MMCCI: Multimodal Cell-Cell Interaction Integrative Analysis of Single Cell and Spatial Data

Levi Hockey,[1,2] Onkar Mulay,[1,2] Zherui Xiong,[1,2] Kiarash Khosrotehrani,[3] Christian M. Nefgzer[2,4] and Quan Nguyen[1,2,*]

[1]Queensland Institute of Medical Research Berghofer, Brisbane, Australia, [2]Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia, [3]Frazer Institute, University of Queensland, Brisbane, Australia and [4]School of Biomedical Sciences, University of Queensland, Brisbane, Australia

*Corresponding author. quan.nguyen@imb.uq.edu.au

Abstract

Cell-cell interaction (CCI) analyses are becoming an indispensable discovery tool for cutting-edge single cell and spatial omics technologies, identifying ligand-receptor (LR) targets in intercellular communications at the molecular, cellular, and microenvironment levels. Different transcriptional-based modalities can add complementary information and provide independent validation of a CCI, but so far no robust methods exist to integrate CCI results together. To address this, we have developed a statistical and computational pipeline, Multimodal CCI (MMCCI), implemented in an open-source Python package, which integrates, analyzes, and visualizes multiple LR-cell-type CCI networks between multiple samples of a single transcriptomic-based modality as well as between multiple modalities. MMCCI implements new and in-depth downstream analyses, including comparison between biological conditions, network and interaction clustering, sender-receiver interaction querying, and pathway analysis. We applied MMCCI to integrate CCIs in our spatial transcriptomics datasets of aging mouse brains (from 10X Visium and BGI STOmics) and melanoma (10X Visium, 10X Xenium and NanoString CosMx) and identified biologically meaningful interactions. Using simulated data, we applied MMCCI integration on four popular CCI algorithms, stLearn, CellChat, Squidpy, and NATMI, and detected highly confident interactions while reducing false interaction discoveries through the integration of multiple samples. With MMCCI, the community will have access to a valuable tool for harnessing the power of multimodal single cell and spatial transcriptomics. MMCCI source code and documentation are available at: `https://github.com/BiomedicalMachineLearning/MultimodalCCI`.

Key words: Cell-Cell Interactions, Multimodal Integration, Multimodal Analysis, Single-Cell Transcriptomics, Spatial Transcriptomics, Ligand-Receptor Interactions, Network Visualization, Aging, Brain, Melanoma

Introduction

Spatial transcriptomics (ST) is a molecular profiling technique that maps RNA sequencing with spatial information to access spatial gene expression patterns that can provide biological insights into many applications including neuroscience and cancer research [1, 2]. A main method to investigate ST data is cell-cell interaction (CCI) analysis, which aims to decipher the intricate communication networks within tissues and understand how cells interact in their spatial context. Multiple computational tools exist for inferring and analyzing intercellular communication networks in single-cell RNA-sequencing (scRNA-seq) and ST data such as stLearn, CellChat, Squidpy, CellPhoneDB, NATMI, and NicheNet [3, 4, 5, 6, 7, 8]. Tools such as stLearn and CellChat take spatial information into account, limiting the inferred communications to between only nearby cells or spots, whereas CellPhoneDB and NATMI infer communications between any cells or spots and are thus prone to false discoveries. Other tools also perform additional downstream analyses such as scDiffcom and COMUNET, which calculate differential LR pairs and LR networks [9, 10]. In all cases, detecting false positive interactions is a major challenge.

For single cell data, integrative analysis of multiple modalities has shown to uncover complex molecular and cellular mechanisms [11]. ST technologies such as CosMX and Xenium provide single cell resolution at the cost of being able to detect fewer genes or proteins, while Visium and STOmics are able to sequence the entire transcriptome, but each Visium spot (55um wide) captures the expression of often one to ten cells [12, 13, 14, 15]. Integrating orthogonal technologies such as these together can give deep insights into biological processes unobtainable from any one technology alone and can provide the benefit of capturing a large number of LR pairs while still retaining the specificity of single cell spatial methods. However, for CCI analysis so far, there are no methods to adequately integrate samples from different modalities together.

Therefore, we introduce MMCCI as a method for multimodal CCI integration and hypothesize that applying MMCCI integration to samples across different technologies will be able to filter false positive interactions while preserving the true interactions in both spatial and non-spatial technologies, thereby constructing a more accurate landscape of the complex intercellular communications than any one sample or modality alone.

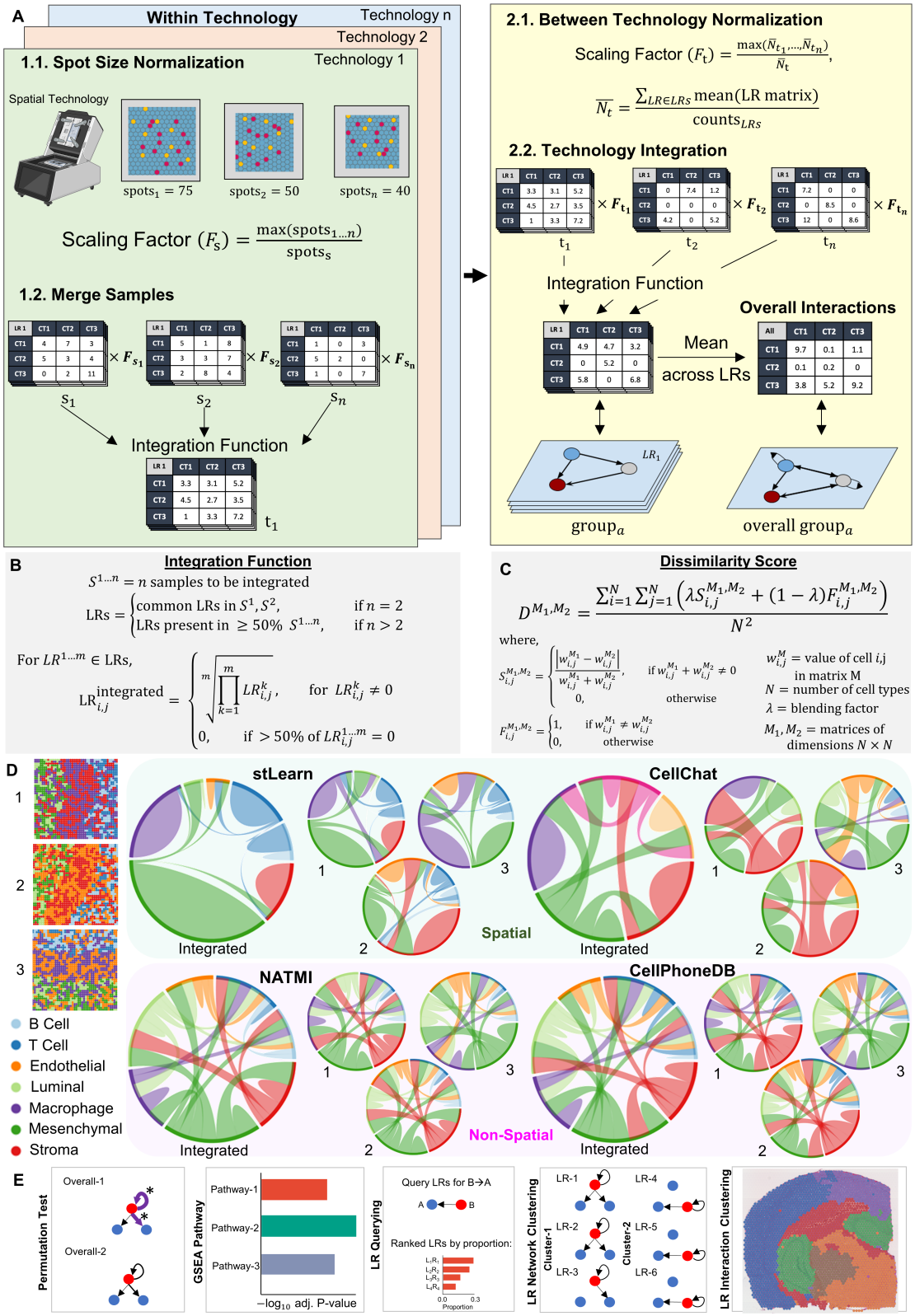


Fig. 1. Overview of MMCCI and benchmarking results. **A.** Overview of the integration pathway showing pipeline for integration within (left) and between (right) technologies. Stacked tables indicate a sample, which is made up of a set of LR matrices. Created with BioRender.com. **B.** The algorithm of the integration function, showing how multiple samples are integrated on the individual LR pair level. **C.** The algorithm of the dissimilarity score, used to rank cell-to-cell networks based on how dissimilar they are from each other. **D.** Benchmarking results of running within-technology integration on simulated spatial breast cancer samples (1, 2, 3) that have been run through different spatial (stLearn and CellChat) and non-spatial (NATMI and Squidpy) CCI methods. All simulated samples have B and T cells spatially separated from luminal and mesenchymal cells (left). Each sample has a unique combination of two out of stromal cells, macrophages, and endothelial cells. Sample one excludes endothelial cells, sample two macrophages, and three stromal cells. Chord plots show the integrated (left) and unintegrated (right) results for each technology and are used to show the cell types that are interacting. Color of chord indicates the cell type that is acting as mostly a sender. **E.** Overview of selected downstream analyses in MMCCI, including permutation testing between two networks, Gene Set Enrichment Analysis (GSEA) on network edges, LR querying on network edges, clustering of LR pairs with similar networks, and clustering of spots with similar LR interactions.

