CrossTalkeR: Analysis and Visualisation of Ligand Receptor Networks

James S. Nagai[1], Nils B.Leimkühler[2], Rebekka K. Schneider[2,3,4], and Ivan G.Costa[1,*]

[1]Institute for Computational Genomics, Joint Research Center for Computational Biomedicine, RWTH Aachen University Medical School, Germany

[2]Department of Hematology, Erasmus Medical Center, the Netherlands

[3]Department of Cell Biology, Institute for Biomedical Engineering, Faculty of Medicine,RWTH Aachen University, Germany

[4]Oncode Institute, Erasmus Medical Center, the Netherlands

[*]ivan.costa@rwth-aachen.de

ABSTRACT

Motivation: Ligand-receptor (LR) analysis allows the characterization of cellular crosstalk from single cell RNA-seq data. However, current LR methods provide limited approaches for prioritization of cell types, ligands or receptors or characterizing changes in crosstalk between two biological conditions.

Results: CrossTalkeR is a framework for network analysis and visualisation of LR networks. CrossTalkeR identifies relevant ligands, receptors and cell types contributing to changes in cell communication when contrasting two biological states: disease vs. homeostasis. A case study on scRNA-seq of human myeloproliferative neoplasms reinforces the strengths of CrossTalkeR for characterisation of changes in cellular crosstalk in disease state.

Availability and Implementation: CrosstalkeR is as a package in R and available as a package in https://github.com/CostaLab/CrossTalkeR.

1 Introduction

Understanding cellular crosstalk is vital for uncovering molecular mechanisms associated with cell differentiation and disease progression. Single cell RNA sequencing (scRNA-seq) allows the characterization of tissue heterogeneity at an unprecedented level, however, information of cellular proximity and crosstalk is not captured. Computational methods, of which search for pairs of cell types expressing compatible ligand-receptor (LR) pairs, have become a powerful approach for dissecting cellular crosstalk from scRNA-seq data (1). However, current LR inference methods usually predicts, hundreds of potential LR pairs for a given scRNA-seq dataset making their interpretation

challenging. Also, most LR methods focus on the analysis of a single scRNA-seq experiment and are not able to characterize changes in cellular crosstalk between pairs of conditions, i.e. disease vs. homeostasis.

Based on published biological network-analysis methods (2), we developed a new package to rank, explore and visualize relevant LR interactions. CrossTalkeR explores network-based measures to rank cell types, receptors, ligands, cell-receptors and cell-ligand pairs regarding their importance in cellular crosstalk in a single or a pair of biological conditions. CrossTalkeR presents the results either graphically or in tabular form summarized as a HTML report. CrossTalkeR is compatible with the input of popular LR prediction methods (CellPhoneDB (3)) and (TalkLR (4)) and stores all results in an R object to foster data sharing and reproducibility of results.

