rCom: A route-based framework for inferencing cell type communication and regulatory network using single cell data

Honglin Wang†  
 Computer Sci. and Engr. Dept., Univ. of Connecticut, Storrs, USA  
 honglin.wang@uconn.edu

xxxxx  
 Computer Sci. and Engr. Dept., Univ. of Connecticut, Storrs, USA [email@email.com](mailto:email@email.com)

Chenyu Zhang  
 Computer Sci. and Engr. Dept., Univ. of Connecticut, Storrs, USA  
 email@email.com

xxxx  
 Computer Sci. and Engr. Dept., Univ. of Connecticut, Storrs, USA  
 honglin.wang@uconn.edu

Chenyu Zhang  
 Computer Sci. and Engr. Dept., Univ. of Connecticut, Storrs, USA [email@email.com](mailto:email@email.com)

Dong-Guk Shin  
 Computer Sci. and Engr. Dept., Univ. of Connecticut, Storrs, USA  
 dong.shin@uconn.edu

ABSTRACT

The mapping of ligand-receptor pairs is the cornerstone of understanding complicated intercellular interactions. With recent advances of single cell RNA (scRNA) sequencing technology, several methods have been proposed to infer cell-cell communication by analyzing ligand-receptor pairs. However, existing methods have limited ways of using what we call “prior knowledge”, i.e., what are already known (albeit incompletely) about the upstream for the ligand and the downstream for the receptor. In this paper, we present a novel framework, called rCom, capable of inferring cell-cell interactions by considering portions of pathways that would be associated with upstream of the ligand and downstream of receptors under examination. The rCom framework integrates knowledge from multiple biological databases including transcription factor-target database, ligand-receptor database and publicly available curated signaling pathway databases. The rCom framework examines combinatoric ways of integrating the partially known relationships against the cohorts of gene expression datasets obtainable through subtyping of cells. We combine both algorithmic methods and heuristic rules to score each putative ligand-receptor matchup between all possible cell subtype pairs. Permutation test is performed to rank the hypothesized cell-cell communication routes. We performed two case studies using single cell transcriptomic data sets of bone biology. Our literature survey suggests that rCom could be effective in discovering novel cell-cell communication relationships that have been only incompletely known in the field.

KEYWORDS

Cell communication, Topology and route-based pathway analysis, Bone marrow, Single cell RNA seq

1 Introduction

Complex intercellular responses start with binding of a ligand to its cognate receptor to activate specific cell signaling pathways. Mapping ligand- receptor pairs is fundamental to understanding how cells respond to signaling from neighboring cells. Single cell RNA-seq technology holds great promise for studying cell-cell communication which was not easy when gene expression data is obtained from cell population-based technologies such as microarray and bulk RNA-seq. Using scRNA-seq data, several methods have been developed to infer ligand-receptor pair communications between interacting cell types. Skelly et al. [1] and Kumar et al. [2] predict ligand-receptor pairs by studying if the two genes are highly co-expressed in different cell types. Zhou et al. [3] suggest identifying ligand-receptor pair signal between cell types is needed above and beyond genetic mutations or intracellular signaling to better understand cancer progression. Vento-Tormo et al. [4] produce a repository of ligand-receptor complexes which could potentially explain interaction between decidua and placenta during the early human pregnancy. More recently, the cell-cell interaction analysis has gone beyond identifying ligand-receptor pairs by attempting to produce gene regulation networks that could model interaction relationships among subpopulations of cells. Wang et al. [7] introduce SoptSC which xxxxx xxxxxxxx. Browaeys et al. [8] report NicheNet which aims to identify not only ligand-receptor pairs of interacting cells but also the genes downstream of them. Hu et al [9] present that CytoTalk can de novo generate a signal transduction network between a pair of cell types, which is basically suggesting ways to interconnect a pair of GRNs, each constructed the cell types under consideration.

Among the more recent reported cell-cell communication network discovery methods based on single cell transcriptomics data, our system, namely, rCom, is similar to NicheNet in the sense that both uses prior knowledge in identifying cell-cell communication network. One key difference is in the way we use curated pathways in the analysis. We have been suggesting that pathway analysis should treat routes of pathways i.e., branches of curated pathways (e.g., KEGG) as the unit for the analysis as opposed to the entire assembled genes of a pathway since each route in a pathway would be meant for some specific biological process. Our previous works include BioTarget which uses the route-based pathway method to extend existing Th1 and GATA pathways using TCGA data sets [10], [11]. It demonstrates the strength of route-based pathway analysis by identifying and scoring transcription factor (TF)-centric routes in pathways [12]. We find that our route-based pathway analysis is very suitable to analyze how two cells interact with each other through ligand-receptor. Specifically, we find that our method can fully take advantage of the siz cohort analysis readily available from scRNA-seq data. For example, if one find 100 cells of one type (immune cells) and 200 cells of another type (pre-osteoblast cells) offers combinatorics to infer the pairing with sufficient statistical power. Some of these pathway routes can lead to regulation of a ligand, and other routes can be directly activated by binding of the ligand. Analyzing these routes which are directly associated with ligand-receptor binding can give us better understanding of cell-cell interactions (CCIs).

We point out that our approach differs from the existing ligand-receptor analysis methods in multiple ways: (i) from However, the existing gene signaling pathway databases have a limited annotations on the ligand-receptor pairs which is the key to analyzing cell type communications. Other databases like CellChat[13] , CellTalkDB[14], namely ligand-receptor pairs databases, only provide the ligand and receptor pairs but never map them into the signaling pathway routes. We postulate that when these two related curated resources are combined and “mined” properly, one can produce an extended ligand and receptor centric cell-cell communication and regulatory routes. In this manuscript, we present a novel computational framework, named rCOM, that uses signaling pathway routes extracted from multiple curated biological sources to identify statistically significant CCIs and visualize them as a network called route-based inter cell type regulatory network (rICRN). In the rICRN, nodes represent cell groups and edges represent the number of interaction routes between the nodes.

