances to ensure they are in a suitable range for analysis. Lastly, the interaction data is extracted and filtered to include only cell-cell contact interactions. This data is then saved to an output CSV file for further analysis.

Table 3.1: Key parameters for CellChatv2 analysis in *computeCommunProb* function.

Parameter	Value	Purpose
method	truncatedMean	Reduces the impact of outliers when computing communication probabilities.
trim	0.1	Indicates the proportion of extreme values to be trimmed from each end of the distribution before computing the mean.
interaction.range	250 µm	Maximum distance within which interactions are considered.
contact.range	10 µm	Maximum distance for contact-dependent interactions ensuring that only close-range interactions are considered.

The CellChatv2 function was applied to four spatial transcriptomics datasets, including VisiumHD, Xenium, CosMx, and Merfish.

Table 3.2: Dataset-specific settings for CellChatv2 analysis,

Dataset	scale.distance	Notes
VisiumHD	5.4	Intersection of cell types and spatial coordinates.
Xenium	5.4	–
CosMx	40	Higher scale.distance needed due to small spatial distances.
MERFISH	5.4	Subsetted to specific brain sections.

As seen in **Tab 3.2** the VisiumHD dataset featured a relatively high spatial resolution (8 µm bin size), but the spots covered multiple cells rather than true single-cell resolution. This mismatch necessitated extra preprocessing steps, specifically intersecting cell and spot names to ensure alignment between the Seurat object and spatial metadata.

On the other hand, CosMx data required specific adjustments. Initial attempts using a *scale.distance* of 5.4 µm did not yield meaningful interactions in CellChatv2. This likely stemmed from CosMx's ultra-high resolution and tightly packed spatial coordinates, which led us to increase the scale.distance to 40 µm.

Both the Xenium and MERFISH datasets allowed straightforward extraction of cell type and positional data, without major preprocessing hurdles.

SpaTalk

Further, we used the SpaTalk R package to infer CCC interactions from the same four spatial transcriptomics datasets. The workflow began with the loading of Seurat objects, from which we extracted key metadata, including x-y spatial coordinates and cell type annotations necessary for spatially resolved CCC analysis.

Next, expression matrices were prepared from the assay data of each dataset to serve as input for SpaTalk. The *createSpaTalk* function was used to build SpaTalk objects, with parameters specifying the species ("mouse") and providing cell type information.

The core analytical step was performed using the *find_lr_path* function, which identifies LR pairs and their associated signaling pathways. This function requires a predefined list of LR pairs and pathway definitions, which were supplied accordingly. Once potential interactions were identified, we applied the *dec_cci_all* function, which decodes and infers global patterns of CCC across the tissue based on the detected LR interactions.

Finally, the results were prepared for downstream analysis and visualization by saving the output in RDS containing all relevant interaction data and facilitating easy integration into subsequent comparative and visualization workflows.

NATMI

The NATMI method is implemented using the LIANA+ Python package to infer CCC interactions across the individual datasets. For each dataset, gene expression data and associated metadata are loaded from a .h5ad file as an AnnData object. Depending on the characteristics of the datasets, the raw expression data were normalized and log-transformed. Normalization scales total counts to a fixed target and Log Transformation applies log1p to the normalized data. The raw data is copied to a layer called *raw* and transformed data is stored in a new layer *lognorm*.

The NATMI method requires a resource of LR pairs, which is selected from the *liana.resource* as mouseconsensus. A filtering step ensures that only LR pairs where both genes are present in the dataset are retained. NATMI is then run on the preprocessed data with specific settings such as grouping column, gene layer, expression proportion, and resource. The results of the NATMI analysis are saved as CSV files.

CellPhoneDB

Next, we applied the CellPhoneDB framework using the Python package. The analytical workflow consisted of the following main steps. First, raw expression matrices were loaded as AnnData objects using *scanpy*. The datasets were normalized to a fixed total count and log-transformed to prepare the expression values for downstream analysis. Since CellPhoneDB operates on human LR pairs, we used the *mousipy* library to map mouse gene symbols to their human orthologs. Genes without clear orthologs were excluded to ensure compatibility with the CellPhoneDB database. Cell-level metadata (cell type annotations and IDs) were extracted into a meta file (meta.txt) and exported the expression matrix in counts format (HDF5), both required inputs for CellPhoneDB. The CellPhoneDB statistical analysis is run with a result of identification of statistically significant LR interactions across cell-type pairs, based on random permutation testing. The results including interaction matrices, mean expression values, and p-values were saved to dataset-specific output directories for further interpretation.

3.4 Use of AI

To enhance analytical accuracy, the AI model Chat GPT-4o[45] was used as a supplementary tool throughout the study. It was used to get feedback on readability and language. In the writing of the literature review part, ChatGPT-4o was used to rephrase complex ideas for improved clarity, summarize academic texts, and suggest the structure of paragraphs. Moreover, during the writing process of the Results and Discussion chapters, the AI was used to support the comparison identified by the researcher, offering a more detached perspective and helping to highlight potential patterns or trends in the data.

4 RESULTS AND DISCUSSION

In this section, we compare the performance and output of four computational methods–namely CellChatv2, SpaTalk, NATMI, and CellPhoneDB–for inferring CCC across four mouse brain SRT datasets–VisiumHD, Xenium, CosMx, and MERFISH. The performance of each method is evaluated based on technical efficiency, as well as biological output. The biological output assessment focuses on the detection of interacting cell type pairs, which are often abbreviated by the respective algorithms. The full names corresponding to these cell type abbreviations can be found in appendices **Tab. A**.

4.1 Results of CCC analysis per dataset

VisiumHD

Starting with the technical features of the VisiumHD dataset analysis, the methods differed considerably in runtime and memory requirements. To complete the analysis, CellChatv2 required 1 hour and 36 minutes, being the slowest method. In comparison, CellPhoneDB completed the analysis in 24 minutes, while NATMI was the fastest, finishing the task in just 22 seconds. SpaTalk, however, failed to complete the analysis due to the large size of the dataset. During execution, it attempted to allocate 55.9 GiB of memory, which exceeded system limitations and caused the process to terminate. Other methods also varied in memory usage. CellChatv2 needed approximately 3.5 GiB of memory, whereas NATMI and CellPhoneDB performed the analysis without any memory-related issues.

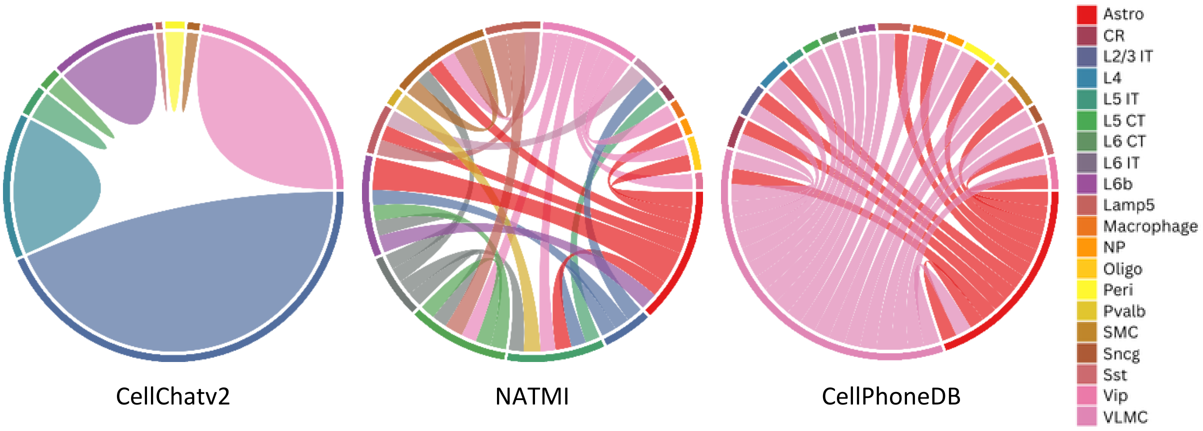

Figure 4.1: Circle plots demonstrating detected cell type pair interactions by CellChatv2, NATMI, and CellPhoneDB for the VisiumHD dataset.

As we see in **Fig. 4.1**, we visualize circle plots, also known as chord diagrams, to show detected interactions across methods, showing interactions among the first 30 most probable cell type pairs. The likelihood of these interactions was measured with the interaction score of each method: CellChatv2 used *prob*, SpaTalk *score*, NATMI *spec_weight*, and CellPhoneDB *lr_means*. In each plot, cell types are represented as segments around the circle, and curved lines between segments indicate detected cell type interactions between sender and receiver cell types.

Although the three circle plots appear very different, these visual discrepancies are largely influenced by differences in interaction scores. The variation in appearance does not necessarily indicate a lack of overlap in underlying interactions. Notably, CellChatv2's highest-scoring interactions are found within the same cell types. On the other hand, NATMI and CellPhoneDB show a wide range of interactions spanning diverse cell types. CellPhoneDB, in particular, detected distinctly more interactions involving the sender cell types astrocytes (Astro) and vascular leptomeningeal cells (VLMC). Interestingly, both NATMI and CellPhoneDB independently identified Astro as major communication hubs, suggesting a degree of biological agreement between these two methods.