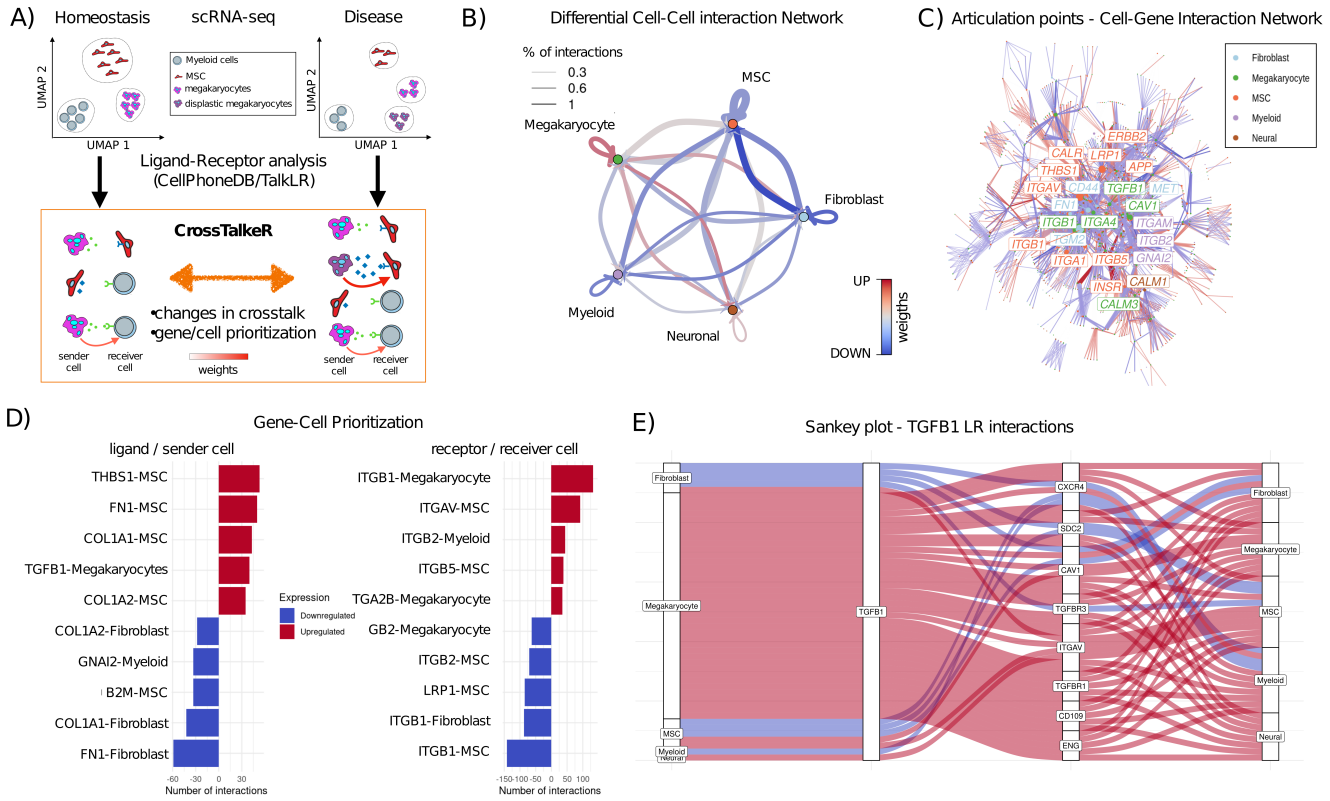

Figure 1. A) - Scheme of CrossTalkeR. B) - Cell-cell Interaction (CCI) and C) Cell Gene Interaction (CGI) differential networks estimated from Human Myelofibrosis scRNA-seq data (5). We highlight at the CGI level all ligand/receptors and respective cell types predicted as articulation points. D) Ranking of Ligand/Source and Receptor/Receivers regarding communication gained or lost within the disease state. E) Sankey plot listing all predicted source, receptor and receiver interactions associated with TGFB1.

2 Overview and implementation

CrossTalkeR requires as input a set of quintuples encoding the LR pair, the sender and receiver cell-pair and a positive score associated with the weight of the interaction. These predictions can be obtained by common LR

analysis methods(3; 4) (Fig. 1A). To measure the weight (or strength) of a LR interaction we use the mean of LR expression as in (3). Next, CrossTalkeR creates distinct network representations and estimates node statistics to prioritize cells or ligand/receptors (Fig. 1B-C). All the analyses can be done using single condition data and a comparative data from two conditions.

2.1 Network Construction

CrossTalkeR constructs two representations of the LR network. On the cell-cell interaction (CCI) network, the nodes are defined by each cell-type and the edges are weighted by characteristics of the interactions, i.e. number of LR pairs and sum of weights of LR pairs. On the cell-gene interaction (CGI) network, the nodes represent gene and cell pairs (ligand and sender cell or receptor and receiving cell) and the weights are amsmath igiven by the mean LR expression. Next, differential CCI and CGI networks are obtained by subtracting the condition state network, e.g. disease, from the control state network. In this way, the interactions with positive values are up-regulated in the condition and the negative valued interactions are down-regulated on the experiment (disease) state.

2.2 Network-based analysis

Node Ranking approaches: We apply distinct network property methods to rank either the CCI and CGI network. By default, measures take the weights of LR pairs into account.

