MMCCI: Multimodal Cell-Cell Interaction Integrative Analysis of Single Cell and Spatial Transcriptomics 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 across multiple samples of the same modality as well as between multiple modalities. MMCCI implements new and in-depth downstream analyses, including comparisons between biological conditions, network and interaction clustering, sender-receiver interaction querying, and biological pathway analyses. We applied MMCCI to statistically 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, piecing together the complex interactions and pathways involved in normal physiology and disease at the molecular level with the statistical confidence of using large, multimodal datasets. 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. This information can provide deep biological insights into many applications including neuroscience and cancer research [1, 2]. A valuable 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 Squidpy, CellPhoneDB and NATMI infer communications between any cells or spots and are thus prone to false discoveries, which is a major challenge in CCI analyses. Other tools also perform additional downstream analyses on interaction networks such as scDiffcom and COMUNET, which calculate differential interactions and networks [9, 10].

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 (bin80) 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, there are currently no methods to adequately integrate CCI results from multiple samples from the same or different transcriptomic modalities.

Therefore, we introduce MMCCI as a statistical framework for multimodal CCI integration. We hypothesize that applying MMCCI integration to CCI results across different spatial transcriptomics technologies will be able to filter false positive interactions while preserving the true interactions, thereby constructing a more accurate landscape of the complex intercellular communications than any one sample or modality alone could. As the first pipeline for statistically integrating CCI results across transcriptomic modalities, MMCCI is able

to provide highly confident CCI inferences through statistical meta-analysis, combining p-values for each interaction across each sample, and batch-effect correction. MMCCI also provides a wide range of downstream analyses to answer specific biological questions, such as the interactions involved in the upregulation of specific pathways, identifying differing interactions between two groups, and LR pairs involved in specific cell-types.

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 finding cell-type compositions of spots which contain 1 - 10 cells, and for the single cell resolution CosMx and Xenium data, we used Seurat V4 Label Transfer for direct single cell to single cell annotation [18].

1.2. stLearn Cell-Cell Interaction Analysis

Cell-cell interactions for the aging brain and melanoma samples were computed using stLearn, a CCI inference method that takes spatial information into account [3]. We first filtered out genes that were not present in more than three spots and performed counts per cell normalization. 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, along with a p-value for each weight. 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. 1C**) 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 CCI 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. 1C**). 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. 1C**).

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

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

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

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

Shown above and in **Fig. 1D**, the CCI 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. The option is given in the package to select other methods for selecting which LR pairs to integrate including using all, the majority of, or only common LR pairs between samples. Each LR pair has a cell-type adjacency matrix which has an interaction score for that LR pair for cell-type i to cell-type j, represented as $\text{LR}_{i,j}$ and referred to in this paper as an interaction. Then for each LR pair in that list, the integrated network for that pair is calculated by combining the networks of each sample, creating a matrix where each interaction is the geometric mean of that interaction in all the samples where the score is not 0. If more than half of the samples have an score of 0 for a given interaction, it will have an integrated score of 0 for that interaction.

This integrated CCI sample then has any non-significant interaction scores filtered out using a given p-value cutoff. The significance of an interaction for a sample is provided from the CCI analysis performed on the sample, and the p-values across all samples for the same interaction are integrated as outlined in section 2.3 of the methods. If between-technology integration is to be performed after, this step should be done only after the between-technology integration is performed to improve the accuracy of the meta analysis by using more p-values.

2.2. Between-Technology Integration

The integrated CCI 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. 1C**). 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. 1C**). 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.

Again, the p-values across all samples used in the integration can be combined as outlined in section 2.3 of the methods.

