Constructing cell-cell interactions by integrating single-cell RNA sequencing and spatial transcriptomics

Institute:	The Delft Bioinformatics Lab
Supervisors:	Ahmed Mahfouz Mohammed Charrout
Examiner:	Marten Postma
Student:	Qirong Mao
Student number:	2660042
Credits:	30EC
Duration:	5 months
Course name:	Major Bioinformatics Research Pr BSB
Course Code:	XM_0070

Abstract:

Cell-cell interactions (CCIs) are essential for arranging cellular behaviors and organizing them in multicellular organisms. It also orchestrates cellular communication precisely and interplays with the cellular spatial organization. However, the relationship between cellular spatial structure and CCIs is still not clear. Here, we tried to explain this relationship by revealing the correlation between interacting ligand-receptor expression and physical distance using integrated the single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics of human cerebellar cortex. First, we identified interacting ligand-receptor pairs between cell clusters based on the expression patterns of ligands and their cognate receptors using scRNA-seq data. Then, we predicted the spatial location of the cells and calculated the physical distance between cell clusters in the integrated data. Next, we assessed the relationship between the physical distances and the expression of ligand-receptor pairs. Finally, we selected the significant negatively and positively distance-correlated ligand-receptor pairs and analyzed them. We noticed that the negatively distance-correlated ligand-receptor pairs are mainly involved juxtracrine signaling with short intercellular distance while the positively distance-correlated ligand-receptor pairs are mostly involved in autocrine or paracrine signaling with longer intercellular distance. Our results elucidated the CCIs encoded the spatial information in the multicellular organisms.

Introduction:

In multicellular organisms, cell-cell interaction (CCI) plays an essential role in coordinating cellular functions and intercellular relationships (Nussinov, 2013). Cells use CCIs to communicate with one another. These communications can sum up to a complex, dynamic CCIs network, which takes part in different physiological processes (Kirouac et al., 2009), like cell development (Akins & Biederer, 2006; Chen, Fitzgerald, Zimmerberg, Kleinman, & Margolis, 2007), immune response (Batista & Dustin, 2013; Cuartero, Ballesteros, Lizasoain, & Moro, 2015), and tissue homeostasis (Ruch, 2002; Zollinger et al., 2018). Evidence also shows the dysfunction of CCI could result in pathogenesis in some diseases or cancer (Brooke, Nitoiu, & Kelsell, 2012; Lindoso et al., 2016). Therefore, a detailed analysis of CCI is crucial to discover the multicellular biological system.

Many molecules and pathways are associated with CCIs, including ligands and receptors. Signals can be transmitted between cells when secreted ligands are selectively binding to their cognate receptors on the receiver cell (Guryanov, Fiorucci, & Tennikova, 2016). These ligand-receptor (L-R) interactions can cross a different range of intercellular distances: For autocrine and paracrine signaling, the interaction acts on itself or the surrounding cells (Graves et al., 1989). For endocrine signaling, the ligand needs to travel via the circulatory system to contact the receptor of the receiver cells (Chrousos, 2007). These dynamic L-R interactions with other CCIs co-modulate the self-assembly of tissue structure (Misteli, 2001; Nussinov, 2013). It has also been shown that the spatial structure of individual cells can be decoded by the L-R interactions (Purvis & Lahav, 2013; Toda, Blauch, Tang, Morsut, & Lim, 2018). However, the relationship between CCI and cellular spatial structure is still unclear and needs further elucidation.

The emergence of the single-cell RNA sequencing (scRNA-seq) techniques enables us to detect the expression profile of individual cells (Buettner et al., 2015). ScRNA-seq helps reveal the heterogeneity in the cellular level and therefore a powerful tool for analyzing CCIs, especially the L-R interactions (E. Armingol, Officer, Harismendy, & Lewis, 2021). The scRNA-seq can infer the expression of ligands and receptors in the individual cells. Compared with the bulk RNA-seq, the scRNA-seq can deepen our understanding of how CCIs play their roles between heterogeneous or rare cell types in highly complex tissue like intra-tumor or brain tissue (Kumar et al., 2018; Sheikh et al., 2019). Databases like HPMR, CellPhoneDB, and newly Omnipath collected annotations of CCI including L-R interactions (Ben-Shlomo, Yu Hsu, Rauch, Kowalski, & Hsueh, 2003; Efremova, Vento-Tormo, Teichmann, & Vento-Tormo, 2020; Turei et al., 2021). With these databases, methods like CellPhonedb can detected the L-R expression by the mutual expression value of the ligands in the sender cells and the cognate receptors in the receiver cells in the expression profile of scRNA-seq data (Kumar et al., 2018; Ramilowski et al., 2015).

Though the scRNA-seq effectively dissects the cellular heterogeneity, the spatial information is lost during tissue dissociation. Recently, the development of spatial transcriptomics techniques has made it possible to record the spatial gene expression across the selected histologic section (Stahl et al., 2016). By adding a spatial dimension, spatial transcriptomics helps discover the CCIs between spatially

proximity regions (Shao, Lu, Liao, Chen, & Fan, 2020). However, compared to the scRNA-seq, the spatial transcriptomic has a trade-off between gene detection sensitivity with spatial resolution, which could not detect large number of gene in single-cell level simultaneously (Abdelaal, Mourragui, Mahfouz, & Reinders, 2020; Zhou, Jia, Pan, Zhao, & Ge, 2020). Since the advantages and limitations of scRNA-seq and spatial transcriptomics are complementary, integrating these two datasets is regarded as a potential avenue of analyzing complex tissues (Almet, Cang, Jin, & Nie, 2021; de Vries, Mahfouz, Koning, & de Miranda, 2020)

In this research, we want to figure out if CCI contains spatial information. We tried to explain this question by finding the relationship between interacting L-R pairs and intercellular distance. First, we integrated scRNA-seq and spatial transcriptomic data and predicted each cell cluster's location using two established pipelines, Novosparc and Cell2location (Kleshchevnikov et al., 2020; Nitzan, Karaiskos, Friedman, & Rajewsky, 2019). Then, we applied CellPhoneDB v2.0 to detect the interacting L-R expression based on the scRNA-seq data (Efremova, Vento-Tormo, Teichmann, & Vento-Tormo, 2020). Next, we calculated the physical distances between cell clusters in the spatial space and measured the correlation between the physical distance and L-R expression. Finally, we conducted the Gene Ontology (GO) enrichment analysis on significant distance-correlated L-R pairs and focused on the strongest distance-correlated L-R pairs for further biological interpretation, which can help us understand how CCIs interplay with spatial structure in the multicellular organisms.

Method:

Data:

In this research, we analyzed the publicly available single-cell nucleus sequencing (snRNA-seq) dataset from the Allen Brain Atlas (https://portal.brain-map.org/atlases-and-data/rnaseq, "Multiple cortical areas-smart-seq (2019)" release) (Hodge et al., 2019). This dataset recorded 47,432 cells with 50,821 transcripts profiles located in multiple cortex areas of the human brain. The metadata of the dataset labeled the cell type, cortex region, cortical layer information of the individual cell.

The spatial transcriptomic data is acquired from the 10X Genomics Visium

Spatial Gene Expression Dataset (https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Human_Brain_Section_2), which corresponds to the human cerebral cortex section. The dataset recorded 4,972 spots with 36,601 genes profiles.

Preprocessing

We preprocessed the snRNA-seq and spatial transcriptomic dataset by Scanpy v1.7.2 (Wolf, Angerer, & Theis, 2018). For snRNA-seq, first we removed all mitochondrial genes and cells that recorded less than 200 genes and genes expressed in less than 3 cells. Then we normalized the data by dividing the counts by the total number of transcripts in each cell. Next, we scaled the data to 10^6 counts per cell, centered the data in each cell by dividing the standard deviation of each cell's transcript number, and log1p transformed the data. After data normalization and transformation, we selected the 750 most variable genes in the expression profile. After that, the shared nearest neighbor (SNN) graph was constructed in all samples with 6 neighbors and we used the Leiden algorithms to obtain 62 cell clusters with a resolution of 1.5 (Traag, Waltman, & van Eck, 2019).

For the spatial transcriptomic data, First, we removed all mitochondrial genes and genes expressed in less than 3 cells. Then, the dataset scaled and centered by subtracting the mean values and dividing by the standard deviation within each spot. Finally, we log1p transformed the data.

Integrating the scRNA-seq and the spatial transcriptomics data

We applied two published methods for combining the scRNA-seq and the spatial transcriptomics data: Novosparc (Nitzan et al., 2019) and Cell2location (Kleshchevnikov et al., 2020).

Novosparc helps to reconstruct the spatial structure of the scRNA-seq data within the framework of optimal transport (Monge, 1781). The model works under the assumption that cells with similar expression profiles should be close to each other. First, Novosparc calculates the graph-based distance matrix for N cells in the scRNA-seq expression space: $D^{exp} \in R^{N \times N}$, and the physical distance matrix for M locations of where cells located:$D^{Phy} \in R^{M \times M}$. Then, Novosparc aims to get the optimal probabilistic embedding matrix $T_{i,j} \in R^{N \times M}$ by minimizing the discrepancy between

D^{exp} and D^{Phy}, with or without the spatial information of the marker genes. Marker genes can help minimize the discrepancy between D^{exp} and D^{Phy} without iteratively training the model. The matrix $T_{i,j}$ recorded the relative probabilities for each cell $i = \{1,2,\dots,n\}$ in the scRNA-seq dataset to each location $j = \{1,2,\dots,n\}$ in the spatial transcriptomic dataset. Here, we ran the Novosparc with the spatial information of intersecting 597 genes between two datasets as marker genes. We ran the Novosparc after the scRNA-seq and spatial transcriptomic data are scaled, centered and log1p transformed.

Cell2location is a Bayesian model which can map the cell clusters in spatial transcriptomic. First, the algorithm derives the reference signature gene expression matrix from the scRNA-seq data based on the average gene expression profile in different cell clusters. For the spatial transcriptomic data, Cell2location then try to decompose the mRNA count $D_{s,g}$of each gene g in each location s. $D_{s,g}$ follows the negative binomial distribution:

$$D_{s,g} \sim NB(\mu_{s,g}, \alpha_{eg}) \quad (1)$$

Where α_{eg} represents the expression variances between genes and $\mu_{s,g}$ is the unobserved expression level:

$$\mu_{s,g} = m_g\left(\sum_f w_{s,f}\, g_{f,g}\right) + l_s + s_g \quad (2)$$

In this equation, m_g as the scaling parameter to adjust the differences of estimated expression with different spatial transcriptomic technologies.$w_{s,f}$ represents the weight of reference signature gene $f = \{1,2,\dots,n\}$ expression in location s, which can be regarded as how many cells in location s express each reference signature f ; $g_{f,g}$ as the reference signature profiles derived in scRNA-seq, consists of f average expression profiles of gene $g = \{1,2,\dots,n\}$. l_s, s_g represent the location and gene-specific background terms. Here, $m_g, w_{s,f}, l_s$, and s_g are derived from the Bayesian gamma prior model. More details can be found in the supplementary method in Cell2location (Kleshchevnikov et al., 2020).

We ran the Cell2location before the scRNA-seq are centered and log1p transformed, and we used the default hyperparameters here based on the tutorial of the Cell2location model as:

Training iterations: 40,000

The expected number of cells per location: 8

The expected number of cell types per location: 7

The expected number of co-located cell type groups per location: 7

Hyperpriors on gene-specific scaling parameter: mean: $\frac{1}{2}$, standard deviation:$\frac{1}{4}$

The ligand-receptor pairs

Here, we detected the interacting L-R pairs by CellphoneDB v2.0 (Efremova, Vento-Tormo, Teichmann, & Vento-Tormo, 2020). CellPhoneDB contains a multi-subunits ligands and receptors database. Based on this database, CellphoneDB search for possible interacting L-R pairs between paired clusters. Then, for each L-R pair, CellPhoneDB calculates their expression value in different paired clusters. Here, the L-R expression is the average of the mean expression values of the ligand in the sender clusters and the receptor in the receiver clusters:

$$Ligand_{sender} - Receptor_{Receiver} = mean(Ligand_{sender}, Receptor_{Receiver}) \quad (3)$$

If the expression of ligand or receptor is 0, the expression value for this L-R pairs is 0.

For each L-R pair, CellPhoneDB also checked if they are significantly expressed in some sender or receiver cluster based on the permutation test, which CellPhoneDB shuffle the cluster labels of all cells in scRNA-seq to generate a null distribution of L-R expression, then CellPhoneDB check where the observed L-R expression located in the null distribution and define whether they are statistically significant or not (Default: 1000 iterations, $p<0.01$).

Probability matrix and Jensen-Shannon Divergence

Novosparc results in the probability matrix $T_{i,j}$, which represents the relative probabilities of each cell i in scRNA-seq located in location j in spatial transcriptomics data. And for Cell2location, we can derive the cell density matrix $w_{s,f}$. Since Cell2location could not provide the location of individual cell but cell clusters, $w_{s,f}$ represents the absolute cell numbers of each cell cluster t in location s.

Here, we unified $T_{i,j}$ and $w_{s,f}$ into the format of probability matrix $P_{s,t}$ in order to compare their results in the downstream analysis. $P_{s,t}$ represents the probabilities of each annotated cell cluster t in each location s. For each cluster t, their probabilities in all locations sum up to 1.

For $T_{i,j}$, we clustered the individual cells into annotated clusters $t = \{1,2,\dots,n\}$. Within each cluster, we averaged the probabilities of cells in location j. Then we

normalized the data to make sure for each cell cluster t , their sum up probabilities in every location equals to one. For $w_{s,f}$, we scaled the cell numbers within each cell cluster t by dividing them by the total cell numbers in cell cluster t.

We used the Jensen-Shannon Divergence for measuring the similarities between probability distributions of different cell clusters computed by Cell2location or Novosparc (Grosse et al., 2002), defined by:

$$JSD(P\|Q) = \frac{1}{2}D(P\|M) + \frac{1}{2}D(Q\|M) \quad (4)$$

$$M = \frac{1}{2}(P + Q) \quad (5)$$

Where D represents the Kullback–Leibler divergence, P, Q represents two different probability distributions.

Correlations between distances and ligand-receptor expression

Based on $T_{i,j}$, we selected locations for each cell cluster t within the top 1% highest probabilities range (50 spots), representing where cell cluster t are mostly located. For two clusters, the distances between them are based on the Euclidean distance:

$$Distance = \frac{\sum_{t_1=1}^{n_1}\sum_{t_2=1}^{n_2}\sqrt{(x_{t1}-x_{t2})^2+(y_{t1}-y_{t2})^2}}{n1\times n2} \quad (6)$$

Where t_1 and t_2 represent two different cell clusters, x, y represents the coordinates of the location, n_1, n_2 represents the numbers of selected locations of each cell cluster ($n = 50$).

