Identification of factors mediating the signaling convergence of multiple receptors following cell-cell interaction

Authors: Ron Sheinin[1,2], Shai Dulberg[2], Ayelet Kaminitz[2], Roded Sharan[1,*] and Asaf Madi[2,*]

Affiliate institutions:

[1]Blavatnik School of Computer Science, Tel Aviv University, 69978 Tel Aviv, Israel.

[2]Department of Pathology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Abstract

Motivation: Cell-cell crosstalk involves simultaneous interactions of multiple receptors and ligands, followed by downstream signaling cascades working through receptors converging at dominant transcription factors which then integrate and propagate multiple signals into a cellular response. Single-cell RNAseq of multiple cell subsets isolated from a defined microenvironment provides us with a unique opportunity to learn about such interactions reflected in their gene expression levels.

Results: We developed the FLOW framework with the intention of mapping the potential ligand-receptor interactions between different cell subsets based on a maximum flow computation in a network of protein-protein interactions (PPIs). The maximum flow approach further allows characterizing the intracellular downstream signal transduction from differentially expressed receptors towards dominant transcription factors. This, therefore enables the association between a set of receptors and their downstream activated pathways. Importantly, we were able to identify key transcription factors toward which the convergence of signaling of multiple receptors occurs. These identified factors have a unique role in the integration and propagation of signaling following cell-cell interactions.

Introduction

Cellular microenvironments are a complicated heterogeneous assembly of cells wherein their crosstalk often drives cellular response. Single-cell RNAseq technology provides us with a snapshot of this transcriptional heterogeneity.

Cell-to-cell interactions require the activation of a receptor by a physical connection of an appropriate ligand. However, it is not trivial to identify such a physical interaction using standard transcriptomic data. Thus, current approaches use heuristic methods to try and assess an "interaction potential" between the different cell populations. Indeed, various efforts have been made in recent years to achieve this goal (Ma et al. 2021), but the lack of gold standards makes the task of evaluating and comparing the different methods challenging.

Importantly, although cell-cell interactions may involve a single agent, in many cases the simultaneous interactions of multiple receptors/ligands are required. In addition, we hypothesize that in a given microenvironment, following cell-cell interaction, downstream signaling progresses via various receptors converge at dominant transcription factors (TFs) that integrate and propagate the multiple signals into a cellular response. To gain insight into potential interactions, we devised an algorithm, FLOW, that maps the expression of ligand/receptor pairs between different cell populations based on single-cell transcriptomes. We then use a framework of maximum flow in a protein-protein interaction (PPI) network to characterize the downstream pathways from a set of differentially expressed receptors toward dominant TFs.

We applied our method to a single cell RNAseq dataset from GL261 brain tumor model which contains both infiltrating and resident cells with some previously described interactions (Yeini et al. 2021). Our approach was able to highlight likely routes of signal transduction which are upregulated with the receptors, without being limited to a single pathway or an assumption on the size of the downstream pathway. The FLOW framework enables us to not only reveal key ligand-receptor pairs with association to the downstream activation signal transduction but also to pinpoint important transcription factors which aggregate the complex signal going through multiple receptors and pathways and propagate it into a cellular response (**Fig. 1**).