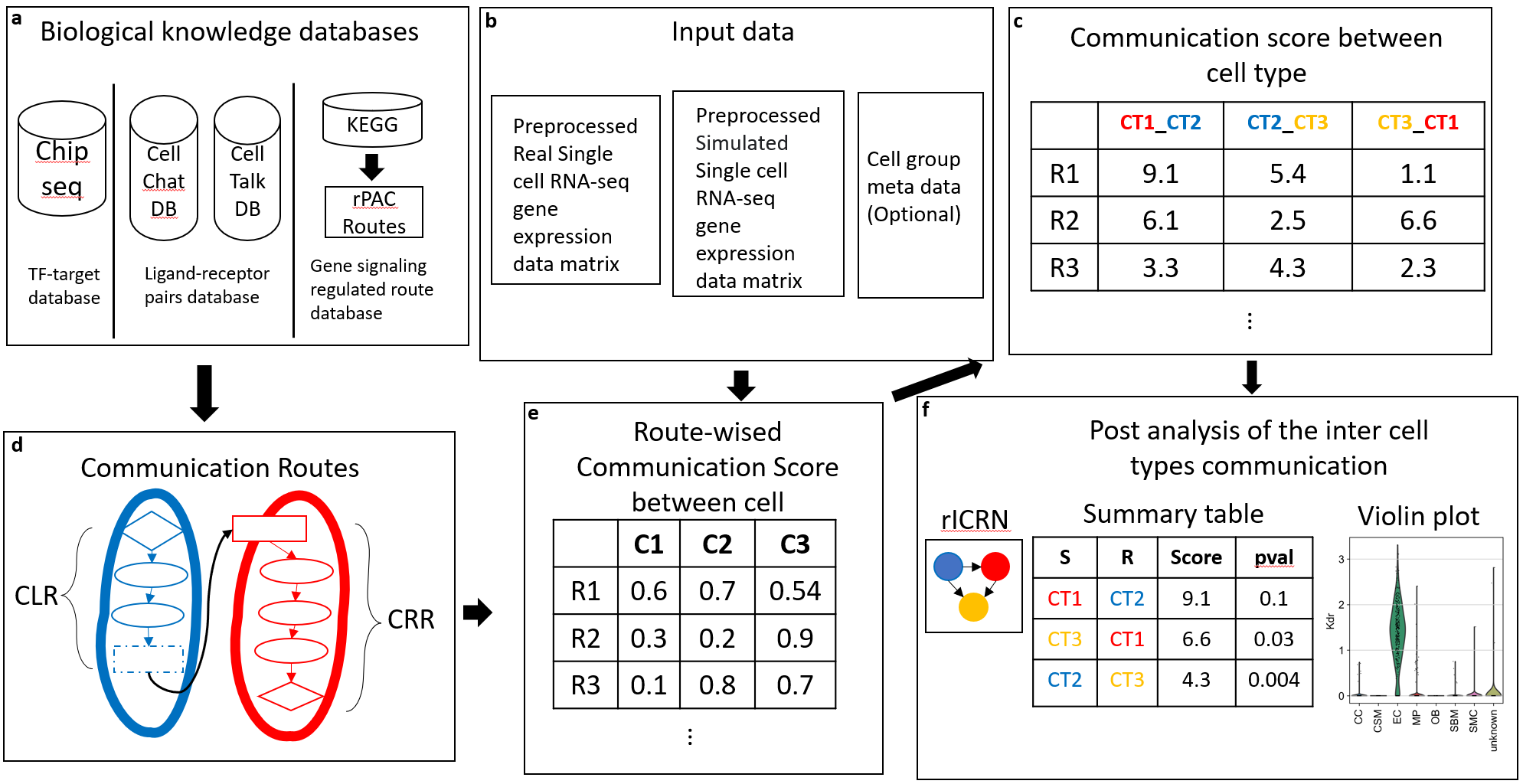


Fig 1. The overview architecture of rCom. **a.** The three categories of biological knowledge databases used in the rCom. **b.** The input data for rCom including: (i). preprocessed scRNA-seq with real cell group label for inferencing inter cell type communication. (ii). preprocessed scRNA-seq with randomized group label for permutation test. (iii). Cell group label meta data file. **c.** The matrix of communication score between cell type. **d.** The communication route pairs extract from 3 type databases. **e.** Route-wised communication scored in each cell. **f.** The post analysis of the cell types communication include: inferenced rICRN, summary table including cell type with permutation p-value, violin plot for the ligand-receptor expression value.

The rest of the paper is organized in the following way. In Section 2, xxxx, In Section 3 xxx, The rCom framework infers cell type communication by mapping preprocessed scRNA-seq transcriptomic data to two types of communication routes: communication ligand route and communication receptor route. Each route in every cell is scored by considering the consistency of regulatory pattern of gene in routes. The inter cell type communication is inferenced by a heuristic rule based on the comparison of the inter and intra cell type communication. In the end, p-value of permutation test is used to identify the statical significance for communication between two cell types through each route. A route based inter cell type regulatory network (rICRN) are built based on significant communication route pairs. The details will be introduced in section 2. We demonstrate how rCOM can be used to analyze single cell datasets in two independent bone related scRNA-seq dataset and discuss the result in the section 3. The conclusion is stated in the last section.

2 System and Methods

The overview of the rCom framework is given in Fig. 1. Fig 1a shows that three different types databases, namely source databases, TF-target database, ligand-receptor pairs database and gene signaling regulated route databases, are examined to build L-R communication candidate routes. The source databases are downloaded in text or KGML format and parsed to build a network graph. Here ligands, receptor, gene, and TF become nodes and gene regulated relationships are formed into edges interconnecting nodes. In the graph structure, nodes regulated interactions are directional and are encoded as either activation or inhibition (Fig.1a, 1c).

Fig 1b shows that three types of input data are used. Single cell gene expression matrix is preprocessed through quality control, normalization, scaled and log transform using SCANPY[15]. In the matrix, columns indicate cells and rows indicate genes. After log transformed, rCom treated gene with negative value as inhibited genes and gene with positive value as activated genes.

2.1 Cell-Cell communication routes

Fig 2 provides the details of the conceptual diagram representing cell-cell communication Fig 1d. The model labels two interacting cells as “Secretor” denoting the cell secreting a ligand and “Receiver” denoting the cell in which the ligand binds to its cognate receptor. This binding is represented as the green dashed line in Fig 2. We call these two cells as L-R pair. we show how we generate communication routes from four different biological knowledge databases including: Chip-seq [16], CellChat , CellTalkDB and rPAC. Unlike other methods that only consider ligand-receptor, the rCom infer cell communication based on the communication routes that are extended from L-R pairs.

Communication route pair has two routes: (i) communication ligand route (CLR) and (ii) communication receptor route (CRR). Communication ligand route captures the signals which are cell-secreted ligands produced when the transcription factor (TF) binds to the target genes. The cell that secrets ligands is known as secretor. CLR usually starts from TF and follow the path to the ligands. Two biological knowledge databases are used to generate CLR, rPAC routes and Chip-seq DB. The CLR is like effector routes in rPAC, but the primary TFs of CLR are not only from KEGG but also from Chip-seq. Communication receptor route captures the signals transduction from cell surface into the nucleus in which each route starts from receptor and follows a path to the TFs. The cell that where ligand binds is known as receiver. CRR are transformed from signaling routes in rPAC (KEGG). The connection of CLR and CRR are ligand and receptor pairs derived from two ligand-receptors pairs databases: CellChatDB, CellTalk and ligand-receptors pairs in signaling routes of rPAC. Totally 104004 communication route pairs are identified and stored in graph structure in rCom.

Diagram

Description automatically generated

Fig 2. The overview of the communication route. The blue route is CLR, and the red route is CRR. The CLR is generated from rPAC routes and Chip-seq database. The CRR is generated from rPAC routes. The ligand and receptor pairs are derived from CellChatDB, CellTablkDB and rPAC.

2.2 Node expected value and evaluation value

*2.2.1 Node expected value.* A node expected value () is either +1 or -1 which indicates whether node should be up-regulated or down-regulated in the route to activate ligand or activated by receptor. A node is assigned an expected value using propagation method which starts from ligands in CLR or receptors in CRR to the end node of the route (TF in both CLR and CRR). The equation of this method is given in Eq 1.

(1)

In this equation, stands for the nodes from the starte node (ligand in CLR and receptor in CRR.). stands the edge expectation value between node and node. The value of can be assign using Eq 2.

*2.2.2 Node evaluation value.* Node evaluation value () is assigned based on the node expected value and node type. Two types of nodes will appear in the communication routes: singleton node and bundle node. A singleton node stands for only one gene in the node which is computed using the Eq 3.

stands for the expression value of gene g in node k after preprocessed. Bundle node is composed of multiple nodes which collectively contribute to regulate the downstream genes

2.3 Communication route score

A communication route () score () measure the probability of cell to communicated with other cells through communication route . A cell with high CLR score means it has a high capability to secret ligand which are highly regulated in this specific route, whereas a cell with a high CRR score indicates a high probability to receive signals through the receptor and regulate it downstream gene and TF. The score is assigned using the equation given in Eq. 4a, 4b.

where is the number of nodes in communication route , and are the node evaluation value of ligand node or receptor node in CLR or CRR. Since the communication behavior is mostly rely on the ligand and receptor, an adjustable hyperparameter is given to emphasize the effect of ligand or receptor. The choice of will be “context-dependent” rang from 0 to 1 and need to be determined empirically like hyperparameters of machine learning models.