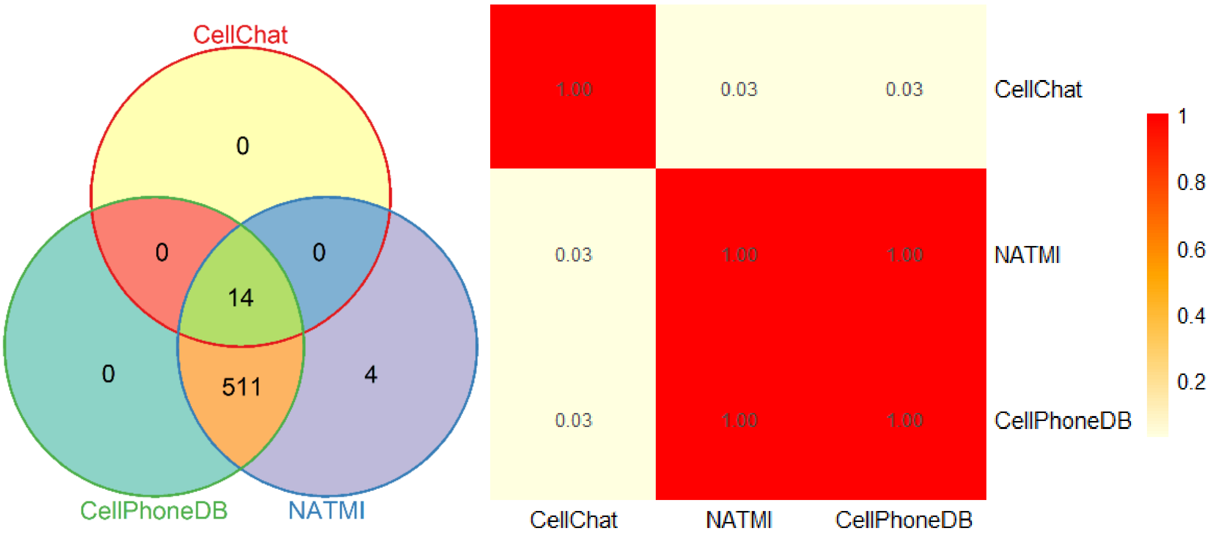

Figure 4.2: Venn diagram and correlation matrix of detected cell type interactions for the VisiumHD dataset across methods: CellChatv2, NATMI, and CellPhoneDB.

For a comparison of inferred interactions, a Venn diagram was constructed based on the amount of cell type pairs identified by each method. According to the diagram in **Fig. 4.2**, CellChatv2 detected only 14 cell type interactions, all of which were also found by NATMI and CellPhoneDB. In contrast, the intersection between NATMI and CellPhoneDB showed a high concordance in both detected interactions and underlying database content. This overlap is further reflected in the correlation matrix of interaction scores, where CellChatv2 shows minimal correlation with other methods. NATMI with CellPhoneDB, on the other hand, showed strong correlation, both in terms of the LR pairs identified and the associated scores.

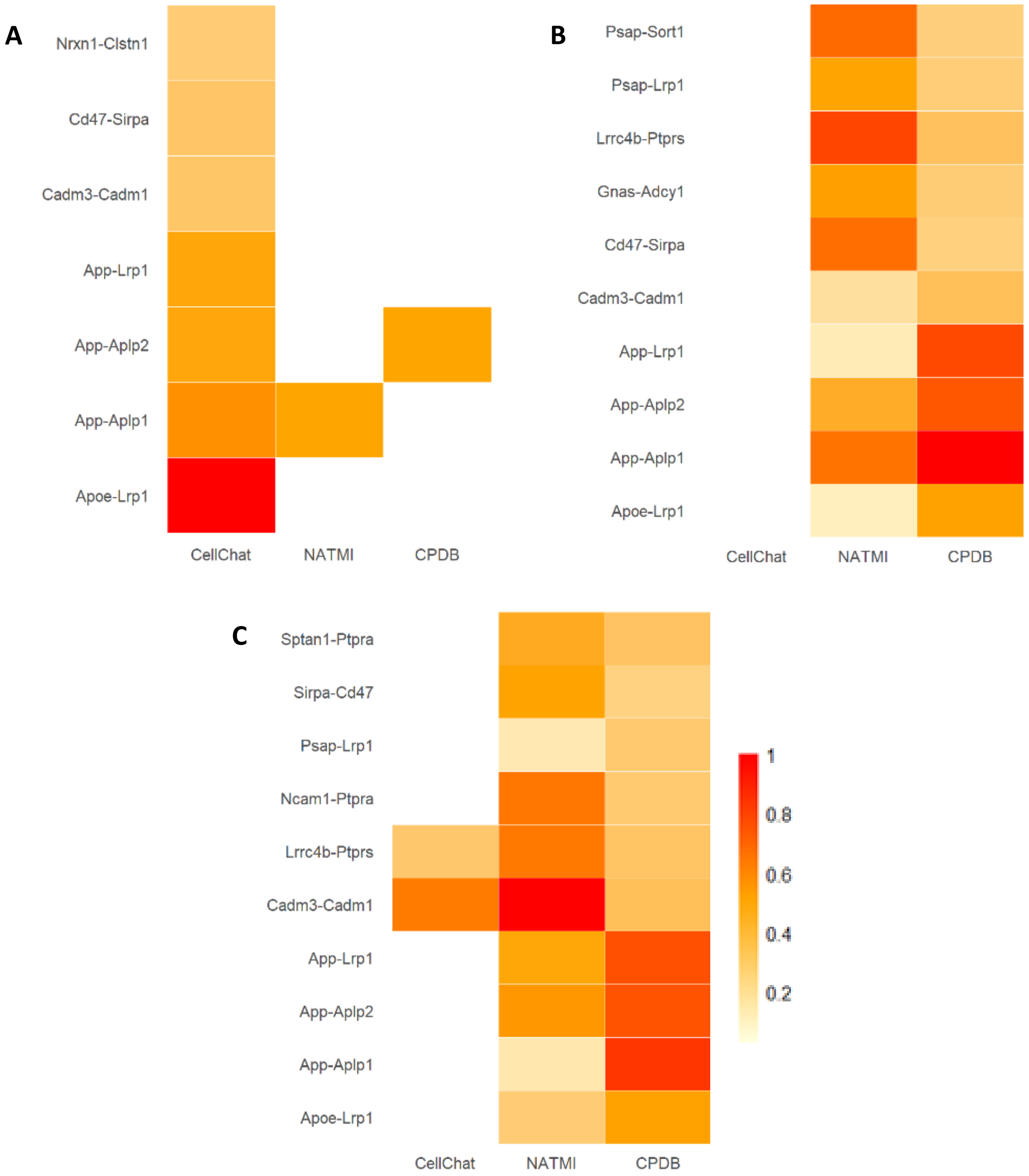

Figure 4.3: Correlation heatmap between detected LR pairs for the VisiumHD dataset across the methods. The cell type interactions chosen were: **A** for CellChatv2, L6 IT-L6IT, **B** for NATMI, Sst-Sst, and **C** CellPhoneDB (CPDB), L2/3 IT-L2/3 IT.

To further explore the correlation between methods, we examine interactions at a more granular, protein-level resolution. For each cell type pair, multiple LR pairs were identified–ranging from single pairs to over 20. To assess the overlap in LR pairs detected by different methods, a heatmap seen in **Fig. 4.3** was constructed based on one representative (highest-scoring) interaction per method. For CellChatv2 interaction L6 IT-L6 IT, an interaction between intratelencephalic neurons, we observe one overlapping LR pair with each NATMI and CellPhoneDB. In the case of NATMI, the selected interaction Sst-Sst, two somatostatin-expressing interneurons, shows full overlap with CellPhoneDB but none with CellChatv2. For CellPhoneDB, the chosen intratelencephalic neuron interaction L2/3 IT-L2/3 IT has complete overlap with NATMI and shares two LR pairs with CellChatv2.

In summary, the results from the analysis of VisiumHD dataset indicates both technical and biological divergence between CCC methods. While NATMI and CellPhoneDB produced fast, consistent results with significant interaction overlap, CellChatv2's results were more limited and methodologically distinct. SpaTalk, due to its failure to produce results, could not be assessed in this context.

Xenium

Furthermore, we evaluated the technical performance and biological interpretation of CellChatv2, SpaTalk, NATMI, and CellPhoneDB on the Xenium dataset.

The runtime varied considerably across methods. SpaTalk was the slowest, requiring approximately 30 minutes to complete. CellChatv2 followed, with a runtime of 15 minutes. In comparison, NATMI and CellPhoneDB were markedly faster, finishing the analysis in 11 seconds and 21 seconds, respectively. None of the methods encountered memory issues, demonstrating good scalability for this dataset.