However, within the same distance, the L-R expression might be different since the effects of ligands and receptors expressed in different cell clusters. Since we are focused on the relationship between the distance with the L-R expression, we averaged each L-R pair expression within the same distance group in order to eliminate the effects of cell clusters.

In the end, we matched the physical distance and the L-R expression based on the sender and the receiver cell clusters. Then we calculated the correlations between them by using the Spearman correlation method.

Statistical Analysis

We performed the permutation test to select the significant distance-correlated L-R pairs. We randomly assigned 50 spots for each cluster and calculated the distances

and the correlations between the L-R expression. After 1000 iterations, we generated a null distribution of correlations of each L-R pair. Then we checked where the correlation located in the null distribution and defined whether they are statistically significant or not. The significant L-R pairs are selected with False Discovery Rate (FDR) <0.05 controlled by the Benjamini-Hochberg methods (Benjamini & Hochberg, 1995).

Gene Ontology Analysis

We used the ClusterProfiler R package for the GO enrichment analysis (Yu, Wang, Han, & He, 2012). The GO enrichment analysis uses defined GO terms to describe the gene products based on the target and the background gene list. We analyzed two target gene lists based on the permutation test result: which are the ligand and receptor genes involved in the significant positive, and the significant negative distance-correlated L-R groups. All ligand and receptor genes are selected as the background genes. The significant GO terms are selected with FDR < 0.05 controlled by the Benjamini-Hochberg methods (Benjamini & Hochberg, 1995).

Code availability

The code is available online at the GitHub repository: https://github.com/QirongMao/InternshipCCI

Results:

Here, we integrated the scRNA-seq and spatial transcriptomics of human cerebrum cortex to explore whether CCIs contain spatial information. The scRNA-seq dataset contains 47,432 cells located in multiple cortical regions. For the spatial transcriptomic data, it encompasses 4,972 encoded spots across the human cerebrum (Figure 1).

First, we obtained the expression of the L-R pair based on CellPhoneDB v2.0, which contains a multi-subunits ligand and receptor database. CellPhoneDB measured the L-R expression based on the average value of the ligand's expression value from the sender cell cluster with the receptor's expression value from the receiver cell cluster based on the scRNA-seq data.

Then, we integrated the scRNA-seq and spatial data using Cell2location and

Novosparc. Cell2location is a Bayesian model mapping the scRNA-seq annotated cell clusters into the spots of spatial transcriptomic data. The model extracts the reference signatures expression of each cell cluster from the scRNA-seq, and Cell2location decomposes the gene expression within every location in the spatial transcriptomic data based on the reference signatures and estimates the cell cluster abundance in every location. While Novosparc assumes cells with similar expression profiles are closer to each other and vice versa. The model reconstructs the spatial structure of scRNA-seq by aligning the graph-based expression space with the physical distance space in the framework of optimal transport.

Next, we transformed the results from Cell2location and Novosparc in the format of the probability matrix separately. Based on each probability matrix, we predicted the locations of each cell cluster and calculated the physical distances between them. we applied the Spearman correlation to measure the relationship between physical distances and each L-R pair expression. We conducted the permutation test for each L-R pair to clarify if their correlation with physical distance is statistically significant. Finally, we selected the significant distance-correlated L-R pairs for biological interpretation by GO enrichment analysis and focus on the most distance-correlated L-R pairs.

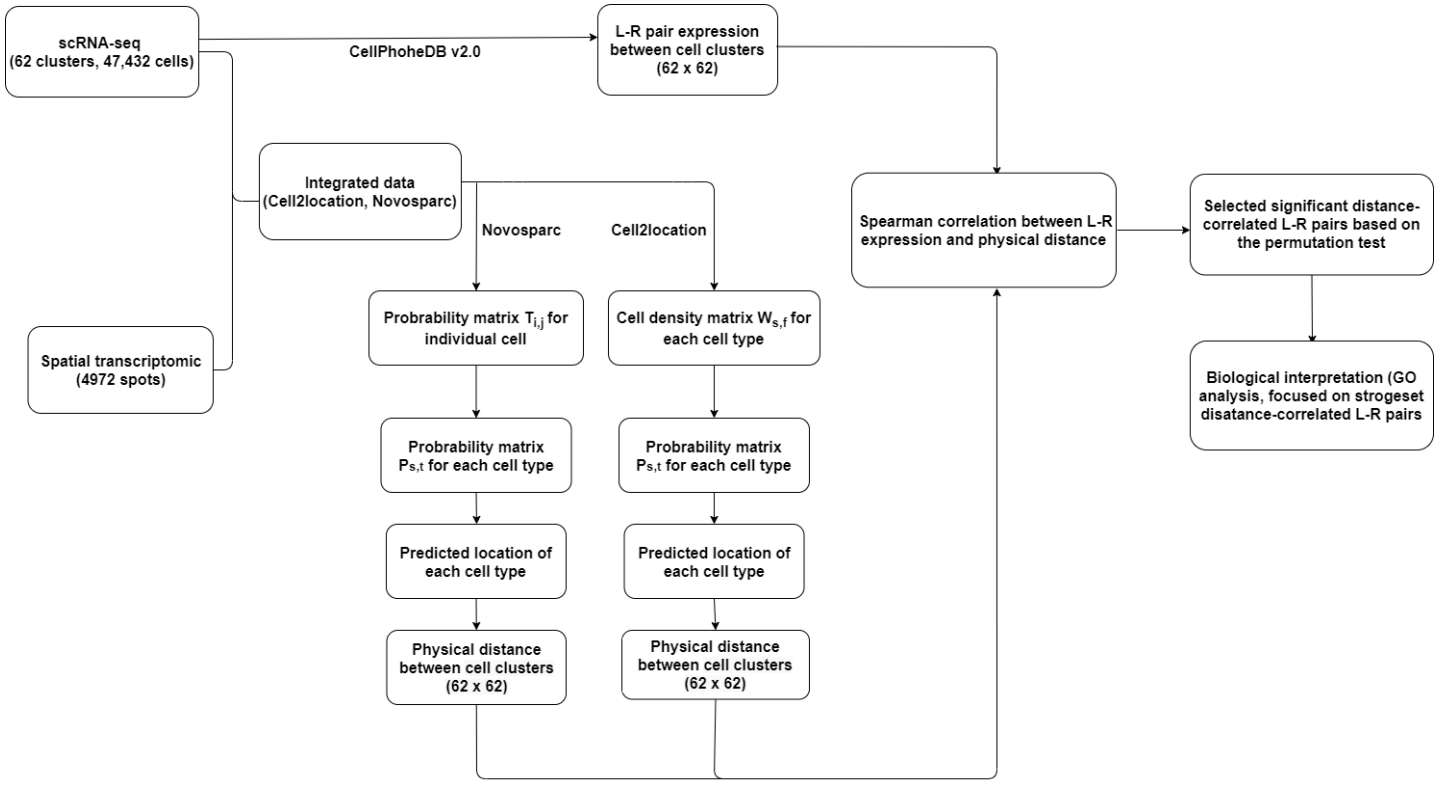


Figure 1: The flowchart of experimental design.

Evaluating the integration of human cerebrum cortex scRNA-seq and spatial transcriptomic data

First, in order to evaluate the performance of Cell2location and Novosparc, we applied annotated cortical layers with the white matter region as input cluster labels to Cell2location and Novosparc, and these clusters are assigned into the spots in spatial transcriptomic data (Figure 2). The results from Cell2location showed defined boundaries of six cortical layers from the inner space extending to the outer area. For Novosparc, the general patterns of some cortical layers can be determined through mapping. However, the results from Novosparc contain more uncertainties compared to the results from Cell2location (Figure 2). Generally, both Cell2location and Novosparc showed capable in integrating the scRNA-seq and spatial transcriptomics.

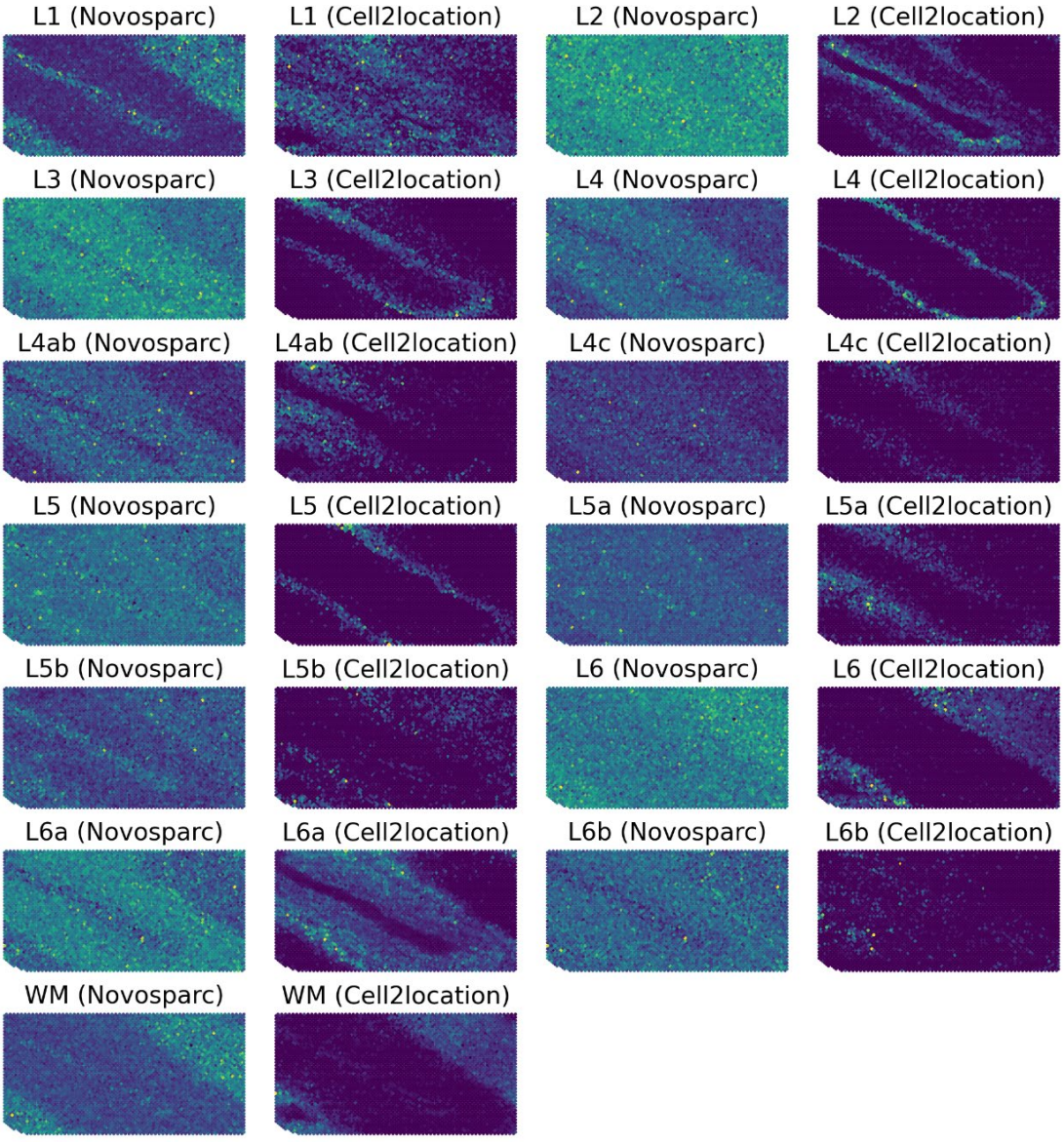


Figure 2: The cortical layers and white matter region assigned by Novosparc and Cell2location based on the probability matrix. Yellow points refer to the probability of the cell cluster located in these spots, points with deeper shade represent higher probability.

Next, we over-clustered the scRNA-seq into 62 cell clusters by the Leiden algorithm (Traag et al., 2019). The purpose of over-clustering the scRNA-seq is to provide more data to calculate the correlation between the L-R expression and the physical distance since both of them are calculated between cluster to cluster. We derived the probability matrix separately from Cell2location and Novosparc, which assigned the occurrence probabilities for each 62 clusters to every spot in the spatial space.

After that, we applied the Jenson-Shannon divergence metric to measure the similarities between the results from Cell2location and Novosparc (Figure 3). In many cases, we noticed that for the same cell cluster, the Jenson-Shannon divergence between their probability spaces from Cell2location or Novosparc are low, which means their results from Cell2location and Novosparc are similar. In the heatmap we also observed that the many probability spaces from Cell2location and Novosparc are clustering together (Figure 3). These results indicated the results from Cell2location and Novosparc shared similarities while integrating the scRNA-seq and spatial transcriptomics.

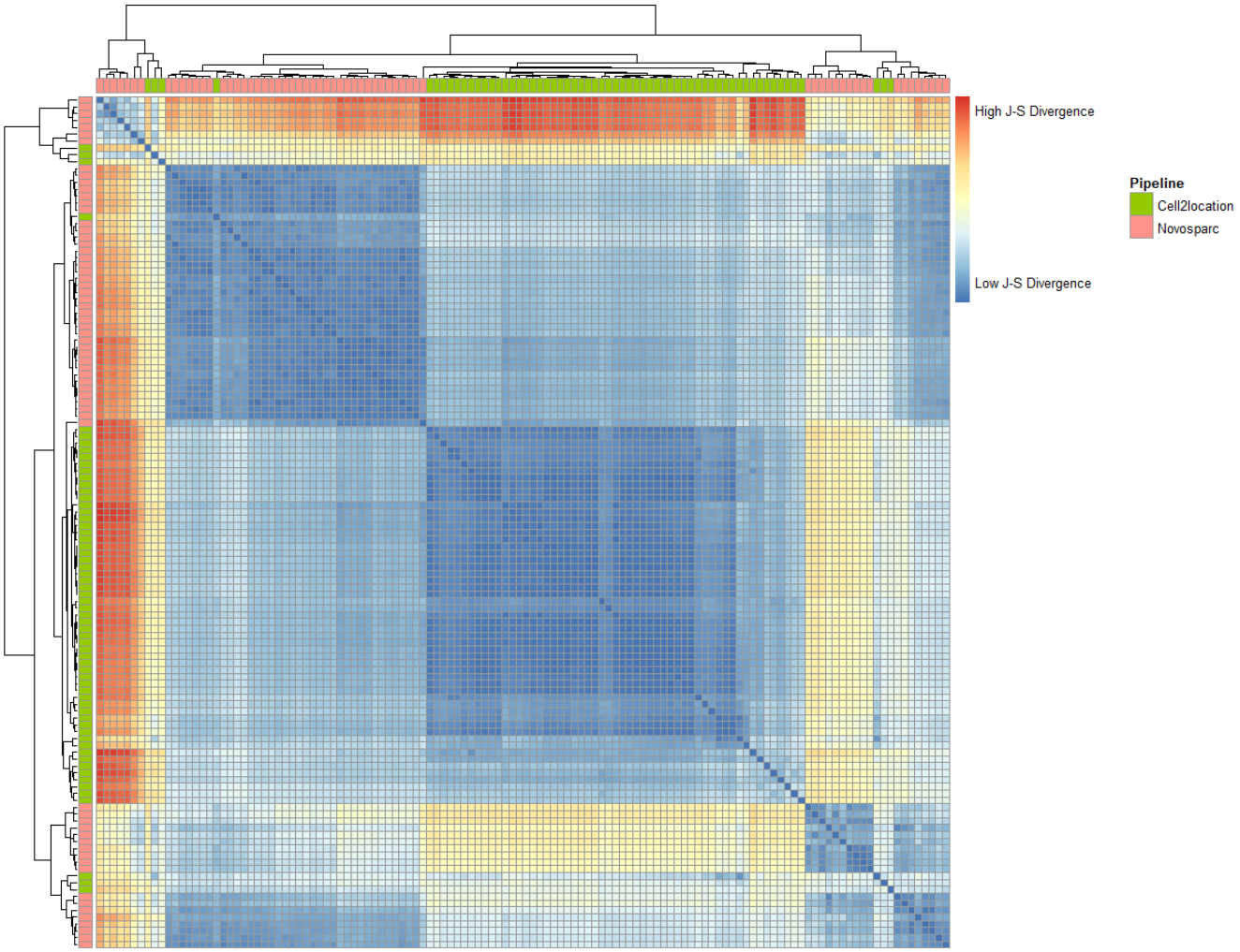


Figure 3: The heatmap of Jensen-Shannon divergence between the probability space of 62 clusters derived by Cell2location or Novosparc separately (label not shown). Both columns and rows represent the clusters and the related pipeline (Cell2location, Novosparc), color from blue to red represent from the low divergence to the high divergence.