2.4 Inferencing cell type communication

To infer the inter- and intra-cellular communications, we introduce two heuristic rules which can model two distinct types of cell-cell communication known as autocrine and paracrine. Since the mass of ligand secreted by a cell type is constant, a strong inter cell type communication will lead to a week intra cell type communication and vice versa. According to this rule, a score of communication between the cell type and through communication route pair is computed using Eq.5.

Here and represent the CLR and CRR score in cell group and were computed using Eq. 6a and 6b. and stands for the score of communication route for cell in cell group .

Diagram

Description automatically generated

Fig 5. The complete view of communication route pairs id 70072.

By checking the communication score , it is possible to infer the cell type communication between all possible pairs of cell types in the data set. The higher score represents for a higher probability of communication between the cells in two cell types through the specific communication routes.

2.5 Identification of statistically significant cell type communication routes

The significant interactions between cell groups are identified using permutation test by randomly permuting the cell type label and then recalculating the communication score between cell type and through communication route pair . The p-value of each communication score is calculated using Eq. 7

Where the score is the communication probability for the m-th permutation. M is the total number permutations (default is 100). The communications with p-value < 0.05 are considered as significant.

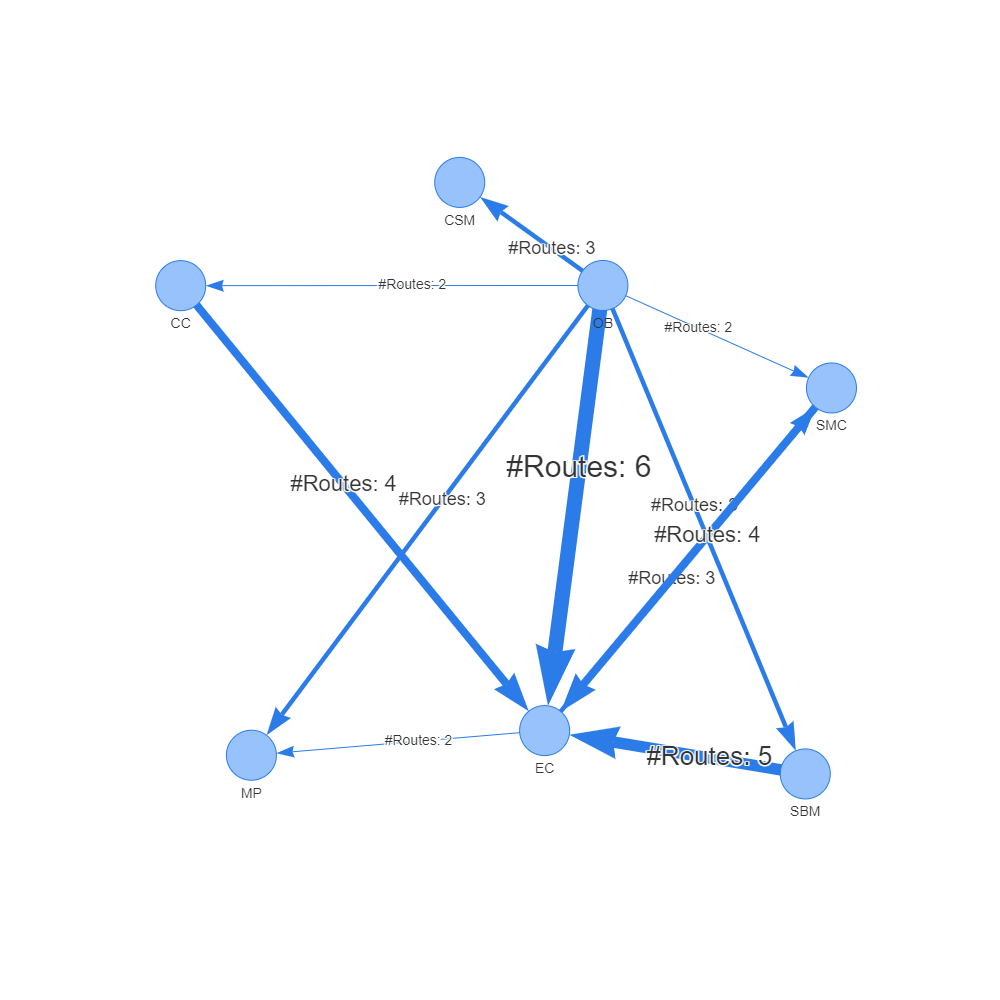
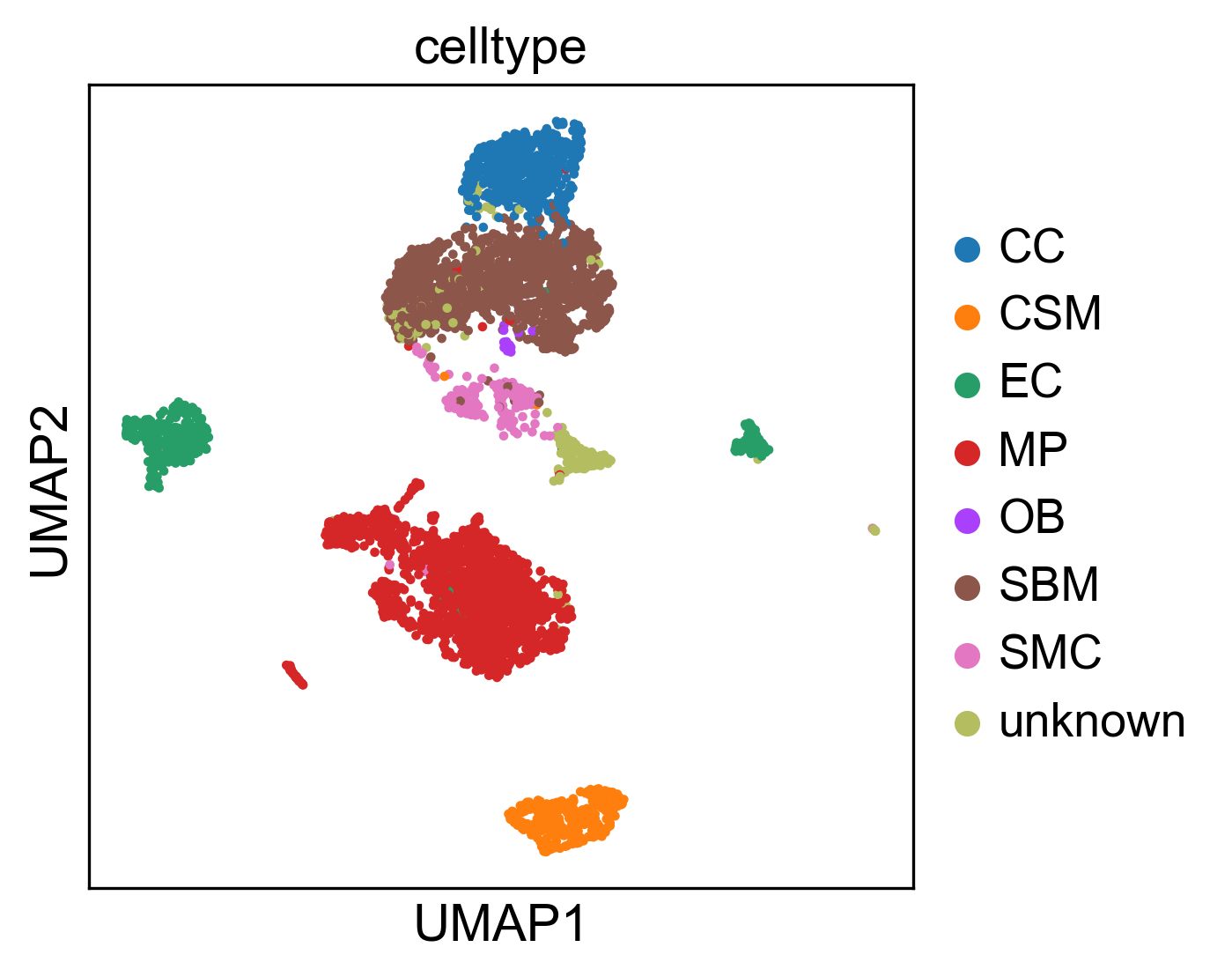


Fig. 4 rICRN generated from rCOM, the weight of edge shows the number of communication routes between two cell group higher than the threshold.