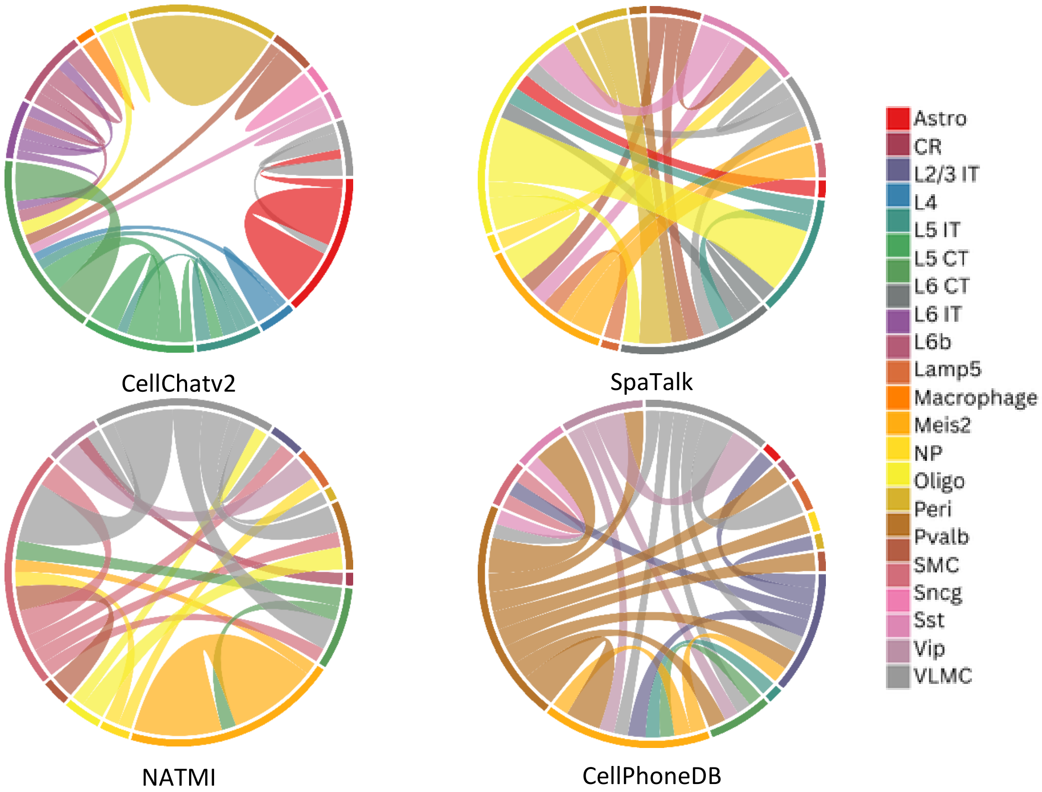

Figure 4.4: Circle plots demonstrating detected cell type interactions by CellChatv2, NATMI, and CellPhoneDB for the Xenium dataset.

The initial biological insights are derived from circle plots in **Fig. 4.4**, again revealing distinct interaction landscapes across methods. All four methods detected a high number of interactions but with varying distributions. CellChatv2 showed a predominance of interactions within the same sender and receiver cell types within the first 30 highest-scoring interactions. However, it captured a broader range of interactions between different cell types than other datasets. We see a high number of interactions involving VLMCs across all methods, particularly in SpaTalk, NATMI and CellPhoneDB. Additionally, the NP cell type is consistently detected as participating in interactions by CellChatv2, SpaTalk, and NATMI.

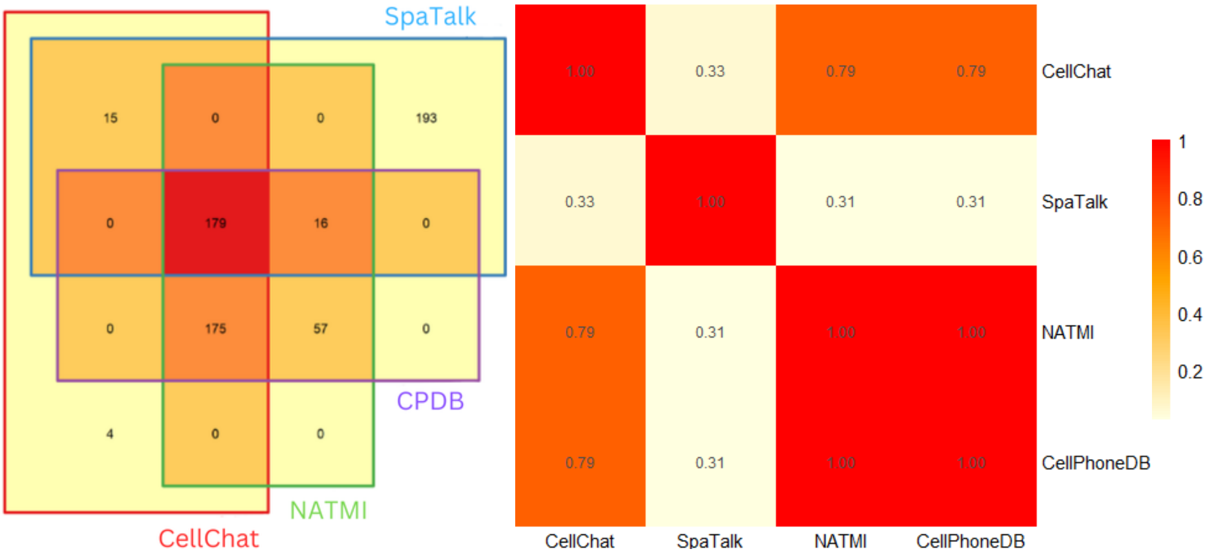

Figure 4.5: Venn diagram and correlation matrix of detected cell type interactions for the Xenium dataset across methods: CellChatv2, NATMI, and CellPhoneDB.

The Venn diagram analysis shown in **Fig. 4.5** shows 179 cell type interactions shared across all four methods. SpaTalk had the highest number of unique interactions (193), while CellChatv2 had only 4. A substantial overlap is detected between CellChatv2, NATMI and CellPhoneDB. NATMI and CellPhoneDB again generated identical interaction sets.

From the correlation matrix, we observe a strong correlation between NATMI and CellPhoneDB. CellChatv2 also correlated strongly with both NATMI and CellPhoneDB (correlation coefficient of 0.79). SpaTalk, although more divergent, still showed moderate correlation (correlation coefficient of 0.33 and 0.31), indicating some level of agreement with other methods.

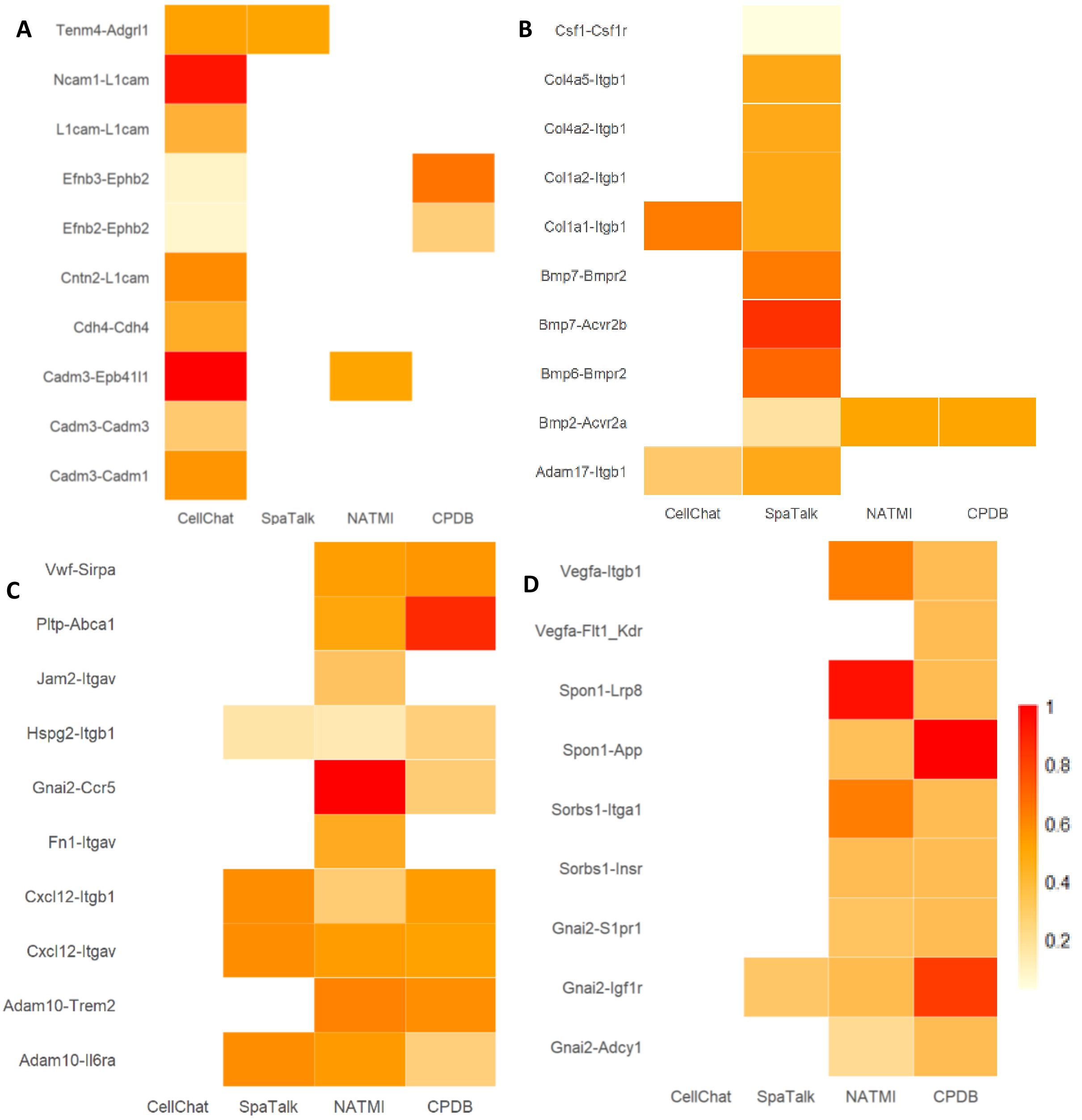

Figure 4.6: Correlation heatmap between detected LR pairs for the Xenium dataset across the methods. The cell type interactions chosen were: **A** for CellChatv2, CR-Sst, **B** for SpaTalk, VLMC-NP, **C** for NATMI, Endo-Macrophage, and **D** CellPhoneDB (CPDB), Astro-Endo.