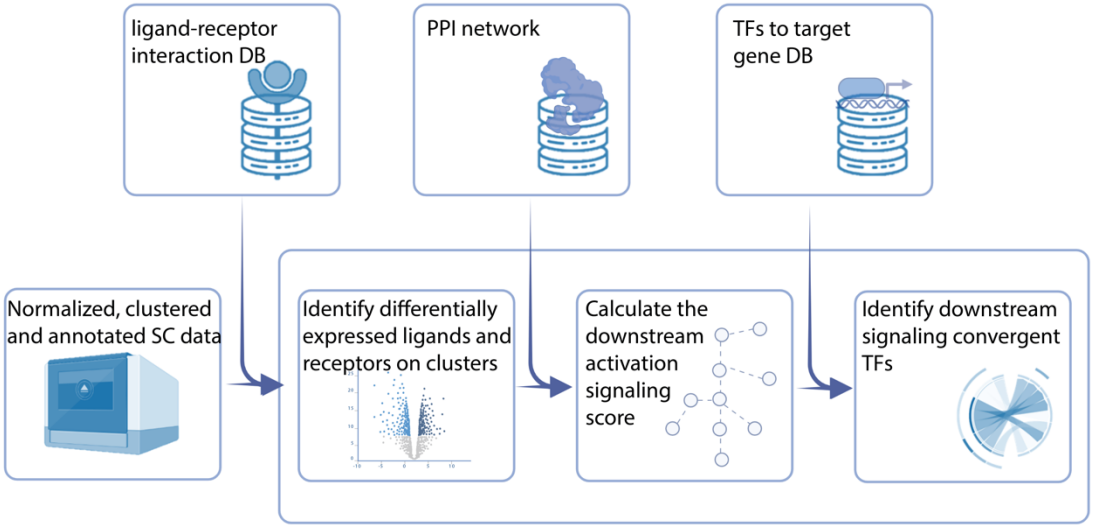


Figure 1. Workflow summary. The FLOW algorithm receives a normalized, clustered and annotated single-cell RNAseq dataset. For each pair of cell clusters, a signal sender and signal receiver clusters are defined and differentially expressed ligands and receptors are identified. To estimate the downstream activation signaling effect of each receptor, we calculate the maximum flow from each receptor to a set of transcription factors using the PPI network. Permutation tests are used to estimate the significance of the receptors and their signaling converging transcription factors. Finally, we calculate an average cluster-to-cluster interaction score and construct a global interaction map per data set.

Methods

Data processing, clustering and annotation

The pipeline accepts Seurat objects following data processing, normalization, clustering and annotation as was previously described (Hao et, al 2021). Briefly, the pipeline consists of the following steps. **LogNormalize**: each feature counts for each cell are divided by the total counts for that cell and multiplied by a scale factor. **Dimensionality reduction:** PCA and tSNE

are calculated from the scale normalized data matrix, where each feature normalized expression is scaled across the cells. The number of PCs for the clustering was manually selected based on an elbow plot showing the gain in variance with each additional vector. **Clustering:** First, we calculated the k-nearest neighbors and constructed the KNN graph, in the reduced PCA space. On that graph, a modularity score is optimized using the Leiden clustering method (Traag, Waltman, and van Eck 2019). **Cluster annotation:** was done manually by the use of known cell population markers and projection of known cell-type gene signatures on the tSNE plots.

Single-cell gene signature scoring

Single-cell gene signature scoring was used to emphasis the differential expression of ligands and receptors on interacting cell subtypes, as described previously (Kurtulus et al. 2019). Briefly, scores were computed by first sorting the normalized scaled gene expression values for each cell followed by summing up the indices (ranks) of the signature genes. A contour plot which takes into account only those cells that have a signature score above the indicated threshold was added on top of the tSNE space, to further emphasize the region of high-scoring cells.

Finding differential expressed receptor-ligands pairs

Ligand and receptor pairs were retrieved from the CellTalkDB database (Shao et al. 2020). As a first step, we identified receptors that are differentially expressed in the signal receiving cluster and their corresponding ligands which are differentially expressed in the signal sending cluster, using the Seurat package "FindMarkers" function. We applied the Wilcoxon Sum Rank with limit testing chosen to detect genes that display an average of at least 0.10-fold difference (log-scale) between the two groups of cells and genes that are detected in a minimum fraction of 0.10 in the upregulated group.

Threshold optimization

To find an optimal threshold, such that is balancing between the number of receptors and their activation score, we suggest plotting the mean downstream activation score (DSA) and the number of DE receptors per cluster for each threshold. Thus, the user can identify a range of thresholds that return a significant subset of receptor with relative high activation score, for all the clusters in the data.

Identifying activated transcription factors

To identify potentially activated transcription factors in each cluster we performed an enrichment test based on the Dorothea TF-gene target database (Garcia-Alonso et al. 2019). For a given gene target list we performed an unpaired Wilcoxon Sum Rank test between the rank distributions of the gene list and the rest of the gene expression vector to check if the mean expression of the genes in the gene list was taken from the same distribution as the background. The resulting p-values were further FDR corrected.

Calculation of downstream activation score (DSA)

In addition to the identification of receptors which are differentially expressed, we further calculated their potential downstream activation signaling. Our working assumption is that such a signaling cascade will be reflected in the transcriptomic profile of the cell, and will be likely to converge at a transcription factor downstream.

Firstly, we found transcription factors whose target genes are enriched in each cluster. To avoid the relatively large number of missing gene expression values associated with single cell data ("zero inflated" data) we calculated these parameters per cluster. For each cluster, we considered only those TFs which received an FDR-corrected p-value of 0.05 or lower. Next, we calculated a maximum flow from each receptor towards the enriched layer of transcription factors in a network of protein-protein interactions (Browaeys, Saelens, and Saeys 2019).