3 Experiment results

We experimented rCom with two sets of independent single cell transcriptomics datas related to bone biology. For the “prior knowledge”, we use four curated biological knowledge databases xx, xxx, xxx and xxx, as suggested in Section 2 System and Methods. The rICRN is build based on the number significant communication route pairs for each dataset. We check the ligand and receptor expression value for the significant cell type communication route along with the upper or downstream gene in the routes. Our literature survey for cell type related genes gives us confidence of the correctness for identified cell type communications.

Fig 3. The reproduce 2-D visualization of UMAP, colored by cell label provided by studies.

**3.1 Experiments with Acta2-lineage cells in osseointegration dataset**

The first single cell transcriptomics datasets we applied rCOM is the specific bone marrow stromal cell populations that contribute to bone formation around metal implants [17]. After preprocess, the scRNA-seq data matrix we collected has 4397 cells and 31053 features (genes). The labels of each cell group are given in this dataset include a). cell of skeletal muscle (CSM) (269 cells), b). chondrocyte (CC) (545 cells), c). endothelia cell (EC) (394 cells), d). macrophage (MP) (1608 cells), e). osteoblast (OB) (19 cells), f). smooth muscle cell (SMC) (191 cells) and g). stromal cell of bone marrow (SBM) (1371 cells). The repoduced 2-D visualization of UMAP with the given labels is shown in the Fig 3. The consistence between cell type label and the cell 2-D location indicate the correctness of our preprocess procedure.

A picture containing map

Description automatically generated

Fig 8. The reproduce 2-D visualization of UMAP, colored by cell label provided by studies.

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Fig 7. The violin plot of COL18A1 and ITGB3 gene expression value in each cell type.

Tab 1 shows the statistically significant communication routes for various of L-R pairs with the threshold of the score larger than 8.5. Fig 4 shows the rICRN of this dataset generated using Pyvis. Each node stands for a cell type, and the direct of edges indicate the signaling direction (From ligand in CLR to receptor in CRR). For example, as highlighted in this Fig, two significant route pairs are identified in between EC and MP. EC is the secretor and the MP is the receiver. The label of edge shows the number of the significant communication pairs between two cell groups. From the networks, we notice that the communication from SBM to EC, from OB to EC and from CC to EC have the greatest number of significant communication routes. Among all the cell type communications, the communications between EC and MP are well studied[18]. Our rCom also identified a communication route with id: 70072 (highlighted in Tab. 1) as a significant route between EC and MP as shown in the Fig 5. In this pair, GATA2 transcripts ICAM2 in EC at first. Then ICAM2 is secreted by EC and binds its receptor ITGB2 in MP. Finally, ITGB2 regulates its down steam genes and actives the MAPK family genes at the end of the routes. The violin plots for gene expression of ICAM2 and ITGB2 in different cell groups are shown in Fig 6. ICAM2 highly expressed in the EC group and poorly expressed in the MP group while on the receptor ends, ITGB2 is highly expressed in the MP cells but poorly expressed in the EC. This indicates a high probability of activation for the ligand and receptor. Our literature survey also indicates that in 2011, Zhang et al. [19] identified the communication between EC and MP through ICAM family gene. Though GATA2, MYC, TAL1 and SCLY are all the TFs that can regulate ICAM2 (Fig 5), only route starts from GATA2 got picked up by rCom. To verify that the rCom find appropriate route, the violin plots of gene expression for upper stream TFs of ICAM2 are shown in Fig 7. Only GATA2 are highly expressed in the EC. The consistence between gene expression value and the output of rCOM validate the correctness of route identification. GATA2🡪ICAM2 as upstream of the ligand ICAM2 and xxx, xxxx} 🡪 … 🡪 {xxxx, xxxx, xxxx} as downstream of ITGB2.



Fig 6. The violin plot of ICAM2 and ITGB2 gene expression value in each cell type.

3.2 Experiments with bone marrow stromal scRNA-seq of 11 weeks old mice dataset

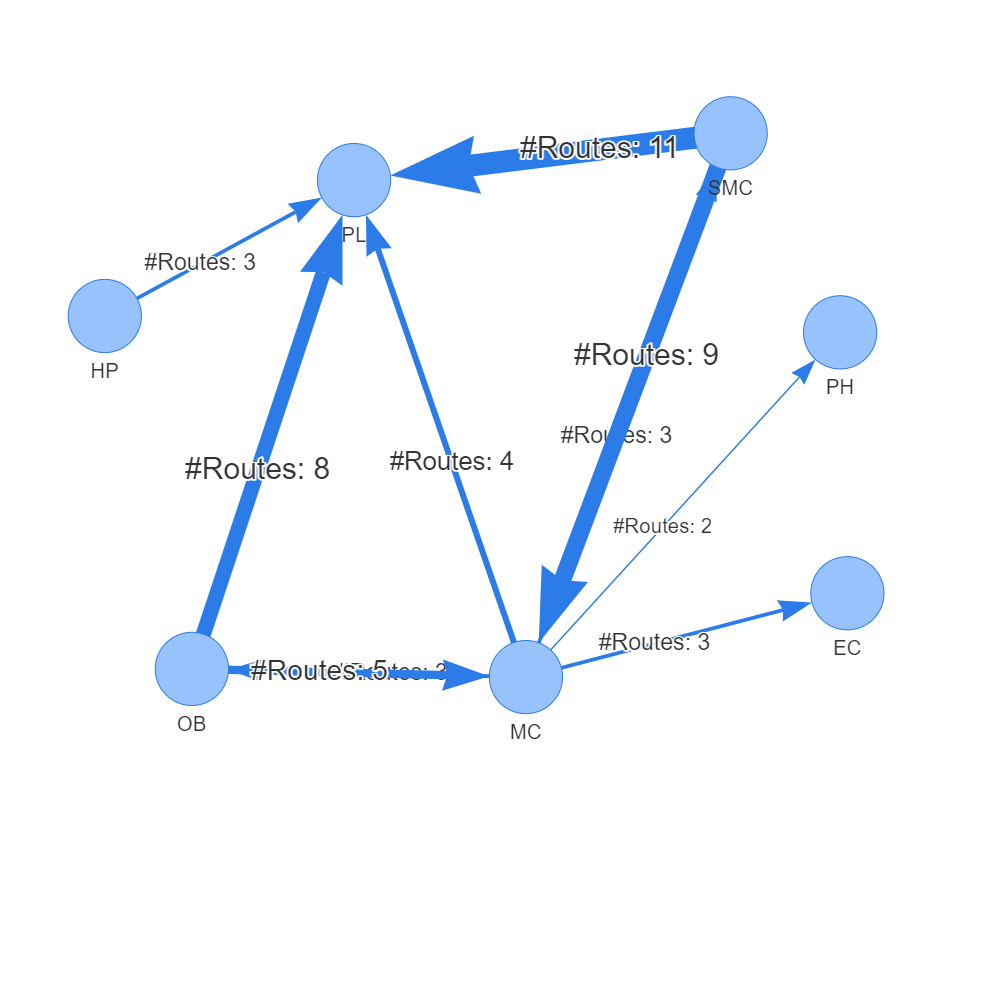


Fig. 9 rICRN generated from rCOM, the weight of edge shows the number of communication routes between two cell group higher than the threshold.