Methods

1. Upstream Data Processing

1.1. Deconvolution

For the aging brain samples, we generated Visium and STOmics data and annotated cell types using RCTD with the Allen Brain Atlas as a single cell reference for deconvolution [16, 17]. For the melanoma samples, we generated Visium, CosMx and Xenium data, followed by automated cell-type annotation by deconvolution and label transferring using an in-house single cell skin cancer reference. Specifically, for Visium data, we used RCTD for deconvolution, and for the single cell resolution CosMx and Xenium data, we used Seurat V4 Label Transfer for single cell annotation [18].

1.2. Baseline Cell-Cell Interactions

Cell-cell interactions for the aging brain and melanoma samples were computed using stLearn, a CCI method that takes spatial information into account [3]. We first filtered out genes that were not present in more than three spots and performed CPM normalization (normalizing counts per cell). stLearn performs ligand-receptor analysis, in which each spot or cell receives an interaction score from its immediate neighboring spots which is used for MMCCI's LR interaction clustering. Importantly, stLearn also predicts significant cell-cell interactions, generating a network of communicating cell-types represented as a matrix per LR pair where the weights are the number of spots or cells interacting significantly. This data is used by MMCCI to perform integration and is the format used by most of the integration, analysis and visualization functions. CCI results obtained from other methods are converted to this format by MMCCI for compatibility with the package. The LR-pair database used in stLearn was from connectomeDB2020 [7].

2. Integration Method

Sample integration (**Fig. 1A**) is split into two sections, within-technology and between technology integration, which can both be run separately.

2.1. Within-Technology Integration

This step performs integration of samples within a technology by accounting for differing sample sizes, reflected by the number of spots in each sample (1.1 of **Fig. 1A**). Each sample's LR pair interaction matrices are scaled based on the number of spots in that sample (sample-size scaling). Each weight in each LR matrix is multiplied by this scaling factor. This is required to control for samples with a disproportionate number of spots, ensuring that each sample has an equal weighting within each technology as a sample with more spots will generally have more interactions.

An integrated sample for a given technology is then created by running our LR-level integration function with each sample of that technology (1.2 of **Fig. 1A**).

$$S^{1\ldots n} = n \text{ samples to be integrated}$$

$$\text{LRs} = \begin{cases} \text{common LRs in } S^1, S^2, & \text{if } n = 2 \\ \text{LRs present in } \geq 50\% \ S^{1\ldots n}, & \text{if } n > 2 \end{cases} \tag{1}$$

For $\text{LR}^{1\ldots m} \in \text{LRs}$,

$$\text{LR}_{i,j}^{\text{integrated}} = \begin{cases} \sqrt[m]{\prod_{k=1}^{m} LR_{i,j}^{k}} & \text{for } \text{LR}_{i,j}^{k} \neq 0 \\ 0 & \text{if} \geq 50\% \ \text{LR}_{i,j}^{1\ldots m} = 0 \end{cases} \tag{2}$$

Shown above and in **Fig. 1B**, the integration works by first creating a list of LR pairs to be included in the integrated sample by either taking the common LR pairs if there are two samples to integrate, or by taking the LRs present in at least half the samples if there are more than two samples to integrate. Then for each LR pair in that list, the integrated network for that pair is calculated by creating a matrix of the geometric means across all the networks for that LR pair, but treating zeros as ones in the multiplicity operation unless more than 50% of the values are zero.

2.2. Between-Technology Integration

The integrated samples for each technology are then normalized by a scaling factor, calculated by averaging the mean of the values across all the LR pairs per technology (2.1 of **Fig. 1A**). This ensures that a technology that detects more interactions is not overrepresented in the integrated sample.

A final integrated sample is then created by running the same LR-level integration function as in the within-technology integration (2.2 of **Fig. 1A**). After this, the overall network of interactions, an overview of the cell-type interactions in the sample, is calculated by scaling the values in each LR matrix so that they sum to one to ensure equal weighting of all LR pairs, and then taking the mean of all the matrices. Creating an overall network can also be done in other points of the integration to easily visualize the overall interaction network of an integrated or non-integrated sample.

3. Downstream Analyses

3.1. Dissimilarity Scoring

$$D^{M_1,M_2} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(\lambda S_{i,j}^{M_1,M_2} + (1-\lambda)F_{i,j}^{M_1,M_2})}{N^2}, \tag{3}$$

where,

$$S_{i,j}^{M_1,M_2} = \begin{cases} \frac{|w_{i,j}^{M_1} - w_{i,j}^{M_2}|}{w_{i,j}^{M_1} + w_{i,j}^{M_2}} & \text{if } w_{i,j}^{M_1} + w_{i,j}^{M_2} \neq 0, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$F_{i,j}^{M_1,M_2} = \begin{cases} 1 & \text{if } w_{i,j}^{M_1} \neq w_{i,j}^{M_2}, \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$N = \text{number of cell types}$$

$$\lambda = \text{blending factor}$$

$$M_1, M_2 = \text{two matrices of dimensions } N \times N$$

The dissimilarity score formula, shown above and in **Fig. 1C**, has been based off of a similar metric introduced in

COMUNET [10]. To compute a dissimilarity score between two matrices, we first ensure that they have the same interacting cell types by padding zeros to the matrix with fewer cell types. For any two LR pair matrices, M_1 and M_2, the continuous dissimilarity score, S, is calculated by taking the absolute difference between the weights of each i_{th} and j_{th} cell of the two matrices divided by the sum of the weights. The binary dissimilarity score, F, is calculated as 1 if the two cells of the matrices have a different number of significantly interacting spots, or otherwise 0. Finally these two dissimilarity scores are added for each LR-pair using a blending factor, λ, and this score is divided by the square of the number of cell-types, giving the dissimilarity score, D.

F helps in increasing D in cases where there are many non-zero interactions, as S will be lower for two non-zero values with the same difference as two values where one is zero. This can lead to D being lower than expected in comparison to two networks each with only a few interactions. By default, we set $\lambda = 0.5$ to equally weigh the two scoring metrics. The effects of changing λ across different LR pairs is shown in **Sup. Fig. 2D**.

3.2. Permutation Testing

The permutation test, represented in **Fig. 1E**, is used to find the significantly different interactions between two networks. The test is performed by creating a matrix which is the absolute of the difference of the two matrices, repeatedly shuffling the values along both rows and columns, and then observing how often the differences in shuffled matrices are bigger than the true differences, returning a p-value.

3.3. LR Network Clustering

For the LR network clustering represented in **Fig. 1E**, we calculated a pairwise dissimilarity score matrix for each LR pair (**Equations 3-5**). Using this matrix we calculate a distance matrix using "pdist" and "squareform" from the SciPy package [19]. The distance matrix was normalized using min-max normalization. We then computed the PCA of the dissimilarity matrix and used the first two components along with the distance matrix for clustering of the LR networks. This function allows the user to perform either hierarchical or k-means clustering, and to either manually choose the number of clusters or automatically compute the optimal number of clusters.

3.4. LR Interaction Clustering

This method is only currently available for individual ST samples processed through stLearn. LR interaction scores from stLearn, which give a score for each LR pair for each spot, are pre-processed and clustered using the standard Scanpy spatial transcriptomics analysis pipeline for Leiden clustering [20]. Clusters are plotted spatially to show anatomically where different types of interactions are occurring. The stacked barplot indicates the proportion of the interacting cell-types in each cluster. An example is shown in **Fig. 1E**.

3.5. Cell-Cell LR Pair Querying

LR pairs can be queried given a sender and receiver cell type and then ranked based on the proportion of the total number of interactions for the given sender and receiver. This is represented in **Fig. 1E**.

3.6. Pathway Analysis

Pathway analysis is performed using GSEApy, a Python package that can perform gene set enrichment analysis [21]. LRs used for analysis are split into the ligand and receiver and are added to a gene list to be run through GSEApy's Enrichr pipeline, which can be used with any specified Enrichr databases. In this pipeline, both mouse and human genes can be inputted and will be automatically accounted for.

4. Visualizations

4.1. Network Visualizations

Network graph plotting is performed using NetworkX, examples of which are in **Fig. 2** [22]. Nodes have a color scale based on whether that cell type is sending (red) or receiving (blue) more interactions. When plotting a network difference plot, p-values from the permutation testing can also be given and edges with significant p values are plotted as purple if they are higher in the first group or green if lower, along with the p-value shown on the edge. As an alternative visualization, network chord plots are included and were adapted from the stLearn package [3].

5. CCI Output Conversions

For broad applications, MMCCI has wrapper functions included in the package that allow conversion of multiple different CCI pipelines. Currently, stLearn, CellPhoneDB, Squidpy, CellChat, and NATMI outputs can be converted to be compatible with MMCCI's pipeline and samples from different CCI methods can be integrated together. The MMCCI pipeline is outlined in **Sup. Fig. 1**.

6. Spatial Transcriptomics Data Generation

6.1. Brain Samples

Young (3 months) and aged (20–24 months) C57BL/6 mice (all female and housed in the same animal facility) were scarified for organ harvest between 8am and 9am to avoid molecular changes related to differences in circadian rhythm. Only healthy animals were processed; mice bearing tumors or displaying any internal alterations (e.g., enlarged spleen, inflamed liver etc.) were excluded.