To further compare the methods, heatmaps depicted in **Fig. 4.6** of LR pairs for selected cell type interactions were constructed. For CellChatv2, the highest-scoring interaction CR-Sst, between Cajal–Retzius cell and somatostatin-expressing interneuron, exhibited minimal overlap with the other methods: SpaTalk and NATMI have one sharing LR pair, and CellPhoneDB shares two LR pairs with CellChatv2, although the associated scores differ notably. For SpaTalk, the interaction VLMC-neurogliafrom neuron (NP) was selected, in which it identified nearly 60 distinct LR pairs. For visualization, only 10 representative LR pairs are shown in the heatmap, along with all LR pairs shared with other methods. In this case, there was an overlap of two pairs with CellChatv2, while NATMI and CellPhoneDB each share one.

For NATMI, the selected interaction is endothelial cell (Endo)-Macrophage. Although the method identified numerous LR pairs, the heatmap displays only those with the highest scores and those intersecting with other methods. We observe almost complete overlap with NATMI, limited overlap with SpaTalk (4 interactions), and none with CellChatv2. Finally, for CellPhoneDB, the Astro-Endo interaction was chosen. The overlap pattern was similar to that observed with NATMI: near-complete overlap with NATMI, one shared LR pair with SpaTalk, and none with CellChatv2.

CosMx

The CosMx spatial transcriptomics dataset was processed using all four computational methods to evaluate their performance and compare inferred CCC results.

From a technical perspective, all methods completed the analysis without any problems with memory space. Nevertheless, their runtimes are very varied. SpaTalk required the longest time, running for 2 hours and 28 minutes, and ultimately didn't find any cell type pairs. In contrast, CellChatv2 took 36 minutes, and CellPhoneDB ran slightly longer than previous datasets, finishing almost 4 minutes. NATMI was again the fastest, completing the analysis in 7 seconds.

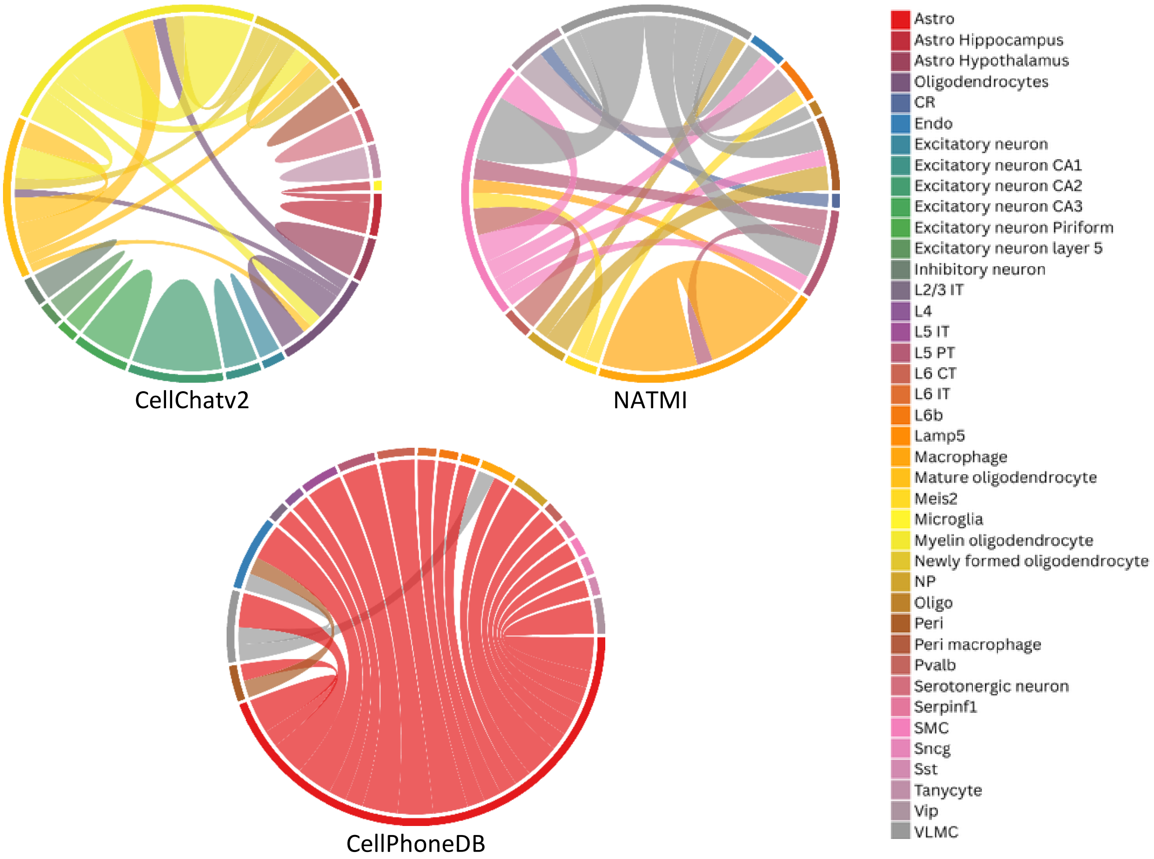

Figure 4.7: Circle plots demonstrating detected interactions by CellChatv2, NATMI, and CellPhoneDB for the CosMx dataset.

Moreover, we examine the biological results using circle plots of the top 30 interactions, based on the interaction scores of each method. According to **Fig. 4.7** CellChatv2 showed a high proportion of interactions within the same cell type. In NATMI's circle plot, we see more interactions with differing source and target cells. CellPhoneDB, on the other hand, detected almost only interactions originating from a single cell type, Astro, with only a few interactions involving other cell types.

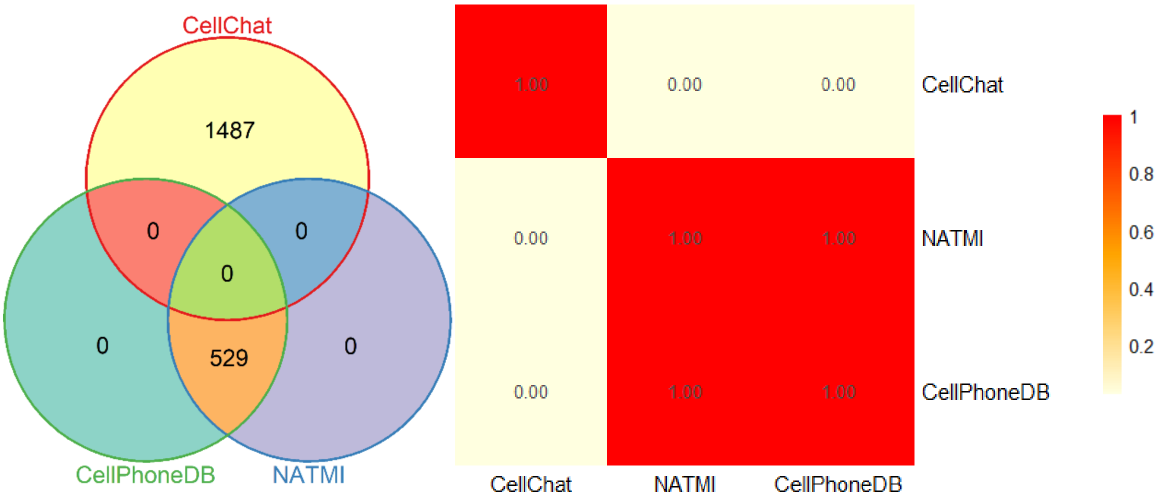

Figure 4.8: Venn diagram and correlation matrix of detected interactions for the CosMx dataset across methods: CellChatv2, NATMI, and CellPhoneDB.