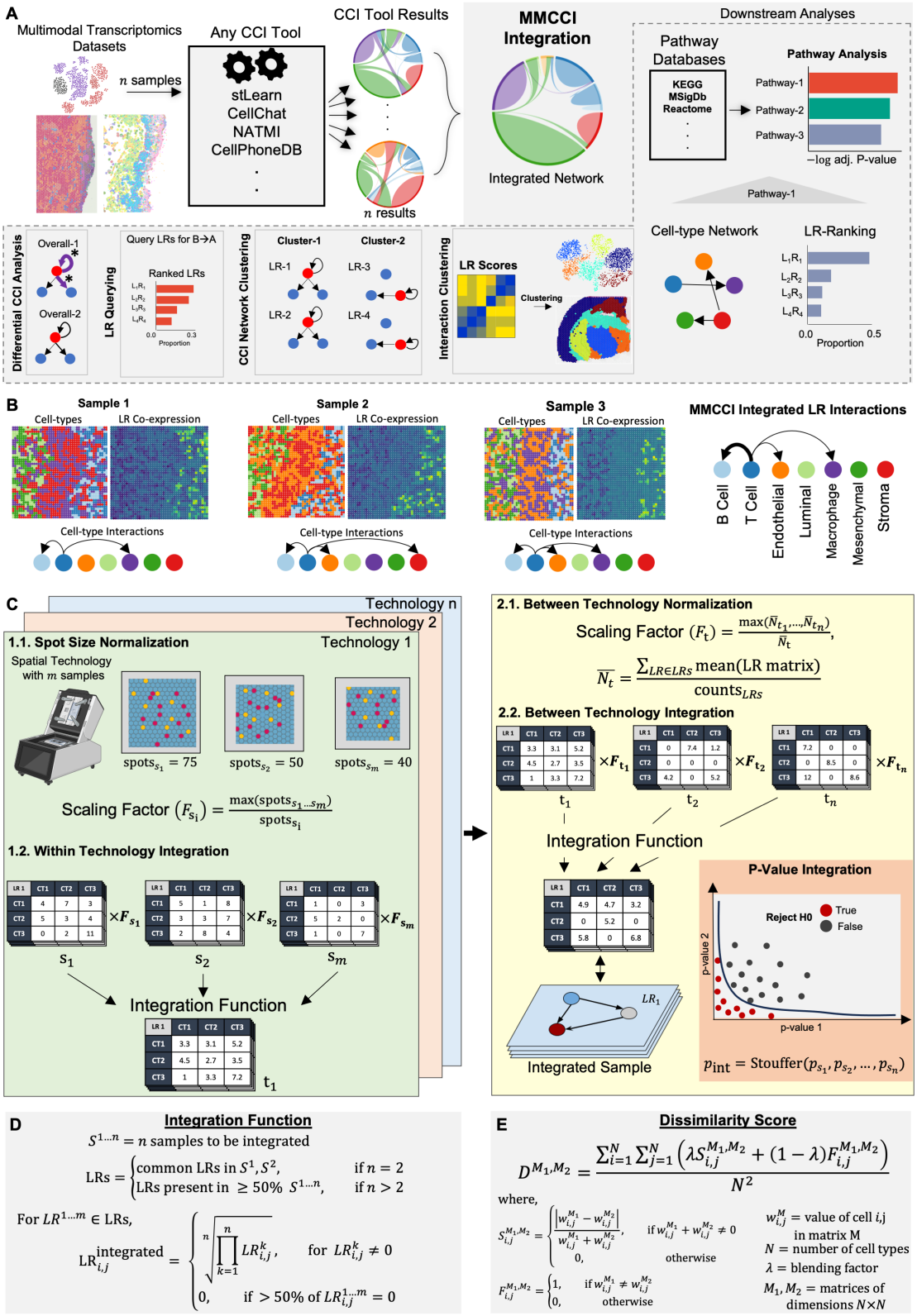


Fig. 1. Overview of MMCCI pipeline and method. **A.** Spatial or single-cell transcriptomics samples are individually processed through one of many different CCI tools. The CCI results from each sample is integrated into one CCI result through MMCCI integration, which is then run through MMCCI's downstream analyses. Shown on the right, GSEA pathway analysis is run on the LR-pairs from the integrated result to identify significant biological pathways and the interactions involved. Shown from the bottom left to right, MMCCI also provides analysis functions including differential CCI analysis with statistical testing, LR pair querying and ranking for a sender-receiver cell-type pair, clustering of LR pairs with similar cell-type networks, and clustering of sample spots with similar interaction scores. **B.** An example of how MMCCI's integration functions. Each spatial transcriptomics sample (simulated from scRNA-seq breast cancer) is run through a CCI algorithm (stLearn in this example). The integration of a single LR pair (HLA-B:KIR2DL3) is shown along with cell-type and LR co-expression spatial plots for each sample. The thicker arrows in the integrated result represents stronger interactions **C.** 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. P-values from the CCI results are integrated using Stouffer's method. Created with BioRender.com. **D.** The algorithm of the integration function, showing how multiple samples are integrated on the individual LR pair level. **E.** The algorithm of the dissimilarity score, used to rank cell-to-cell networks based on how dissimilar they are from each other.

These p-values can be used to filter insignificant interactions in the integrated sample.

2.3. Meta-analysis: A Statistical Framework for Integration

Meta-analysis is a statistical technique to combine the results from multiple different studies addressing a similar question. Most of the CCI algorithms that MMCCI is able to integrate provide p-values for every interaction score, which are integrated to calculate a final p-value for each interaction in the integrated results. We used Stouffer's method to integrate p-values from CCI results of different samples, which converts p-values to z-scores before combining them together. The z-scores are calculated using the inverse cumulative distribution function of p-values (**Equ. 3**), integrated (**Equ. 4**), and then converted back into a p-value using the cumulative distribution function. This meta-analysis improves the robustness and accuracy of MMCCI's integration method.

$$Z_i = \phi^{-1}(1 - p_i) \tag{3}$$

$$Z \sim \frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}} \tag{4}$$

3. Downstream Analyses

3.1. Dissimilarity Scoring

$$D^{M_1,M_2} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(\lambda S_{i,j}^{M_1,M_2} + (1-\lambda)F_{i,j}^{M_1,M_2})}{N^2}, \tag{5}$$

where,

$$S_{i,j}^{M_1,M_2} = \begin{cases} \frac{|w_{i,j}^{M_1} - w_{i,j}^{M_2}|}{w_{i,j}^{M_1} + w_{i,j}^{M_2}} & \text{if } w_{i,j}^{M_1} + w_{i,j}^{M_2} \neq 0, \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$F_{i,j}^{M_1,M_2} = \begin{cases} 1 & \text{if } w_{i,j}^{M_1} \neq w_{i,j}^{M_2}, \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

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

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

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

The dissimilarity score formula, shown above and in **Fig. 1E**, 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 cell-type adjacency matrices, M_1 and M_2, the continuous dissimilarity score, S, is calculated by taking the absolute difference between the weights of each sender-receiver cell-type pair 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. D is a representation of approximately the proportion of the network that is different, and for most cases this difference is around $0 - 0.2$ due to many edges having a weight of 0 in both networks as it is rare to find LR pairs that interact between all cell-types in a sample.

For S, the denominator adds the $w_{i,j}^{M_1}$ and $w_{i,j}^{M_2}$ and in numerator it is subtracted, due to which S is lower for interactions with same numerator, where both values are non-zero compared to one of them being zero. So, 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. 2C**. For networks with many edges that change slightly, the dissimilarity is higher when F is weighted higher, whereas in networks where there are some edges in one and none in the other, there is no difference between F and S.

3.2. Differential CCI Permutation Testing

CCI results provide a network of the different interacting cell-types. The edge represents the number of cells/spots involved in the interaction. Here, the permutation test, represented in **Fig. 1A**, is used to find the significantly different interactions between two networks. The overall interactions from two samples, integrated or unintegrated, are normalized so that the matrix sums to 1 and then subtracted. This could also be done instead for a specific LR pair between two groups. Next, permutation testing is performed by 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 to identify which specific interactions between two cell-types are statistically higher than the background.

3.3. LR Network Clustering

For the LR network clustering represented in **Fig. 1A**, we calculated a pairwise dissimilarity score matrix for each LR pair (**Equations 5-7**). Using this matrix, we calculated 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, or alternatively hierarchical or k-means 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. 3F**.

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. 1A**.

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 receptor 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. Individual pathways can be further analyzed by subsetting a CCI result to only LR pairs that have at least one or both genes involved in the pathway, and then a cell-type network can be plotted as well as the LRs involved and their relative proportion.

4. Visualizations

4.1. Network Visualizations

Network graph plotting is performed using NetworkX, examples of which are in **Fig. 2** [22]. Nodes have an outer ring 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 **Fig. 1** and **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 Stereo-seq Permeabilization User 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 (**Sup. Fig 2B**). All CCI tools used connectomeDB2020 as the LR database [7]. The individual and stLearn integrated cell-type interaction networks were generated for the HLA-B:KIR2DL3 LR pair to show how the integration method works at the LR pair level (**Fig. 1B**).