Selecting distance-correlated ligand-receptor pairs

Then, we focused on the correlation between the expression of interacting L-R pair with physical distance. We detected interacting L-R pairs by applying CellPhoneDB v2.0. CellPhoneDB can detect interacting L-R pairs by matching expressed transcripts in scRNA-seq with their own L-R pairs repository (Efremova et al., 2020). After selecting L-R pairs, CellPhoneDB then calculate the expression value of each L-R pair between paired cell clusters base on the average expression value of the ligand from the sender cluster and receptors from the receiver cluster (Method,

Equation 3).

Next, we assigned 50 spots with highest probabilities for each cell cluster based on the probability matrix from Cell2location and Novosparc separately. For cell clusters, these spots are the most likely locations that contain them. Then we calculated the distance between paired cell clusters, using the average Euclidean distance between every spot we assigned between these two cell clusters (Method, Equation 6).

After that, we calculated the Spearman correlation between physical distances with the expression value of interacting L-R pairs. Since for each L-R pair, the probability space from Cell2location and Novosparc are different, so we derived two sets of correlations for each L-R pair corresponding to these two methods.

Here, we detected 1347 interacting L-R pairs from the scRNA-seq. For each L-R pair, we performed the permutation test which generate the null distribution of correlations to test if their correlations derived by Cell2location and/or Novosparc are statistically significant or not. First, we observed that the correlations of these L-R pairs derived by Cell2location and Novosparc are mostly in the range of -0.5 ~ 0.5, and we also noticed for many L-R pairs, their expression do significantly correlate to the distance (Figure 4). Moreover, Comparing to Cell2location, the results from Novosparc showing more L-R pairs expression are significant correlated to the distance. Interestingly, we also noticed the correlations from Cell2location and Novosparc are highly matched to each other (Figure 4).

Finally, we selected the negatively and positively distance-correlated L-R pairs from the mutual significant group, which means their correlations calculated by Cell2location and Novosparc are both shown statistically significant in the permutation test (Figure 4).

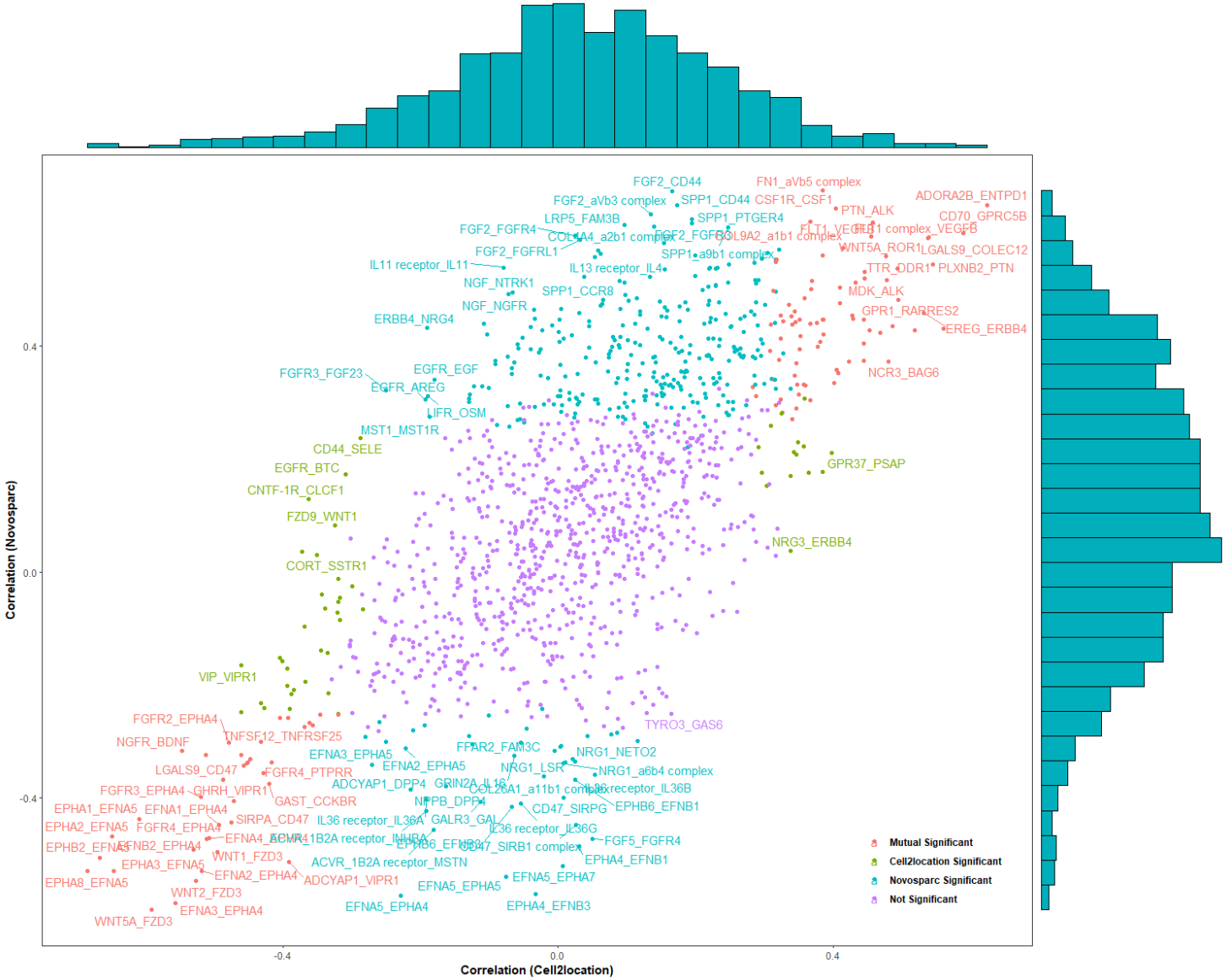


Figure 4: The Spearman correlation between L-R pairs expression and the physical distance, each point represents an L-R pair. The X-axis represents the range of correlation obtained from Cell2location. The bar plot on the top shows the density of the L-R pair's correlations from Cell2location (aligned with the x-axis). Y-axis represents the range of correlation obtained from Novosparc. The bar plot on the right shows the density of the L-R pair's correlations (aligned with the y-axis). Points with red color represent the correlations from both pipelines are statistically significant based on the permutation test. Points with green color represent the correlation from Cell2location is statistically significant. Points with blue color represent the correlation from Novosparc is statistically significant, and purple points represent none of the correlations from both methods are statistically significant.)

Gene Ontology enrichment analysis of distance-related ligand-receptor groups

Then, we conducted the Gene Ontology (GO) enrichment analysis for the ligand and receptor genes in the positively and negatively distance-correlated L-R groups. The GO enrichment analysis provides significant GO terms to describe gene products in three categories: biological process, molecular function, and cellular component (Ashburner et al., 2000). These GO terms enable us to understand further the biological functions of positively and negatively distance-correlated L-R pairs.

For the positively distance-correlated L-R pairs, the GO enrichment analysis showed the significant biological process GO terms are focused on anatomical structure morphogenesis, regulation of cell motility, extracellular matrix organization, cell proliferation, and division. As for the cellular component category, GO terms are significant in the intracellular membrane-bounded organelle, vesicle, organelle lumen, and extracellular matrix. The molecular function category shows significant GO terms like carbohydrate derivative binding, anion binding, etc. (Figure 5).

As for the negatively distance-correlated L-R pairs, the significant GO terms of the biological process are focused on transmembrane receptor protein kinase signaling pathway, neuron projection morphogenesis, and neuron projection guidance. Cellular component GO terms are cell junction, synapse, and presynapse. Significant GO terms of kinase activity, ATP binding, and ephrin receptor activity are mentioned in the molecular function category (Figure 6).

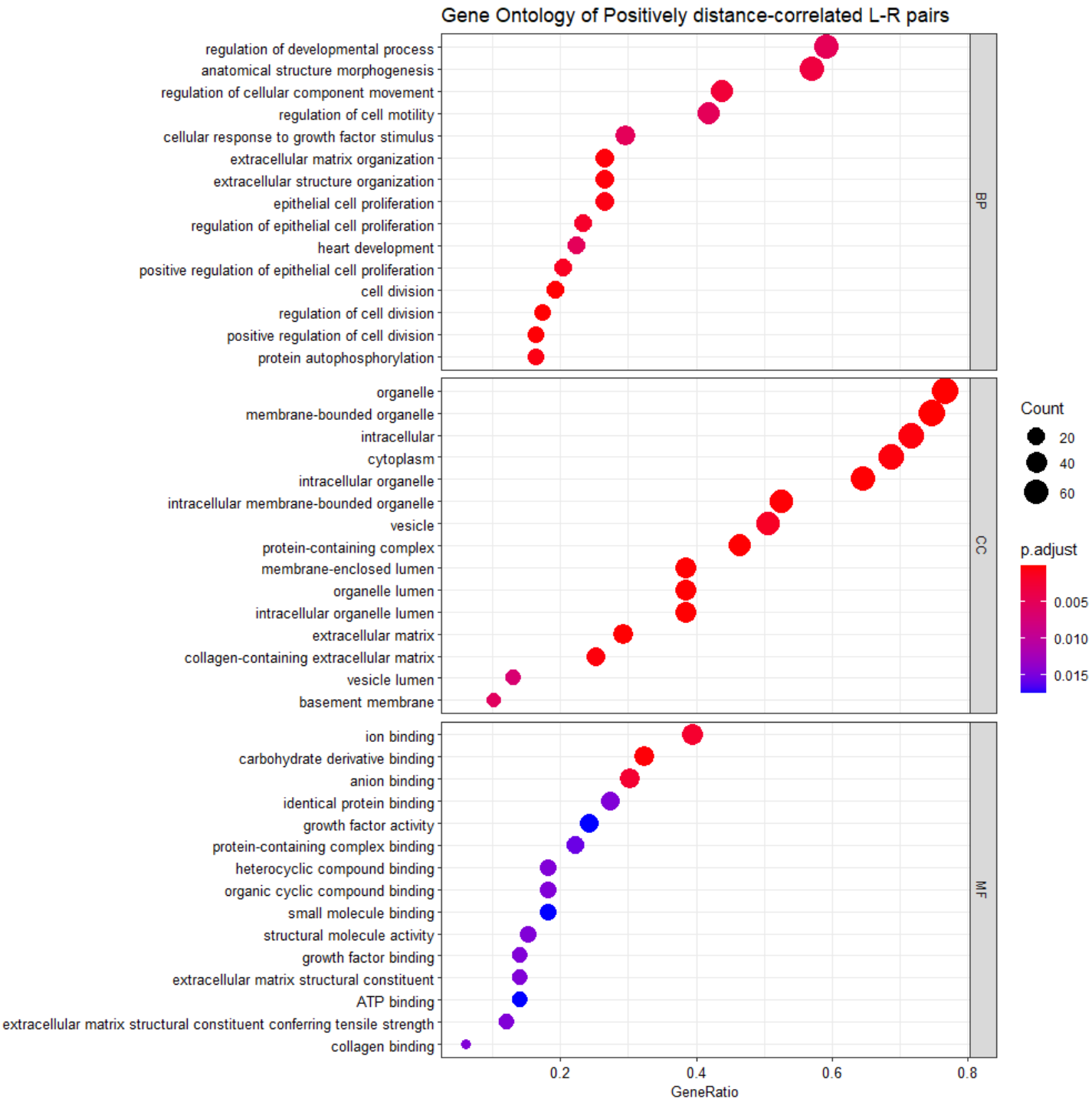


Figure 5: Dot plots of significant GO terms of positively distance-correlated L-R pairs. Most significant 15 GO terms are selected in each category. For each GO term, the size of plot represents the number of annotated genes, the color of plot represents the adjust P-value of GO term.