- *In/Out Degree:* Number of in-going (out-going) edges
- *Closeness Centrality:* Higher closeness centrality indicates cells/genes that can communicate directly with other cells/genes.
- *Betweeness Centrality:* Indicates cells/genes that are moderating most cell communication (or bottleneck nodes).
- *Eigenvector Centrality:* Ranks cells/genes that are connected to important neighbors.
- *Pagerank:* Rank the cells/genes according to the number of the interactions and interaction power.
- *Articulation Point:* Cell(s)/gene(s) that break the communication between other cells/genes modules.
- *PC analysis*: Principal component (PC) analysis to combine the previous measures.

See supplementary material for formalization of networks and details of the metrics.

HTML report and graphical representations: CrossTalkeR creates an graphical and tabular representations in the form of a HTML report. This includes representations of the CCI and CGI networks, where edge thickness and color represent their characteristics and node colors represent distinct cell-types (Fig. 1B-C). We also produce bar plots to show top-ranked cells or gene/cell pairs (Fig. 1D). Moreover, it is possible to list all interactions associated within a particular ligand or receptor via a Sankey plot (Fig. 1E). The report also includes tabular representation of rankings, which can be sorted by distinct criteria or searched for particular genes and PC analysis. All results can be exported to spreadsheet formats. The package also produces an R object, which can be used for generation of further graphs, i.e. generation of Sankey plots of ligand or receptors of interest or simplification of plots by filtering ligand, receptors or cells.

The software is provided as an open source package (`https://github.com/CostaLab/CrossTalkeR`). This includes a tutorial to integrate the tool with ligand-receptor analysis tools such as CellPhoneDB and TalkR, as well as example code to reproduce the analysis of stromal cells in human myelofibrosis (5).

3 Case study: Cellular crosstalk in myelofibrosis

We revisit here an analysis of scRNA-seq of stromal cells on human primary myelofibrosis (PMF) (5). The CCI network graph (Fig.1B) indicates that megakaryocytes have the highest number of interactions up-regulated on disease towards itself and mesenchymal stromal cells (MSC). Articulation points and ranking of the CGI network (Fig. 1C-D) indicates importance of matrix related proteins (FN1 and collagens), which are secreted by MSC cells, and TGFB1, which is secreted by megakaryocytes. We use a a Sankey plot to further investigate TGFB-related interactions (Fig. 1E). It indicates that TGFb signalling is mostly provided by megakaryocytes towards MSC cells via multiple receptors. These results are in accordance with known mechanisms associated to myelofibrosis, e.g. the crosstalk between megakaryocytes and MSCs via TGFb signalling; and the deposition on matrix via MSC cells (6). This examples illustrates how CrossTalkeR can characterize changes in cellular crosstalk from scRNA-seq data.

4 Conclusion and Future Work

CrossTalkeR is a unique framework for network analysis of LR interactions and comparison of changes between two biological conditions. CrossTalkeR provides results on both graphical, tabular and R objects leveraging both interpretability and reproducibility of results. Future work includes the extension of network-based method analysis as well as the analysis of more complex experimental designs as time series or the comparison of several biological conditions.

Funding

This project has been funded by the clinical research unit CRU344 supported by the German Research Foundation (DFG) and the E:MED Consortia Fibromap funded by the German Ministry of Education and Science (BMBF).

References

1. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell-cell interactions and communication from gene expression. *Nature reviews. Genetics* epub (2020). URL http://www.nature.com/articles/s41576-020-00292-xhttp://www.ncbi.nlm.nih.gov/pubmed/33168968http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7649713.

2. Pavlopoulos, G. A. *et al.* Using graph theory to analyze biological networks. *BioData mining* **4**, 10 (2011).

3. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB v2.0: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes (2019).

4. Wang, Y. talklr uncovers ligand-receptor mediated intercellular crosstalk. *BioRxiv* (2020).

5. Leimkühler, N. B. *et al.* Heterogeneous bone-marrow stromal progenitors drive myelofibrosis via a druggable alarmin axis. *Cell Stem Cell* (2020). URL http://www.sciencedirect.com/science/article/pii/S1934590920305427.

6. Gleitz, H. F., Kramann, R. & Schneider, R. K. Understanding deregulated cellular and molecular dynamics in the haematopoietic stem cell niche to develop novel therapeutics for bone marrow fibrosis. *The Journal of Pathology* **245**, 138–146 (2018). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/path.5078. https://onlinelibrary.wiley.com/doi/pdf/10.1002/path.5078.