For Visium Spatial Gene Expression, The tissue sections were placed on the pre-equilibrated Visium Spatial Tissue Optimisation Slide (10X Genomics, cat no.3000394) and Visium Spatial Gene Expression Slide (10X Genomics, cat no.2000233). The Visium library was constructed according to the Visium Spatial Gene Expression User Guide (CG000239 Rev B, 10X Genomics).

For STOmics, the tissue sections were placed on Stereo-seq Chips (BGI STOmics, cat no. 211SP118). The tissue permeabilization was optimized according to STOmics tereo-seq Permeabilization Unser Manual (Version A0). The STOmics library preparation was carried out as described in STOmics Stereo-seq Transcriptomics User Manual (Version A0).

6.2 Melanoma Samples

Formalin-fixed, paraffin-embedded (FFPE) melanoma samples from a case-case based series of thin melanoma were selected for downstream spatial transcriptomics [23]. Three samples from the fatal and the non-fatal melanoma cases were selected based on their Breslow Depth, sex and age. The tissue preparation for Xenium in situ gene expression was carried out as described in Tissue Preparation Guide (CG000578 Rev C, 10X Genomics). In short, six 5 μm FFPE sections were multiplexed on one

Xenium Slide (10X Genomics, cat no.3000941) to maximize the usage of the capture area while minimizing the batch effect. The detailed onboard image processing, decoding and cell segmentation have been described in Janesick. et al., 2023 [13]. Melanoma CosMX samples were provided and generated by NanoString.

7. Benchmarking

The simulated spatial breast cancer data was generated using a scRNA-seq breast cancer dataset. The gene expression distribution in each of the cell types was estimated by fitting a negative binomial distribution, and this was used to generate 10,000 simulated cells of each cell type [24]. An empty spatial tissue was generated and split into pools of 2-3 spots, which were randomly assigned a cell-type based on a set of rules specifying which cell types were not allowed to neighbor each other. These spatial spots were populated with a mixture of the simulated cells so that the spatial spots had similar properties to Visium data, in terms of counts and number of cells per spot. These three simulated spatial samples, each with a set of cell types that did and did not neighbor, were then run through stLearn, CellChat, NATMI, and Squidpy CCI analysis and then integrated using MMCCI. All CCI tools used connectomeDB2020 as the LR database [7]. Related plots are in **Sup. Fig. 2**.

Results

Benchmarking

To investigate MMCCI's performance for within-technology integration, we integrated samples within a single ST modality using three simulated breast cancer spatial samples that were run through four different CCI methods - stLearn, CellChat, NATMI and Squidpy (**Fig. 1D**). The simulated data was designed such that in all samples, B and T cells (group 1) neighbor each other, as well as luminal and mesenchymal cells (group 2) (**Fig. 1D**). Group 1 and 2 are separated spatially in all samples, ensuring that there would be no biological interaction between these two groups. In each sample, there is also a unique combination of two cell types amongst endothelial, macrophage and stromal cells (group 3) that is positioned between group 1 and 2. This means each sample will have unique interactions within group 3. It was expected that the integrated result should filter these interaction results out as they are not present in the majority of the samples.

For all CCI methods, no interactions between endothelial, macrophage and stromal cells are observed in the integrated CCI network (**Fig. 1D**). stLearn and CellChat, which take into account spatial information, both correctly exclude interactions between group 1 and 2 in the integrated CCI networks, while retaining the interactions seen in the majority of the samples. Unlike CellChat, stLearn detected interactions between B and T cells, which is an expected biological interaction, suggesting that stLearn was the best performing algorithm in terms of both detecting expected interactions while also filtering false positive interactions. These benchmarking results show that MMCCI can remove interactions uncommon to the majority of samples while preserving significant interactions present in the majority of samples across different CCI methods, both spatial and non-spatial. These results also show how an integrated interaction network, based on multiple samples of the same modality type, is more significant than a single sample run through either a spatial or non-spatial CCI algorithm.

Aging Brain

Next, MMCCI integration and downstream analyses were performed on eight Visium mouse brain samples (four aged and four young) and four STOmics mouse brain samples (two aged and two young) (**Fig. 2**) . Replicate samples were first integrated by technology and age group (**Fig. 2A**, **Fig. 2D**). These four combined samples were again integrated by age group, resulting in one aged and one young mouse integrated sample (**Fig. 2C** and **Fig. 2F**). Prior to integration, many cell types were observed interacting in individual samples, making it unclear which interactions were falsely identified and which were biological (**Fig. 2A** and **Fig. 2D**). After integration, only highly confident interactions were retained (i.e. the interactions present in the majority of the samples and across technologies) (**Equation 1**, **Equation 2**, and **Fig. 2**).

Overall Network Comparison

The overall interaction networks for aged and young were constructed and the difference between these overall networks was calculated and run through permutation testing to identify cell-type pair interactions that were significantly different between the aged and young samples (**Fig. 3A**). The dissimilarity score between the overall aged and young networks was 0.239, which indicates that nearly a quarter of the network edges are significantly different between the age groups (**Equations 3-5**).

Interactions within ependymal cells and oligodendrocytes were shown to decrease significantly with age, while interactions within astrocytes and from astrocytes to neurons significantly increased with age. This reflects the current understanding that the ependymal layer thins during aging and reactive astrocytes proliferate and interpose themselves within the ependymal cell layer [25, 26]. Oligodendrocyte numbers and function are also known to decrease with age while astrocyte interactions increase, which is shown in our findings [27, 28, 29].

Dissimilar LR Pairs

To establish the specific LR interactions that changed significantly in different sender and receiver cell types between age groups, MMCCI's LR pair network dissimilarity ranking was run on the integrated aged and young samples (**Fig. 3B**, (**Equations 3-5**)). The most dissimilar pair, Vimentin (Vim)-Cd44, showed interactions mostly within and between astrocytes and oligodendrocytes in the aged brain, while in the young brain, the interactions occurred within and between ependymal cells, oligodendrocytes, and neurons (**Fig. 3B**). Vim is a known astrocyte marker, expressed in ependymal cells in healthy brains and in reactive astrocytes in aging [30, 31]. Cd44 is an astro-mesenchymal marker, so this interaction was expected to change with age, with Vim release shifting from ependymal to astrocytes during aging [32].

The second pair, ApoE-Lrp8, similarly showed interactions within and from astrocytes in the aged brain that were not present in the young and interactions within and from ependymal cells in the young brain that were not present in the aged (**Fig. 3B**). This was expected as astrocytes are a major producer of ApoE, which are highly active in the aged brain, while in the young brain, ApoE is synthesized in ependymal cells, but to a lesser extent [33, 34, 35].

Cell-Cell LR Pair Querying

Since the interactions from astrocytes to neurons were significantly increased in the aged brain, this interaction was