To increase our confidence in the identified signaling pathway, we further normalized the generated network. Each edge with a confidence score above 0.7 received a weight that represented the mutual information between the expression's distribution in the cluster of genes that made up the edge:

let (X, Y) be a pair of random variables, we denote $P(x, y)$ their joint distribution and P_X , P_Y the margin distribution of X, Y . The mutual information of $I(X{:}\,Y)$ is defined as:

$$I(X{:}\,Y) = \sum_{X}\sum_{Y} P_{X,Y}(x, y) \cdot \log\left(\frac{P_{X,Y}(x, y)}{P_X \cdot P_Y(y)}\right)$$

using the mutual information between each pair of genes in the cluster, we marked the edge weight as follows:

$$\forall < g_1, g_2 > \in E{:}\, weight(g_1, g_2) = \frac{I(e_1, e_2) - \min\limits_{<g_i, g_j>} I(e_i, e_j)}{\max\limits_{<g_i, g_j>} I(e_i, e_j) - \min\limits_{<g_i, g_j>} I(e_i, e_j)}$$

.

where $< g_1 , g_2 >$ is an edge is the PPI network, and e_1, e_2 are the expression vectors in the signal receiver cluster of g_1 , g_2 respectively $I(e_1{:}\, e_2\ )$ is the mutual information of e_1, e_2 , and $cs(g_1, g_2)$ is the reliability of the edge in the PPI network.Thus, the weight of a given edge reflects the co-expression of the genes in the cluster, as a proxy for their co-activation.

Following the generation of a weighted signaling network, we added a virtual node to the graph that represents the sink. Each of the identified TFs was then connected to the sink, with an edge weight of infinity. Each edge weight within the resulting normalized network represents its flow capacity. Next, Dinitz's algorithm (Dinitz, Y. 2006) was applied to find the maximum flow from each receptor via the signaling pathway toward the transcription factors ending at the virtual sink node.

To further assess the significance of the flow value, we performed random degree preserving permutation on the signaling network. For each permutation, we randomly shuffled the edges 10*|E| times, where |E| is the number of edges in the graph, such that each edge will be replaced by an edge with a similar weight. For that, we used the switching algorithm as discussed in (Biran et al. 2019). For each permutation we calculate the max flow, as in the original network, enabling us to provide an empirical statistical value that represents the significance of the observed flow (after FDR correction), compared to flow gained on random networks, while reducing the effect of node centrality in the network.

Generating a global interactions map

Finally, we calculated the cluster-to-cluster (c1, c2) interaction score for every pair of clusters in the data, using the next formula:

$$Score(c_1, c_2) = \sum_{r \in Receptors} \sum_{L \in Ligands(r)} DSA(r) \cdot \overline{e_{1r}} \cdot \overline{e_{2l}}$$

Where $Ligands(r)$ is the set of ligands in c_2 that corresponds to the receptor r , the $DSA(r)$ is the flow score of receptor r, and $e_{1r}$is the normalized expression of receptor r in cluster c_1 , and e_{2l} is the normalized expression of ligand l in cluster c_2 .

Time complexity

For each receptor, we calculated the maximum flow values using the Dinitz algorithm in a strongly polynomial time complexity of $O(V^2 \cdot E)$ where V is the number of nodes in the graph and E is the number of edges. Thus, the time complexity for calculating all flow values between a pair of clusters is $O(|Receptors| \cdot V^2 \cdot E)$. To calculate the permutation test we ran the Dinitz algorithm 1000 times per receptor.

Calculations of all the interactions between all pairs of clusters in the data took combinatorial time complexity in the number of clusters in the data set. To obtain a relatively effective performance on real-world data sets (commonly with a bounded number of clusters) our implementation of the framework used multiprocessor computing.

Result

Ligand receptor interactions in the tumor microenvironment

We applied our method to a single cell RNAseq dataset from GL261a murine brain tumor model which contained both infiltrating and resident cells with some previously described interactions (Yeini et al. 2021). As a first step, we identified differentially expressed receptor-ligand pairs. While this approach assured us that we are uncovering ligand-receptors pairs that best represent the interaction between those clusters, it excluded autocrine interactions or interactions between cells in the same cluster. For example, an analysis of the crosstalk between macrophages and T cell clusters identified 37 ligands and 39 corresponding receptors, receptively altogether forming 102 different receptor-ligand interactions (**Fig. 2**). As expected, the signal receiving cluster is seen to be strongly associated with the receptors while the signal sender cluster is seen to be associated with the corresponding ligands.