Results

Aging Brain

Next, MMCCI integration and downstream analyses were performed on the stLearn CCI results from 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. These four combined samples were again integrated by age group, resulting in one aged and one young mouse integrated sample. Prior to integration, many cell types were observed interacting in individual samples, making it unclear which interactions were falsely identified and which were biological. After integration, only statistically significant interactions were retained (**Equations 1-4** 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 5-7**).

Interactions within ependymal cells and oligodendrocytes were shown to decrease significantly with age, while interactions between astrocytes and neurons significantly increased with

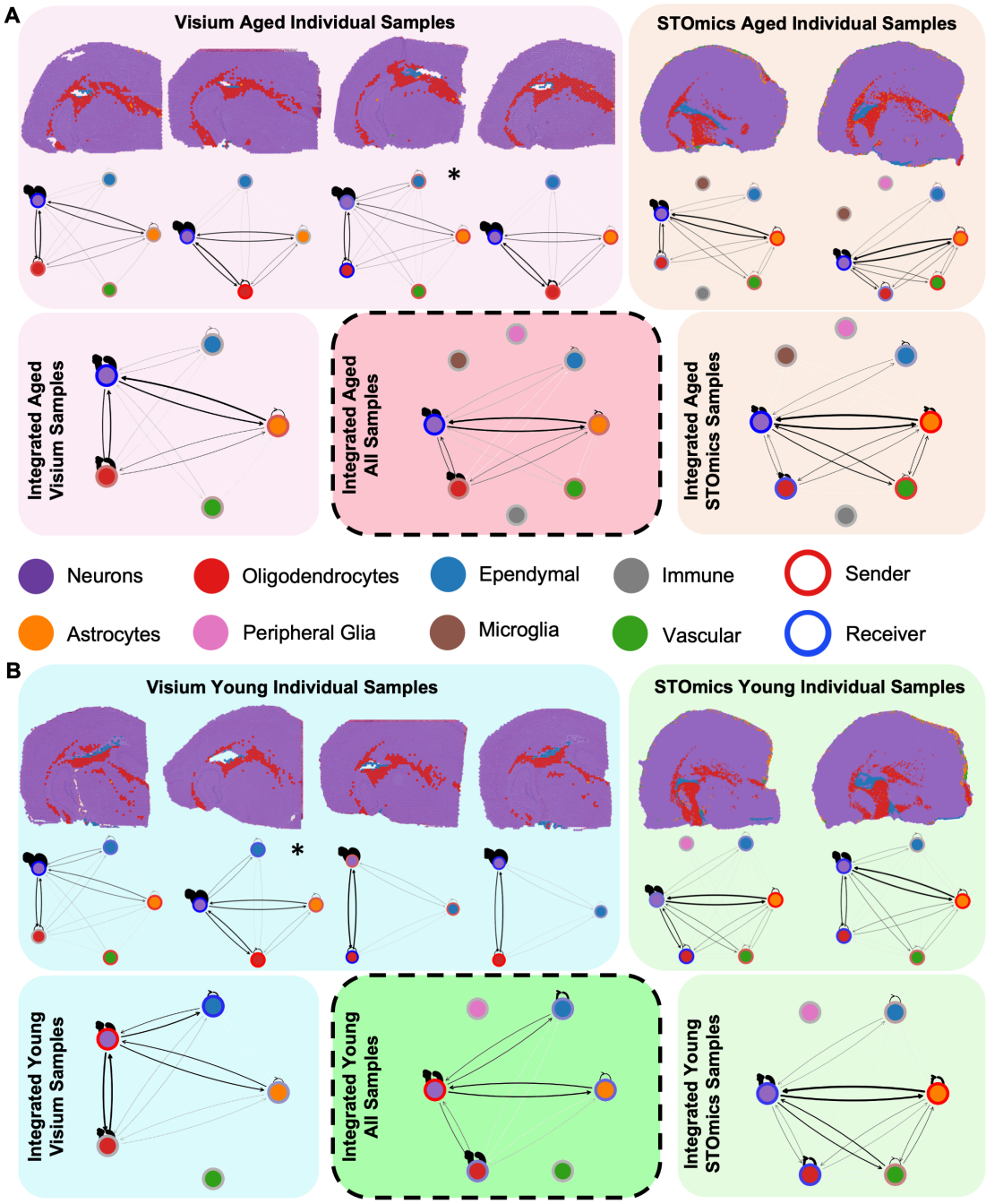


Fig. 2. Integration of stLearn CCI results from aged and young mouse brain spatial transcriptomics samples. **A.** Cell-type spatial plots and overall stLearn interaction network plots of individual Visium (left) and STOmics (right) aged mouse brain results. The network with a (*) is the aged sample used in Fig. 3F. Below, the overall within-technology integration network of aged Visium (left) and aged STOmics (right) is shown along with the overall between-technology integration network of all aged samples (middle). This sample was used as the aged brain sample for downstream analyses. In the network plots, the inner circle of the node shows the cell type and the outer ring shows if the cell type is sending more interactions (red) or receiving more interactions (blue) **B.** Similarly, young samples are shown.

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 5-7**)). The most dissimilar pair, Vimentin (Vim)-Cd44, showed interactions mostly within and between neurons,