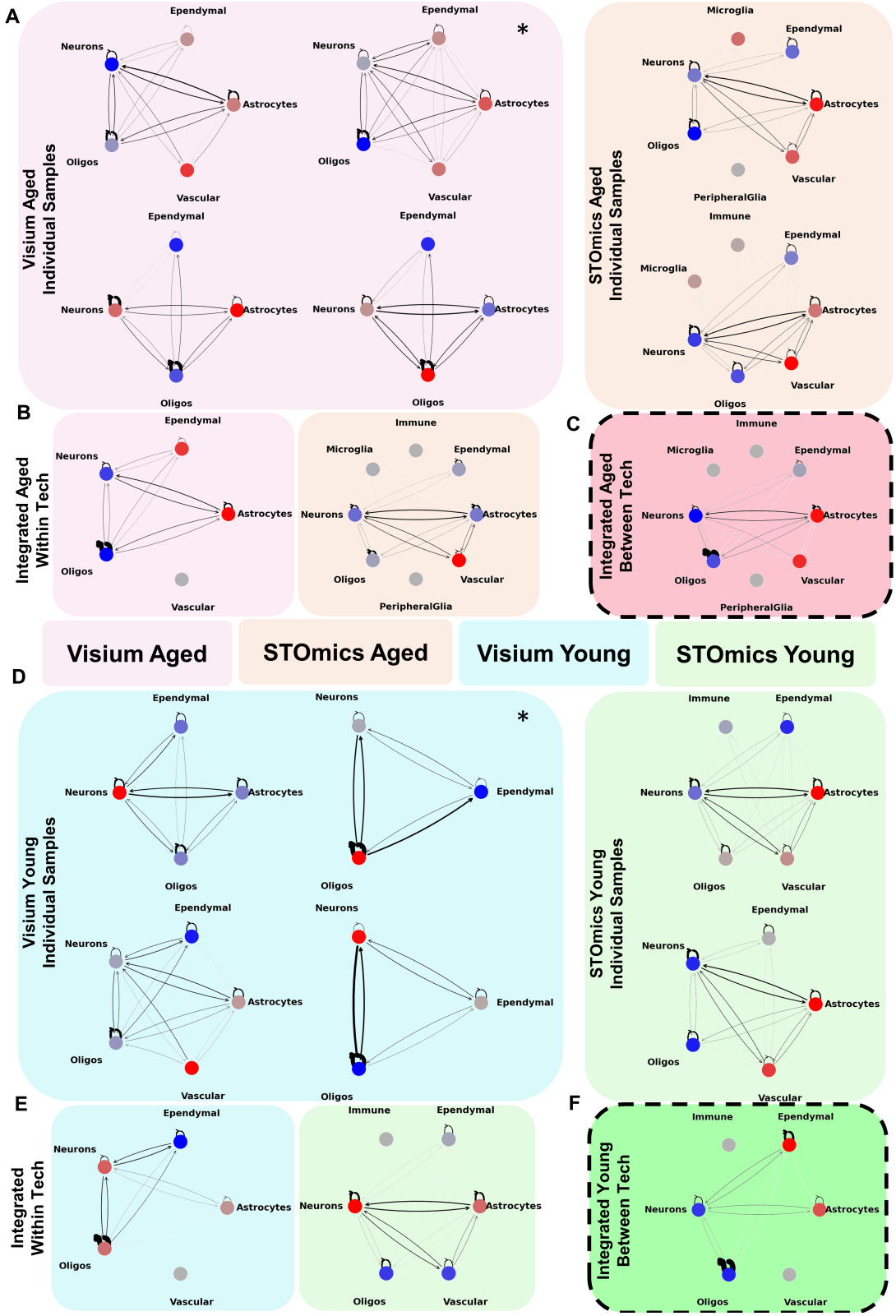


Fig. 2. Integrative CCI analysis of aged and young mouse brains. **A.** Overall networks of individual Visium (left) and STOmics (right) aged mouse brain results. The network with a (*) is the aged sample used in Fig. 3E. **B.** Overall within-technology integration network of aged Visium (left) and aged STOmics (right). **C.** Overall between-technology integration network of all aged samples. This sample was used as the aged brain sample for downstream analyses. **D.** Overall networks of individual Visium (left) and STOmics (right) young mouse brain results. The network with a (*) is the young sample shown in Fig. 3E for spatial distribution. **E.** Overall within-technology integration network of young Visium (left) and young STOmics (right). **F.** Overall between-technology integration network of all young samples. This sample was used as the young brain sample for downstream analyses.

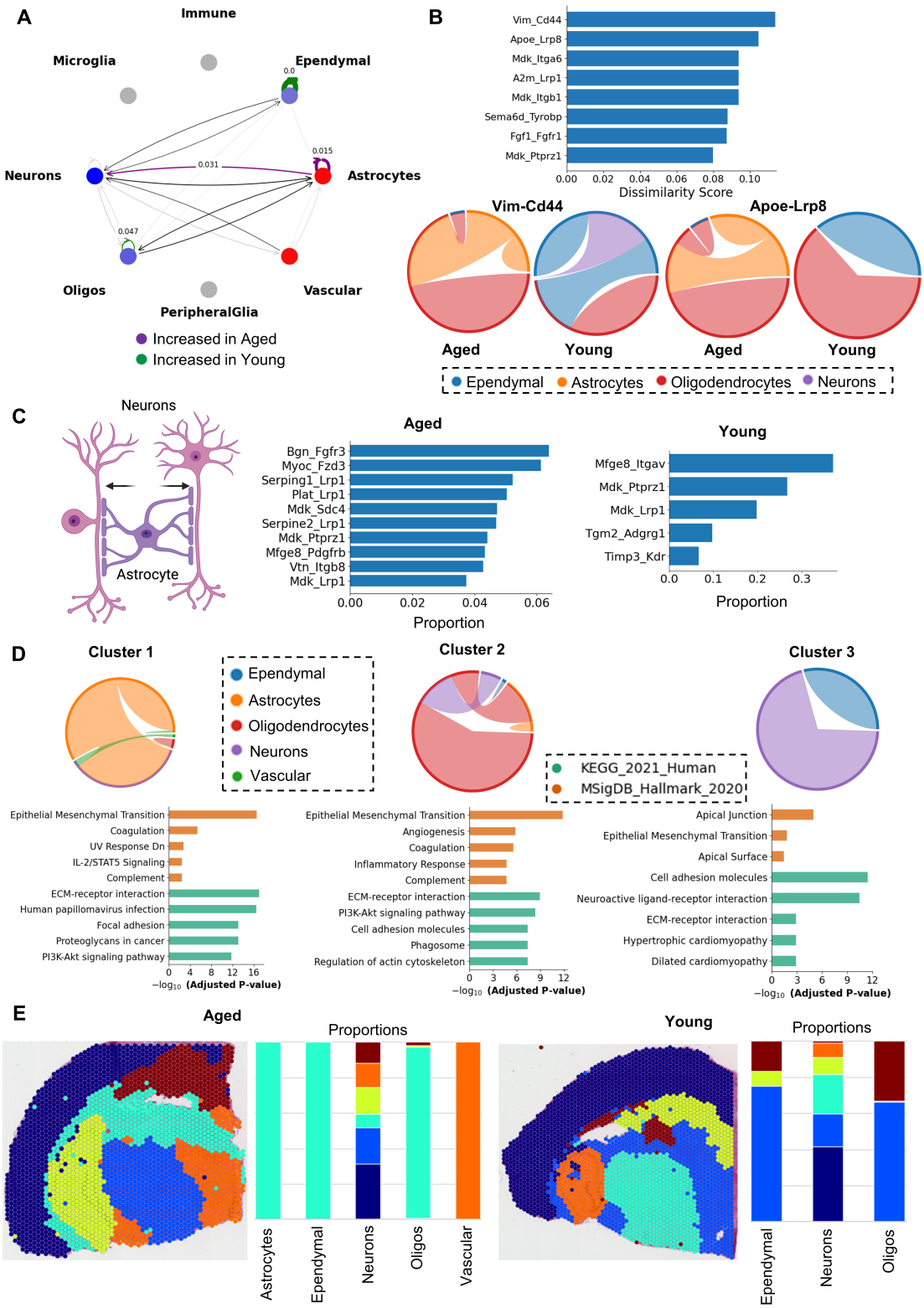


Fig. 3. Downstream analysis of aged and young mouse brain CCI results. **A.** Overall network difference plot between aged and young, where significant edges are coloured and labeled with their p-value. The color of the edge indicates the age group that the interaction is significantly increased in. **B.** Barplot of LR pairs with the most dissimilar networks between aged and young with the top two LR pair's networks for the integrated aged and young samples shown as chord plots. Chord plots are used to show the interacting cell types and the color of the chord indicates the cell type that is acting as mostly a sender. **C.** Barplot of top LR pairs interacting from astrocytes to neurons and their proportion compared to total interactions in that sender-receiver pair, shown for both aged and young. Created with BioRender.com. **D.** Three clusters of LR pairs from the integrated aged sample, where LR pairs with similar networks are clustered together, along with GSEA pathway analysis using EnrichR with the KEGG 2021 Human and MSigDB Hallmark 2020 databases. **E.** LR Interaction results for an aged and a young Visium brain sample with spatial and bar plots, showing the cell type contents of each cluster.

queried to find the LR pairs involved along with their relative strength, shown by their interaction count proportion in that sender-receiver cell-type pair (**Fig. 3C**). These results show that not only did the overall strength of the interactions between the cell types increase in age, but so did the number of LR pairs as well. On top of this, there were mostly different LR pairs interacting in the aged compared to the young.

In many of these interactions, predominantly in the aging brain, Lrp1 was involved, a multifunctional cell surface receptor [36]. In the young brain, Midkine (Mdk), a neurotrophic growth factor that is involved in growth and proliferation during embryogenesis, was the active ligand, while in aging, the serpin family became the predominant ligands acting on Lrp1 [37]. Other ligands involved in the young brain include Timp3 and Tgm2, which are both involved in neurogenesis and maintenance [38, 39].

LR Network Clustering and Pathway Analysis

To summarise the many interactions, we applied LR network clustering on the integrated aged sample, which clustered LR pairs with similar networks together. This analysis was performed to identify LR pairs with similar cell-type networks, grouping LR pairs with similar biological roles. Further, GSEA analysis was run on each cluster to find the pathways involved in the interactions occurring between cell types in a particular cluster (**Fig. 3D**). Cluster 1 in **Fig. 3D** grouped LRs interacting mostly between astrocytes and neurons and we found pathways related to ECM-receptor interaction and focal adhesion, which are significant components of the astrocyte to neuron interaction [40]. Cluster 3, which consisted of LRs self-interacting in neurons and ependymal cells also displayed expected pathways such as neuroactive LR interactions. These results show how analyzing a specific subset of LR pairs can reveal specific and relevant biological pathways.

LR Interaction Clustering

We also explored a new way of clustering the spatial transcriptomics data based on LR interaction scores. The LR network clustering forms a shared network of interacting cell-types, and the LR interaction clustering adds onto this information by indicating the spatial location of the interacting cell-types through the partitioning of LRs based on interaction scores. The stacked bar plots show the proportion of the interacting cell-types across different clusters (**Fig. 3E**). In the aged brain samples, we found astrocytes, ependymal and oligodendrocytes spatially clustered together, indicating they were interacting together. Pathway analysis on the LR-pairs interacting between those three cell-types in the aged revealed pathways related to neurodegeneration not present in the young brain. The plots in **Fig. 3E** correspond to the two marked samples in **Fig. 2**.

Melanoma

Cell-Cell Interactions

The melanoma dataset consists of four Visium samples, eight CosMX samples, and six Xenium samples. The Visium and CosMX samples were matched across four patients, so the samples were integrated by patient, which was able to remove many of the likely false positive interactions that were not common between the two technologies (**Fig. 4A**). After integration, all the patients gave similar interaction networks, despite the individual samples having high intervariability, particularly the CosMX samples, highlighting the strengths of the integration method. Samples were also integrated based on technology and each integrated technology sample was then integrated into one final integrated sample (**Fig. 4B**). This integrated sample was highly similar to the integrated patient samples, which show that the Xenium samples were able to support the other technologies' findings. This between-technology integrated sample was used for all the further downstream analyses.