The second single cell transcriptomics data sets is from a study [21] which explored osteogenesis in adult mice and show that a bone-targeting, high affinity without causing adverse effects in other organs, which are known to reply on intact Notch signaling interactions. After preprocessing, the scRNA-seq data matrix has 17015 cells and 15046 features(genes). The labels of each cell groups are given in the meta file of the dataset including, CC (2270 cells), EC (1677 cells), Hematopoietic (HP) (849 cells), MSC (9168 cells), OB (971 cells), platelets (PL) (232 cells), proliferating Hem (PH) (1532 cells), SMC (238 cells), MC (78 cells). We show in Fig 9 that our UMAP analysis approximately reproduces the authors’ single cell 2-D plot visualization. The pattern those cells sharing same group are close to each other stands for the correctness of our preprocess.

Tab.2 shows the significant communication routes identified by rCom with threshold of score larger than 9. The corresponding xxxx i.e., rICRN, is shown in Fig.10. Among all the communication between each two cell types. The communication between xxxx SMC and xxxx PL has the largest number of significant routes pairs, and the communication from MC to PH only have 2 route pairs. In the network, a communication route pair (id: 95696), shown in Fig.11, between SMC and PL draws our attention as SMC and PL are known to communicate to each other [22]. As highlighted in this Fig 9, SMC as secretors sending ligands to their receivers, PL. In all these route pairs, the pair id: 95696 gain the highest communication score, as highlighted in the Tab 2. As shown in Fig 10, TBX3 transcripts COL18A1 in SMC at first. Then COL18A1 is secreted by SMC and binds its receptor, ITGB3 in MP. Finally, ITGB3 regulates its down steam genes and actives the PIK family genes at the end of the routes. The gene expression value of COL18A1 and ITGB3 ligand and receptor pairs are compared using violin plot in Fig. 11. In all cell groups, COL18A1 only highly expressed in the SMC and its receptor, ITGB3 are highly expressed in the PL and only few MC cells have ITGB3 expressed. This inferred that SMC cells have a high chance of sending COL18A1 and communicate with PL cells through ITGB3. Our literature shows that in 2018, Misra et al. [23] indicates that ITGB3 plays and important role for smoothing muscle-derived atherosclerotic plaque cells including: SMC and PL. We demonstrate the violin plot of gene expression value in the dataset of all the TFs that can regulate the COL18A1 include: TFAP2A, TBX3, NUCKS1, SOX11 and RCOR2 as shown in Fig 13. Only TBX3 are highly expressed in the SMC and rCOM picks the TBX3 as the upper stream for the COL18A1 instead of other. The consistence between gene expression value and the output of rCOM validate the correctness of route identification.

Chart, diagram

Description automatically generated

Fig 11. The complete view of communication route pairs id 95696

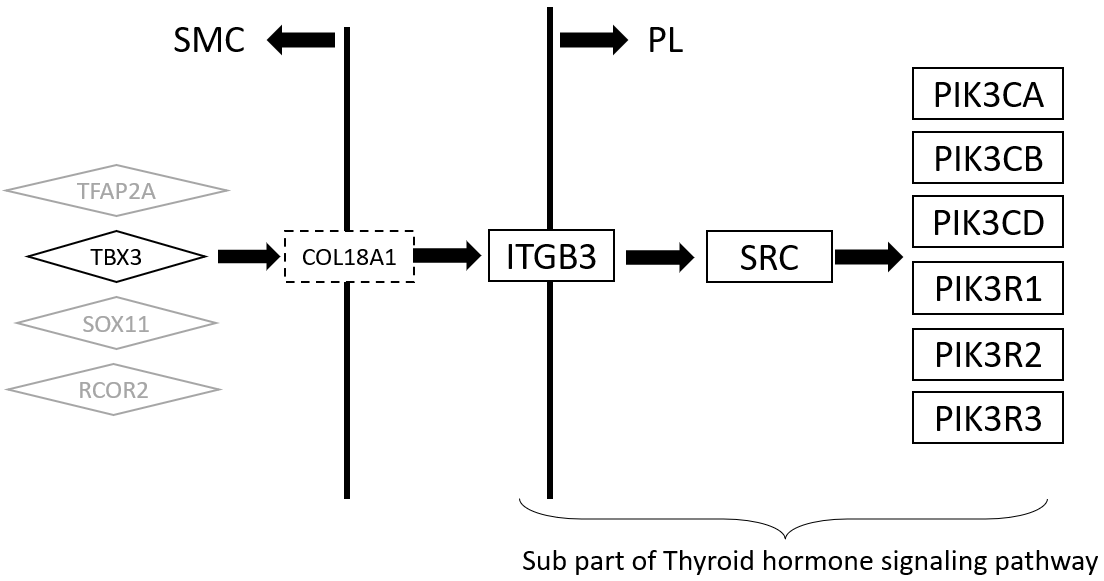
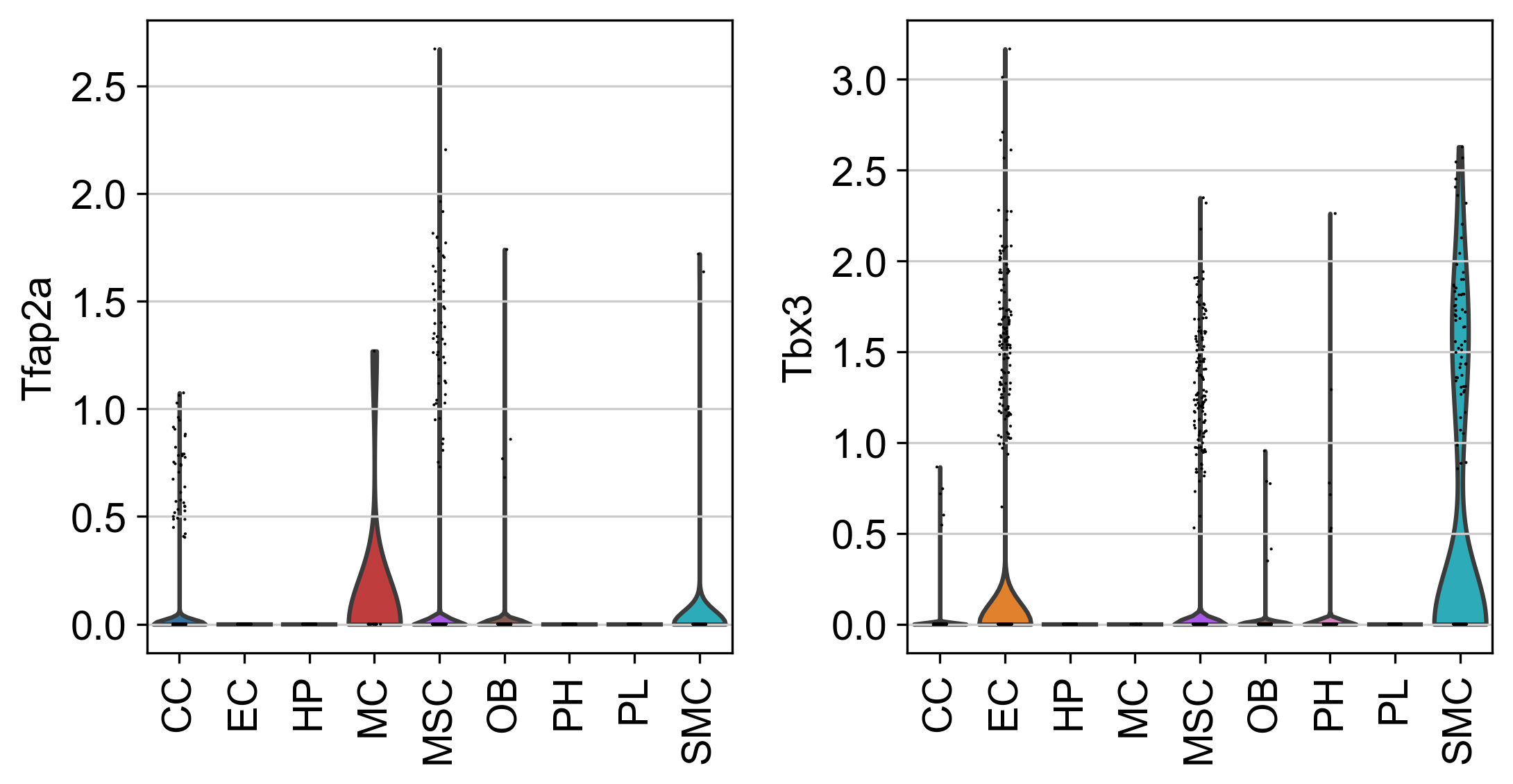