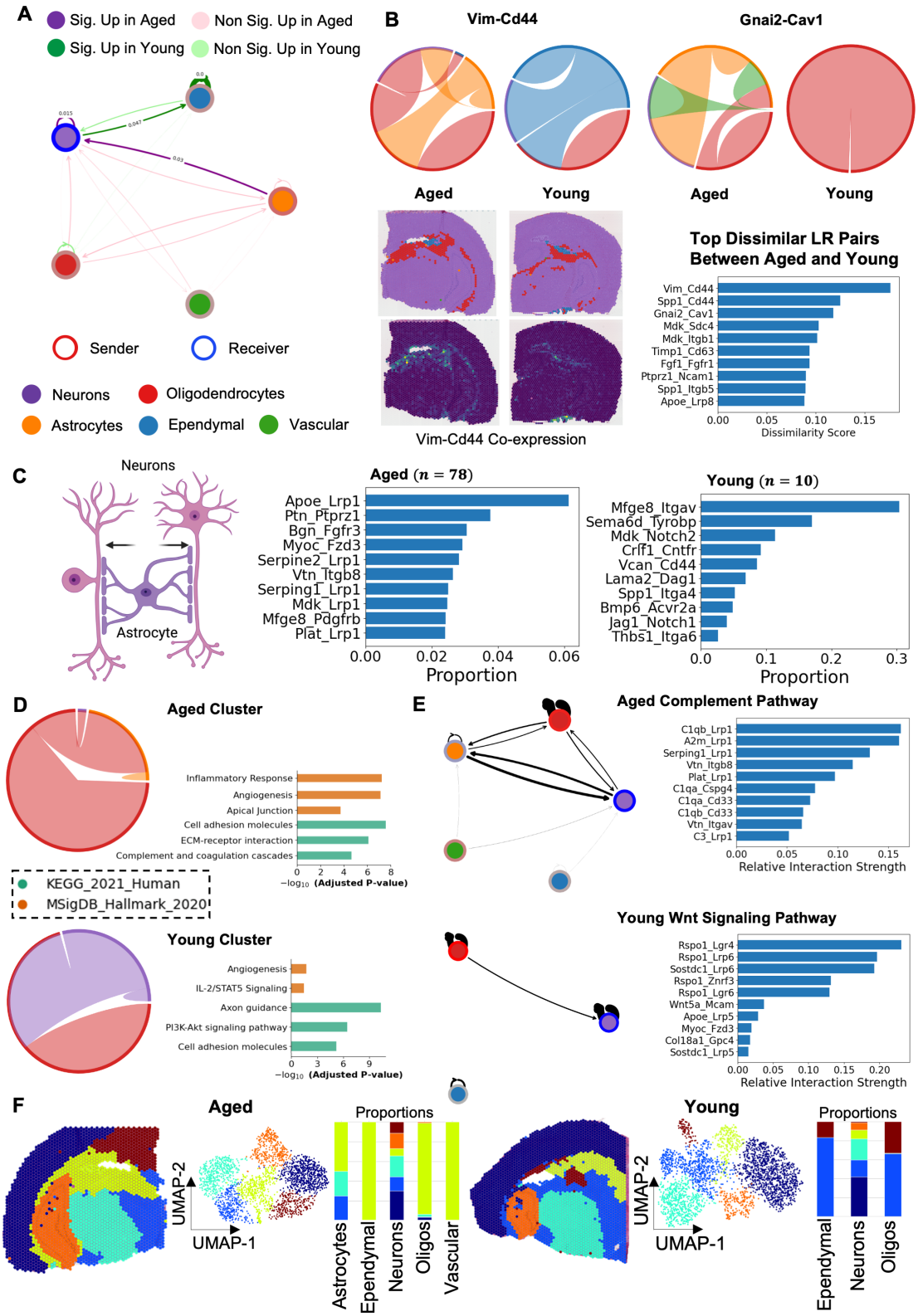


Fig. 3. Downstream analysis of aged and young mouse brain CCI results. **A.** Overall network difference plot between aged and young, where significant edges are coloured and labeled with their p-value. The color of the edge indicates the age group that the interaction is increased in. Edges with a p-value < 0.05 are darker in colour. **B.** Barplot of LR pairs (bottom right) with the most dissimilar cell-type interaction networks between aged and young with the 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. Spatial cell-type and Vim-Cd44 expression plots for an aged and young Visium sample are shown. **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. The number of LR pairs interacting from astrocytes to neurons in both aged groups is shown above the barplots Created with BioRender.com. **D.** A pair of summed networks of clustered LR pairs from the integrated aged and young samples, 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 pairs with their summed networks involved in the complement pathway in the integrated aged sample and the Wnt signaling pathway in the young integrated sample. **F.** LR interaction score clustering results for an aged and a young Visium brain sample with spatial, UMAP and bar plots, showing the cell type contents of each cluster.

astrocytes, and oligodendrocytes in the aged brain, while in the young brain, there were no interactions with astrocytes (**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, Gnαi2-Cav1, interacted only within oligodendrocytes in the young brain, but in the aged, interactions occurred between vascular cells, neurons, astrocytes, and oligodendrocytes. Cav1 is a ligand known to be involved with neuroinflammation at the blood brain barrier, (BBB), and Gαi2 regulates tight junction integrity of BBB [33, 34].

Cell-Cell LR Pair Querying

Since the interactions from astrocytes to neurons were significantly increased in the aged brain, this interaction was 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 [35]. 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 [36]. In the young, LR pairs known to be involved in neurodevelopment are seen, such as Crlf1-Cntfr [37].

LR Network Clustering

To summarise the many interactions, we applied LR network clustering on the integrated aged and young samples separately, 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**). The selected aged cluster in **Fig. 3D** grouped LRs interacting mostly between astrocytes and oligodendrocytes, and we found pathways related to inflammatory response and ECM-receptor interaction, which are significant components of the astrocyte to neuron and oligodendrocyte interaction [38]. The selected young cluster grouped interactions in neurons and oligodendrocytes and discovered neurodevelopmental pathways such as angiogenesis and axon guidance. These results show how analyzing a specific subset of LR pairs can reveal specific and relevant biological pathways.

Pathway Analysis

To look deeper into specific biological pathways, the LR pair networks involved in the complement and Wnt signaling pathways were extracted from the aged and young integrated samples respectively and summed, allowing the visualisation of the cell types involved and the relative strengths of the LR pairs involved. In the aged complement pathway, interactions between astrocytes, neurons, oligodendrocytes were the strongest, with neurons acting as more of a receiver. Lrp1, a protein found predominantly on neurons and reactive astrocytes, was the predominant receptor in the top complement interactions [39]. This receptor, particularly through interactions with A2m, is known to be involved in Alzheimer's disease pathogenesis [40]. The young Wnt/β-catenin pathway had interactions predominantly in oligodendrocytes and neurons, which is expected as this pathway is known to have an important role in oligodendrocyte development and myelin formation [41, 42].