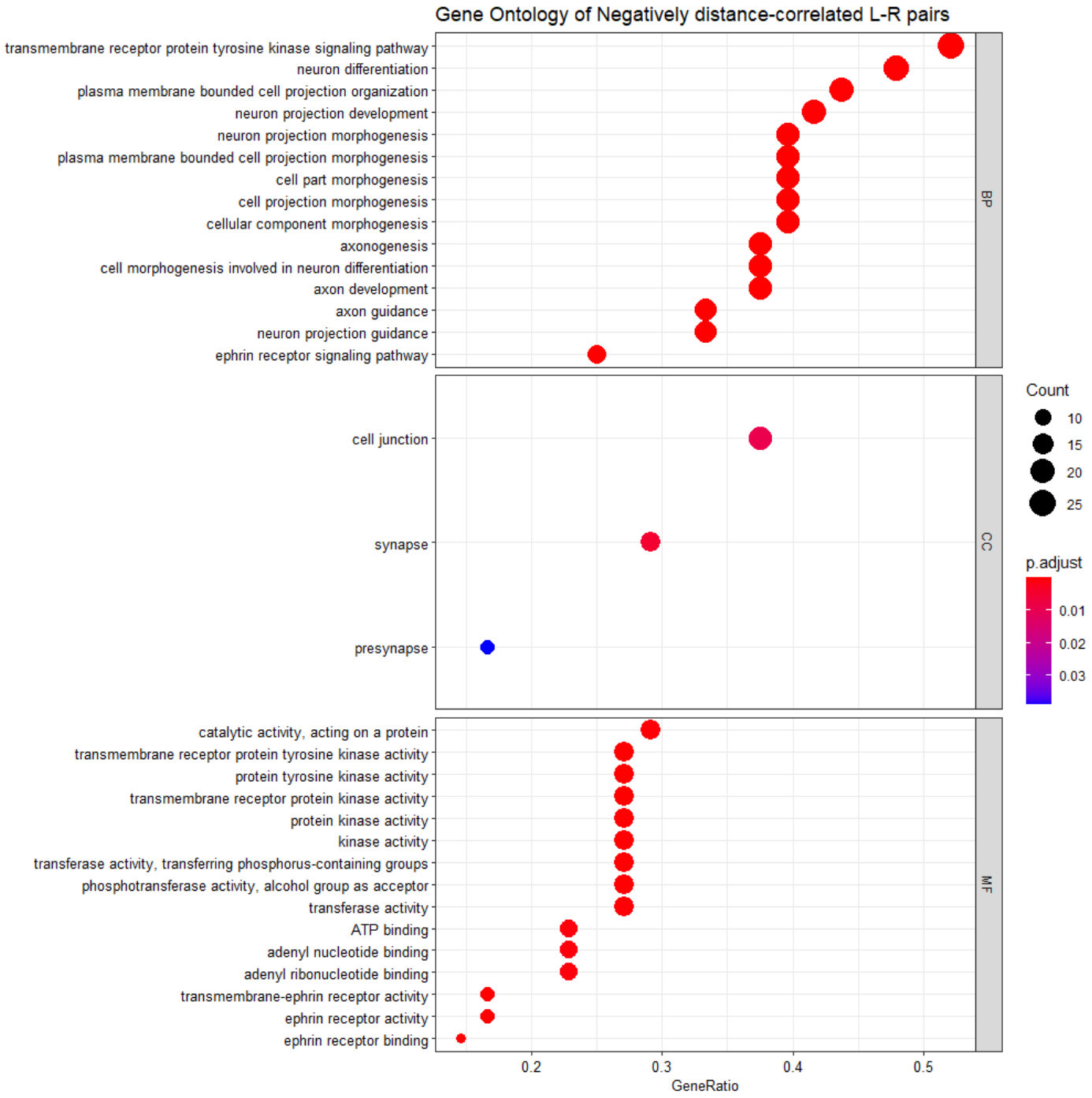


Figure 6: Dot plots of significant GO terms of negatively distance-correlated L-R pairs. Most significant 15 GO terms are selected in each category. For each GO term, the size of plot represents the number of annotated genes, the color of plot represents the adjust P-value of GO term.

Focusing on most distance-correlated ligand-receptor pairs

We also focused on specific distance-correlated L-R pairs and answering two questions: How do they function in the biological system and if they are activating between specific cell types? We selected the most positively and negatively distance-correlated L-R pairs (Table 1). Based on the permutation test of CellPhoneDB, we knew if the L-R pairs are significantly expressed between specific cell clusters. By mapping the annotations in the scRNA-seq metadata, we also derived the composition of brain cell types in every 62 clusters (Figure 7).

Table 1: Most negatively and positively distance-correlated L-R pairs (correlation for Cell2location and Novosparc are both <=-0.5 or >= 0.5)

Group	L-R pairs
Most positively distance-correlated	FLT1 complex-VEGFB LGALS9-COLEC12 PLXNB2-PTN CD70-GPRC5B ADORA2B-ENTPD1
Most negatively distance-correlated	WNT5A-FZD3 EPHA4-EFNA2 EPHA4-EFNA3 EPHB2-EFNA5 EPHA3-EFNA5 EPHA8-EFNA5

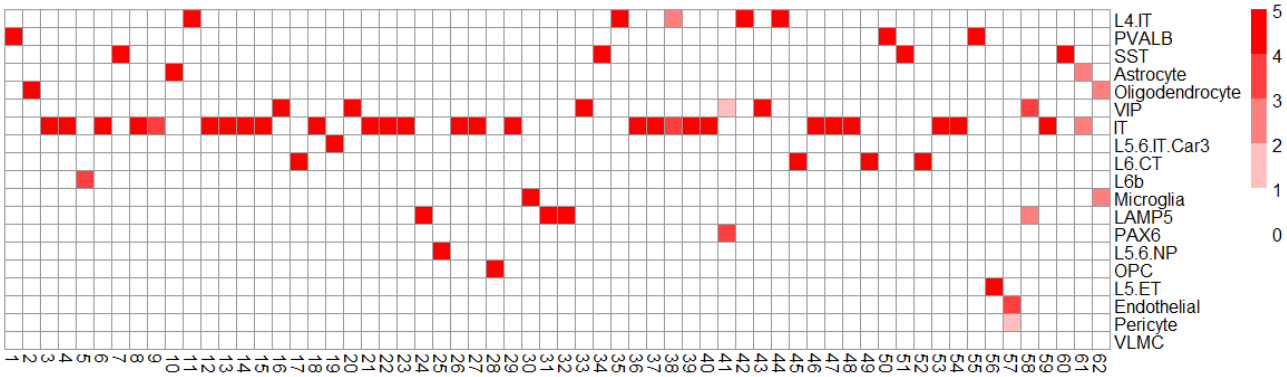


Figure 7: Compositions of brain cell types in 62 clusters. The x-axis represents the cell cluster, the y-axis represents the cell type label. Color represents the scaled cell type composition in each cluster, cell type with deeper shades represents a higher composition in the cluster.

In the positively distance-correlated groups, we selected five L-R pairs (Table 1). Within these L-R pairs, VEGFB displayed protective effects on neuronal cells and involved in angiogenesis and cell migration, which enable interacting with FLT1 (Dhondt et al., 2011), Here, FLT1 complex-VEGFB is mainly received by the endothelial cell (Figure 7, 8A). For LGALS9-COLEC12 pairs, LGALS9 coding Galectin 9 (Gal-9), Gal-9 has been reported to modulate the immune response during neuroinflammatory (Steelman & Li, 2014). LGALS9-COLEC12 is mainly sending by the endothelial and received by the microglia (Figure 7, Supplementary Figure 1A), which are potent immune effector cells in the central nervous system (Yang, Han, Kaur, Crane, & Parsa, 2010) . PTN encoded pleiotrophin is related to neuron differentiation (Henry, Sugino, Tozer, Branco, & Sternson, 2015). PLXNB2-PTN is mainly interacting with astrocyte, endothelial and OPC (Figure 7, Supplementary Figure 1B). We are not sure about the functions of LGALS9-COLEC12 and PLXNB2-PTN in neurons. In CD70-GPRC5B, GPRC5B is related to neuron differentiation and inflammatory response (Kurabayashi, Nguyen, & Sanada, 2013; Zambrano et al., 2019). Here, they are mainly interacting with Oligodendrocyte and IT cell (Figure 7, Supplementary Figure 1C), but the function of CD70-GPRC5B is unknown. We are also not sure about how the ADORA2B-ENTPD1 pair functions in the cortex. Here, they are mainly interacting with microglia and astrocyte (Figure 7, Supplementary Figure 1D).

In the negatively distance-correlated groups, we selected six L-R pairs (Table 1). Five of them are related to ephrin-Eph signaling (EPHA4-EFNA2, EPHA4-EFNA3, EPHB2-EFNA5, EPHA3-EFNA5, EPHA8-EFNA5). Eph receptors are the biggest subfamily of receptor tyrosine kinases, they are separated into EphAs and EphBs (Shiuan & Chen, 2016). Both Eph receptors and their cognate ephrin ligands are membrane-bounded, so the signaling generally depends on close contact with neighboring cells (Taylor, Campbell, & Nobes, 2017). WNT5A is a Wingless-type (Wnt) ligand, usually binds to the Frizzled Class Receptor (Fzd). WNT5A is regarded as an autocrine or paracrine ligand, its secretion is controlled by precise concentration gradients (Pourreyron et al., 2012). We noticed that L-R pairs related to Ephrin-Eph signaling are commonly sent by IT L6 CT , and L4 IT cells, and WNT5A-FZD3 is received by IT and L6CT cells (Figure 7, 9A-F). All of these ligands play an important role in the development of the nervous system, including cell migration and

morphogenesis (Cramer & Miko, 2016; Stuebner, Faus-Kessler, Fischer, Wurst, & Prakash, 2010)

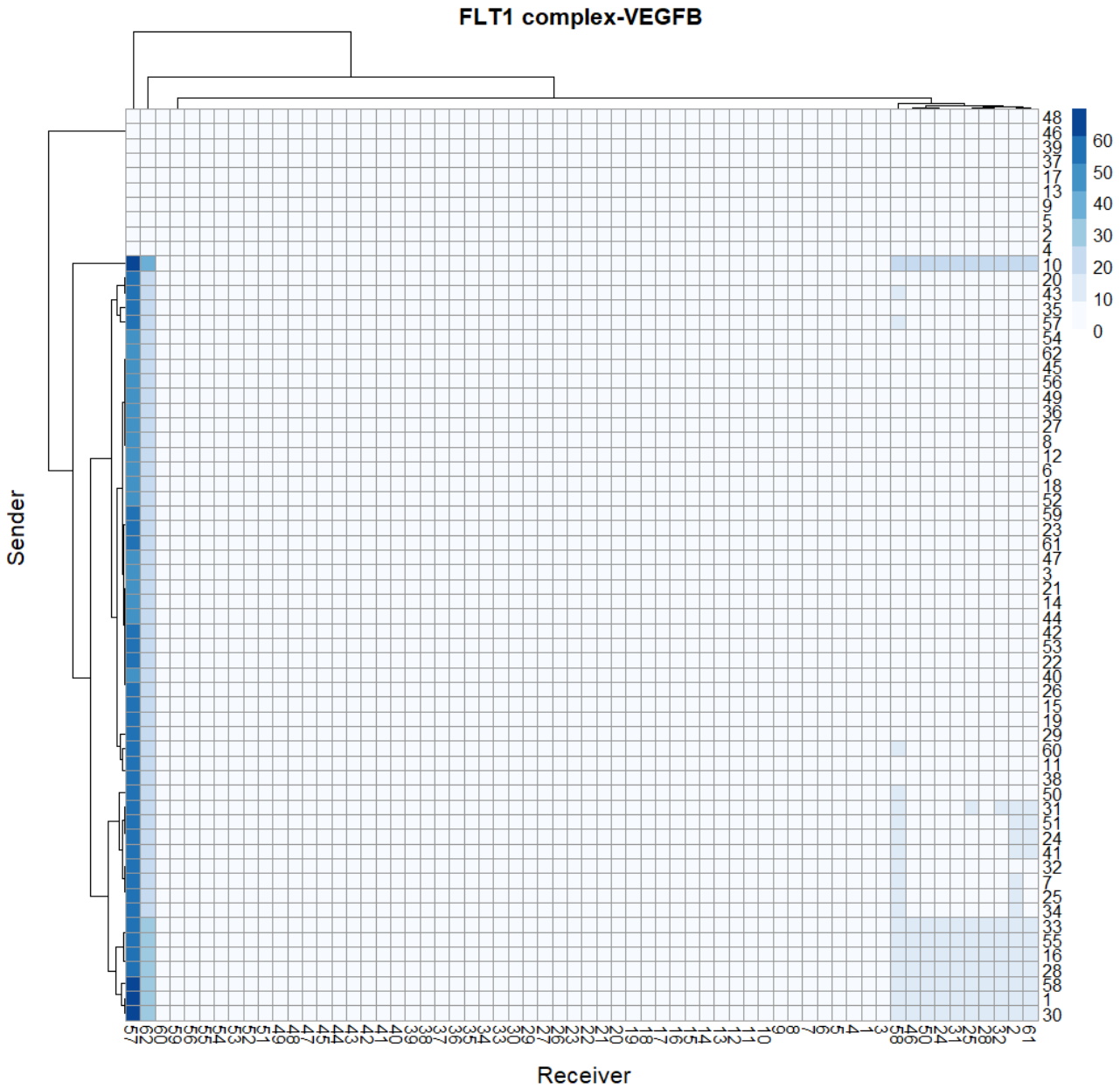


Figure 8: Heatmaps of significant expression of FLT1 complex-VEGFB. The x-axis represents sender cell clusters; the y-axis represents receiver cell clusters, blue color represents the expression value of L-R pair, deeper shades represent higher L-R expression between the corresponding clusters.

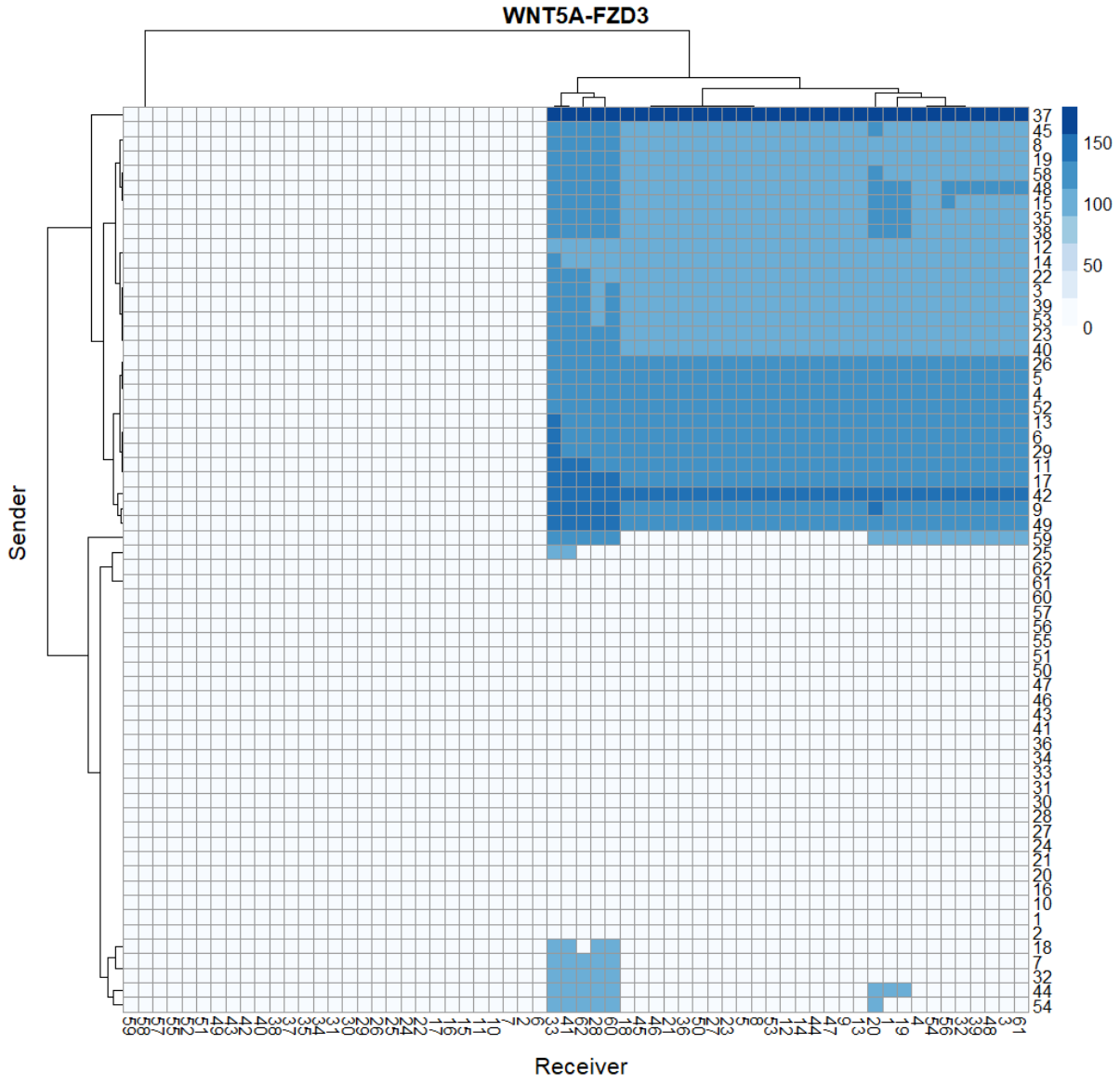


Figure 9: Heatmaps of significant expression of WNT5A-FZD3. The x-axis represents sender cell clusters; the y-axis represents receiver cell clusters, blue color represents the expression value of L-R pair, deeper shades represent higher L-R expression between the corresponding clusters.