Additionally, when looking at the Venn diagram in **Fig. 4.8**, we observe that CellChatv2 identified a large number of cell type interactions (1487), however, none overlapped with NATMI or CellPhoneDB. On the other hand, NATMI and CellPhoneDB shared 529 cell type pairs. Nonetheless, this overlap does not imply identical results, as this Venn diagram reflects only unique sender-receiver pairs and doesn't take different LR pairs into account.

This distinction is further highlighted in the correlation matrix in **Fig. 4.8**, where we see a strong correlation between NATMI and CellPhoneDB, while CellChatv2's results show no correlation with either method. This lack of overlap between CellChatv2 and the other methods is unlikely due to incorrect predictions but rather results from differences in database and cell type naming convention, which hinder direct comparison across methods. The correlation matrix mirrored the observed patterns in the Venn diagram.

Due to the absence of overlapping inferred cell type interactions between CellChatv2 and the other methods, correlation heatmaps based on LR pairs were not constructed.

MERFISH

Next, we evaluated the performance of CellChatv2, SpaTalk, NATMI, and CellPhoneDB on the MERFISH dataset, comparing both technical and biological aspects.

All methods completed the analysis in a comparable timeframe. Spatalk ran for five minutes but failed to detect any interactions, likely due to overly strict internal thresholds or filtering criteria. The remaining methods were executed smoothly within minutes or seconds: CellChatv2 in 3 minutes, NATMI in 3 seconds, CellPhoneDB in 3 minutes and 17 seconds.

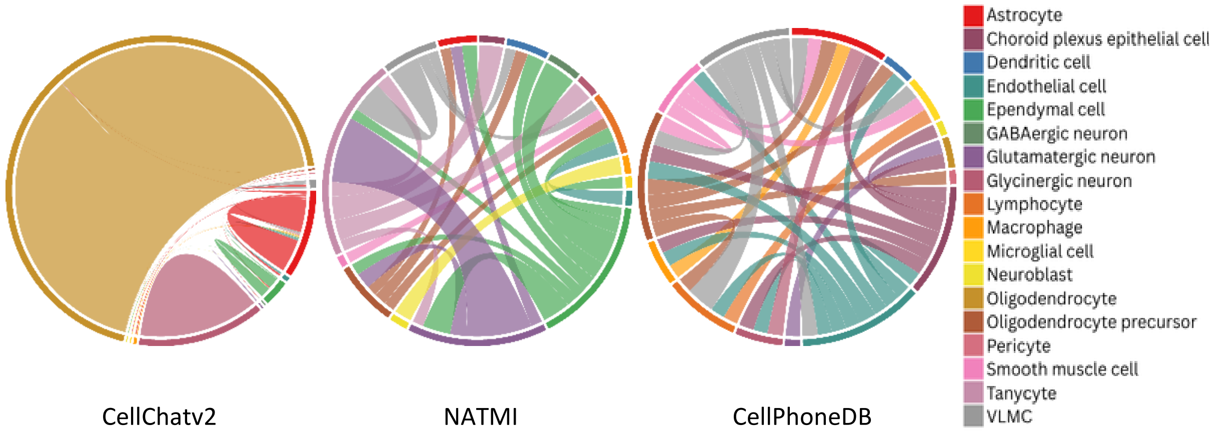

Figure 4.9: Circle plots demonstrating detected cell type interactions by CellChatv2, NATMI, and CellPhoneDB for the MERFISH dataset.

To explore the biological relevance of the predicted interactions, we again visualized the 30 first interactions, ranked by interaction score, using circle plots as shown in **Fig. 4.9**. CellChatv2 detected predominantly same-cell-type communication, with one main cell type (oligodendrocyte) contributing to most of the detected and most probable interactions.

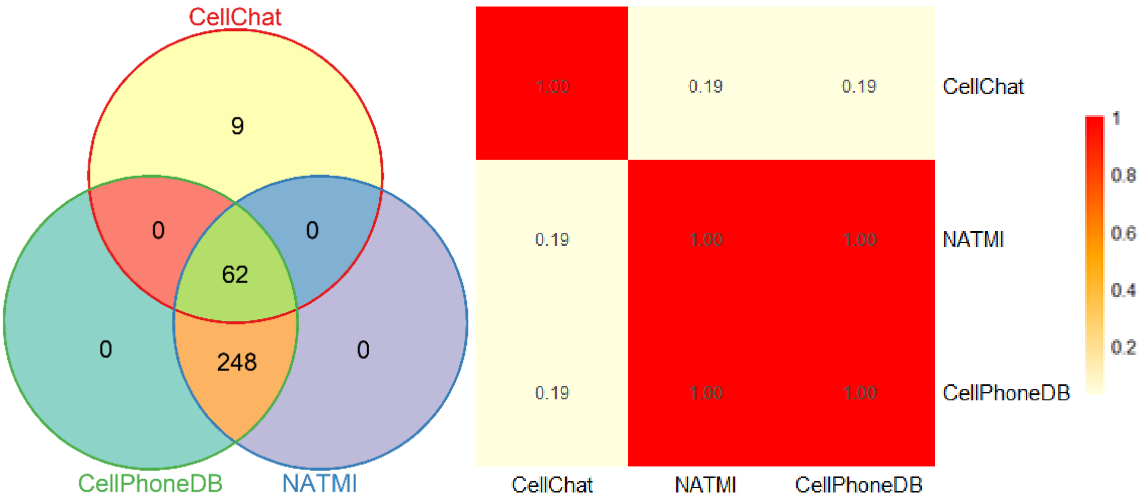

Figure 4.10: Venn diagram and correlation matrix of detected interactions for the MERFISH dataset across methods: CellChatv2, NATMI, and CellPhoneDB.

The Venn diagram and correlation matrix in **Fig. 4.10** both further illustrate these patterns. The three methods do share 62 source-target interactions of the same cell type. NATMI and CellPhoneDB share all of their detected interactions. SpaTalk's failure to detect results again limited the ability to assess the broader methodological agreement. In the correlation matrix, a strong correlation was observed between NATMI and CellPhoneDB, while CellChatv2 showed minimal correlation with either.

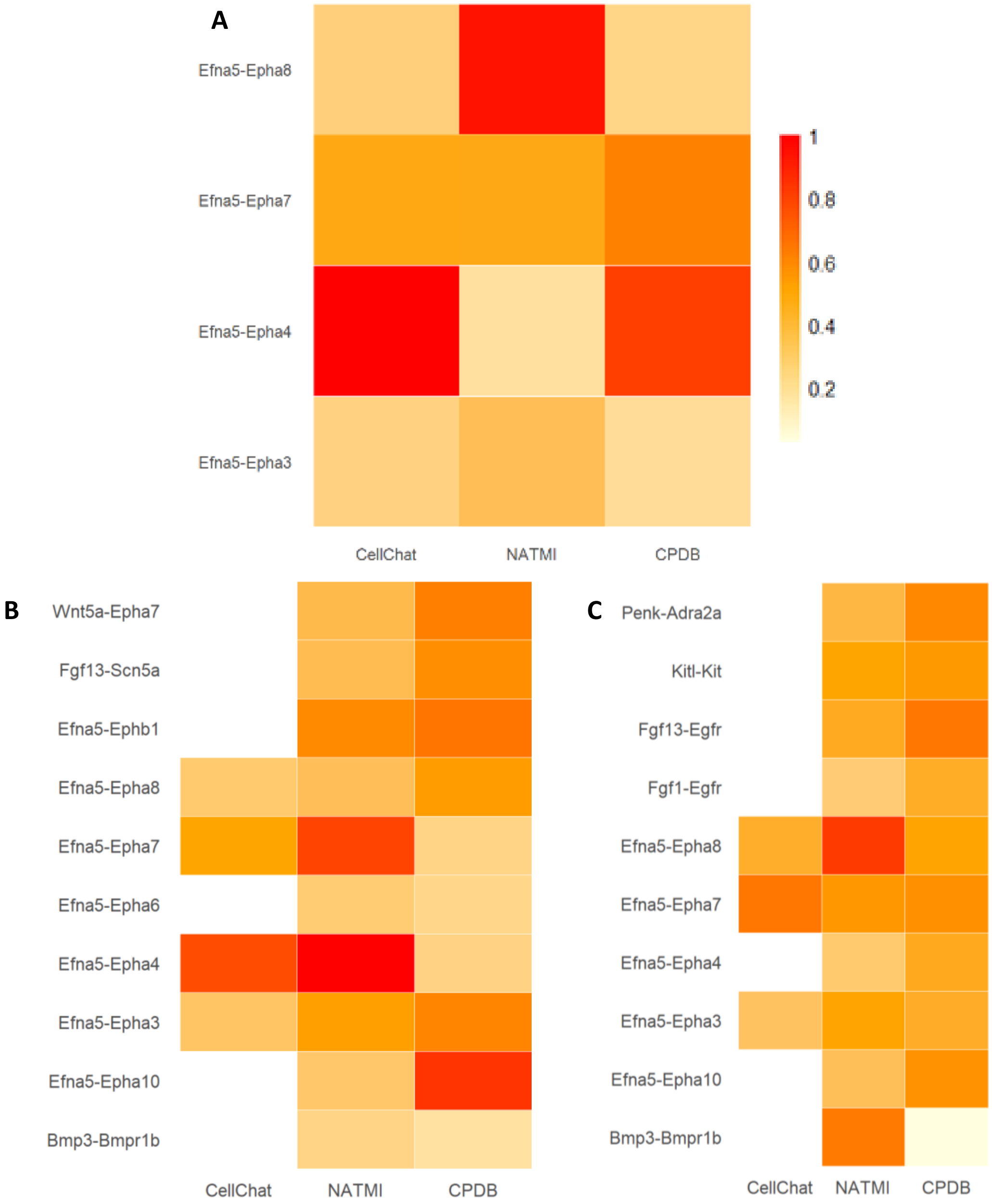

Figure 4.11: Correlation heatmap between detected LR pairs for the MERFISH dataset across the methods. The cell type interactions chosen were: **A** for CellChatv2, glutamatergic neuron-GABAergic neuron, **B** for NATMI, glutamatergic neuron-glutamatergic neuron, and **C** CellPhoneDB (CPDB), glutamatergic neuron-VLMC.