LR Interaction Clustering

Similar to gene expression clustering of ST data, we perform LR interaction clustering, which is based on the LR interaction scores. LR interaction scores were computed using stLearn and it considers the interactions of LRs within a spatial distance up to 250μm.

$$LR_{score} = \frac{1}{2}(mean(Expr_{L,S|N} \times [E_{R,S} > 0])) + mean(Expr_{R,S|N} \times [E_{L,S} > 0]) \quad (8)$$

Shown in **Equ. 8**, stLearn calculates a location-specific LR score for each spot, S with spatial neighbours, N [3]. LR interaction clustering helps to identify the group of cells that are interacting through similar LR pairs. The plots in **Fig. 3F** correspond to the two marked samples in **Fig. 2A, 2B**, where we found the cell-types interacting, and in **Fig 3F** we can visualise the spatial location where those cell-types are interacting. The stacked bar plots show the involvement of each cell-type in interactions in different clusters (**Fig. 3F**). The LR clustering revealed the green cluster as the most distinct between aged and young as it involved interactions between astrocytes, oligodendrocytes, vascular and ependymal cells in aged but not in young. Pathway analysis on the LR-pairs interacting between the four cell-types in green cluster from the aged brain revealed pathways related to neurodegeneration, highlighting the aging process. No aging-related pathways were seen for the interactions in the green cluster in young samples.

Melanoma

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.

Cell-Cell LR Pair Analysis

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. 4D**). For fibroblasts to endothelial cells, most of the interactions involved COL1A1

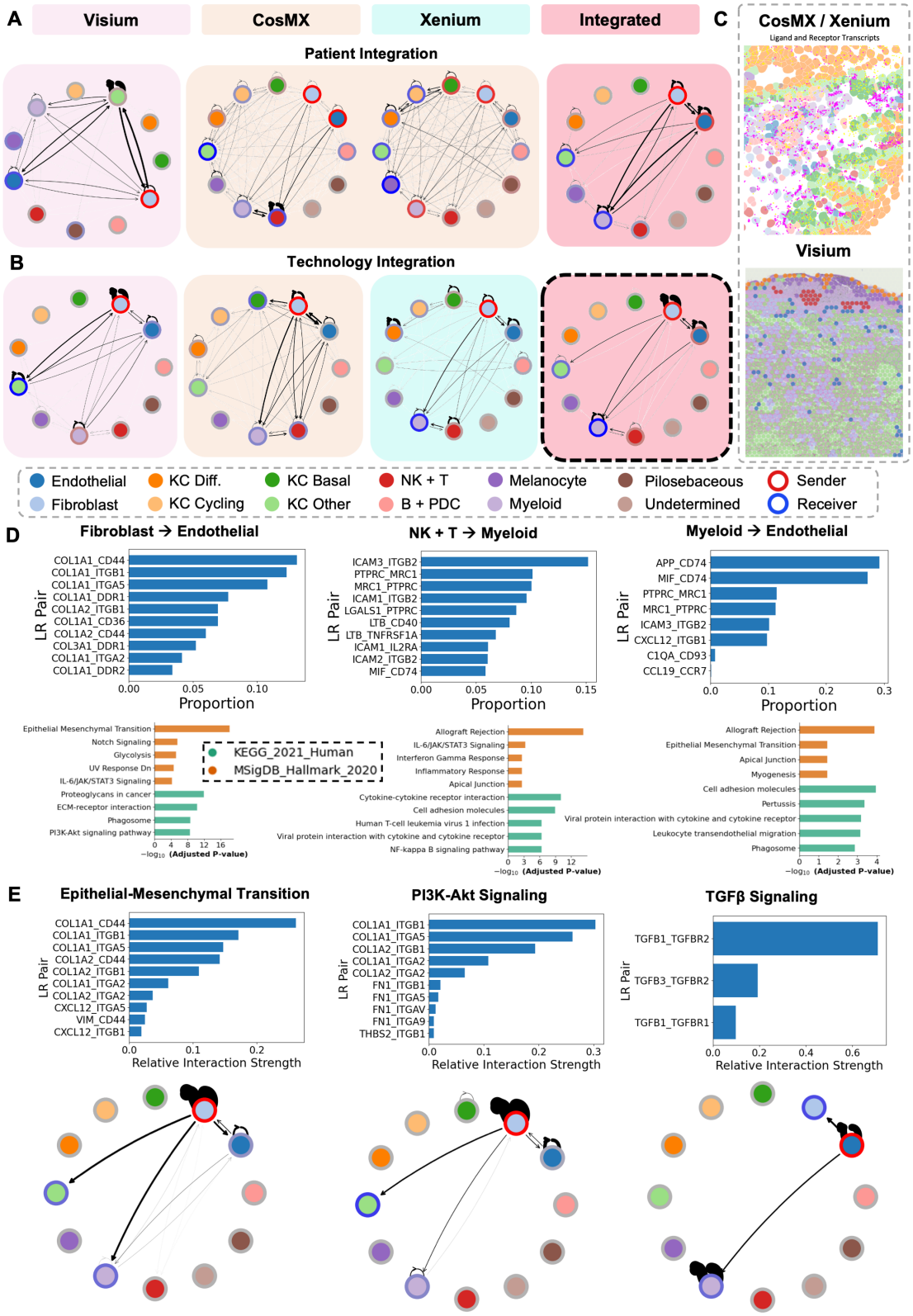


Fig. 4. Integration and analysis of multimodal human melanoma CCI results. **A.** CCI integration results for a patient, where one Visium (pink) and two CosMX (yellow) samples of the same patient were integrated together (red). **B.** Within-technology integration results for four Visium (pink), eight 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.** Two example melanoma spatial plots, showing the difference between the single-cell resolution technologies, CosMX and Xenium, and the non-single cell spatial technologies, Visium. For the CosMX sample, the transcripts for COL1A1 (magenta) and CD44 (yellow) are plotted to show the resolution of these imaging-based technologies. **D.** 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. **E.** LR pairs from the integrated sample involved in different cancer-related pathways and the cell-type network of these LRs.

ligands, which are considered to influence tumor invasion and progression [43]. Pathways such as focal adhesion and ECM-receptor interaction appeared in the pathway analysis (**Fig. 4D**). 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 [44]. This is reflected in the pathway analysis, which had many immune and CAM related pathways (**Fig. 4D**). 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 [45]. Many relevant immune and tumor microenvironment alteration pathways to this receptor's function were discovered in the pathway analysis (**Fig. 4D**). These results show how analyzing specific cell-type interactions provide informative biological insights that may be obscured in global interaction analyses.