Discussion:

This research illustrated the relationship between physical distances and CCIs by integrating scRNA-seq and spatial transcriptomic of the human cerebellum. We showed that both Cell2location and Novosparc could successfully integrate the scRNA-seq and spatial transcriptomic, retaining both the transcript profile from scRNA-seq and the spatial information. When mapping the cortex layers to the spatial transcriptomic, we observed Cell2location showed a more apparent spatial structure than Novosparc while mapping these cell clusters based on the visual inspection (Figure 1). One possible reason is that Cell2location has the prior information of cell cluster annotation from scRNA-seq and aims to map these cell clusters to the spatial space based on scRNA-seq reference signatures. In comparison, Novosparc focuses on the global alignment of the expression space and the physical space in scRNA-seq, without any prior information of cell clusters. After we over-clustered the data and compared the probability space of Cell2location and Novosparc using Jenson-Shannon divergence, we noticed that in most cases, results from Cell2location and Novosparc share a similar probability space for the same cell cluster (Figure 2). These results showed that Cell2location and Novosparc could reconstruct the general spatial information of cell clusters. The results from Cell2location and Novosparc share some similarities while mapping these cell clusters.

Then, for each L-R pair, we calculated the Spearman correlation between their expression and the physical distance. We revealed their relationship by discovering the L-R pairs whose expression is negatively or positively correlated with the physical distance (Figure 3). Also, we noticed that the correlations derived by Cell2location and Novosparc are highly correlated (Figure 3), which further demonstrated the similarities between the results from these two methods. The overall correlations are between -0.5 ~ 0.5, which is week. Possibly because we only focused on the general relationship between distances and L-R expression but ignored that some ligands and receptors are only activated in some specific cell clusters (Qiao et al., 2014), these L-R pairs might have a strong correlation with distance in specific cell clusters but submerged by the overall data here.

Next step, we explored the biological functions of significant distance-correlated L-R pairs by selecting the ligand and receptor genes within the two groups to conduct the GO enrichment analysis. For the positively distance related L-R pairs, the GO

terms of the biological process related to cell development, morphogenesis, cell proliferation, and division. As for the negatively distance-correlated L-R group, the GO terms of the biological process are mainly focused on neuron morphogenesis and axon development. The GO terms in the cellular component category showed the positively distance-correlated L-R pairs are significant in the cytoplasm, vesicle, and lumen, while the negatively distance-correlated L-R pairs are significant in cell junction, synapse, and pre-synapse. For the positively distance-correlated L-R pairs, the GO terms of vesicle and lumen indicated these L-R pairs might involve in the secretion of the signaling molecules (Schwarz & Blower, 2016). Within the gene list in the positively distance-correlated L-R pairs, we discovered ligands in Fibroblast growth factors (FGF) family and Vascular Endothelial Growth Factor A (VEGF-A) ligand. These are all secretory ligands and associate with the endocrine signaling (Cao, 2014; Itoh, Ohta, & Konishi, 2015). On the other hand, for the negatively distance-correlated L-R pairs, the GO terms of cell-cell junction, synapse and pre-synapse indicates they are more likely involving in contact-based signaling or synapse signaling events. Within the gene list in the negatively distance-correlated L-R pairs, we discovered many Ephrin ligands and Erythropoietin-producing hepatocellular carcinoma (Eph) receptors, they are involving in the contact-base Ephrin-Eph signaling (Mosch, Reissenweber, Neuber, & Pietzsch, 2010). These results implied that the negatively distance-correlated L-R pairs are more involved in short-distance interactions while the positively distance-correlated L-R pairs are more involved in long-distance interactions.

Further, we analyzed the most distance-correlated L-R pairs in order to check if they fit the results from the GO enrichment analysis. For the positively distance-correlated L-R pairs, we selected five L-R pairs (Table 1). They are related to cell migration, neuron protection, neuron differentiation, and immune response (Dhondt et al., 2011; Henry et al., 2015; Zambrano et al., 2019). In the negatively correlated group, we selected six L-R pairs (Table 1), which related to neuron differentiation and morphogenesis (Cramer & Miko, 2016; Stuebner et al., 2010). The biological functions of these selected L-R pairs are consistent with the GO enrichment analysis.

We also explored what signaling events are these most distance-correlated L-R pairs involving. In the negatively correlated group, five of them belong to the contact-based Ephrin-Eph signaling. Both of the ephrin ligands and Eph receptors are

membrane-bound proteins and they are mostly interacting in a short distance (Kania & Klein, 2016; Mosch, Reissenweber, Neuber, & Pietzsch, 2010). Besides, WNT5A is regarded as the paracrine signaling factor (Borcherding et al., 2015), also meaning this L-R pair interaction happened in short intercellular distance. In the positively correlated group, the ligand genes are mainly coding proteins like cell surface enzymes or growth factors and most of them are reported involving autocrine or paracrine signaling (de Jong, Gabius, & Baron, 2020; Qin et al., 2017; Zafar et al., 2017). It suggested that the negatively distance-correlated L-Rs are interacting in shorter distances than the positively distance-correlated L-Rs.

Finally, we identified these distance-correlated L-R pairs are significantly expressed within specific cell clusters based on their function. Like LGALS9-COLEC12 related to immune response and they are significantly interact with microglia, which are potent immune effector cells in the central nervous system (Yang et al., 2010), and FLT1 complex-VEGFB is significantly interact with endothelial cell, which is consistent with VEGFB controls endothelial cell uptakee of fatty acids (Hagberg et al., 2010). These results indicate L-R pairs would focus on specific cell clusters based on their function. Future studies should take into account the connection between spatial information and these cell type specific CCIs.

Here, we demonstrated CCIs containing spatial information by discovering the correlation between L-R pair expression and physical distance. These distance-correlated L-R pairs are further proved to be related to short-distance or long-distance CCIs. However, there's still room for further improvement to strengthen our conclusion. More research is needed into the relationship between cell type specific CCIs and physical distance. Furthermore, we would expect our conclusions to be verified in different datasets corresponding to different tissues or different spatial transcriptomic techniques. Besides, we assumed a more comprehensive CCI database like Omnipath can help us identify our conclusion with more CCIs (Turei et al., 2021)

We believe CCIs can be further applied to reconstruct the spatial architecture in multicellular organisms (Wu, Liu, Zhang, Li, & Wang, 2021). Novel pipelines like CSOmap can combine the scRNA-seq and a database of ligand-receptor database in order to reconstruct the spatial location of individual cell *de novo* (Ren et al., 2020). CSOmap assumes ligand-receptor interactions mediated cellular self-assembly, and the spatial location of cells depends on the effect of ligand-receptor affinity.

Compared to another de novo spatial reconstruction pipeline Novosparc, CSOmap applies extra ligand-receptor interactions. That might fill in the blank that ligand-receptor interactions could play an essential role in arranging spatial location in some physiological processes, like cell morphogenesis, while Novosparc's assumption cannot indicate the effect of ligand-receptor interaction (Ren et al., 2020). On the other hand, pipelines like SVCA tried to quantify how CCIs contribute to the spatial variance of gene expression based on the spatial molecular data, and SVCA revealed CCIs is the major driver of protein expression heterogeneity in their breast cancer dataset (Arnol, Schapiro, Bodenmiller, Saez-Rodriguez, & Stegle, 2019). These examples show the detailed interpretation of CCIs is needed to reveal the heterogeneity of spatial structure in multicellular organisms.

Reference:

Abdelaal, T., Mourragui, S., Mahfouz, A., & Reinders, M. J. T. (2020). SpaGE: Spatial Gene Enhancement using scRNA-seq. *Nucleic Acids Res, 48*(18), e107. doi:10.1093/nar/gkaa740

Akins, M. R., & Biederer, T. (2006). Cell-cell interactions in synaptogenesis. *Curr Opin Neurobiol, 16*(1), 83-89. doi:10.1016/j.conb.2006.01.009

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

Armingol, E., Joshi, C. J., Baghdassarian, H., Shamie, I., Ghaddar, A., Chan, J., . . . Lewis, N. E. (2020). Inferring the spatial code of cell-cell interactions and communication across a whole animal body. 2020.2011.2022.392217. doi:10.1101/2020.11.22.392217 %J bioRxiv

Armingol, E., Officer, A., Harismendy, O., & Lewis, N. E. (2021). Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet, 22*(2), 71-88. doi:10.1038/s41576-020-00292-x

Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J., & Stegle, O. (2019). Modeling Cell-Cell Interactions from Spatial Molecular Data with Spatial Variance Component Analysis. *Cell Rep, 29*(1), 202-211 e206. doi:10.1016/j.celrep.2019.08.077

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet, 25*(1), 25-29. doi:10.1038/75556

Batista, F. D., & Dustin, M. L. (2013). Cell:cell interactions in the immune system. *Immunol Rev, 251*(1), 7-12. doi:10.1111/imr.12025

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *57*(1), 289-300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Borcherding, N., Kusner, D., Kolb, R., Xie, Q., Li, W., Yuan, F., . . . Zhang, W. (2015). Paracrine WNT5A Signaling Inhibits Expansion of Tumor-Initiating Cells. *Cancer Res, 75*(10), 1972-1982. doi:10.1158/0008-5472.CAN-14-2761

Brooke, M. A., Nitoiu, D., & Kelsell, D. P. (2012). Cell-cell connectivity: desmosomes and disease. *J Pathol, 226*(2), 158-171. doi:10.1002/path.3027

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., . . . Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol, 33*(2), 155-160. doi:10.1038/nbt.3102

Cao, Y. (2014). VEGF-targeted cancer therapeutics-paradoxical effects in endocrine organs. *Nat Rev Endocrinol, 10*(9), 530-539. doi:10.1038/nrendo.2014.114

Chen, S. S., Fitzgerald, W., Zimmerberg, J., Kleinman, H. K., & Margolis, L. (2007). Cell-cell and cell-extracellular matrix interactions regulate embryonic stem cell differentiation. *Stem Cells, 25*(3), 553-561. doi:10.1634/stemcells.2006-0419

Chrousos, G. P. (2007). Organization and Integration of the Endocrine System. *Sleep Med Clin, 2*(2), 125-145. doi:10.1016/j.jsmc.2007.04.004

Cramer, K. S., & Miko, I. J. (2016). Eph-ephrin signaling in nervous system development. *F1000Res, 5*. doi:10.12688/f1000research.7417.1

Cuartero, M. I., Ballesteros, I., Lizasoain, I., & Moro, M. A. (2015). Complexity of the cell-cell interactions in the innate immune response after cerebral ischemia. *Brain Res, 1623*, 53-62. doi:10.1016/j.brainres.2015.04.047

de Jong, C., Gabius, H. J., & Baron, W. (2020). The emerging role of galectins in (re)myelination and its potential for developing new approaches to treat multiple sclerosis. *Cell Mol Life Sci, 77*(7), 1289-1317. doi:10.1007/s00018-019-03327-7

de Vries, N. L., Mahfouz, A., Koning, F., & de Miranda, N. (2020). Unraveling the Complexity of the Cancer Microenvironment With Multidimensional Genomic and