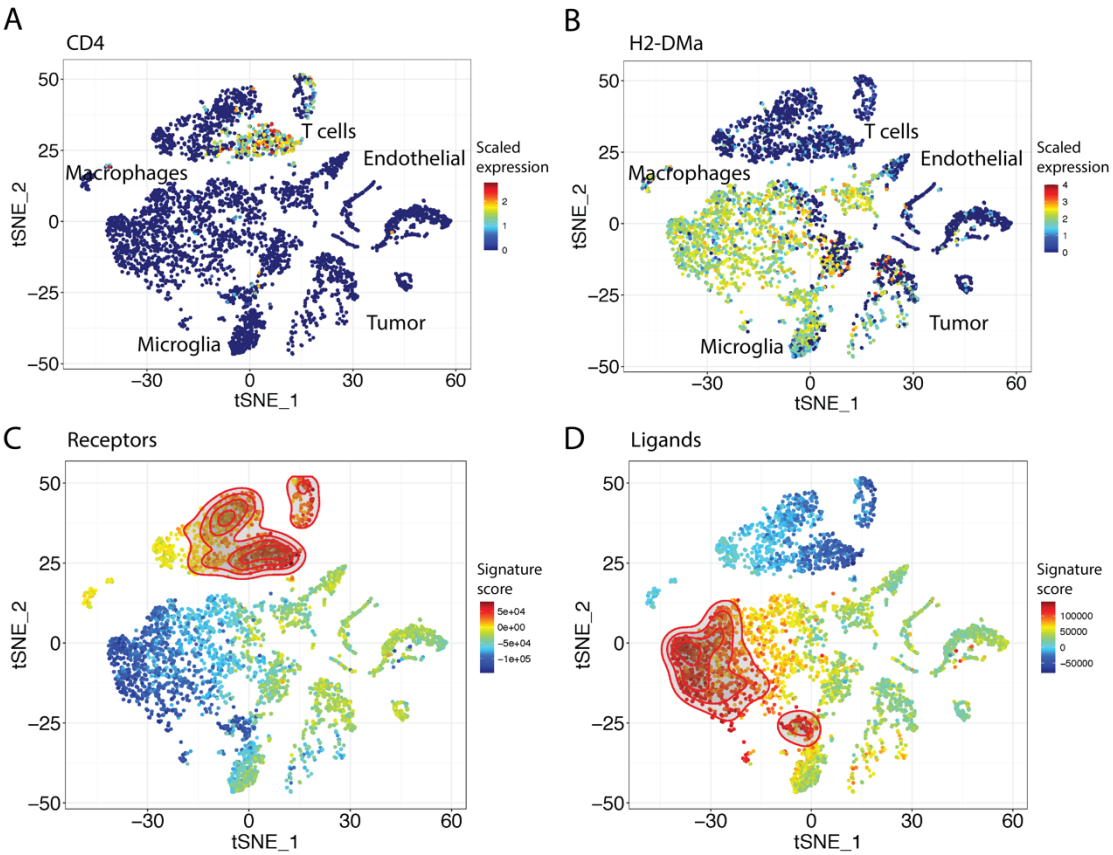


Figure 2. Identification of potential ligand-receptor pairs between macrophages and CD4 T cell clusters. (A-B) The normalized expression projection on the tSNE plot of potential interaction between the MHCII molecule (H2-DMa) on the macrophage cluster and its corresponding receptor CD4 molecule on the T cell cluster. (C-D) Signature projection of 39 receptors identified on the T cell cluster (left) and their 39 corresponding ligands identified on the macrophage cluster (right). The identified receptors are observed to be strongly associated with the T cell cluster while their corresponding ligands are associated with the macrophage and part of the microglia clusters.

Integrative visualization of microglia macrophage interactions

Following the identification of ligand-receptor pairs and further calculation of the downstream activation (DSA) score for each of the receptors, the algorithm generates a visualization plot which can be further interrogated for specific interaction between two clusters of choice. This plot provides the following information: the many to many ligand-receptor potential interactions, expression level, significance of differential expression, indication for the downstream signal activation and its statistical significance (**Fig. 3**).