To understand the biological processes happening between particular cell types of interest, we looked into the GSEA pathways of the LR pairs interacting across three different sender-receiver cell-type pairs. (**Fig. 4C**). For fibroblasts to endothelial cells, most of the interactions involved COL1A1 ligands, which are considered to influence tumor invasion and progression [41]. Pathways such as focal adhesion and ECM-receptor interaction appeared in the pathway analysis (**Fig. 4C**). For natural killer and T cells to myeloid cells, many cell adhesion molecules (CAMs) were involved in the interactions, which are known to be dysregulated in melanoma, particularly members of the integrin and immunoglobulin family [42]. This is reflected in the pathway analysis, which had many immune and CAM related pathways (**Fig. 4C**). For myeloid to endothelial cells, CD74 was the dominant target, which is a receptor known to be highly involved in immunomodulation in various types of cancers, affecting activities of infiltrated cells in the tumor microenvironment [43]. Many relevant immune and tumor microenvironment alteration pathways to this receptor's function were discovered in the pathway analysis (**Fig. 4C**). These results show how analyzing specific cell-type interactions provide informative biological insights that may be obscured in global interaction analyses.

Discussion

We built MMCCI, as a fast, complete, and open-source Python package for integrating, analyzing and visualizing the LR interactions between cell types in spatial transcriptomics and scRNA-seq data across multiple modalities. MMCCI is the first platform for integrating CCI networks both within and between multiple modalities in order to strengthen the confidence of the inferred interactions. It is compatible with multiple CCI tool outputs, including stLearn, CellPhoneDB, Squidpy, CellChat, and NATMI.

Applying MMCCI's integration and analysis pipeline on simulated breast cancer ST data and multimodal brain and melanoma data, we show the effectiveness of using MMCCI to integrate CCIs both within and between technologies. MMCCI provides insightful biological analyses with the benefit of producing confident results that come with using multiple samples across multiple modalities. Data from each individual sample harbours some insight into the true biological cell-cell communications, but analysing a single sample is prone to producing false discoveries. Through the integration of multiple samples with MMCCI, we can gain a deeper understanding with higher confidence about the intercellular interactions occurring in normal physiology and disease.

The benchmarking results show the benefits of integrating multiple samples, even within the same technology (**Fig. 1D**). Each sample had cell types with neighbors common across all the samples as well as neighbors unique to that sample. This allowed us to confirm whether integration could recover the known ground truth, which were the cell types that neighboured each other in the majority of the samples (B and T cells etc.).

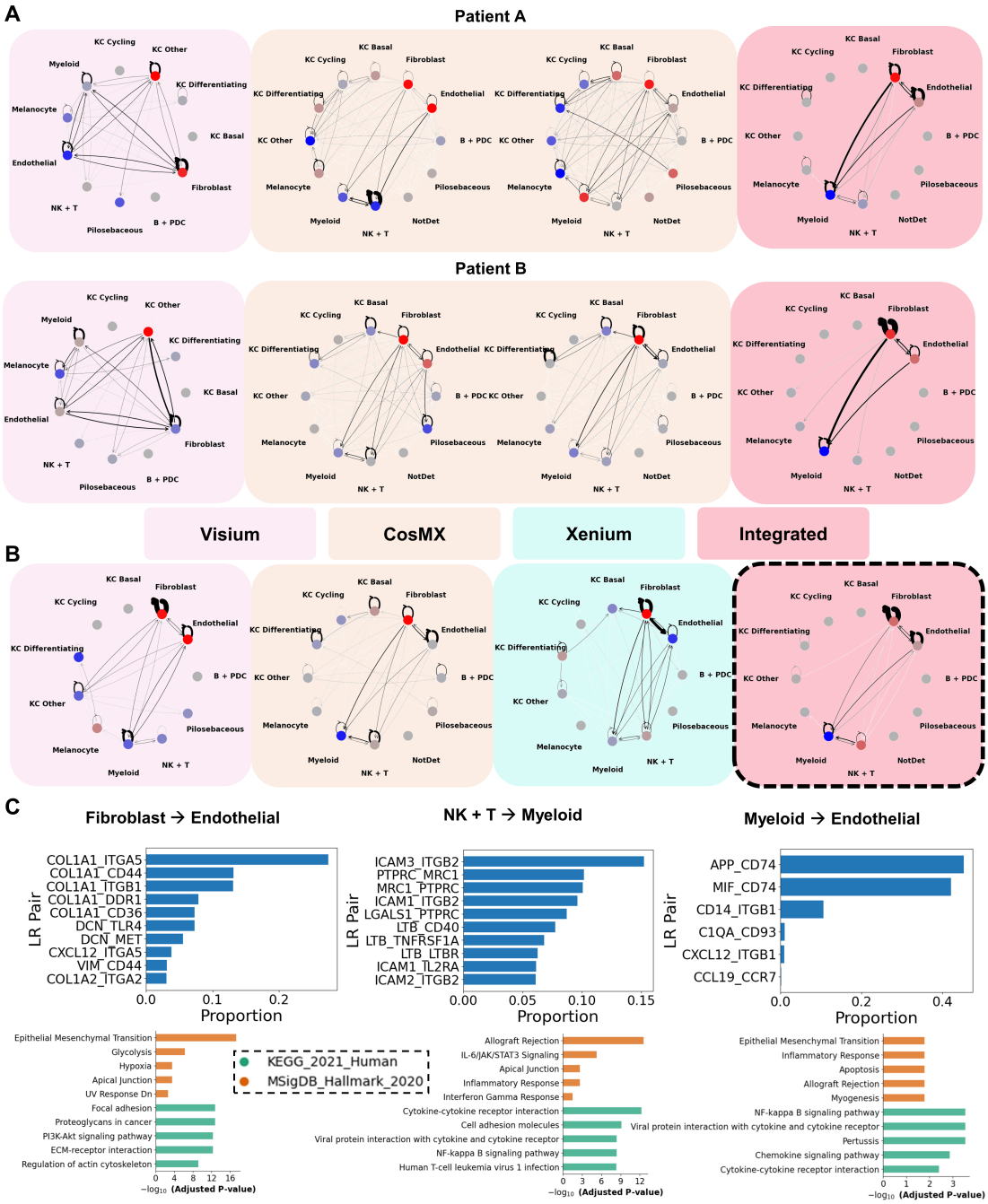


Fig. 4. Integration and analysis of multimodal human melanoma CCI results. **A.** CCI integration results for two patients, where one Visium (pink) and two CosMX (yellow) samples for each patient were integrated together (red). **B.** Within-technology integration results for four Visium (pink), eight CosMX (yellow), and six Xenium (blue) samples, along with between-technology integration of all samples (red). The integrated sample (bordered) is the sample that the downstream analysis is performed on. **C.** Three examples of LR pairs and their proportions that are interacting from a selected sender to a receiver cell type pair and GSEA pathway analysis of the LR pairs using EnrichR with the KEGG 2021 Human and MSigDB Hallmark 2020 databases.

This approach also enables us to correctly filter false positive interactions, which were the cell types that neighboured each other in the minority of the samples (endothelial and stromal cells etc.). In both simulated and biological data, we found that MMCCI successfully removed the false positive interactions while maintaining the true interactions across different spatial and non-spatial CCI methods. We also found that using stLearn CCI on multiple samples and integrating with MMCCI had the highest performance with detecting expected biological interactions while filtering false interactions [3]. stLearn uses spatial information in the analysis, and thus finds more confident interactions.

MMCCI was able to provide relevant biological insights into the multimodal aging mouse brain and melanoma datasets

that may not have been observable if only one sample or modality was analyzed. In the aging brain, astrocytes were shown to become more active, not only increasing their interaction strength but also the types of LR interactions, which is consistent with the current understanding of the roles of astrocytes in aging [27, 28, 29, 30, 31, 32, 33, 34, 35]. The thinning of the ependymal layer expected in aging was also shown by the observed decrease in ependymal interactions in aging [25, 26].

With the melanoma samples, MMCCI was able to take advantage of the different characteristics of multiple ST technologies to provide clear insights into the intercellular interactions between cell types. Many important interactions were captured by this integration, while less significant interactions were filtered out. One example of an important interaction found in the integrated sample was the Macrophage migration inhibitory factor (MIF)-CD74 LR pair. MIF is produced by immune cells and is known to interact with CD74, which is found on a number of different cell types, to promote tumor survival by activating the PI3K/AKT pathway in melanoma, which has made it a therapeutic target and a significant predictive measure for prolonged survival [43, 44, 45]. An example like this shows how MMCCI's integration and analyses are able to reveal biologically significant CCIs with more confidence than what can be inferred from any one modality alone.

MMCCI would be limited with regards to detecting fewer interactions as a result of finding consistent patterns across samples and technologies. However, this problem is markedly less important, given the fact that high false positive detection is the main issue in CCI analysis. Moreover, based on MMCCI's quantitative scoring system for all LR pairs and cell type pairs, it is possible to rank CCIs from high to low confidence, and thereby enabling the assessment of interactions that are not significant by default selection. In addition, several technologies are constrained by the targeted set of genes in the panel (e.g., Xenium and CosMX), and this would limit the detection of CCIs outside panels but are covered in transcriptome-wide technologies (e.g., Visium and STomics). In such cases, imputation of missing genes or applying CCI analysis based on cell type neighbourhood information only would mitigate the technological issues.