Cancer Pathway Analysis

To understand the LR pairs and cell-types involved in different cancer-related pathways, we subsetted the integrated CCI results to only the LR pairs involved in each pathway, which were sourced from the GSEA pathway analysis. For LR pairs involved in epithelial-mesenchymal transition (EMT), COL1A1 and COL1A2 ligands were dominant and CD44, ITGA1, ITGA5, and ITGB1 receptors were the most dominant (**Fig. 4E**). Most of the interactions were from and within fibroblasts to keratinocytes, myeloid cells, and endothelial cells, which is expected as cancer-associated fibroblasts (CAFs) are known to be the main contributors to EMT signaling in the tumour microenvironment (**Fig. 4E**) [46, 47]. PI3K-Akt signaling had similar top LR pairs and a similar cell-type network, but without interactions between endothelial and myeloid cells, which is also expected as CAFs are also involved as senders in the PI3K-Akt signaling pathway [48, 49]. TGF-β signaling was predominantly from TGFB1 to TGFBR2 and from endothelial cells to fibroblasts and myeloid cells as well as within endothelial and myeloid cells, which is also expected as TGF-β signaling is known to occur from endothelial cells to epithelial cells as well as having significant effects on myeloid cells in the tumor microenvironment [50, 51].

Benchmarking

To investigate MMCCI's performance across multiple CCI tools, both spatial and non-spatial, we integrated three simulated breast cancer spatial samples that were run through four different CCI methods - stLearn, CellChat, NATMI and Squidpy (**Sup. Fig. 2B**). For each CCI tool, the integrated result was able to compile a clearer picture of the interactions as a whole across the samples, meaning common interactions are made stronger and rarer interactions are made weaker in the integrated results.

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.). 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], since stLearn uses spatial information in the analysis to find more confident interactions.

Discussion

We built MMCCI, as a fast, complete, and open-source Python package for integrating cell-cell interactions from multimodal transcriptomics datasets, along with a downstream analysis toolset for exploring the biological interactions in single sample as well as MMCCI integrated results. MMCCI is the first platform for integrating CCI networks from different CCI tools both within and between multiple modalities in order to strengthen the confidence of the inferred interactions. MMCCI integrates, analyzes and visualizes outputs from multiple CCI tools, including stLearn, CellPhoneDB, Squidpy, CellChat, and NATMI. 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 [52]. Similarly, LIANA is a tool that runs multiple CCI methods on a single sample and calculates a consensus rank for interactions [53]. However, this is also different to MMCCI in that it only processes a single scRNA-seq sample rather than integrating different samples together.

Applying MMCCI's integration and downstream analyses pipeline on stLearn CCI results from simulated breast cancer ST data and multimodal brain and melanoma data, we show the effectiveness of using MMCCI to statistically integrate CCIs both within and between technologies. MMCCI provides insightful biological analyses, uncovering the interactions involved in specific biological pathways on CCI results integrated across multiple samples and modalities. Data from each individual sample harbours some insight into the true biological cell-cell communications, but analyzing a single sample is prone to producing false discoveries or missing some important information. This advantage of MMCCI integration is clearly seen in the melanoma single-patient integration, shown in **Fig. 4A**. The CosMX samples have many interactions between all cell-type, many of which are likely falsely detected or not biologically relevant. On the other hand, the Visium sample lacks some expected interactions, such as with natural killer (NK) cells, T cells, and melanocytes. After integration, a clearer and more statistically confident picture of the interactions occurring in this tissue is generated. Through the integration of multiple CCI results from different samples using MMCCI, we can gain a deeper understanding with higher confidence about the intercellular interactions occurring in normal physiology and disease.

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 transcriptional modality was analyzed. A clear example of this is that biologically important interactions with vascular cells are missing in many of the Visium brain samples, but after integration with STOmics samples, these are preserved (**Fig 2**). Many important interactions in the aging brain were observed in the integrated results. Mainly, 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 becoming more reactive in aging, contributing to neuroinflammation and the complement pathways [27, 28, 29, 30, 31, 32, 54, 55, 56]. The thinning of the ependymal layer expected in aging was also shown by the observed decrease in ependymal interactions in aging [25, 26]. Overall, MMCCI allows for deep insight to the molecular mechanisms involved in aging in the central nervous system.

With the melanoma samples, MMCCI was able to take advantage of the different ST technologies to provide clear insights into the intercellular interactions between cell types. Many important interactions were captured by the integration of stLearn CCI results across three spatial transcriptomics technologies, while statistically insignificant 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 [45, 57, 58]. MMCCI allowed for a detailed insight into the complicated network of interactions within the tumour microenvironment, particularly with the roles of fibroblasts in EMT and PI3K-Akt signaling. An example like this shows how MMCCI's integration and analyses are able to reveal biologically significant interactions with a higher degree of confidence than what can be inferred from any one sample from a single modality alone.

MMCCI would be limited with regards to detecting fewer or statistically insignificant interactions that would be expected biologically as a result of using only technologies with a lower number of detectable genes, as MMCCI's integration is limited to the quality of the CCI algorithm and the dataset. 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 that would be covered in transcriptome-wide technologies (e.g., Visium and STOmics). In such cases, a future direction of CCI result integration could explore the imputation of missing genes. However, this is not always a significant limitation as the gene panels selected are often selected based on biological relevance to the sample, so the integrated results still provide valuable insight into the biological pathways.

Overall, MMCCI is a novel tool that provides integration methods not seen in any other tools. 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 is the first tool for statistical integration and meta-analysis of multimodal spatial transcriptomics and scRNA-seq CCI results from different tools.
- 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 applied MMCCI integrative analysis to uncover the cell-types and LR pairs involved in complex aging and cancer related pathways.
- We made the MMCCI software open-source and provided detailed usage instructions, aiming to facilitate the CCI analyses by the broad community.

Data Availability

Installation instructions, tutorials, 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, Maksim V. Plikus, and Qing Nie. Cellchat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. *bioRxiv*, 2023.