Cytometric Technologies. *Front Oncol, 10*, 1254. doi:10.3389/fonc.2020.01254
Dhondt, J., Peeraer, E., Verheyen, A., Nuydens, R., Buysschaert, I., Poesen, K., . . . Lambrechts, D. (2011). Neuronal FLT1 receptor and its selective ligand VEGF-B protect against retrograde degeneration of sensory neurons. *FASEB J, 25*(5), 1461-1473. doi:10.1096/fj.10-170944
Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc, 15*(4), 1484-1506. doi:10.1038/s41596-020-0292-x
Graves, D. T., Valentin-Opran, A., Delgado, R., Valente, A. J., Mundy, G., & Piche, J. (1989). The potential role of platelet-derived growth factor as an autocrine or paracrine factor for human bone cells. *Connect Tissue Res, 23*(2-3), 209-218. doi:10.3109/03008208909002419
Grosse, I., Bernaola-Galvan, P., Carpena, P., Roman-Roldan, R., Oliver, J., & Stanley, H. E. (2002). Analysis of symbolic sequences using the Jensen-Shannon divergence. *Phys Rev E Stat Nonlin Soft Matter Phys, 65*(4 Pt 1), 041905. doi:10.1103/PhysRevE.65.041905
Guryanov, I., Fiorucci, S., & Tennikova, T. (2016). Receptor-ligand interactions: Advanced biomedical applications. *Mater Sci Eng C Mater Biol Appl, 68*, 890-903. doi:10.1016/j.msec.2016.07.072
Hagberg, C. E., Falkevall, A., Wang, X., Larsson, E., Huusko, J., Nilsson, I., . . . Eriksson, U. (2010). Vascular endothelial growth factor B controls endothelial fatty acid uptake. *Nature, 464*(7290), 917-921. doi:10.1038/nature08945
Henry, F. E., Sugino, K., Tozer, A., Branco, T., & Sternson, S. M. (2015). Cell type-specific transcriptomics of hypothalamic energy-sensing neuron responses to weight-loss. *Elife, 4*. doi:10.7554/eLife.09800
Hodge, R. D., Bakken, T. E., Miller, J. A., Smith, K. A., Barkan, E. R., Graybuck, L. T., . . . Lein, E. S. (2019). Conserved cell types with divergent features in human versus mouse cortex. *Nature, 573*(7772), 61-68. doi:10.1038/s41586-019-1506-7
Itoh, N., Ohta, H., & Konishi, M. (2015). Endocrine FGFs: Evolution, Physiology, Pathophysiology, and Pharmacotherapy. *Front Endocrinol (Lausanne), 6*, 154. doi:10.3389/fendo.2015.00154
Kania, A., & Klein, R. (2016). Mechanisms of ephrin-Eph signalling in development, physiology and disease. *Nat Rev Mol Cell Biol, 17*(4), 240-256. doi:10.1038/nrm.2015.16
Kirouac, D. C., Madlambayan, G. J., Yu, M., Sykes, E. A., Ito, C., & Zandstra, P. W. (2009). Cell-cell interaction networks regulate blood stem and progenitor cell fate. *Mol Syst Biol, 5*, 293. doi:10.1038/msb.2009.49
Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H. W., Li, T., . . . Bayraktar, O. A. (2020). Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. 2020.2011.2015.378125. doi:10.1101/2020.11.15.378125 %J bioRxiv
Kumar, M. P., Du, J., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D. C., . . . Raue, A. (2018). Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep, 25*(6), 1458-1468 e1454. doi:10.1016/j.celrep.2018.10.047
Kurabayashi, N., Nguyen, M. D., & Sanada, K. (2013). The G protein-coupled receptor GPRC5B contributes to neurogenesis in the developing mouse neocortex. *Development, 140*(21), 4335-4346. doi:10.1242/dev.099754
Lindoso, R. S., Sandim, V., Collino, F., Carvalho, A. B., Dias, J., da Costa, M. R., . . . Vieyra, A. (2016). Proteomics of cell-cell interactions in health and disease. *Proteomics, 16*(2), 328-344. doi:10.1002/pmic.201500341
Misteli, T. (2001). The concept of self-organization in cellular architecture. *J Cell Biol, 155*(2), 181-185. doi:10.1083/jcb.200108110
Monge, G. (1781). Memoire sur la theorie des deblais et des remblais. *Histoire de l'Academie*

Royale des Sciences de Paris.
Mosch, B., Reissenweber, B., Neuber, C., & Pietzsch, J. (2010). Eph receptors and ephrin ligands: important players in angiogenesis and tumor angiogenesis. *J Oncol, 2010*, 135285. doi:10.1155/2010/135285
Nitzan, M., Karaiskos, N., Friedman, N., & Rajewsky, N. (2019). Gene expression cartography. *Nature, 576*(7785), 132-137. doi:10.1038/s41586-019-1773-3
Nussinov, R. (2013). The spatial structure of cell signaling systems. *Phys Biol, 10*(4), 045004. doi:10.1088/1478-3975/10/4/045004
Pourreyron, C., Reilly, L., Proby, C., Panteleyev, A., Fleming, C., McLean, K., . . . Foerster, J. (2012). Wnt5a is strongly expressed at the leading edge in non-melanoma skin cancer, forming active gradients, while canonical Wnt signalling is repressed. *PLoS One, 7*(2), e31827. doi:10.1371/journal.pone.0031827
Purvis, J. E., & Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. *Cell, 152*(5), 945-956. doi:10.1016/j.cell.2013.02.005
Qiao, W., Wang, W., Laurenti, E., Turinsky, A. L., Wodak, S. J., Bader, G. D., . . . Zandstra, P. W. (2014). Intercellular network structure and regulatory motifs in the human hematopoietic system. *Mol Syst Biol, 10*, 741. doi:10.15252/msb.20145141
Qin, E. Y., Cooper, D. D., Abbott, K. L., Lennon, J., Nagaraja, S., Mackay, A., . . . Monje, M. (2017). Neural Precursor-Derived Pleiotrophin Mediates Subventricular Zone Invasion by Glioma. *Cell, 170*(5), 845-859 e819. doi:10.1016/j.cell.2017.07.016
Ramilowski, J. A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V. P., . . . Forrest, A. R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun, 6*, 7866. doi:10.1038/ncomms8866
Ren, X., Zhong, G., Zhang, Q., Zhang, L., Sun, Y., & Zhang, Z. (2020). Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res, 30*(9), 763-778. doi:10.1038/s41422-020-0353-2
Ruch, R. J. (2002). Intercellular communication, homeostasis, and toxicology. *Toxicol Sci, 68*(2), 265-266. doi:10.1093/toxsci/68.2.265
Schwarz, D. S., & Blower, M. D. (2016). The endoplasmic reticulum: structure, function and response to cellular signaling. *Cell Mol Life Sci, 73*(1), 79-94. doi:10.1007/s00018-015-2052-6
Shao, X., Lu, X., Liao, J., Chen, H., & Fan, X. (2020). New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein Cell, 11*(12), 866-880. doi:10.1007/s13238-020-00727-5
Sheikh, B. N., Bondareva, O., Guhathakurta, S., Tsang, T. H., Sikora, K., Aizarani, N., . . . Akhtar, A. (2019). Systematic Identification of Cell-Cell Communication Networks in the Developing Brain. *iScience, 21*, 273-287. doi:10.1016/j.isci.2019.10.026
Shiuan, E., & Chen, J. (2016). Eph Receptor Tyrosine Kinases in Tumor Immunity. *Cancer Res, 76*(22), 6452-6457. doi:10.1158/0008-5472.CAN-16-1521
Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., . . . Frisen, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science, 353*(6294), 78-82. doi:10.1126/science.aaf2403
Steelman, A. J., & Li, J. (2014). Astrocyte galectin-9 potentiates microglial TNF secretion. *J Neuroinflammation, 11*, 144. doi:10.1186/s12974-014-0144-0
Stuebner, S., Faus-Kessler, T., Fischer, T., Wurst, W., & Prakash, N. (2010). Fzd3 and Fzd6 deficiency results in a severe midbrain morphogenesis defect. *Dev Dyn, 239*(1), 246-260. doi:10.1002/dvdy.22127
Taylor, H., Campbell, J., & Nobes, C. D. (2017). Ephs and ephrins. *Curr Biol, 27*(3), R90-R95. doi:10.1016/j.cub.2017.01.003
Toda, S., Blauch, L. R., Tang, S. K. Y., Morsut, L., & Lim, W. A. (2018). Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science, 361*(6398), 156-162. doi:10.1126/science.aat0271
Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep, 9*(1), 5233. doi:10.1038/s41598-019-41695-z

Waylen, L. N., Nim, H. T., Martelotto, L. G., & Ramialison, M. (2020). From whole-mount to single-cell spatial assessment of gene expression in 3D. *Commun Biol, 3*(1), 602. doi:10.1038/s42003-020-01341-1

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol, 19*(1), 15. doi:10.1186/s13059-017-1382-0

Wu, D., Liu, X., Zhang, J., Li, L., & Wang, X. (2021). Significance of single-cell and spatial transcriptomes in cell biology and toxicology. *Cell Biol Toxicol, 37*(1), 1-5. doi:10.1007/s10565-020-09576-8

Yang, I., Han, S. J., Kaur, G., Crane, C., & Parsa, A. T. (2010). The role of microglia in central nervous system immunity and glioma immunology. *J Clin Neurosci, 17*(1), 6-10. doi:10.1016/j.jocn.2009.05.006

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS, 16*(5), 284-287. doi:10.1089/omi.2011.0118

Zafar, M. I., Zheng, J., Kong, W., Ye, X., Gou, L., Regmi, A., & Chen, L. L. (2017). The role of vascular endothelial growth factor-B in metabolic homoeostasis: current evidence. *Biosci Rep, 37*(4). doi:10.1042/BSR20171089

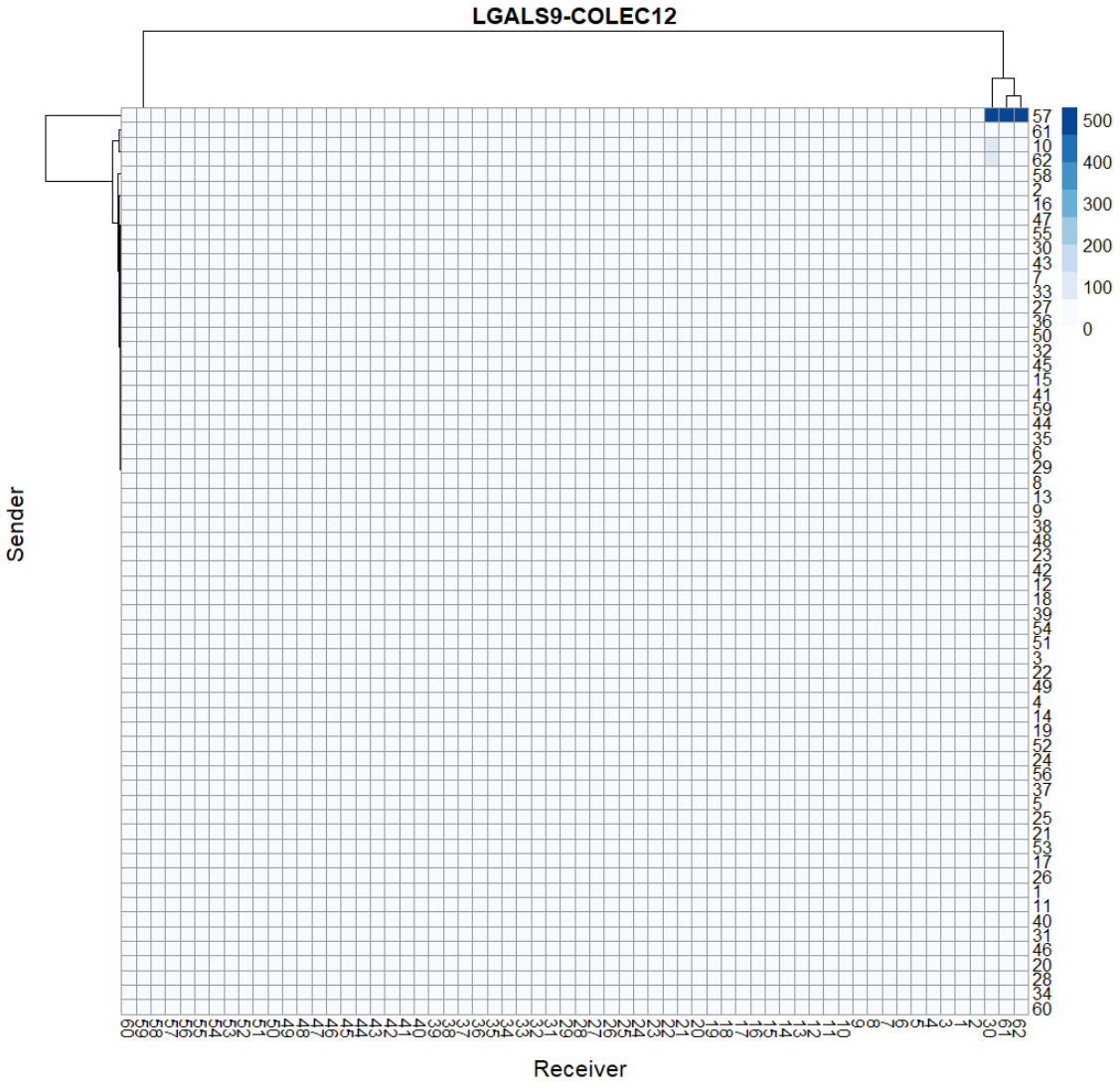
Zambrano, S., Moller-Hackbarth, K., Li, X., Rodriguez, P. Q., Charrin, E., Schwarz, A., . . . Patrakka, J. (2019). GPRC5b Modulates Inflammatory Response in Glomerular Diseases via NF-kappaB Pathway. *J Am Soc Nephrol, 30*(9), 1573-1586. doi:10.1681/ASN.2019010089

Zhou, Y., Jia, E., Pan, M., Zhao, X., & Ge, Q. (2020). Encoding Method of Single-cell Spatial Transcriptomics Sequencing. *Int J Biol Sci, 16*(14), 2663-2674. doi:10.7150/ijbs.43887

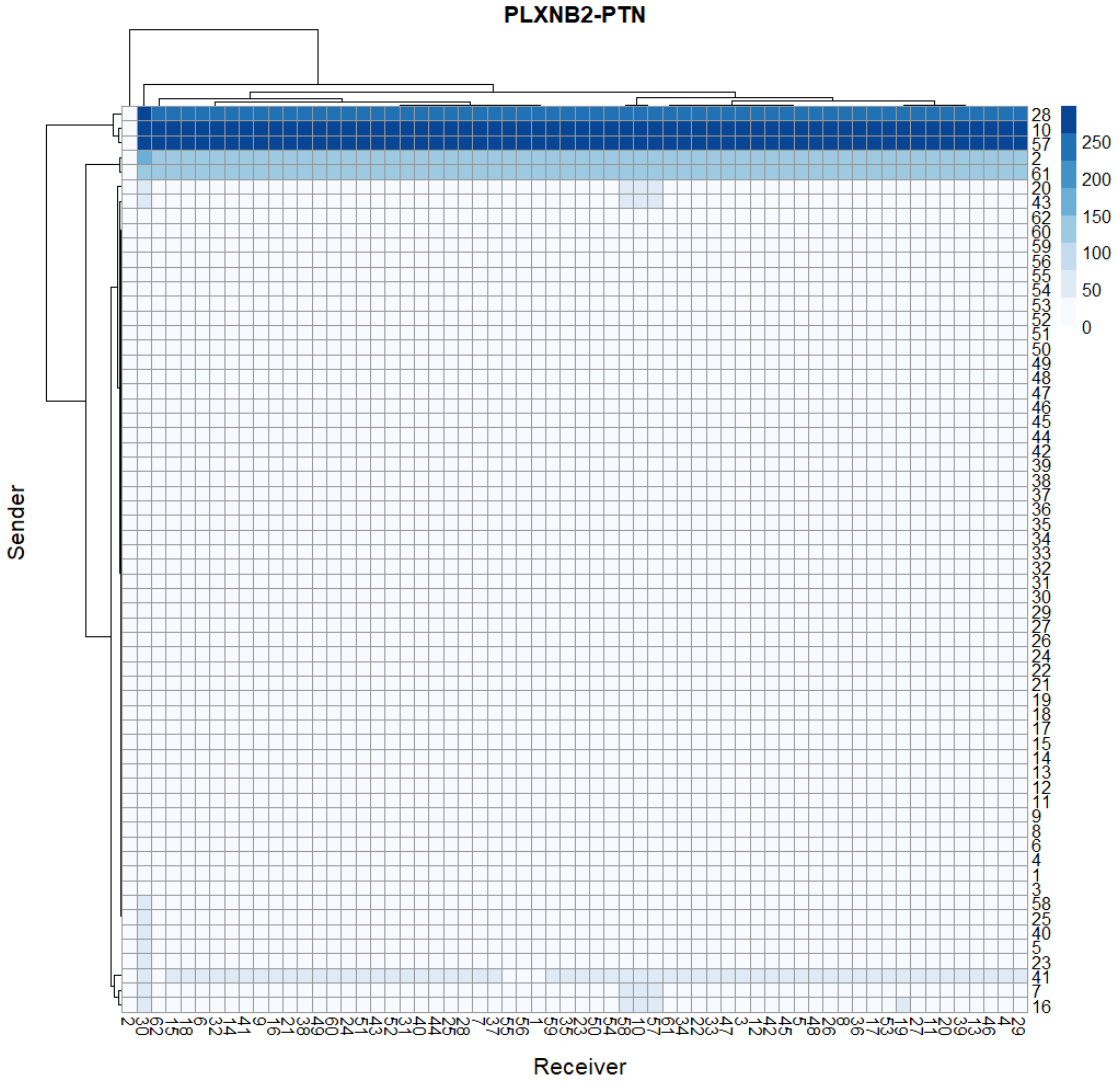
Zollinger, A. J., Xu, H., Figueiredo, J., Paredes, J., Seruca, R., Stamenovic, D., & Smith, M. L. (2018). Dependence of Tensional Homeostasis on Cell Type and on Cell-Cell Interactions. *Cell Mol Bioeng, 11*(3), 175-184. doi:10.1007/s12195-018-0527-x