Overall, MMCCI is a novel tool that provides integration methods not seen in any other CCI tools. Recently, CellCommuNet was developed to integrate scRNA-seq samples before running CCI analysis, but unlike MMCCI, it does not integrate the CCI results themselves and is also limited to non-spatial scRNA-seq samples [46]. Similarly, LIANA is a tool that runs multiple CCI methods on a single sample and calculates a consensus rank for interactions [47]. However, this is also different to MMCCI in that it only processes a single scRNA-seq sample rather than integrating different samples together. In summary, MMCCI is the first tool that can integrate both spatial and non-spatial CCI results from different transcriptomic modalities and CCI methods together, a much needed method for the rapidly evolving spatial transcriptomics field.

Key Points

- We have developed a new algorithm, MMCCI, which would be the first tool for integrating multimodal spatial transcriptomics and scRNA-seq CCI results.
- MMCCI includes novel downstream analyses for ligand-receptor and cell-cell interactions, allowing users to summarise complex interactions into biologically relevant patterns.
- We found that applying MMCCI integration on samples processed through spatially-aware CCI analysis outperforms existing CCI methods.
- We applied MMCCI integration to multimodal brain and melanoma spatial transcriptomics datasets and were able to significantly improve CCI results by filtering out false interactions.
- We made the MMCCI software open-source and provided detailed usage instructions, aiming to facilitate the CCI analyses by the broad community.

Data Availability

The CCI outputs used to produce all manuscript figures and source code are publicly available via: `https://github.com/BiomedicalMachineLearning/MultimodalCCI`.

Author contributions statement

Q.H.N., L.H., O.M., conceived experiments and developed the algorithms. L.H. and O.M., wrote the software. A.X. conducted experiments. L.H., O.M., and Q.H.N. analyzed data. K.K. and C.M.N. provided samples and helped with data generation. Q.N., K.K., and C.M.N. provided funding for the work. L.H., O.M., and Q.H.N. wrote the manuscript. All authors have reviewed and approved the manuscript.

Acknowledgments

We thank our collaborators at the Genome Innovation Hub, Nanostring, BGI and the Frazer Institute for assistance with data generation. We thank all members in Nguyen's Genomics and Machine Learning Lab. This work was supported by the Australian Research Council [grant number DE190100116 to Q.H.N.], National Health & Medical Research Council [grant number 2001514 to Q.H.N.], NHMRC Investigator Grant [grant number GNT2008928 to Q.H.N.], and the UQ Genome Innovation Hub, and a MGI DNBSEQ MPS Research Grant awarded to C.M.N.. We thank the University of Queensland for providing a scholarship to O.M.

References

1. Bilal N. Sheikh, Olga Bondareva, Sukanya Guhathakurta, Tsz Hong Tsang, Katarzyna Sikora, Nadim Aizarani, Sagar, Herbert Holz, Dominic Grün, Lutz Hein, and Asifa Akhtar. Systematic identification of cell-cell communication networks in the developing brain. *iScience*, 21:273–287, 2019.
2. Xinyi Wang, Axel A. Almet, and Qing Nie. The promising application of cell-cell interaction analysis in cancer from single-cell and spatial transcriptomics. *Semin Cancer Biol*, 95:42–51, 2023.
3. Duy Pham, Xiao Tan, Brad Balderson, Jun Xu, Laura F. Grice, Sohye Yoon, Emily F. Willis, Minh Tran, Pui Yeng Lam, Arti Raghubar, Priyakshi Kalita-de Croft, Sunil Lakhani, Jana Vukovic, Marc J. Ruitenberg, and Quan H. Nguyen. Robust mapping of spatiotemporal trajectories

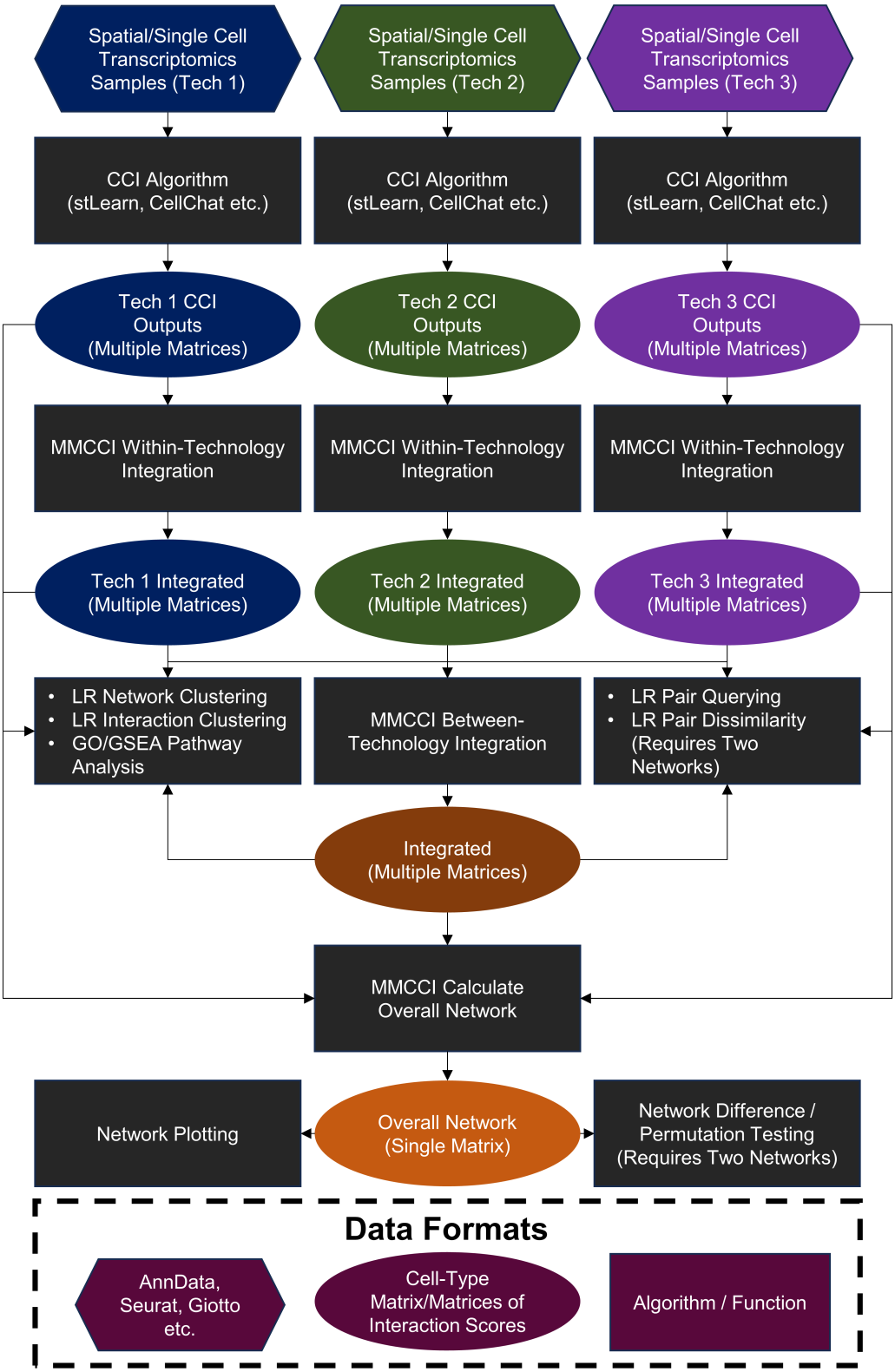
and cell–cell interactions in healthy and diseased tissues. *Nature communications*, 14(1):7739–7739, 2023.