Supplementary Material

Network Construction

Definitions

Lets define the quintuple $x = (s, g_{lig}, r, g_{rec}, w)$, where $s \in \{1, .., k\}$ is the sender cell, $r \in \{1, .., k\}$ is the receiver cell, g_{lig} represents a ligand gene, g_{rec} is a receptor gene and $w > 0$ is the weight of the interaction. Cell types (s and r) are defined as one of the k clusters detected in a given scRNA-seq data. The weight w of the interaction indicate the strength in the ligand-receptor signal and is estimated by the mean expression of the ligand and receptor (MeanLR) as in CellPhoneDB. CrossTalkeR receives a list of quintuples $X = \{x_1, ..., x_n\}$ representing n significant LR interactions predicted by a LR tool from a single cell RNA-seq data set.

Gene Cell Interaction based Network

The Gene Cell Interaction network has vertices V_{GCI} represented by all possible pairs of ligand and sender cells ((s, g_{lig})) and receptor and receiver cells ((r, g_{rec})). We can encode a weighted directed graph in the vertice space V_{GCI} via a matrix $\mathbf{W_{GCI}} = \{w_{i,j}\}^{n \times n}$, where $|V_{GCI}| = n$. For a tuple $(s, g_{lig}, r, g_{rec}, w) \in X$, $w_{(g_{lig}, s), (g_{rec}, r)} = w$. For pairs $(s, g_{lig}), (r, g_{rec})$ not present in $\mathbf{X}$, i.e. no significant interaction was detected by CellPhoneDB, $w_{(s, g_{lig}), (r, g_{rec})} = 0$.

Cell Cell Interaction (CCI) based Network

We obtain a CCI network by converting X into a list of triples. Each triple of that list is defined by $t = (s, r, w_{s,r})$, where s represents the sender cell, r the receiver cell and $w_{s,r}$ represents the weight in the interactions from cell s to r. Let $W_{s,r}$ be the set of all weights, where s is the sender and r is the receiver cell,

$$W_{s,r} = \{w | (s', g'_{lig}, r', g'_{rec}, w) \in X, s' = s, r' = r\}. \tag{1}$$

We define

$$w'_{s,r} = \sum_{x \in W_{s,r}} w \tag{2}$$

From these, we can define the weighted directed graph, where V_{CCI} is the set of cell types $\{1, .., k\}$ and $\mathbf{W'} = \{w'_{s,r}\}^{k \times k}$.

Differential GCI and CCI Networks

Given two *GCI* networks, $\mathbf{W^{exp}}$ (experimental condition) and $\mathbf{W^{cont}}$ defined over the same vertex space V, we can obtain a differential network by $\mathbf{X^{exp}} - \mathbf{X^{cont}}$. Here positive values indicate interactions with higher weights in the

experiment and negative values interactions with higher weights in the control. A similar operation is performed for CCI networks.

It is important to note that interactions predicted in the experiment X^{exp} and controls X^{cont} should be performed on integrated and normalized single cell RNA-seq experiments to guarantee comparability of weights and cell/gene spaces. An example of a valid integration and normalisation strategy using Seurat and based on the PMF data is found in our tutorial. In short, we used algorithms NormalizeData from Seurat package with the following parameters normalization.method = "LogNormalize" and scale.factor = 10000.

Implementation of Network measures

We calculated the follow state-of-art centrality measures implemented by igraph library: (1) degree (in/out); (2) betweeness centrality; (3) closeness centrality; (4) eigenvector centrality and (5) page rank. With the exception of degree measures, all statistics were calculated using the absolute value of the edges (e.g MeanLR). Because of the GCI presents, the measures were calculated for each component and normalized by its node size, the total number of component nodes.

PC analysis of network properties

CrossTalkerR employs PCA as a mean to combine distinct network measures. By default, it provides scores of two first PCs as a mean to rank either gene-cell pairs in GCI networks or cells in CCI networks. CrossTalkeR also provide importance of variables within the PCs to make then interpretable. For example, in the PMF data and GCI network (Sup. Figure 1), the first PC is associated to high Pagerank, Eigenvector and closeness, while the 2nd PC is associated to high in/out degree and between. We interpret that the first PC capture global interactions, while the second PC local interactions. Top up-regulated cell-gene pairs predicted by PC2 include matrix (FN1/MSC, COL1A1/MSC) and TGFB signalling associated genes (TGFB1/Megakaryocyte), as previously discussed.