Supplementary Materials:

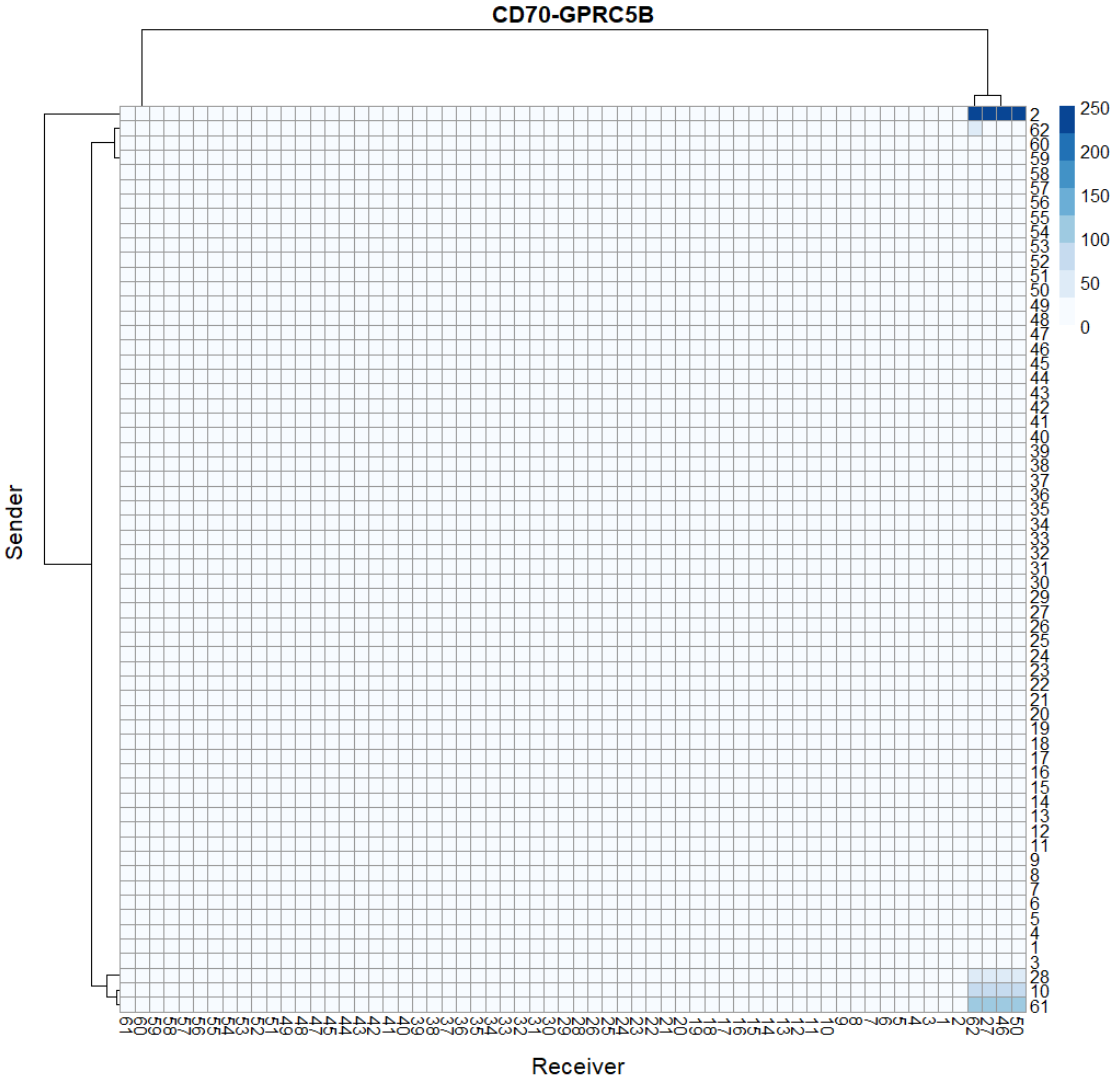
A.



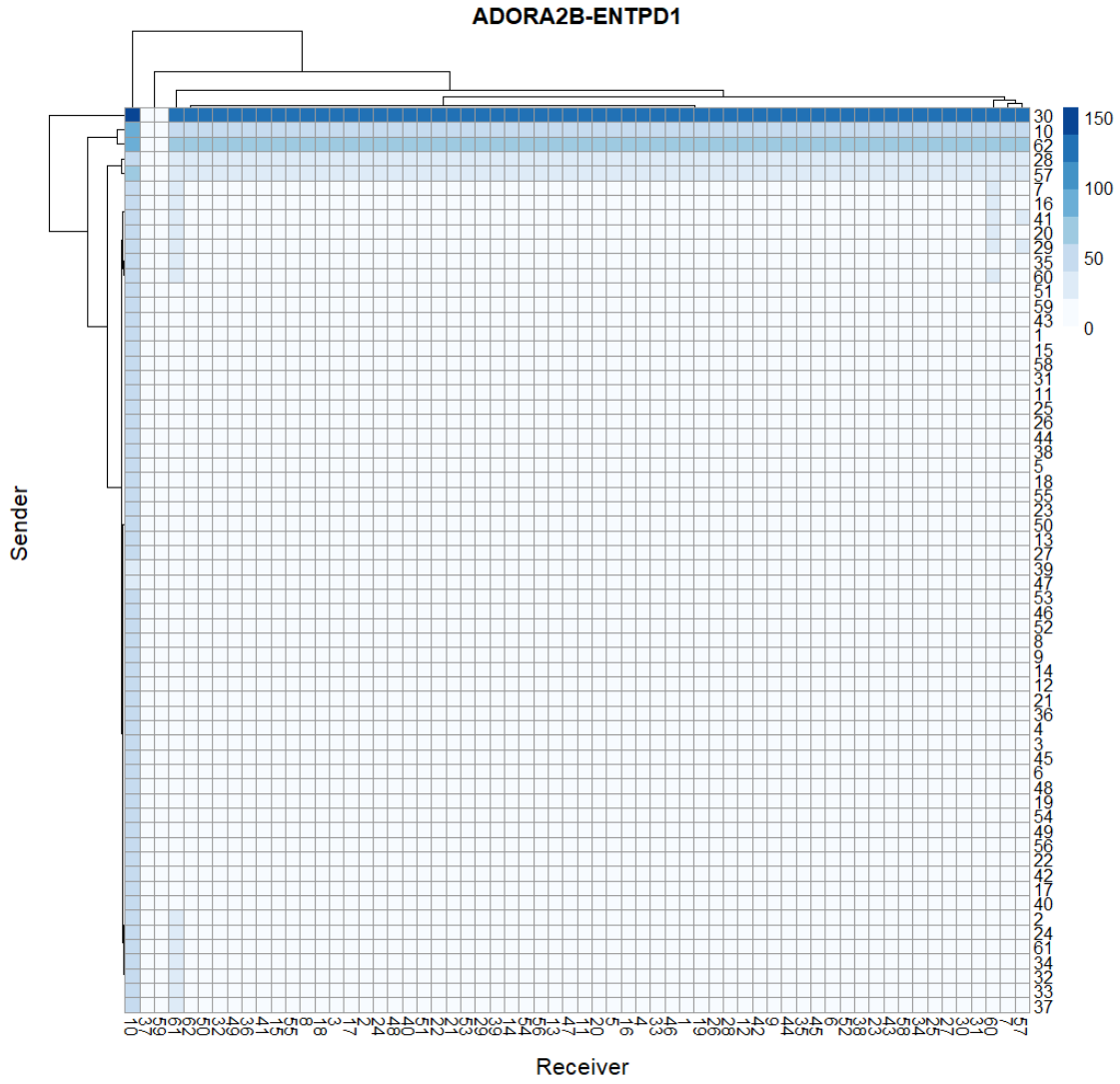
B.



C.

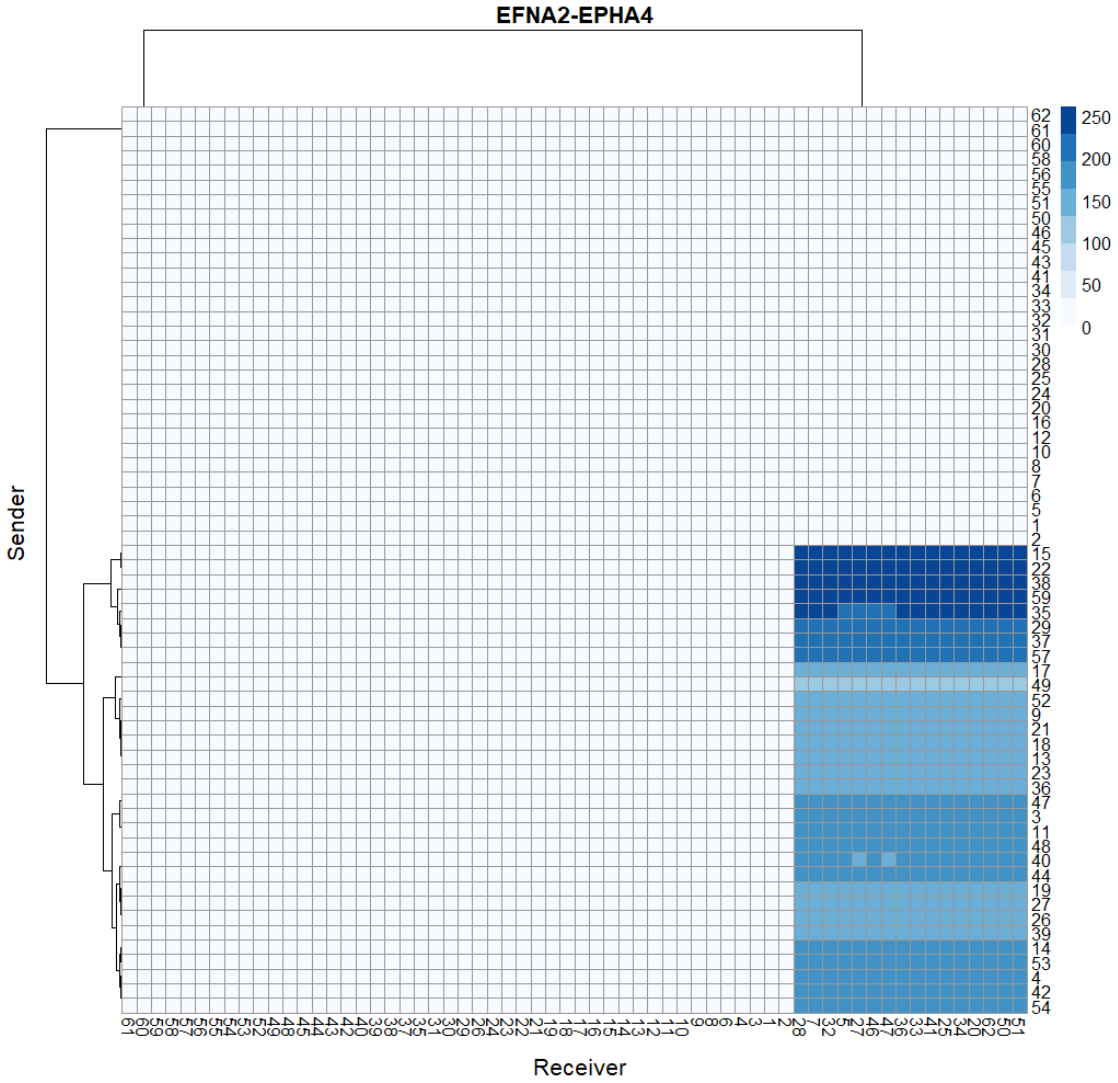


D.