Fig 10. The complete view of communication route pairs id 95696

4. Discussion and conclusion

Developing rCom was motivated by inferencing the cell type communication at route-based level through gene signaling pathway route with ligand-receptor pairs mapped. Among all the existing methods, Ctyotalk returns a signal transduction network between cell types which is similar to rCOM. The difference between rCom and Cytoalk is how we borrow the strength of prior knowledge. The network inferred by Cytotalk is data driven while rCom try to identify the meaningful network by connecting routes which are already well studied and supported by wet lab experiments. While we acknowledge that there are number of existing methods which focus on inferring cell type communication using ligand-receptor expression patterns, our survey revealed that none of these methods use pathway topology (activation/inhibition) and regulatory consistency in their model. The rCom framework provides a new way of using well-established prior knowledge to translate the communication behavior.



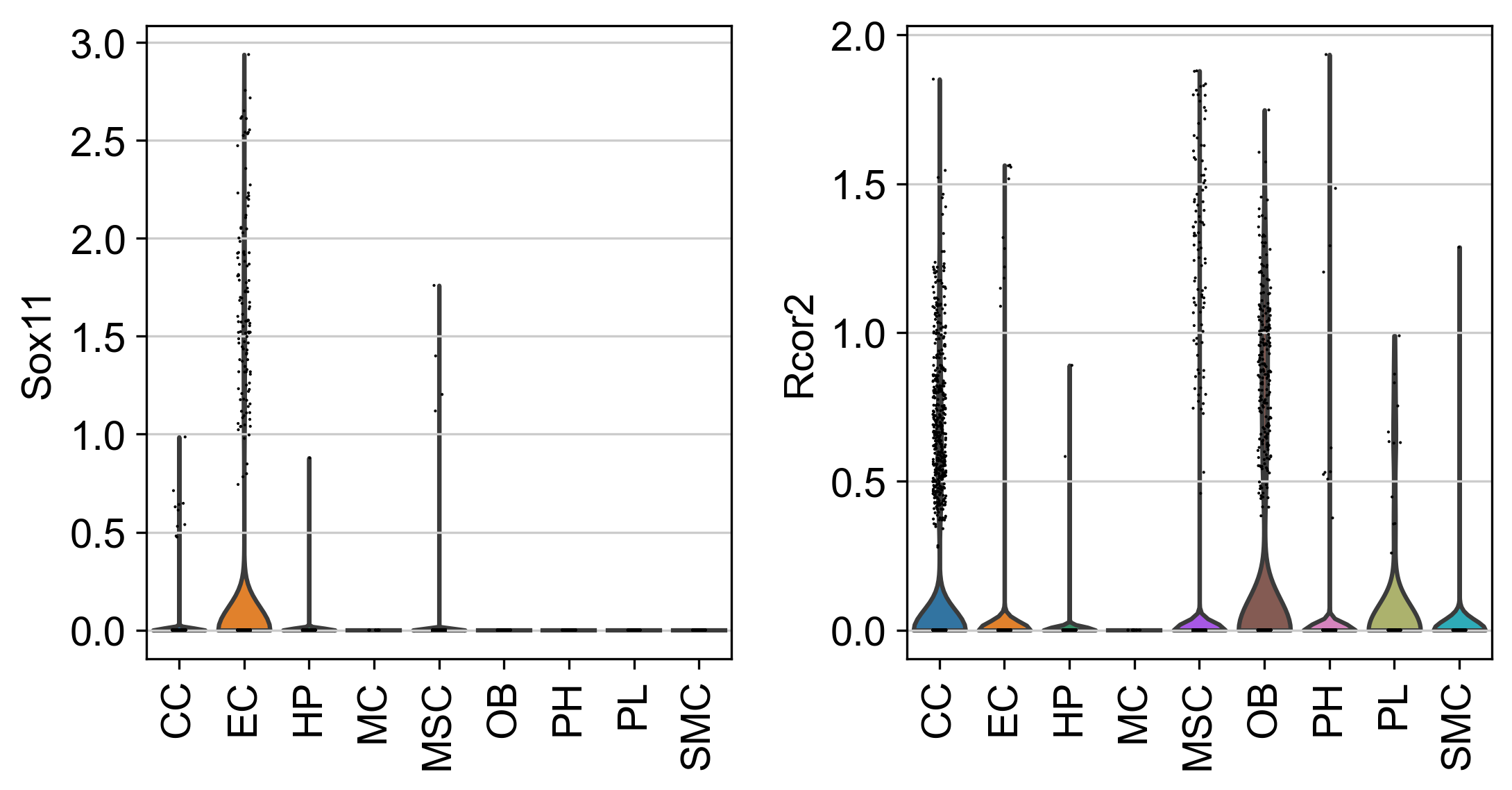


Fig 12. The violin plot for upper stream gene expression value of COL18A1 and ITGB3.

rCOM also provide a new way for combining biological knowledge databases to study pathway crosstalk. Identifying crosstalk between interacting pathways has been an active research area. Nodes-in-Common (NIC) predicts crosstalk by checking if two pathways share any proteins [24]–[26]. Hsu et.al. estimate pathway crosstalk based on similarities in Gene Ontology annotations[27]. Given a protein-protein interaction network, Li et al. links two pathways A and B if more edges connect the proteins in A to the proteins in B than expected by chance in a randomly wired network [28]. In our previous work, ctBuilder [29] is developed to identify a subnetwork interconnecting two pathways. However, with combination of ligand-receptor pairs databases and curated gene signaling pathway databases, the rCom framework can provide a novel aspect of exploring and identifying pathway crosstalk, which is the connection between ligand and receptor pairs. More specifically, without finding NIC, signal originating from CLR of pathway A in one cell can impact its downstream gene in CRR which is derived from a different pathway B in another cell. These two routes can be connected through the ligand-receptor pairs which is not originally annotated in pathways but supported by curated ligand-receptor databases. By analyzing the communication route pairs, unknown crosstalk between two independent pathways might be identified.

The current experiment with rCom was done using Chip-seq, CellChat, CellTalkDB and KEGG to utilize the causal information curated in the database. But it can be easily extended to use other databases like BioGrid [30]. An extendable work of rCom framework is to use spatial genomics data to validate result by checking if two cell types are co-localizing two each other. With the spatial genomic data, the interaction between cells instead of cell types may be identified and validated. How to utilize the spatial data and combine to rCom is an interesting future problem.

ACKNOWLEDGMENTS

Research reported in this work was supported in part by NIH/NICHD Grant No.1R01HD098636-01. Its contents are solely the responsibility of the authors and donot necessarily represent the official views of the NIH.

**REFERENCES**

[1] D. A. Skelly et al., “Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart,” Cell Reports, vol. 22, no. 3, pp. 600–610, Jan. 2018, doi: 10.1016/J.CELREP.2017.12.072.

[2] M. P. Kumar et al., “Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics,” Cell Reports, vol. 25, no. 6, pp. 1458-1468.e4, Nov. 2018, doi: 10.1016/J.CELREP.2018.10.047.