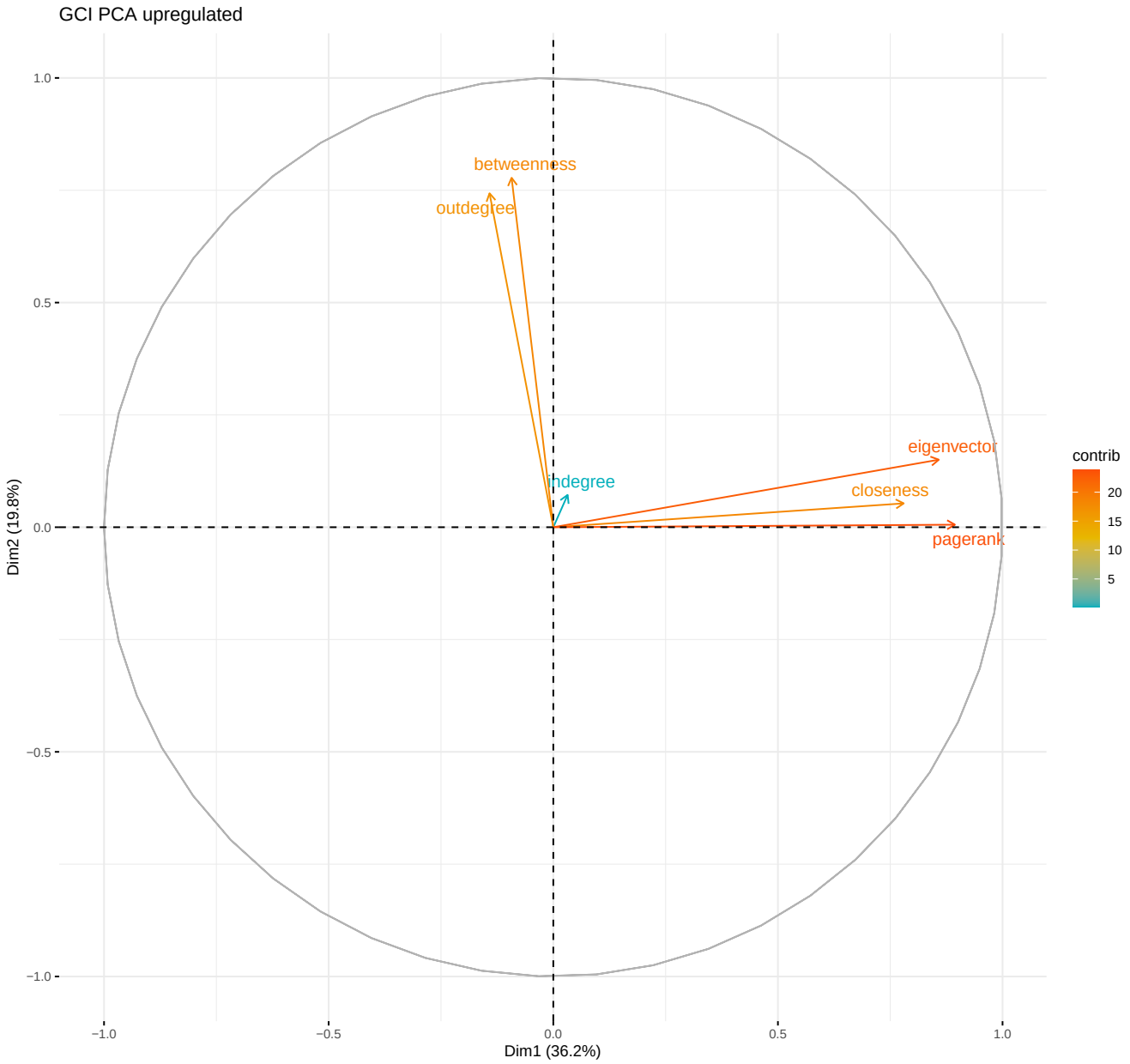

Figure 2. Weights of 2 first PCs of the GCI network from human Myelofibrosis.