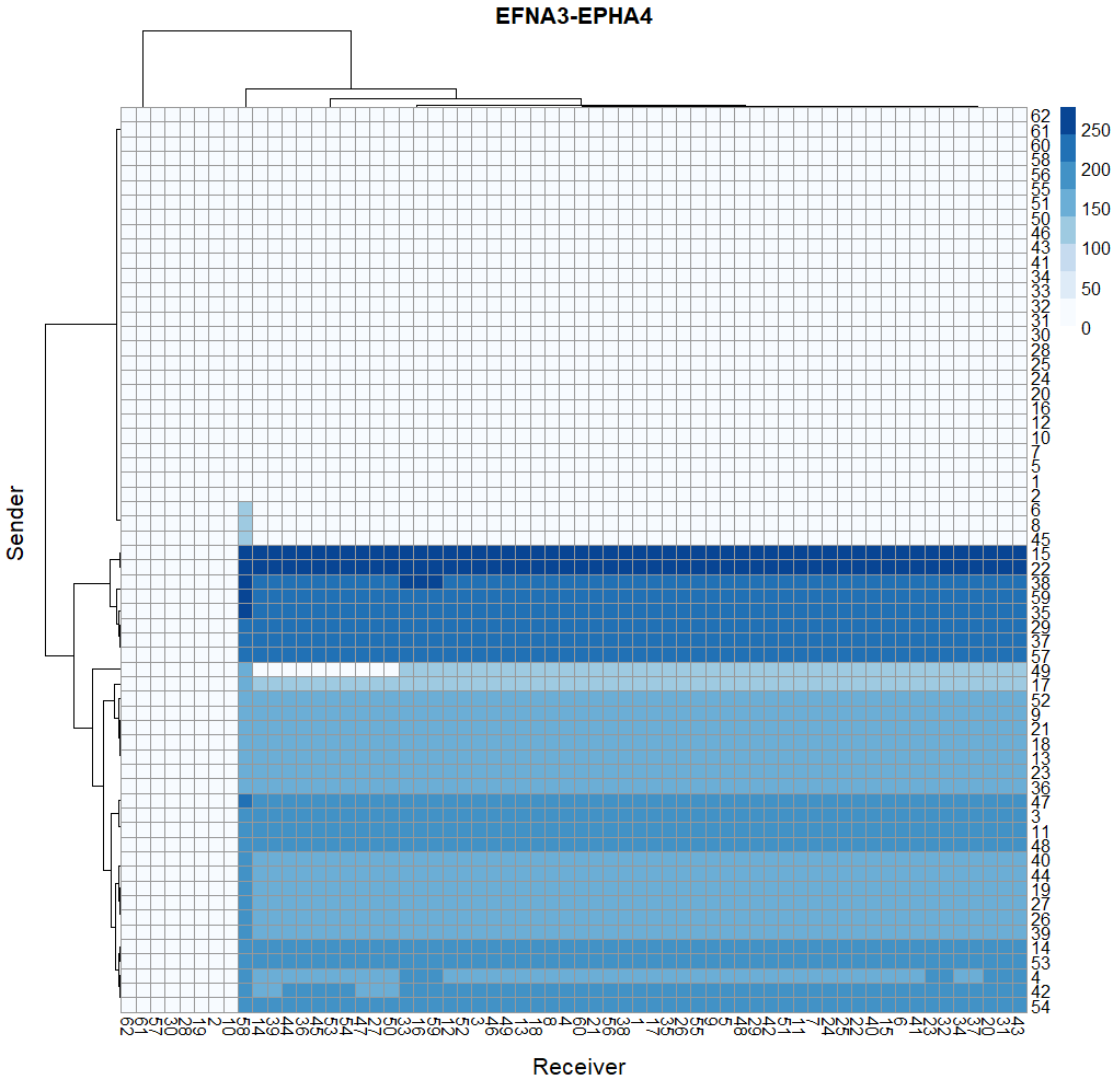


Supplementary Figure 1: Heatmaps of significant expression of positively distance-correlated **(A)** LGALS9-COLEC12 **(B)** PLXNB2-PTN **(C)** CD70-GPRC5B **(D)** ADORA2B-ENTPD1. The x-axis represents sender cell clusters; the y-axis represents receiver cell clusters, blue color represents the expression value of L-R pair, deeper shades represent higher L-R expression between the corresponding clusters.

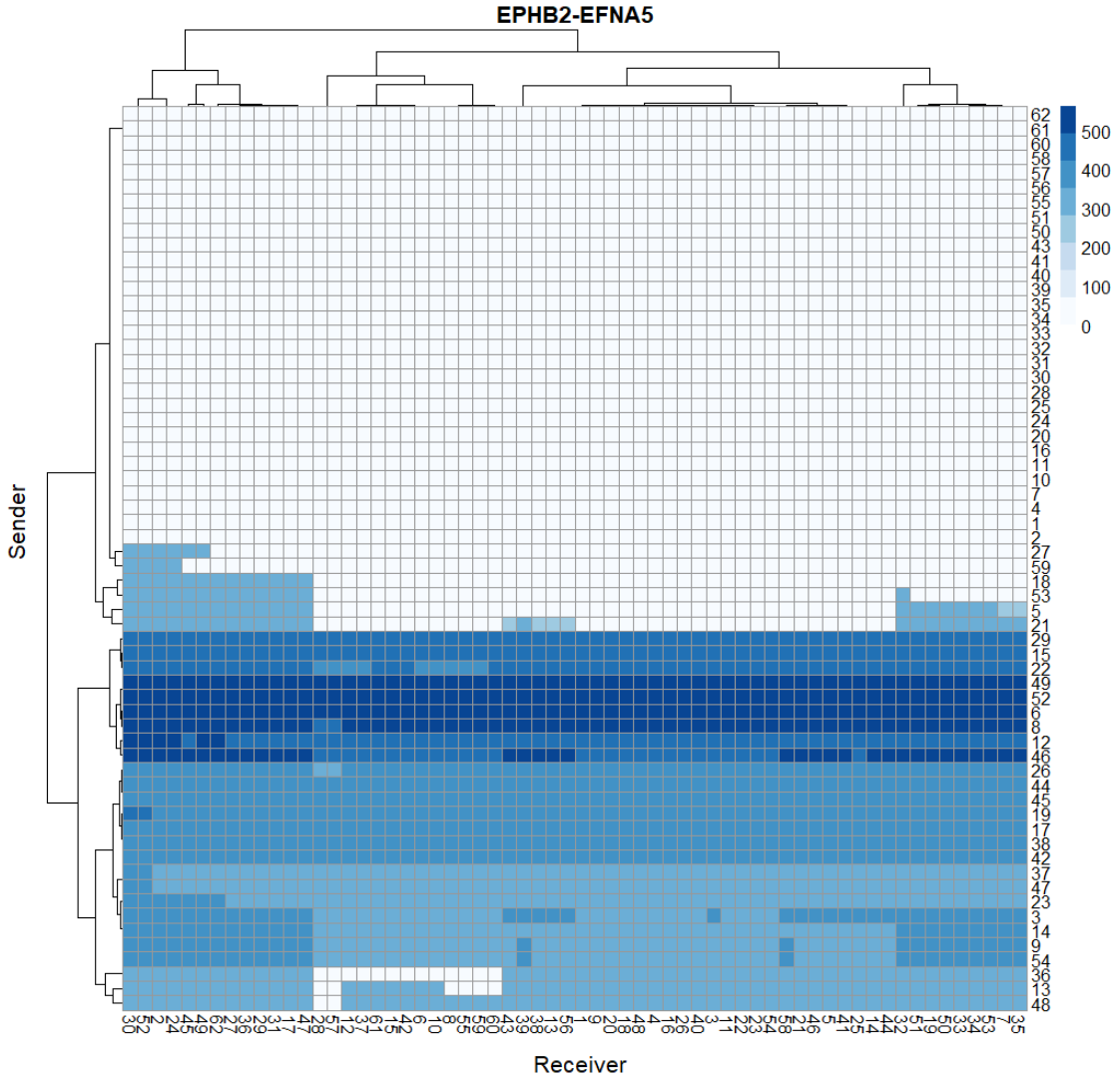
A.



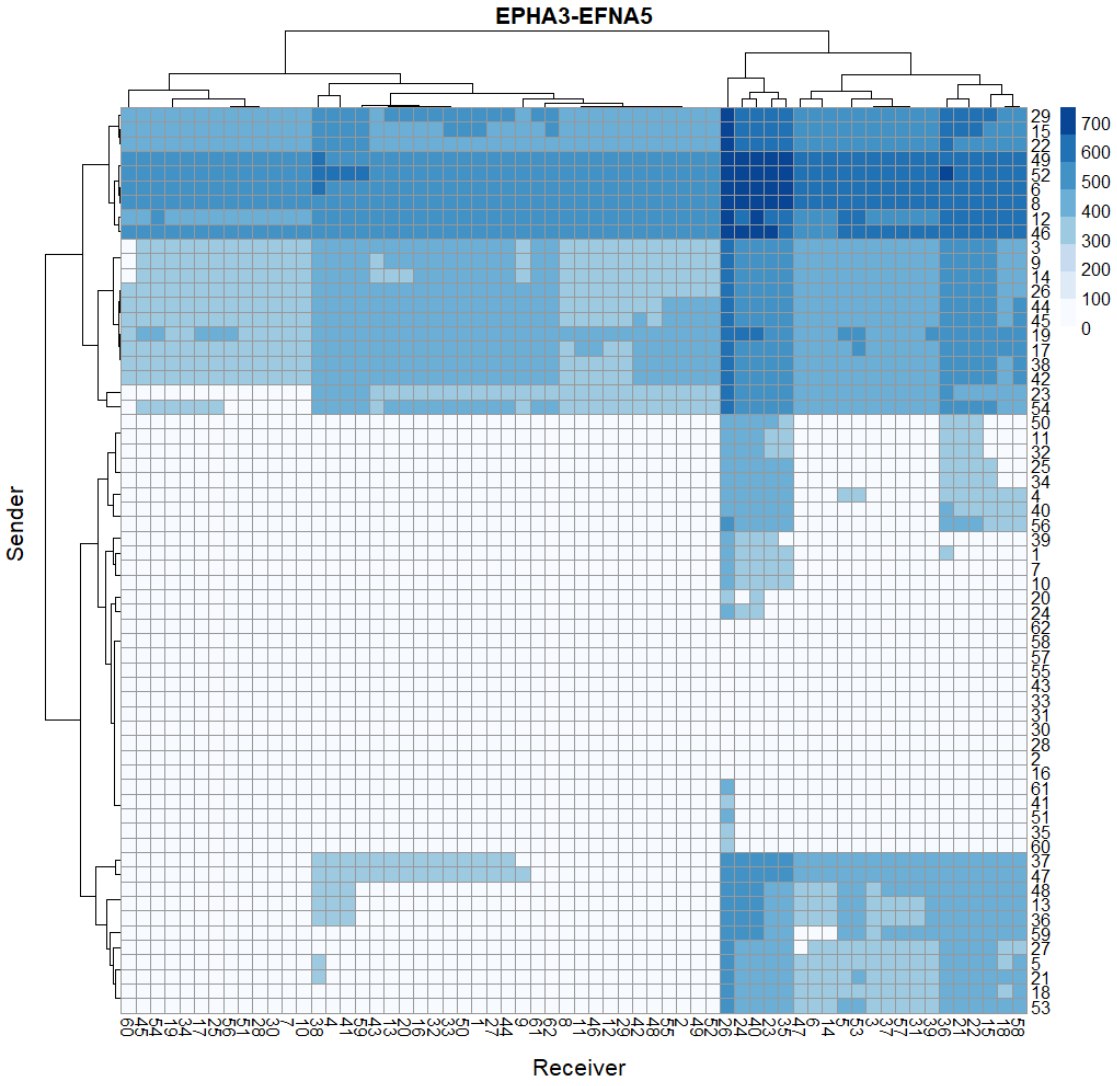
B.

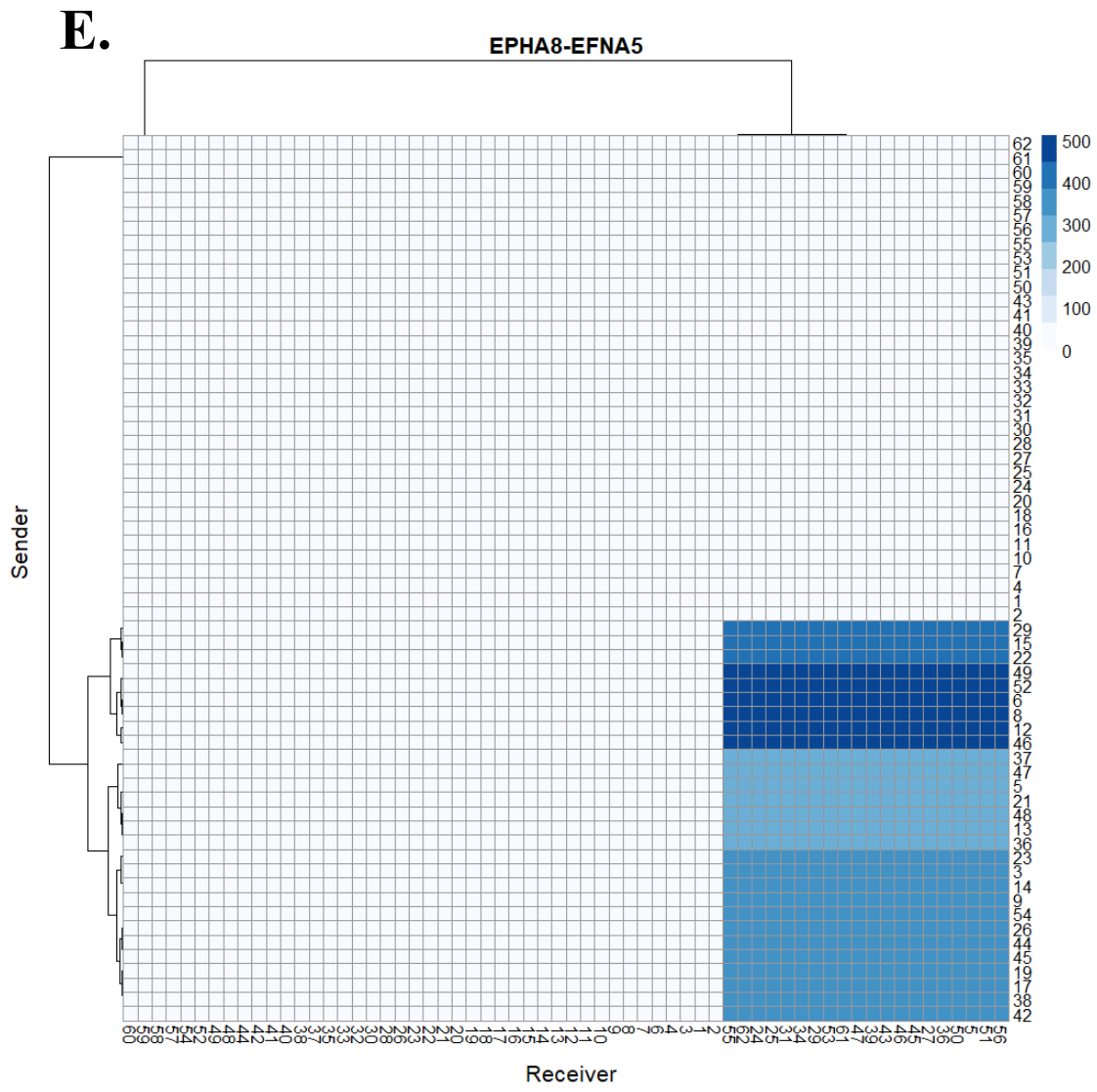


C.



D.





Supplementary Figure 2: Heatmaps of significant expression of negatively distance-correlated (A) EPHA4-EFNA2 (B) EPHA4-EFNA3 (C) EPHB2-EFNA5 (D) EPHA3-EFNA5 (E) EPHA8-EFNA5. The x-axis represents sender cell clusters; the y-axis represents receiver cell clusters, blue color represents the expression value of L-R pair, deeper shades represent higher L-R expression between the corresponding clusters.