Finally, we constructed heatmaps shown in **Fig. 4.11** demonstrating the correlation of LR pairs identified in shared interactions across different methods. For CellChatv2, the selected interaction is between glutamatergic and GABAergic neurons. Although CellChatv2 found only four LR pairs, all were also detected by NATMI and CellPhoneDB. In the case of NATMI, we chose the glutamatergic neuron-glutamatergic neuron interaction. Which provided more than 20 LR pairs. All of these were also found by CellPhoneDB, and four LR pairs were shared with CellChatv2. For CellPhoneDB, the interaction between glutamatergic neurons and VLMC was selected. This interaction includes three LR pairs shared with CellChatv2 and a complete overlap with NATMI.

4.2 Summary of results

The following **Tab 4.1** summarizes the comparison across all four methods, highlighting the number of inferred cell type interactions, same-cell-type interactions, and detected LR pairs for each dataset.

Table 4.1: Summary of results.

Dataset	**Method**	**Number of cell type interactions**	**Same cell type interactions**	**Number of LR interactions**
VisiumHD	CellChatv2	54	14	54
	SpaTalk	0	0	0
	NATMI	529	23	3237
	CellPhoneDB	525	23	2923
Xenium	CellChatv2	373	23	9163
	SpaTalk	403	0	15211
	NATMI	427	21	4023
	CellPhoneDB	427	21	4027

Table 4.1 - continued

Dataset	Method	Number of cell type interactions	Same cell type interactions	Number of LR interactions
CosMx	CellChatv2	1487	44	7830
	SpaTalk	0	0	0
	NATMI	529	21	101047
	CellPhoneDB	529	23	91168
MERFISH	CellChatv2	62	15	90
	SpaTalk	0	0	0
	NATMI	310	18	3273
	CellPhoneDB	310	18	4105

4.3 Discussion

In this study, we evaluate four CCC inference methods–CellChatv2, SpaTalk, NATMI, and CellPhoneDB–across four SRT datasets: VisiumHD, Xenium, CosMx, and MERFISH. The aim of this research is to assess the performance of each tool by comparing both technical performance and biological interpretability and identifying recurring trends and discrepancies to the methods or datasets.

Technical performance

From a technical perspective, method setup, runtime and memory usage, and compatibility were key differentiators across tools. CellChatv2 offered a robust and user-friendly R pipeline with extensive visualization options that weren't used in this thesis. However, it was the slowest method in all datasets, requiring up to 3.5 hours on large datasets, and had the highest memory demands. Its sensitivity to parameters such as *scale.distance* also made tuning essential, especially for spatially dense data.

SpaTalk was the least stable tool, successfully completing the analysis only on Xenium. On larger datasets, it failed due to memory limitations. This is likely caused by its *dist* function, which calculates spatial distances between all cell pairs—a task that becomes computationally unmanageable when dealing with tens of thousands of cells (e.g., approximately 2.5 billion combinations in CosMx or VisiumHD). Subsetting the dataset or tuning parameters can partially mitigate this, but the method remains impractical for large-scale applications.

In contrast, NATMI, implemented via the LIANA framework in Python, was found to be the most computationally efficient method across all datasets, completing analyses in under 30 seconds with minimal memory requirements, even for large datasets.

CellPhoneDB also ran reliably with moderate runtime and memory usage. A notable limitation is its reliance on human gene nomenclature, requiring prior conversion from mouse to human orthologs. This disadvantage may result in gene loss and affect the analysis.

Number of interactions

In regard to biological insights, the number of inferred cell type interactions varied widely across methods and datasets.

CellChatv2 generally produced fewer interactions, many of which were intra-cell-type, than NATMI or CellPhoneDB, implying that the method may have a more conservative detection threshold or bias in spatial distance scaling. In the VisiumHD dataset, CellChatv2 detected only 14 interactions compared to 529 by NATMI and CellPhoneDB, while in CosMx, it reported approximately three times more. For Xenium, the number of interactions was comparable and for MERFISH, CellChatv2 inferred at least three times fewer interactions than the others.

When SpaTalk analysis was successful–as in the case of Xenium–it detected a high number of interactions. Specifically, the number of LR pairs was almost double that of CellChatv2 and three times higher than those identified by NATMI and CellPhoneDB. Despite the high number of interactions, SpaTalk did not detect any intra-cell-type interactions, which contrast with the results from CellChatv2.

NATMI and CellPhoneDB performed consistently throughout the analysis with a similarly high number of detected interactions, reporting around 300 to 600 interactions across all datasets. Interestingly, in the CosMx dataset, both NATMI and CellPhoneDB reported an especially high number of LR pairs-each detected 529 cell type pairs and around 100000 LR pairs. This highlights their scalability and robustness in analyzing high-resolution spatial transcriptomics data.

Overlap Between Methods

All in all, the overlap between interactions and cell type pairs inferred by the methods was generally low, possibly reflecting differences in databases, scoring systems, thresholds, cell type naming conventions, and distinct handling of spatial data.

Notably, the biologically meaningful interactions from CellChatv2 differed substantially from those detected by other methods. This divergence in interaction profile may be caused by CellChatv2's explicit incorporation of spatial information. SpaTalk, even when successful, showed largely non-overlapping results with other methods, further showing the impact of spatial modeling on inferred interactions. In NATMI and CellPhoneDB, however, we observed a consistent broad overlap. Both of these do not use spatial distances explicitly, causing more similar and intersecting results.

This overall variability highlights a high dependency on the method and reflects a lack of standardization in the field. Therefore, a careful method selection based on dataset characteristics, as well as cross-method validation, is important when studying biological processes.

Dataset-specific trends

The VisiumHD dataset, with its pseudo-single-cell resolution and large file size, exposed technical bottlenecks in SpaTalk and processing challenges in CellChatv2. The methods that rely on full pairwise cell comparisons have difficulties with such large datasets.

Thanks to the well-annotated single-cell resolution data that were offered by the Xenium dataset, all methods processed it successfully, demonstrating high compatibility and reliable results.

CosMx dataset, with high spatial resolution, showed a very low overlap in inferred interactions across all tools. CellChatv2 required careful tuning of spatial parameters to produce interpretable results that were still highly sparse. SpaTalk failed to produce any output, not due to problems with data, but rather because of the limitations within an internal function of the method responsible for computing distances between cells. A potential solution is to subset the data prior to analysis to reduce computational demands and avoid function failure.

MERFISH dataset revealed some biological convergence, indicating that despite different algorithms, certain biological signals are robust across tools.

Strengths and limitations of each method

Table 4.2: Summary of strengths and limitations of each method.

Method	Strengths	Limitations

CellChatv2	Integrates spatial context;
	customizable;
	extensive visualization
		Long runtime; memory-intensive;
		sensitive to spatial parameters

SpaTalk	Incorporates pathway-level proximity scoring	Unstable on large data; high memory use
NATMI	Fast; scalable; easy to implement in LIANA+; method flexibility	No built-in spatial modeling; minimal filtering
CellPhoneDB	Robust; scalable; consistent across datasets	Requires gene name conversion

This variability suggests that no single method performs optimally across all dataset types. Instead, the method suitability depends on dataset resolution and spatial complexity.

To summarize, NATMI and CellPhoneDB emerged as the most consistent methods, offering a balance of speed, scalability, and interpretability. CellChatv2, while more conservative, may be useful for focused analyses with stringent interaction criteria. SpaTalk requires further optimization to ensure reliability.

LIANA+

The implementation of NATMI through the LIANA+ framework has proven to be a reliable and flexible resource for CCC inference. It offers a complementary perspective by integrating outputs from multiple methods and harmonizing LR resources, making it particularly suitable for cross-validation and comparative analysis. By providing a unified framework, LIANA+ facilitates the running and comparison of various CCC methods. Nonetheless, a potential limitation is that LIANA+ may not reflect the most recent updates, and algorithmic changes of the native methods, which can lead to outdated results.

NicheNet

Although initially considered, NicheNet was excluded from this comparison. Its output lacks source-target cluster pairing, making it incompatible with the comparative framework of this study. While it could be run on selected cluster pairs for a uniform analysis, such selective analysis would cause a cross-method inconsistency.