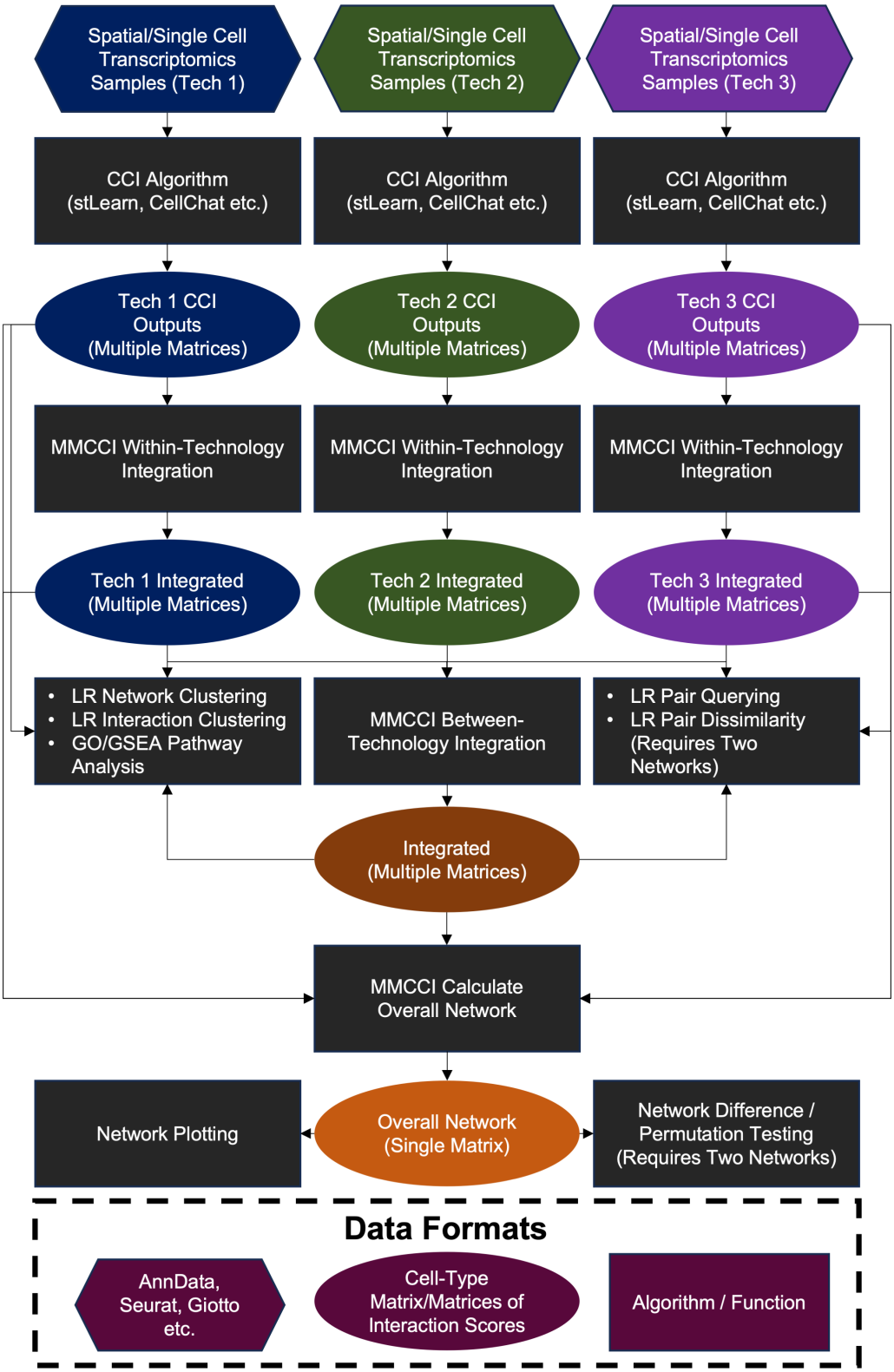
5. Giovanni Palla, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L. Ibarra, Olle Holmberg, Isaac Virshup, Mohammad Lotfollahi, Sabrina Richter, and Fabian J. Theis. Squidpy: a scalable framework for spatial omics analysis. *Nat Methods*, 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 (≤1.00 mm) melanomas: A nested case–case study from queensland, australia. *J Eur Acad Dermatol Venereol*, 37(9):e1168–e1172, 2023.