4. Suoqin Jin, Christian F. Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V. Plikus, and Qing Nie. Inference and analysis of cell-cell communication using cellchat. *Nat Commun*, 12(1):1088–1088, 2021.
5. Giovanni Palla, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L. Ibarra, Olle Holmberg, Isaac Virshup, Mohammad Lotfollahi, Sabrina Richter, and Fabian J. Theis. Squidpy: a scalable framework for spatial omics analysis. *Nat Methods*, 19(2):171–178, 2022.
6. Mirjana Efremova, Miquel Vento-Tormo, Sarah A. Teichmann, and Roser Vento-Tormo. Cellphonedb: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc*, 15(4):1484–1506, 2020.
7. Rui Hou, Elena Denisenko, Huan Ting Ong, Jordan A. Ramilowski, and Alistair R. R. Forrest. Predicting cell-to-cell communication networks using natmi. *Nature communications*, 11(1):5011–5011, 2020.
8. Robin Browaeys, Wouter Saelens, and Yvan Saeys. Nichenet: modeling intercellular communication by linking ligands to target genes. *Nat Methods*, 17(2):159, 2020.
9. Cyril Lagger, Eugen Ursu, Anaïs Equey, Roberto A. Avelar, Angela Oliveira Pisco, Robi Tacutu, and João Pedro de Magalhães. scdiffcom: a tool for differential analysis of cell-cell interactions provides a mouse atlas of aging changes in intercellular communication. *Nat Aging*, 3(11):1446–1461, 2023.
10. Maria Solovey and Antonio Scialdone. Comunet: a tool to explore and visualize intercellular communication. *Bioinformatics*, 36(15):4296–4300, 2020.
11. Katy Vandereyken, Alejandro Sifrim, Bernard Thienpont, and Thierry Voet. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet*, 24(8):494–515, 2023.
12. Shanshan He, Ruchir Bhatt, Carl Brown, Emily A. Brown, Derek L. Buhr, Kan Chantranuvatana, Patrick Danaher, Dwayne Dunaway, Ryan G. Garrison, Gary Geiss, Mark T. Gregory, Margaret L. Hoang, Rustem Khafizov, Emily E. Killingbeck, Dae Kim, Tae Kyung Kim, Youngmi Kim, Andrew Klock, Mithra Korukonda, Alecksandr Kutchma, Zachary R. Lewis, Yan Liang, Jeffrey S. Nelson, Giang T. Ong, Evan P. Perillo, Joseph C. Phan, Tien Phan-Everson, Erin Piazza, Tushar Rane, Zachary Reitz, Michael Rhodes, Alyssa Rosenbloom, David Ross, Hiromi Sato, Aster W. Wardhani, Corey A. Williams-Wietzikoski, Lidan Wu, and Joseph M. Beechem. High-plex multiomic analysis in ffpe at subcellular level by spatial molecular imaging. *bioRxiv*, page 2021.11.03.467020, 2022.
13. Amanda Janesick, Robert Shelansky, Andrew D. Gottscho, Florian Wagner, Stephen R. Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A. Morrison, Michelli F. Oliveira, Jordan T. Sicherman, Andrew Kohlway, Jawad Abousoud, Tingsheng Yu Drennon, Seayar H. Mohabbat, and Sarah E. B. Taylor. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat Commun*, 14(1):8353–8353, 2023.
14. Kristen R. Maynard, Leonardo Collado-Torres, Lukas M. Weber, Cedric Uytingco, Brianna K. Barry, Stephen R. Williams, Joseph L. Catallini, Matthew N. Tran, Zachary Besich, Madhavi Tippani, Jennifer Chew, Yifeng Yin, Joel E. Kleinman, Thomas M. Hyde, Nikhil Rao, Stephanie C. Hicks, Keri Martinowich, and Andrew E. Jaffe. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci*, 24(3):425–436, 2021.
15. Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, Xin Wang, Huifang Lu, Xi Chen, Xing Liu, Xin Huang, Zhao Li, Yan Hong, Yujia Jiang, Jian Peng, Shuai Liu, Mengzhe Shen, Chuanyu Liu, Quanshui Li, Yue Yuan, Xiaoyu Wei, Huiwen Zheng, Weimin Feng, Zhifeng Wang, Yang Liu, Zhaohui Wang, Yunzhi Yang, Haitao Xiang, Lei Han, Baoming Qin, Pengcheng Guo, Guangyao Lai, Pura Muñoz-Cánoves, Patrick H. Maxwell, Jean Paul Thiery, Qing-Feng Wu, Fuxiang Zhao, Bichao Chen, Mei Li, Xi Dai, Shuai Wang, Haoyan Kuang, Junhou Hui, Liqun Wang, Ji-Feng Fei, Ou Wang, Xiaofeng Wei, Haorong Lu, Bo Wang, Shiping Liu, Ying Gu, Ming Ni, Wenwei Zhang, Feng Mu, Ye Yin, Huanming Yang, Michael Lisby, Richard J. Cornall, Jan Mulder, Mathias Uhlén, Miguel A. Esteban, Yuxiang Li, Longqi Liu, Xun Xu, and Jian Wang. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792.e21, 2022.
16. Dylan M. Cable, Evan Murray, Luli S. Zou, Aleksandrina Goeva, Evan Z. Macosko, Fei Chen, and Rafael A. Irizarry. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol*, 40(4):517–526, 2022.
17. Elaine H. Shen, Caroline C. Overly, and Allan R. Jones. The allen human brain atlas: Comprehensive gene expression mapping of the human brain. *Trends Neurosci*, 35(12):711–714, 2012.
18. Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, Aaron J. Wilk, Charlotte Darby, Michael Zager, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P. Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar M. Fleming, Bertrand Yeung, Angela J. Rogers, Juliana M. McElrath, Catherine A. Blish, Raphael Gottardo, Peter Smibert, and Rahul Satija. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.e29, 2021.
19. Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*, 17(3):261–272, 2020.
20. F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. Scanpy: Large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15–15, 2018.
21. Zhuoqing Fang, Xinyuan Liu, and Gary Peltz. Gseapy: a comprehensive package for performing gene set enrichment analysis in python. *Bioinformatics*, 39(1), 2023.
22. Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using

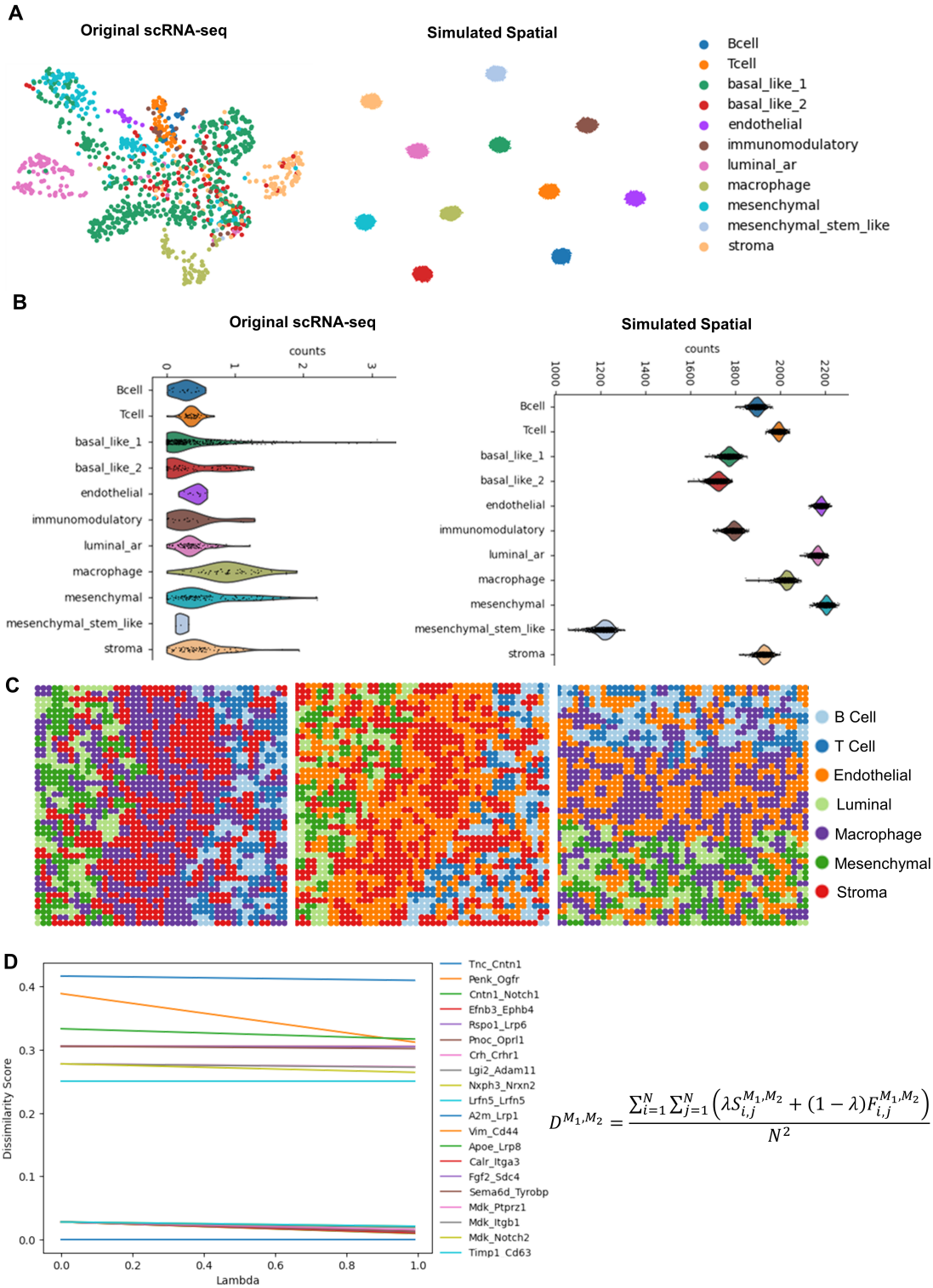
networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