4.4 Research limitations

Given the complexity of CCC inference across diverse spatial transcriptomics platforms, several methodological and practical limitations were encountered during this study.

First, a limitation lies in incongruities across different datasets, that can impact the resulting inferred interactions. Dataset-specific preprocessing challenges present inconsistencies in method application and results interpretation. Varying cell type annotations, gene symbol formats, and different levels of data sparsity all contribute to a difficulty in standardizing workflows across methods. Furthermore, the differences in scoring metrics across methods complicate the comparative analysis. Each tool applies a particular approach to quantify the interaction confidence or strength, necessitating the normalization of interaction scores (e.g., via z-score).

This study would also benefit from a more systematic parameter optimization approach. The use of default settings may not fully reflect the optimal performance of each method.

Finally, a key difficulty in this research is the absence of experimental validation. Without biological evidence or a known truth, it remains a challenge to assess the accuracy of predicted interactions and to draw definitive conclusions about their biological significance.

4.5 Suggestions for future research

Building upon the limitations identified in this study, several areas remain open for future research in the field of CCC inference in spatial transcriptomics.

From the technical standpoint, the development of benchmark datasets and standardized cell type annotations would be beneficial for the reduction of inconsistencies and improvements in tool comparisons. This unification of annotations and preprocessing pipeline could further improve reproducibility. Expanding LR databases to support more than human or mouse organisms would also broaden the applicability of CCC tools in biological contexts.

On the other hand, looking at conceptually biological advancements, while most current CCC methods infer static interactions, it would be interesting for future approaches to explore dynamic modeling to capture context-dependent signals.

5 CONCLUSION

CCC is essential for the coordination of cellular activity and maintaining tissue organization, influencing both healthy physiology and disease progression. The introduction of SRT opened new ways for exploring these interactions within their native spatial context. Nevertheless, the outcomes of such analyses are highly dependent on the computational tools employed.

In this study, we systematically compare four computational methods–CellChatv2, SpaTalk, NATMI, and CellPhoneDB–for inferring CCC from SRT datasets generated using VisiumHD, Xenium, CosMx, and MERFISH platforms.

The results were described and visualized using circle plots, and Venn diagrams to illustrate method overlap and correlation matrices of LR pair detection. The comparison analysis focused on both technical performance and biological interpretability.

Substantial variability was observed across tools, underscoring the current lack of standardization in CCC pipelines. These findings align with previous literature, highlighting the influence of spatial resolution, annotation discrepancies, and database differences in shaping CCC outputs.

CellChatv2 demonstrated strong performance, however its detections had minimal overlap with those of other methods. SpaTalk encountered technical issues with large datasets, revealing the internal problem with the method's functions. NATMI, implemented via the flexible and user-friendly LIANA+ framework, proved to be a reliable and accessible method. CellPhoneDB produced results that overlapped significantly with NATMI, offering complementary insights.

This work contributes to the growing body of comparative CCC tools assessment and emphasizes the importance of method selection tailored to dataset characteristics. Future research is suggested to continue refining both the computation frameworks and the biological assumptions underlying CCC inference, ultimately enhancing the robustness and interpretability of spatial communication analyses.

REFERENCES

(1) Zhou, X.; Franklin, R. A.; Adler, M.; Jacox, J. B.; Bailis, W.; Shyer, J. A.; Flavell, R. A.; Mayo, A.; Alon, U.; Medzhitov, R. Circuit Design Features of a Stable Two-Cell System. *Cell* **2018**, *172* (4), 744-757.e717. DOI: 10.1016/j.cell.2018.01.015 (acccessed 2024/11/07).

(2) Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B. B.; Siddiqui, A.; et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **2009**, *6* (5), 377-382. DOI: 10.1038/nmeth.1315.

(3) Ståhl, P. L.; Salmén, F.; Vickovic, S.; Lundmark, A.; Navarro, J. F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J. O.; Huss, M.; et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **2016**, *353* (6294), 78-82. DOI: 10.1126/science.aaf2403 From NLM.

(4) Jin, S.; Guerrero-Juarez, C. F.; Zhang, L.; Chang, I.; Ramos, R.; Kuan, C.-H.; Myung, P.; Plikus, M. V.; Nie, Q. Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **2021**, *12* (1), 1088. DOI: 10.1038/s41467-021-21246-9.

(5) Shao, X.; Li, C.; Yang, H.; Lu, X.; Liao, J.; Qian, J.; Wang, K.; Cheng, J.; Yang, P.; Chen, H.; et al. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nature Communications* **2022**, *13* (1), 4429. DOI: 10.1038/s41467-022-32111-8.

(6) Hou, R.; Denisenko, E.; Ong, H. T.; Ramilowski, J. A.; Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nature Communications* **2020**, *11* (1), 5011. DOI: 10.1038/s41467-020-18873-z.

(7) Efremova, M.; Vento-Tormo, M.; Teichmann, S. A.; Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols* **2020**, *15* (4), 1484-1506. DOI: 10.1038/s41596-020-0292-x.

(8) Bongrand, P. Ligand-receptor interactions. *Rep. Prog. Phys.* **1999**, *62* (6), 921-968, Review. DOI: 10.1088/0034-4885/62/6/202.

(9) Su, J.; Song, Y.; Zhu, Z.; Huang, X.; Fan, J.; Qiao, J.; Mao, F. Cell–cell communication: new insights and clinical implications. *Signal Transduction and Targeted Therapy* **2024**, *9* (1), 196. DOI: 10.1038/s41392-024-01888-z.

(10) Gilula, N. B. Cell-to-Cell Communication and Development. In *The Cell Surface: Mediator of Developmental Processes*, Subtelny, S. Ed.; Academic Press, 1980; pp 23-41.

(11) Bloemendal, S.; Kück, U. Cell-to-cell communication in plants, animals, and fungi: a comparative review. *Naturwissenschaften* **2013**, *100* (1), 3-19. DOI: 10.1007/s00114-012-0988-z.

(12) Bongrand, P. Ligand-receptor interactions. *Rep. Prog. Phys.* **1999**, *62.6*, p. 921.

(13) Dhanasekaran, N. Cell signaling: An overview. *Oncogene* **1998**, *17* (11), 1329-1330, Editorial Material. DOI: 10.1038/sj.onc.1202170.

(14) Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **2000**, *103* (2), 211-225. DOI: 10.1016/S0092-8674(00)00114-8 (acccessed 2025/04/03).

(15) Dorsam, R. T.; Gutkind, J. S. G-protein-coupled receptors and cancer. *Nature Reviews Cancer* **2007**, *7* (2), 79-94. DOI: 10.1038/nrc2069.

(16) Zhou, B.; Lin, W.; Long, Y.; Yang, Y.; Zhang, H.; Wu, K.; Chu, Q. Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduction and Targeted Therapy* **2022**, *7* (1), 95. DOI: 10.1038/s41392-022-00934-y.

(17) Duchartre, Y.; Kim, Y.-M.; Kahn, M. The Wnt signaling pathway in cancer. *Critical Reviews in Oncology/Hematology* **2016**, *99*, 141-149. DOI: https://doi.org/10.1016/j.critrevonc.2015.12.005.

(18) Armingol, E.; Officer, A.; Harismendy, O.; Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics* **2021**, *22* (2), 71-88. DOI: 10.1038/s41576-020-00292-x.

(19) Sinha, S.; Hembram, K. C.; Chatterjee, S. Chapter Four - Targeting signaling pathways in cancer stem cells: A potential approach for developing novel anti-cancer therapeutics. In *International Review of Cell and Molecular Biology*, Mukherjee, S., Chatterjee, K. Eds.; Vol. 385; Academic Press, 2024; pp 157-209.

(20) Dimitrov, D.; Türei, D.; Garrido-Rodriguez, M.; Burmedi, P. L.; Nagai, J. S.; Boys, C.; Ramirez Flores, R. O.; Kim, H.; Szalai, B.; Costa, I. G.; et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nature Communications* **2022**, *13* (1), 3224. DOI: 10.1038/s41467-022-30755-0.

(21) Freytag, S.; Tian, L.; Lönnstedt, I.; Ng, M.; Bahlo, M. Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. *F1000Res* **2018**, *7*, 1297. DOI: 10.12688/f1000research.15809.2 From NLM.

(22) Zheng, G. X. Y.; Terry, J. M.; Belgrader, P.; Ryvkin, P.; Bent, Z. W.; Wilson, R.; Ziraldo, S. B.; Wheeler, T. D.; McDermott, G. P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **2017**, *8* (1), 14049. DOI: 10.1038/ncomms14049.

(23) Kolodziejczyk, Aleksandra A.; Kim, J. K.; Svensson, V.; Marioni, John C.; Teichmann, Sarah A. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell* **2015**, *58* (4), 610-620. DOI: 10.1016/j.molcel.2015.04.005 (acccessed 2024/09/29).

(24) Saliba, A.-E.; Westermann, A. J.; Gorski, S. A.; Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research* **2014**, *42* (14), 8845-8860. DOI: 10.1093/nar/gku555 (acccessed 9/29/2024).