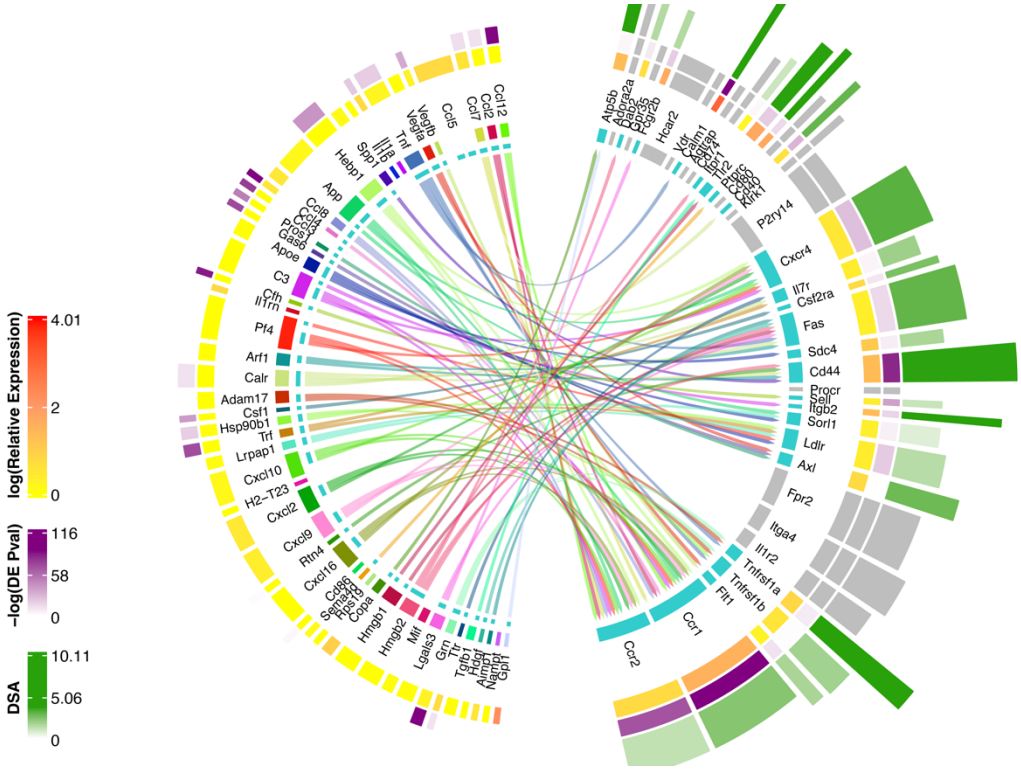


Figure 3. Ligand-receptor interaction with integrated receptor downstream activation signaling score. Detailed analysis of the differentially enriched ligand-receptor interactions between microglia and macrophage clusters. Inner lines indicate potential ligand-receptor connections and the width of the inner circle ribbon indicates the number of potential connections. The second circle ribbon reflects the expression level, the third outer ribbon indicates the p-value of upregulation compared to all other clusters in the dataset and the outer ribbon indicates the downstream activation score (DSA), receptors that did not show significant value in the permutation test are colored in grey.

Signaling converging transcription factors in macrophages

As part of the framework, FLOW highlights significant transcription factors that are likely to be activated by the signals received from multiple receptors. To this end, we applied the multi-source max flow algorithm, from all the receptors down to the defined set of transcription factors and calculated the amount of flow that is going through each of the transcription factors. Next, to rank the contribution of the different TFs, we calculated the importance coefficient of the TF (**Fig. 4A**). Specifically, we removed each of the TFs from the network, and recalculated the multi- sourced max flow in order to obtain the difference in the maximum flow in the network. A subnetwork that represents the flow pathways from multiple receptors toward a specific transcription factor can be retrieved from the original network, allowing us to further investigate a specific signal transduction pathway and detect potential hubs (**Fig. 4B**).

Global interactions in the tumor microenvironment

By running the previously described method on each pair of clusters in the data set we have generated a global map of interactions between all cell populations in the dataset (**Fig. 5A**). We hypothesized that generating such an interaction map could help us to better understand

the global dynamics between different cell populations in the data set, and uncover the key interaction axes common and cell-type specific to a studied biological system (**Fig. 5B**).

As can be observed in Figure 5B, there are almost no significant interactions between non-immune cell populations (endothelial, vascular, tumor and oligodendrocyte). However they do show relatively strong interactions with antigen-presenting myeloid cells (macrophages and microglia). In addition, we detected expected interactions between CD4 helper T cells and the antigen-presenting populations.

Finally, our global analysis can also be used to highlight common ligand-receptor interactions between different cell types. Indeed, we found many such interactions which are common to communications between multiple cell types.