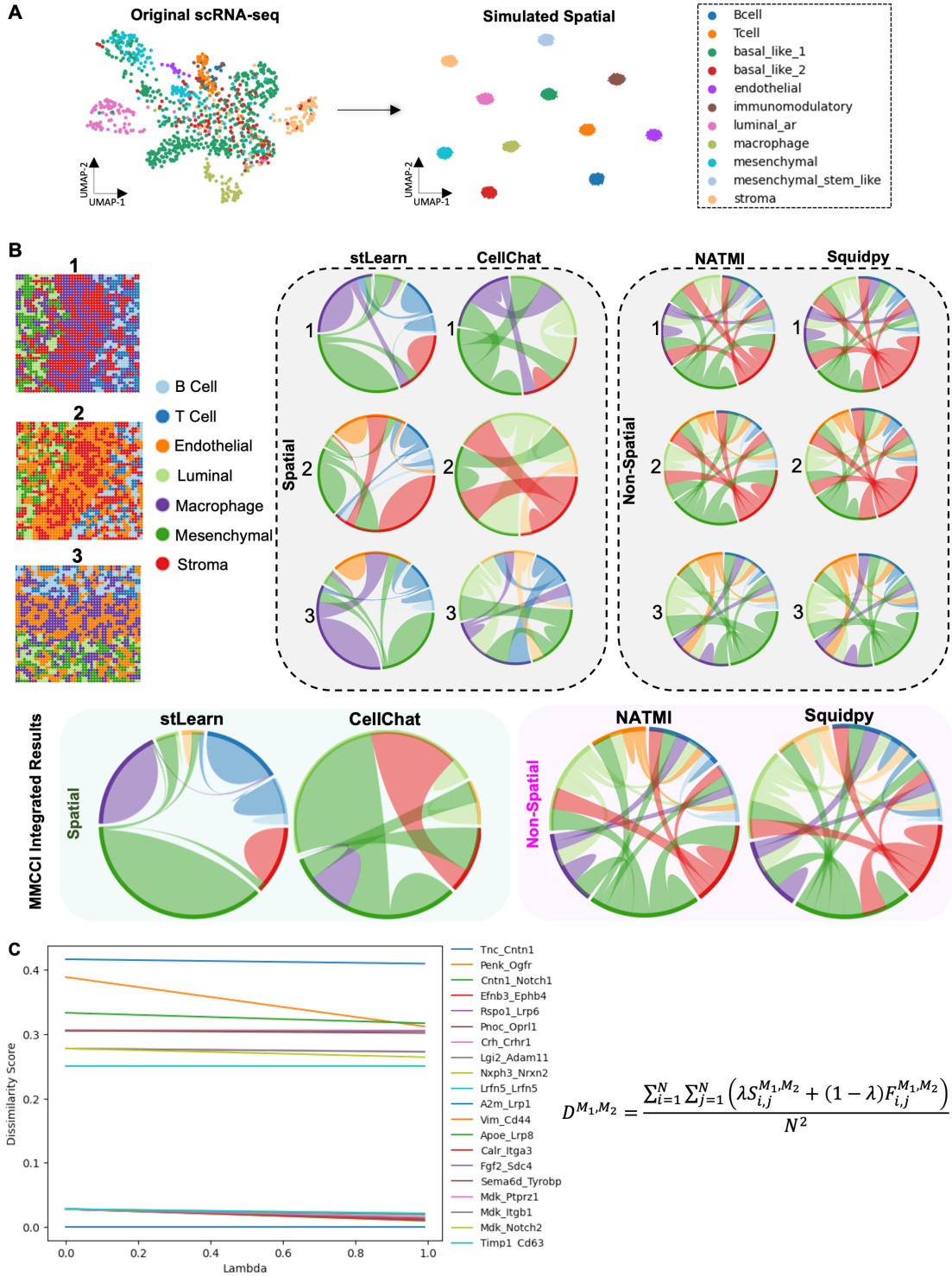
24. Mihriban Karaayvaz, Simona Cristea, Shawn M. Gillespie, Anoop P. Patel, Ravindra Mylvaganam, Christina C. Luo, Michelle C. Specht, Bradley E. Bernstein, Franziska Michor, and Leif W. Ellisen. Unravelling subclonal heterogeneity and aggressive disease states in tnbc through single-cell rna-seq. *Nat 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. Troy N. Trevino, Ali A. Almousawi, KaReisha F. Robinson, Avital B. Fogel, Jake Class, Richard D. Minshall, Leon M. Tai, Justin M. Richner, and Sarah E. Lutz. Caveolin-1 mediates blood-brain barrier permeability, neuroinflammation, and cognitive impairment in sars-cov-2 infection. *Journal of neuroimmunology*, 388:578309–578309, 2024.
34. Anny-Claude Luissint, Christian Federici, François Guillonneau, Fabrice Chrétien, Luc Camoin, Fabienne Glacial, Kayathiri Ganeshamoorthy, and Pierre-Olivier Couraud. Guanine nucleotide-binding protein gi2: A new partner of claudin-5 that regulates tight junction integrity in human brain endothelial cells. *Journal of cerebral blood flow and metabolism*, 32(5):860–873, 2012.
35. 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.
36. 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.
37. Laura Crisponi, Insa Buers, and Frank Rutsch. Crlf1 and clcf1 in development, health and disease. *International journal of molecular sciences*, 23(2):992, 2022.
38. Christabel X. Tan and Cagla Eroglu. Cell adhesion molecules regulating astrocyte–neuron interactions. *Curr Opin Neurobiol*, 69:170–177, 2021.
39. Ulrike Beisiegel, Wilfried Weber, Gudrun Ihrke, Joachim Herz, and Keith K Stanley. The ldl–receptor–related protein, lrp, is an apolipoprotein e-binding protein. *Nature*, 341(6238):162–164, 1989.
40. Quan Yuan, Fen Wang, Sufang Xue, and Jianping Jia. Association of polymorphisms in the lrp1 and a2m genes with alzheimer's disease in the northern chinese han population. *Journal of Clinical Neuroscience*, 20(2):253–256, 2013.
41. Zhong Min Dai, Shuhui Sun, Chunyang Wang, Hao Huang, Xuemei Hu, Zunyi Zhang, Qing Richard Lu, and Mengsheng Qiu. Stage-specific regulation of oligodendrocyte development by wnt/-catenin signaling. *The Journal of neuroscience*, 34(25):8467–8473, 2014.
42. Oligodendrocytes development and wnt signaling pathway. *International journal of human anatomy*.
43. 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.
44. Cian D'Arcy and Christina Kiel. Cell adhesion molecules in normal skin and melanoma. *Biomolecules (Basel, Switzerland)*, 11(8):1213, 2021.
45. 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.
46. Peter M. Szabo, Amir Vajdi, Namit Kumar, Michael Y. Tolstorukov, Benjamin J. Chen, Robin Edwards, Keith L. Ligon, Scott D. Chasalow, Kin Hoe Chow, Aniket Shetty, Mohan Bolisetty, James L. Holloway, Ryan Golhar, Brian A. Kidd, Philip Ansumana Hull, Jeff Houser, Logan Vlach, Nathan O. Siemers, and Saurabh Saha. Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment. *Scientific reports*, 13(1):3051–3051, 2023.
47. Hideyuki Takahashi, Susumu Rokudai, Reika Kawabata-iwakawa, Koichi Sakakura, Tetsunari Oyama, Masahiko Nishiyama, and Kazuaki Chikamatsu. Akt3 is a novel regulator of cancer-associated fibroblasts in head and neck squamous cell carcinoma. *Cancers*, 13(6):1–17, 2021.
48. Yan He, Miao Miao Sun, Guo Geng Zhang, Jing Yang, Kui Sheng Chen, Wen Wen Xu, and Bin Li. Targeting

pi3k/akt signal transduction for cancer therapy. *Signal transduction and targeted therapy*, 6(1):425–425, 2021.

49. Fanglong Wu, Jin Yang, Junjiang Liu, Ye Wang, Jingtian Mu, Qingxiang Zeng, Shuzhi Deng, and Hongmei Zhou. Signaling pathways in cancer-associated fibroblasts and targeted therapy for cancer. *Signal transduction and targeted therapy*, 6(1):218–218, 2021.
50. Yun Yang, Wen-Long Ye, Ruo-Nan Zhang, Xiao-Shun He, Jing-Ru Wang, Yu-Xuan Liu, Yi Wang, Xue-Mei Yang, Yu-Juan Zhang, and Wen-Juan Gan. The role of tgf- signaling pathways in cancer and its potential as a therapeutic target. *Evidence-based complementary and alternative medicine*, 2021:1–16, 2021.
51. Paola A Guerrero. Tgf- activation and signaling in angiogenesis. In *InTech eBooks*, pages 3–23. IntechOpen, RIJEKA, 2017.
52. 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.
53. 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.
54. 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.
55. 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.
56. 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.
57. 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.
58. 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.



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.** Spatial plots of simulated spatial breast cancer samples (left). 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. **C.** 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.