[3] J. X. Zhou, R. Taramelli, E. Pedrini, T. Knijnenburg, and S. Huang, “Extracting Intercellular Signaling Network of Cancer Tissues using Ligand-Receptor Expression Patterns from Whole-tumor and Single-cell Transcriptomes,” Scientific Reports 2017 7:1, vol. 7, no. 1, pp. 1–15, Aug. 2017, doi: 10.1038/s41598-017-09307-w.

[4] R. Vento-Tormo et al., “Single-cell reconstruction of the early maternal–fetal interface in humans,” Nature 2018 563:7731, vol. 563, no. 7731, pp. 347–353, Nov. 2018, doi: 10.1038/s41586-018-0698-6.

[5] Y. E. Antebi, N. Nandagopal, and M. B. Elowitz, “An operational view of intercellular signaling pathways,” Current Opinion in Systems Biology, vol. 1, pp. 16–24, Feb. 2017, doi: 10.1016/J.COISB.2016.12.003.

[6] M. Billmann, V. Chaudhary, M. F. ElMaghraby, B. Fischer, and M. Boutros, “Widespread Rewiring of Genetic Networks upon Cancer Signaling Pathway Activation,” Cell Systems, vol. 6, no. 1, pp. 52-64.e4, Jan. 2018, doi: 10.1016/J.CELS.2017.10.015.

[7] S. Wang, M. Karikomi, A. L. Maclean, and Q. Nie, “Cell lineage and communication network inference via optimization for single-cell transcriptomics,” Nucleic Acids Research, vol. 47, no. 11, pp. e66–e66, Jun. 2019, doi: 10.1093/NAR/GKZ204.

[8] R. Browaeys, W. Saelens, and Y. Saeys, “NicheNet: modeling intercellular communication by linking ligands to target genes,” Nature Methods 2019 17:2, vol. 17, no. 2, pp. 159–162, Dec. 2019, doi: 10.1038/s41592-019-0667-5.

[9] Y. Hu, T. Peng, L. Gao, and K. Tan, “CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data,” Science Advances, vol. 7, no. 16, Apr. 2021, doi: 10.1126/SCIADV.ABF1356/SUPPL\_FILE/ABF1356\_TABLE\_S6.XLSX.

[10] A. J. Bass et al., “Comprehensive molecular characterization of gastric adenocarcinoma,” Nature, vol. 513, no. 7517, p. 202, Sep. 2014, doi: 10.1038/NATURE13480.

[11] D. M. Muzny et al., “Comprehensive molecular characterization of human colon and rectal cancer,” Nature, vol. 487, no. 7407, p. 330, Jul. 2012, doi: 10.1038/NATURE11252.

[12] P. Joshi, B. Basso, H. Wang, S. H. Hong, C. Giardina, and D. G. Shin, “rPAC: Route based pathway analysis for cohorts of gene expression data sets,” Methods, Oct. 2021, doi: 10.1016/J.YMETH.2021.10.002.

[13] S. Jin et al., “Inference and analysis of cell-cell communication using CellChat,” Nature Communications 2021 12:1, vol. 12, no. 1, pp. 1–20, Feb. 2021, doi: 10.1038/s41467-021-21246-9.

[14] X. Shao, J. Liao, C. Li, X. Lu, J. Cheng, and X. Fan, “CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice,” Briefings in Bioinformatics, vol. 22, no. 4, Jul. 2021, doi: 10.1093/BIB/BBAA269.

[15] F. A. Wolf, P. Angerer, and F. J. Theis, “SCANPY: Large-scale single-cell gene expression data analysis,” Genome Biology, vol. 19, no. 1, pp. 1–5, Feb. 2018, doi: 10.1186/S13059-017-1382-0/FIGURES/1.

[16] “Chromatin Immunoprecipitation Sequencing (ChIP-Seq).” https://www.illumina.com/techniques/sequencing/dna-sequencing/chip-seq.html (accessed May 11, 2022).

[17] A. Vesprey et al., “Tmem100- and Acta2-Lineage Cells Contribute to Implant Osseointegration in a Mouse Model,” J Bone Miner Res, vol. 36, no. 5, pp. 1000–1011, May 2021, doi: 10.1002/JBMR.4264.

[18] H. He et al., “Endothelial cells provide an instructive niche for the differentiation and functional polarization of M2-like macrophages,” Blood, vol. 120, no. 15, p. 3152, Oct. 2012, doi: 10.1182/BLOOD-2012-04-422758.

[19] J. Zhang et al., “Regulation of Endothelial Cell Adhesion Molecule Expression by Mast Cells, Macrophages, and Neutrophils,” PLoS ONE, vol. 6, no. 1, 2011, doi: 10.1371/JOURNAL.PONE.0014525.

[20] H. Mayer, H. Bertram, W. Lindenmaier, T. Korff, H. Weber, and H. Weich, “Vascular endothelial growth factor (VEGF-A) expression in human mesenchymal stem cells: autocrine and paracrine role on osteoblastic and endothelial differentiation,” J Cell Biochem, vol. 95, no. 4, pp. 827–839, Jul. 2005, doi: 10.1002/JCB.20462.

[21] C. Xu et al., “Induction of osteogenesis by bone-targeted Notch activation,” Elife, vol. 11, Feb. 2022, doi: 10.7554/ELIFE.60183.

[22] R. Ross and L. Harker, “Platelets, endothelium, and smooth muscle cells in atherosclerosis,” Adv Exp Med Biol, vol. 102, pp. 135–141, 1978, doi: 10.1007/978-1-4757-1217-9\_8.

[23] F. Hartmann et al., “SMC-Derived Hyaluronan Modulates Vascular SMC Phenotype in Murine Atherosclerosis,” Circ Res, vol. 129, no. 11, pp. 992–1005, Nov. 2021, doi: 10.1161/CIRCRESAHA.120.318479.

[24] D. M et al., “Analysis and correction of crosstalk effects in pathway analysis,” Genome Res, vol. 23, no. 11, pp. 1885–1893, Nov. 2013, doi: 10.1101/GR.153551.112.

[25] K. H and K. MR, “Abiotic stress signalling pathways: specificity and cross-talk,” Trends Plant Sci, vol. 6, no. 6, pp. 262–267, 2001, doi: 10.1016/S1360-1385(01)01946-X.

[26] T. CM, E. B, and K. CR, “Critical nodes in signalling pathways: insights into insulin action,” Nat Rev Mol Cell Biol, vol. 7, no. 2, pp. 85–96, Feb. 2006, doi: 10.1038/NRM1837.

[27] C.-L. Hsu and U.-C. Yang, “Discovering pathway cross-talks based on functional relations between pathways,” BMC Genomics 2012 13:7, vol. 13, no. 7, pp. 1–15, Dec. 2012, doi: 10.1186/1471-2164-13-S7-S25.

[28] L. Y, A. P, and R. D, “A global pathway crosstalk network,” Bioinformatics, vol. 24, no. 12, pp. 1442–1447, Jun. 2008, doi: 10.1093/BIOINFORMATICS/BTN200.

[29] H. Wang, P. Joshi, S. H. Hong, D. J. Shin, and D. G. Shin, “CtBuilder: A framework for building pathway crosstalks by combining single cell data with bulk cell data,” Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, pp. 270–273, 2021, doi: 10.1109/BIBM52615.2021.9669811.