23. M. Claeson, S. X. Tan, D. Lambie, S. Brown, M. D. Walsh, P. D. Baade, N. Pandeya, K. J. Whitehead, H. P. Soyer, B. M. Smithers, D. C. Whiteman, and K. Khosrotehrani. The association between braf-v600e mutations and death from thin ($\leq$1.00 mm) melanomas: A nested case–case study from queensland, australia. *J Eur Acad Dermatol Venereol*, 37(9):e1168–e1172, 2023.
24. Mihriban Karaayvaz, Simona Cristea, Shawn M. Gillespie, Anoop P. Patel, Ravindra Mylvaganam, Christina C. Luo, Michelle C. Specht, Bradley E. Bernstein, Franziska Michor, and Leif W. Ellisen. Unravelling subclonal heterogeneity and aggressive disease states in tnbc through single-cell rna-seq. *Nat Commun*, 9(1):3588–10, 2018.
25. Shiyu Deng, Lin Gan, Chang Liu, Tongtong Xu, Shiyi Zhou, Yiyan Guo, Zhijun Zhang, Guo-Yuan Yang, Hengli Tian, and Yaohui Tang. Roles of ependymal cells in the physiology and pathology of the central nervous system. *Aging Dis*, 14(2):468–483, 2023.
26. Diana G. Nelles and Lili-Naz Hazrati. Ependymal cells and neurodegenerative disease: outcomes of compromised ependymal barrier function. *Brain communications*, 4(6):fcac288–fcac288, 2022.
27. Eleanor Catherine Sams. Oligodendrocytes in the aging brain. *Neuronal signaling*, 5(3):NS20210008–NS20210008, 2021.
28. Lilach Soreq, Jamie Rose, Eyal Soreq, John Hardy, Daniah Trabzuni, Mark R. Cookson, Colin Smith, Mina Ryten, Rickie Patani, and Jernej Ule. Major shifts in glial regional identity are a transcriptional hallmark of human brain aging. *Cell Rep*, 18(2):557–570, 2017.
29. Sergey V. Gudkov, Dmitriy E. Burmistrov, Elena V. Kondakova, Ruslan M. Sarimov, Roman S. Yarkov, Claudio Franceschi, and Maria V. Vedunova. An emerging role of astrocytes in aging/neuroinflammation and gut-brain axis with consequences on sleep and sleep disorders. *Ageing Res Rev*, 83:101775–101775, 2023.
30. Sreedevi Raman, Gayathri Srinivasan, Nicholas Brookhouser, Toan Nguyen, Tanner Henson, Daylin Morgan, Joshua Cutts, and David A. Brafman. A defined and scalable peptide-based platform for the generation of human pluripotent stem cell-derived astrocytes. *ACS Biomater. Sci. Eng*, 6(6):3477–3490, 2020.
31. Charles P. Couturier, Shamini Ayyadhury, Phuong U. Le, Javad Nadaf, Jean Monlong, Gabriele Riva, Redouane Allache, Salma Baig, Xiaohua Yan, Mathieu Bourgey, Changseok Lee, Yu Chang David Wang, V. Wee Yong, Marie-Christine Guiot, Hamed Najafabadi, Bratislav Misic, Jack Antel, Guillaume Bourque, Jiannis Ragoussis, and Kevin Petrecca. Single-cell rna-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nature communications*, 11(1):1–19, 2020.
32. Natallia Makarava, Olga Mychko, Jennifer Chen-Yu Chang, Kara Molesworth, and Ilia V. Baskakov. The degree of astrocyte activation is predictive of the incubation time to prion disease. *Acta Neuropathol Commun*, 9(1):87–87, 2021.
33. Rosalía Fernández-Calle, Sabine C. Konings, Javier Frontiñán-Rubio, Juan García-Revilla, Lluís Camprubí-Ferrer, Martina Svensson, Isak Martinson, Antonio Boza-Serrano, José Luís Venero, Henrietta M. Nielsen, Gunnar K. Gouras, and Tomas Deierborg. Apoe in the bullseye of neurodegenerative diseases: impact of the apoe genotype in alzheimer's disease pathology and brain diseases. *Molecular neurodegeneration*, 17(1):1–62, 2022.
34. Manuel Buttini, Eliezer Masliah, Gui-Qiu Yu, Jorge J. Palop, Shengjun Chang, Aubrey Bernardo, Carol Lin, Tony Wyss-Coray, Yadong Huang, and Lennart Mucke. Cellular source of apolipoprotein e4 determines neuronal susceptibility to excitotoxic injury in transgenic mice. *Am J Pathol*, 177(2):563–569, 2010.
35. Angela Marie Jablonski, Lee Warren, Marija Usenovic, Heather Zhou, Jonathan Sugam, Sophie Parmentier-Batteur, and Bhavya Voleti. Astrocytic expression of the alzheimer's disease risk allele, apoeε4, potentiates neuronal tau pathology in multiple preclinical models. *Sci Rep*, 11(1):3438–18, 2021.
36. Petra May, Astrid Rohlmann, Hans H. Bock, Kai Zurhove, Jamey D. Marth, Eike D. Schomburg, Jeffrey L. Noebels, Uwe Beffert, J. David Sweatt, Edwin J. Weeber, and Joachim Herz. Neuronal lrp1 functionally associates with postsynaptic proteins and is required for normal motor function in mice. *Mol Cell Biol*, 24(20):8872–8883, 2004.
37. Hisako Muramatsu, Kun Zou, Nahoko Sakaguchi, Shinya Ikematsu, Sadatoshi Sakuma, and Takashi Muramatsu. Ldl receptor-related protein as a component of the midkine receptor. *Biochem Biophys Res Commun*, 270(3):936–941, 2000.
38. Lingyan Fang, Takaaki Kuniya, Yujin Harada, Osamu Yasuda, Nobuyo Maeda, Yutaka Suzuki, Daichi Kawaguchi, and Yukiko Gotoh. Timp3 promotes the maintenance of neural stem-progenitor cells in the mouse subventricular zone. *Front Neurosci*, 17:1149603–1149603, 2023.
39. Ruo-Xi Shi, Cong Liu, Ya-Jie Xu, Ying-Ying Wang, Bao-Dong He, Xuan-Cheng He, Hong-Zhen Du, Baoyang Hu, Jianwei Jiao, Chang-Mei Liu, and Zhao-Qian Teng. The role and mechanism of transglutaminase 2 in regulating hippocampal neurogenesis after traumatic brain injury. *Cells*, 12(4):558, 2023.
40. Christabel X. Tan and Cagla Eroglu. Cell adhesion molecules regulating astrocyte–neuron interactions. *Curr Opin Neurobiol*, 69:170–177, 2021.
41. Peiliang Lin, Peng Tian, Jiaqi Pang, Lan Lai, Gui He, Yang Song, and Yiqing Zheng. Clinical significance of col1a1 and col1a2 expression levels in hypopharyngeal squamous cell carcinoma. *Oncology letters*, 20(1):803–809, 2020.
42. Cian D'Arcy and Christina Kiel. Cell adhesion molecules in normal skin and melanoma. *Biomolecules (Basel, Switzerland)*, 11(8):1213, 2021.
43. Shengchao Xu, Xizhe Li, Lu Tang, Zhixiong Liu, Kui Yang, and Quan Cheng. Cd74 correlated with malignancies and immune microenvironment in gliomas. *Frontiers in molecular biosciences*, 8:706949–706949, 2021.
44. Keiji Tanese, Yuuri Hashimoto, Zuzana Berkova, Yuling Wang, Felipe Samaniego, Jeffrey E. Lee, Suhendan Ekmekcioglu, and Elizabeth A. Grimm. Cell surface cd74–mif interactions drive melanoma survival in response to interferon-γ. *J Invest Dermatol*, 135(11):2775–2784, 2015.
45. Yasunari Fukuda, Matias A. Bustos, Sung-Nam Cho, Jason Roszik, Suyeon Ryu, Victor M. Lopez, Jared K. Burks, Jeffrey E. Lee, Elizabeth A. Grimm, Dave S. B. Hoon, and Suhendan Ekmekcioglu. Interplay between soluble cd74 and macrophage-migration inhibitory factor drives tumor growth and influences patient survival in melanoma. *Cell death & disease*, 13(2):117–117, 2022.

46. Qinfeng Ma, Qiang Li, Xiao Zheng, and Jianbo Pan. Cellcommunet: an atlas of cell-cell communication networks from single-cell rna sequencing of human and mouse tissues in normal and disease states. *Nucleic Acids Res*, 52(D1), 2024.
47. Daniel Dimitrov, Dénes Türei, Martin Garrido-Rodriguez, Paul L. Burmedi, James S. Nagai, Charlotte Boys, Ricardo O. Ramirez Flores, Hyojin Kim, Bence Szalai, Ivan G. Costa, Alberto Valdeolivas, Aurélien Dugourd, and Julio Saez-Rodriguez. Comparison of methods and resources for cell-cell communication inference from single-cell rna-seq data. *Nat Commun*, 13(1):3224–3224, 2022.



Sup. Fig. 1. MMCCI pipeline workflow. Square nodes are functions/algorithms and round and hexagon nodes are samples/outputs as outlined at the bottom. scRNA-seq and ST samples stored as AnnData, Seurat, Giotto, or other objects are first processed through one or more desired CCI algorithms such as stLearn or CellChat. The output of the CCI analysis can then be converted using MMCCI to be in the correct data structure for use with MMCCI functions, which is a set of sender cell-types (rows) to receiver cell-types (columns) matrices, one per LR. Each set of outputs within a technology can be run through within-technology integration, and then the integrated sample can then be integrated between technologies with other integrated technology samples. Any samples, integrated or not, can be run through MMCCI downstream analysis and also converted to an overall network for use with plotting and network difference testing.



Sup. Fig. 2. Simulated benchmarking data and dissimilarity score visualization. **A.** UMAP plots showing cell types in the original scRNA-seq and the simulated spatial breast cancer data. **B.** Violin plots of counts of cell types for original scRNA-seq and the simulated spatial breast cancer data. **C.** Spatial plots of simulated spatial breast cancer samples. **D.** Line plot showing how changing the value of lambda in the dissimilarity score function impacts the value of the dissimilarity score across different LR pair examples.