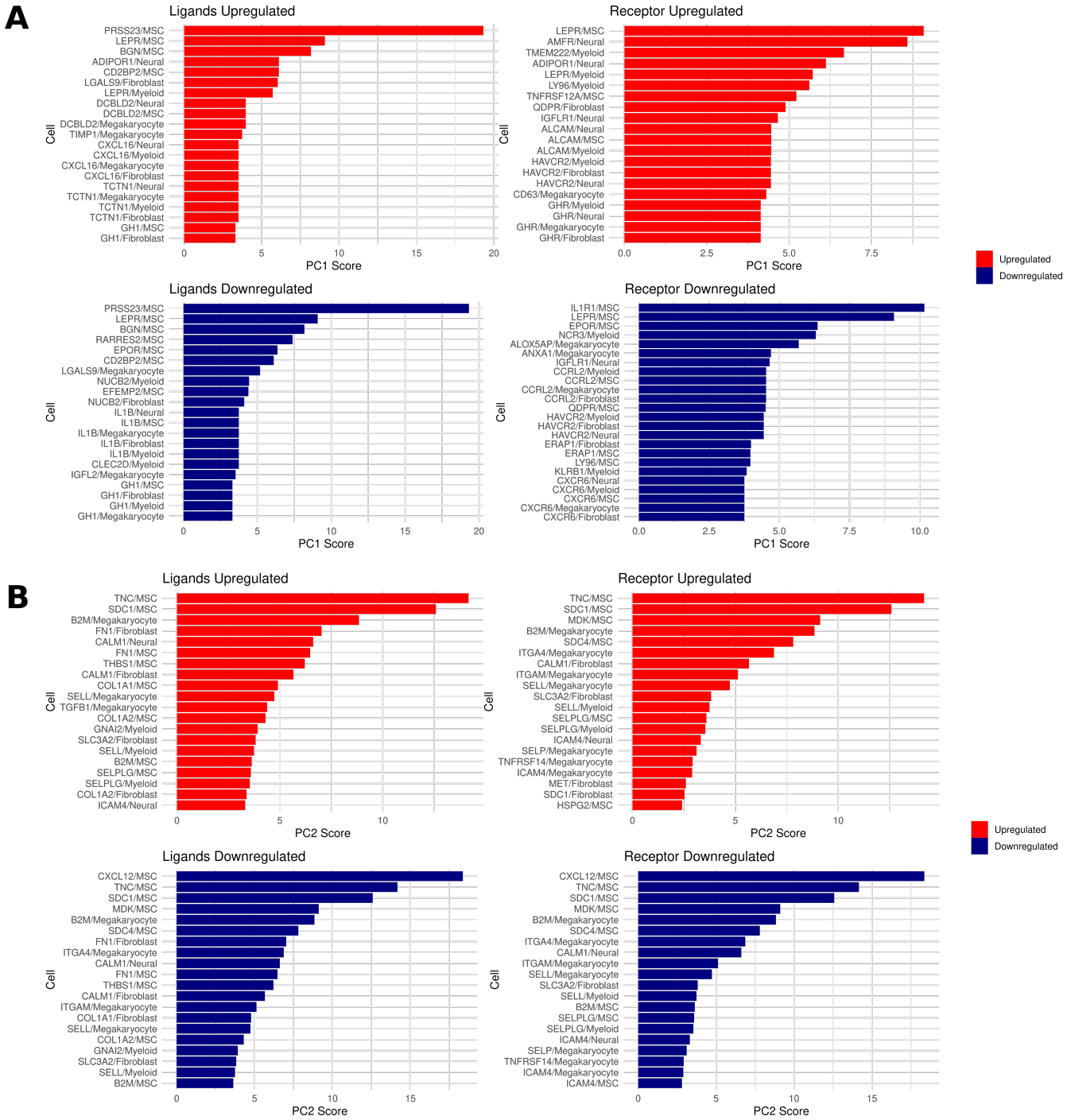

Figure 3. We shown the top 10 up (red) and down (blue) regulated ligand-sender cell (left) or receptor-sender cell pairs using either PC1 (A) and PC2 (B).