(25) Shao, X.; Lu, X.; Liao, J.; Chen, H.; Fan, X. New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein & Cell* **2020**, *11* (12), 866-880. DOI: 10.1007/s13238-020-00727-5 (acccessed 9/12/2024).

(26) Longo, S. K.; Guo, M. G.; Ji, A. L.; Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics* **2021**, *22* (10), 627-644. DOI: 10.1038/s41576-021-00370-8.

(27) Du, J.; Yang, Y.-C.; An, Z.-J.; Zhang, M.-H.; Fu, X.-H.; Huang, Z.-F.; Yuan, Y.; Hou, J. Advances in spatial transcriptomics and related data analysis strategies. *Journal of Translational Medicine* **2023**, *21* (1), 330. DOI: 10.1186/s12967-023-04150-2.

(28) Wang, Y.; Liu, B.; Zhao, G.; Lee, Y.; Buzdin, A.; Mu, X.; Zhao, J.; Chen, H.; Li, X. Spatial transcriptomics: Technologies, applications and experimental considerations. *Genomics* **2023**, *115* (5), 110671. DOI: 10.1016/j.ygeno.2023.110671 From NLM.

(29) Williams, C. G.; Lee, H. J.; Asatsuma, T.; Vento-Tormo, R.; Haque, A. An introduction to spatial transcriptomics for biomedical research. *Genome Medicine* **2022**, *14* (1), 68. DOI: 10.1186/s13073-022-01075-1.

(30) Domanskyi, S.; Srivastava, A.; Kaster, J.; Li, H.; Herlyn, M.; Rubinstein, J. C.; Chuang, J. H. Nextflow pipeline for Visium and H&E data from patient-derived xenograft samples. *Cell Rep Methods* **2024**, *4* (5), 100759. DOI: 10.1016/j.crmeth.2024.100759 From NLM.

(31) Macosko, E. Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A. R.; Kamitaki, N.; Martersteck, E. M.; et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **2015**, *161* (5), 1202-1214. DOI: 10.1016/j.cell.2015.05.002 From NLM.

(32) Ma, Y.; Zhou, X. Accurate and efficient integrative reference-informed spatial domain detection for spatial transcriptomics. *Nat Methods* **2024**, *21* (7), 1231-1244. DOI: 10.1038/s41592-024-02284-9 From NLM.

(33) Lu, Q.; Ding, J.; Li, L.; Chang, Y. Graph contrastive learning of subcellular-resolution spatial transcriptomics improves cell type annotation and reveals critical molecular pathways. *Briefings in Bioinformatics* **2025**, *26* (1), bbaf020. DOI: 10.1093/bib/bbaf020 (acccessed 5/7/2025).

(34) He, S.; Bhatt, R.; Brown, C.; Brown, E. A.; Buhr, D. L.; Chantranuvatana, K.; Danaher, P.; Dunaway, D.; Garrison, R. G.; Geiss, G.; et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nature Biotechnology* **2022**, *40* (12), 1794-1806. DOI: 10.1038/s41587-022-01483-z.

(35) Mennillo, E.; Lotstein, M. L.; Lee, G.; Johri, V.; Ekstrand, C.; Tsui, J.; Hou, J.; Leet, D. E.; He, J. Y.; Mahadevan, U.; et al. Single-cell spatial transcriptomics of fixed, paraffin-embedded biopsies reveals colitis-associated cell networks. *bioRxiv* **2024**. DOI: 10.1101/2024.11.11.623014 From NLM.

(36) Armingol, E.; Baghdassarian, H. M.; Lewis, N. E. The diversification of methods for studying cell-cell interactions and communication. *Nat Rev Genet* **2024**, *25* (6), 381-400. DOI: 10.1038/s41576-023-00685-8 From NLM.

(37) Browaeys, R.; Saelens, W.; Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nature Methods* **2020**, *17* (2), 159-162. DOI: 10.1038/s41592-019-0667-5.

(38) Dimitrov, D.; Schäfer, P. S. L.; Farr, E.; Rodriguez-Mier, P.; Lobentanzer, S.; Badia-i-Mompel, P.; Dugourd, A.; Tanevski, J.; Ramirez Flores, R. O.; Saez-Rodriguez, J. LIANA+ provides an all-in-one framework for cell–cell communication inference. *Nature Cell Biology* **2024**, *26* (9), 1613-1622. DOI: 10.1038/s41556-024-01469-w.

(39) Almet, A. A.; Cang, Z.; Jin, S.; Nie, Q. The landscape of cell–cell communication through single-cell transcriptomics. *Current Opinion in Systems Biology* **2021**, *26*, 12-23. DOI: https://doi.org/10.1016/j.coisb.2021.03.007.

(40) Almet, A. A.; Cang, Z.; Jin, S.; Nie, Q. The landscape of cell-cell communication through single-cell transcriptomics. *Curr Opin Syst Biol* **2021**, *26*, 12-23. DOI: 10.1016/j.coisb.2021.03.007 From NLM.

(41) Jin, S.; Plikus, M. V.; Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. *bioRxiv* **2023**, 2023.2011.2005.565674. DOI: 10.1101/2023.11.05.565674.

(42) Hou, R.; Denisenko, E.; Ong, H. T.; Ramilowski, J. A.; Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nat Commun* **2020**, *11* (1), 5011. DOI: 10.1038/s41467-020-18873-z From NLM.

(43) Wang, S.; Zheng, H.; Choi, J. S.; Lee, J. K.; Li, X.; Hu, H. A systematic evaluation of the computational tools for ligand-receptor-based cell-cell interaction inference. *Brief Funct Genomics* **2022**, *21* (5), 339-356. DOI: 10.1093/bfgp/elac019 From NLM.

(44) Wang, S.; Zheng, H.; Choi, J. S.; Lee, J. K.; Li, X.; Hu, H. A systematic evaluation of the computational tools for ligand-receptor-based cell–cell interaction inference. *Briefings in Functional Genomics* **2022**, *21* (5), 339-356. DOI: 10.1093/bfgp/elac019 (acccessed 11/12/2024).

(45) ChatGPT, ver 4.0. In *OpenAI*, San Fransisco, 2023.

LIST OF ABBREVIATIONS

CCC	Cell-Cell Communication
CCI	Cell-Cell Interaction
CPDB	CellPhoneDB
CosMx SMI	CosMx spatial molecular imager
FFPE	Formalin-Fixed Paraffin-Embedded
GPCR	G-protein coupled receptor
LR	ligand-receptor
MERFISH	Multiplexed Error-Robust Fluorescence in Situ Hybridization
NATMI	Network Analysis Toolkit for Multicellular Interactions
RTK	Receptor Tyrosine Kinase
scRNA-seq	single cell RNA sequencing
SRT	Spatially Resolved Transcriptomics

APPENDICES

Table A: The following table lists the cell type abbreviations used throughout this thesis alongside their corresponding full names. These terms are commonly used in the single-cell transcriptomics literature.

Abbreviation of cell type	**Full cell type name**
Astro	Astrocyte
Astro Hippocampus	Hippocampal Astrocyte
Astro Hypothalamus	Hypothalamic Astrocyte
Astro Oligodendrocytes	Astrocyte-like Oligodendrocyte
CR	Cajal–Retzius Cell
Endo	Endothelial Cell
Excitatory neuron CA1	CA1 Region Excitatory Neuron
Excitatory neuron CA2	CA2 Region Excitatory Neuron
Excitatory neuron CA3	CA3 Region Excitatory Neuron
Excitatory neuron Piriform	Piriform Cortex Excitatory Neuron
Excitatory neuron layer 5	Layer 5 Excitatory Neuron
Inhibitory neuron	Inhibitory Neuron
L2/3 IT	Layer 2/3 Intratelencephalic Neuron
L4	Layer 4 Neuron
L5 IT	Layer 5 Intratelencephalic Neuron
L5 PT	Layer 5 Pyramidal Tract Neuron
L5 CT	Layer 5 Corticothalamic Neuron
L6 CT	Layer 6 Corticothalamic Neuron
L6 IT	Layer 6 Intratelencephalic Neuron
L6b	Layer 6b Neuron
Lamp5	Lysosomal Associated Membrane Protein Family Member 5+
Mature oligodendrocyte	Mature Oligodendrocyte
Meis2	Meis Homeobox 2+ Neuron
Microglia	Microglial Cell
Myelin oligodendrocyte	Myelinating Oligodendrocyte

Table A – continued

Abbreviation of cell type	**Full cell type name**
Newly formed oligodendrocyte	Newly Formed Oligodendrocyte
NP	Neurogliaform/NP (Neurophysin-expressing) Neuron
Oligo	Oligodendrocyte
Peri	Pericyte
Peri macrophage	Perivascular Macrophage
Pvalb	Parvalbumin-Expressing Interneuron
Serotonergic neuron	Serotonergic Neuron
Serpinf1	Serpin Family F Member 1+ Cell
SMC	Smooth Muscle Cell
Sncg	Synuclein Gamma-Expressing Neuron
Sst	Somatostatin-Expressing Interneuron
Vip	Vasoactive Intestinal Peptide-Expressing Interneuron
VLMC	Vascular and Leptomeningeal Cell