[30] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, “BioGRID: a general repository for interaction datasets”, doi: 10.1093/nar/gkj109.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S** | **R** | **R id** | **Score** | **CLR** | **Ligand** | **Receptor** | **CRR** | **Pval** |
| SMC | PL | 95696 | 12.64 | TBX3 | COL18A1 | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| SMC | MC | 80134 | 11.58 | NFIB | LGALS1 | PTPRC | LCK, ZAP70, LAT, GRB2, SOS1, SOS2, RASGRP1, HRAS, KRAS,  NRAS, RAF1, MAP2K1, MAP2K2, MAPK1, MAPK3, FOS, JUN | <0.0005 |
| SMC | PL | 95739 | 11.03 | PPARG | COL4A2 | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| SMC | PL | 95421 | 10.34 | NR3C1 | MFGE8 | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| SMC | PL | 95367 | 9.7 | PPARG | NID1 | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| MC | SMC | 72456 | 9.63 | RELA,  RELB | IL1B | IL1R1 | IRAK1, IRAK4, MYD88, TRAF6, MAP3K7, TAB1, TAB2, TAB3,  IKBKG, CHUK, IKBKB, NFKBIA, NFKB1, RELA | <0.0005 |
| OB | PL | 82280 | 9.6 | RUNX2 | IBSP | ITGB3 | PLCD3, PLCB1, PLCE1, PLCB2, PLCB3, PLCB4, PLCD1, PLCG1, PLCG2, PLCD4, PLCZ1, C00165, PRKCA, PRKCB, PRKCG, MAPK1, MAPK3, STAT1 | <0.0005 |
| OB | PL | 82619 | 9.45 | KLF4 | FN1 | ITGB3 | PLCD3, PLCB1, PLCE1, PLCB2, PLCB3, PLCB4, PLCD1, PLCG1, PLCG2, PLCD4, PLCZ1, C00165, PRKCA, PRKCB, PRKCG, MAPK1, MAPK3, STAT1 | <0.0005 |
| MC | EC | 72456 | 9.44 | RELA,  RELB | IL1B | IL1R1 | IRAK1, IRAK4, MYD88, TRAF6, MAP3K7, TAB1, TAB2, TAB3, IKBKG, CHUK, IKBKB, NFKBIA, NFKB1, RELA | <0.0005 |
| CC | PL | 90899 | 9.42 | STAT3 | TGM2 | ITGB3 | HRAS, KRAS, NRAS, RAF1, MAP2K1, MAP2K2, MAPK1, MAPK3, TP53 | <0.0005 |
| OB | PL | 82963 | 9.41 | CCN1 | CCN1 | ITGB3 | PLCD3, PLCB1, PLCE1, PLCB2, PLCB3, PLCB4, PLCD1, PLCG1, PLCG2, PLCD4, PLCZ1, C00165, PRKCA, PRKCB, PRKCG, MAPK1, MAPK3, STAT1 | <0.0005 |
| SMC | PL | 95503 | 9.39 | TCF4 | FBN1 | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| OB | MC | 69430 | 9.36 | CCN1 | CCN1 | ITGB2 | PTK2B, RAC1, RAC2, RAC3, PAK1, MAP2K1, MAP2K2, MAPK1, MAPK3 | <0.0005 |
| SMC | PL | 95268 | 9.35 | CCND1 | VTN | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| HP | PL | 95560 | 9.34 | MYB | HSP | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| SMC | MC | 69220 | 9.3 | MEF2A | JAM3 | ITGB2 | PTK2B, RAC1, RAC2, RAC3, PAK1, MAP2K1, MAP2K2, MAPK1, MAPK3 | <0.0005 |
| MC | CC | 72456 | 9.11 | RELA,  RELB | IL1B | IL1R1 | IRAK1, IRAK4, MYD88, TRAF6, MAP3K7, TAB1, TAB2, TAB3, IKBKG, CHUK, IKBKB, NFKBIA, NFKB1 | <0.0005 |
| SMC | PL | 95290 | 9.08 | EP300 | TGFB3 | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| SMC | PL | 96031 | 9.08 | TRIM28 | ITGAV | ITGB3 | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| MC | OB | 72456 | 9.04 | RELA,  RELB | IL1B | IL1R1 | IRAK1, IRAK4, MYD88, TRAF6, MAP3K7, TAB1, TAB2, TAB3, IKBKG, CHUK, IKBKB, NFKBIA, NFKB1 | <0.0005 |
| OB | PL | 82885 | 9.04 | KLF4 | SPP1 | ITGB3 | PLCD3, PLCB1, PLCE1, PLCB2, PLCB3, PLCB4, PLCD1, PLCG1, PLCG2,  PLCD4, PLCZ1, C00165, PRKCA, PRKCB, PRKCG, MAPK1, MAPK3, STAT1 | <0.0005 |
| SMC | MC | 69430 | 9 | CCN1 | CCN1 | ITGB2 | PTK2B, RAC1, RAC2, RAC3, PAK1, MAP2K1, MAP2K2, MAPK1, MAPK3 | <0.0005 |
| SMC | PL | 95963 | 9 | CCN1 | CCN1 | ITGB3 | PTK2B, RAC1, RAC2, RAC3, PAK1, MAP2K1, MAP2K2, MAPK1, MAPK3 | <0.0005 |

Table 2. The significant communication routes pairs with score higher than threshold (9). S stands for secretor and R stands for receiver.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S** | **R** | **R id** | **Score** | **CLR** | **Ligand** | **Receptor** | **CRR** | **Pval** |
| OB | EC | 82280 | 9.83 | RUNX2 | IBSP | ITGB3 | PLCD3, PLCB1, PLCE1, PLCB2, PLCB3, PLCB4, PLCD1, PLCG1, PLCG2, PLCD4, PLCZ1, C00165, PRKCA, PRKCB, PRKCG, MAPK1, MAPK3, STAT1 | <0.0005 |
| OB | MP | 96177 | 9.53 | RUNX2 | IBSP | ITGAV | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| OB | SMC | 96177 | 9.18 | RUNX2 | IBSP | ITGAV | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| OB | SBM | 96177 | 9.05 | RUNX2 | IBSP | ITGAV | SRC, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3 | <0.0005 |
| **EC** | **MP** | **70072** | **9.04** | **GATA2** | **ICAM2** | **ITGB2** | **ITGAL, VAV3, VAV1, VAV2, RAC1, RAC2, RAC3, PAK1, MAP2K1, MAP2K2, MAPK1, MAPK3** | <0.0005 |
| OB | EC | 6452 | 8.68 | SMAD3 | TNC | PTPRB | CTNNB1, CTNNA1, CTNNA2, CTNNA3, ACTB, ACTG1 | <0.0005 |
| EC | OB | 77882 | 8.66 | CREM | GNAI2 | LPAR3 | GNAS, ADCY1, ADCY2, ADCY3, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, ADCY4, C00575, RAPGEF3, RAPGEF4, RAP1A, RAP1B | <0.0005 |
| OB | EC | 103917 | 8.65 | RUNX2 | VEGFA | KDR | PLCG1, PLCG2, C00076, PRKCA, PRKCB, PRKCG, SPHK2,  SPHK1, HRAS, KRAS, NRAS, RAF1, MAP2K1, MAP2K2,  MAPK1, MAPK3 | <0.0005 |
| OB | EC | 4562 | 8.64 | RUNX2 | SEMA5A | MET | MAPK1, MAPK3, SNAI2, SNAI1 | <0.0005 |
| OB | EC | 103718 | 8.63 | CTNNB1 | PDGFC | KDR | PLCG1, PLCG2, C00076, PRKCA, PRKCB, PRKCG, SPHK2,  SPHK1, HRAS, KRAS, NRAS, RAF1, MAP2K1, MAP2K2,  MAPK1, MAPK3 | <0.0005 |
| OB | EC | 89677 | 8.55 | RUNX2 | IBSP | ITGAV | HRAS, KRAS, NRAS, RAF1, MAP2K1, MAP2K2, MAPK1,  MAPK3, STAT1 | <0.0005 |
| SMC | EC | 103686 | 8.53 | PBX1 | VTN | KDR | PLCG1, PLCG2, C00076, PRKCA, PRKCB, PRKCG,  SPHK2, SPHK1, HRAS, KRAS, NRAS, RAF1, MAP2K1, MAP2K2,  MAPK1, MAPK3 | <0.0005 |

Table 1. The significant communication routes pairs with score higher than threshold (8.5). S stands for secretor and R stands for receiver.