Taken together, our global analysis was well adapted to capture some of the known cell-cell interactions in this microenvironment and had highlighted unique and more common interactions that must to be taken into account in perturbation applications.

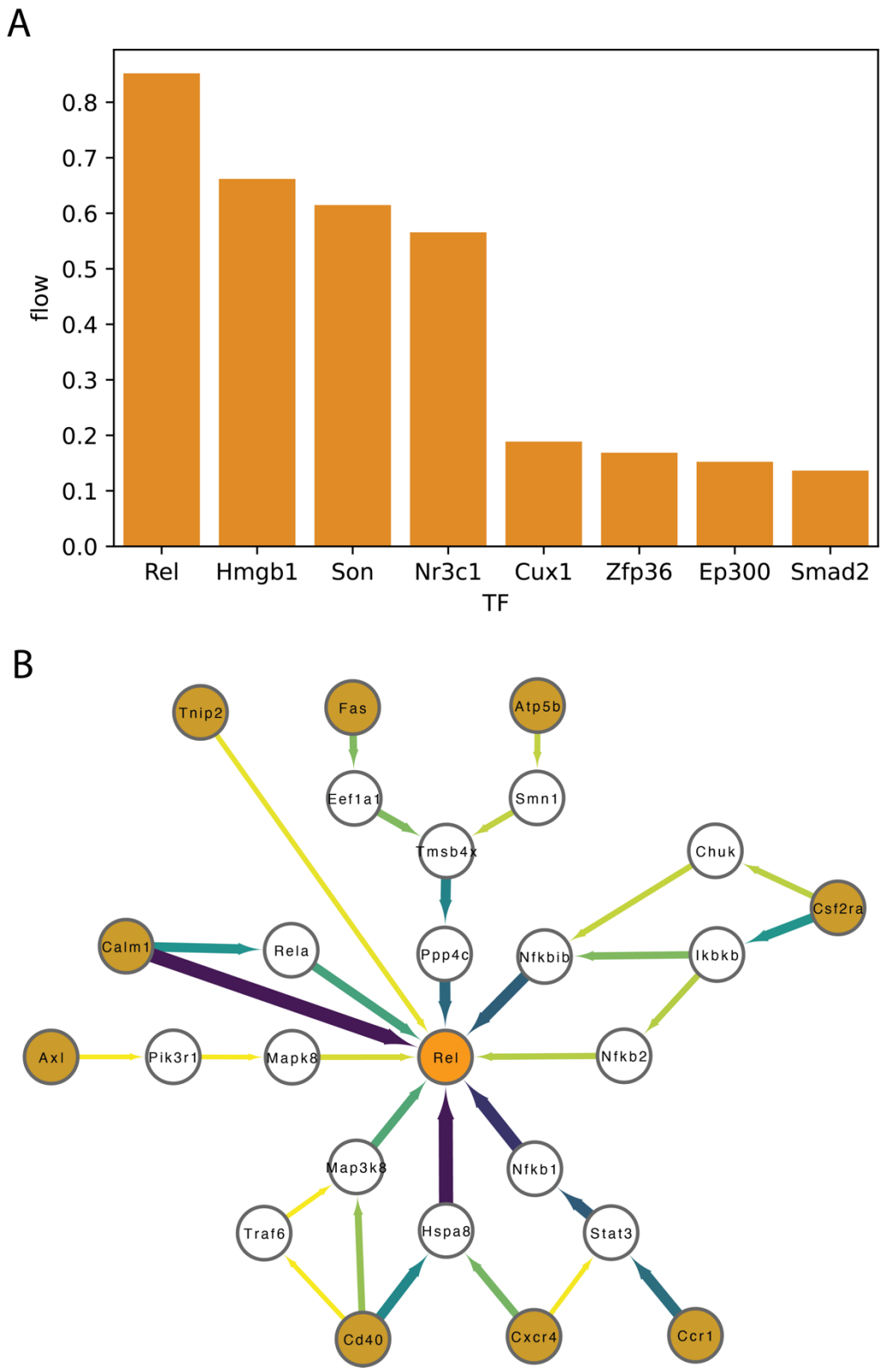


Figure 4. Identification of downstream signaling converging transcription factors. (A) bar plot showing the top-ranking TFs in the interaction between microglia and macrophage clusters shown in Fig. 3. (B) Subnetwork of the flow from multiple receptors (orange) to the Proto-Oncogene, NF-KB Subunit (Rel) transcription factor. The edge color and width represent the amount of flow that is passing through the edge as a proxy for the significance of the pathway in the subnetwork.

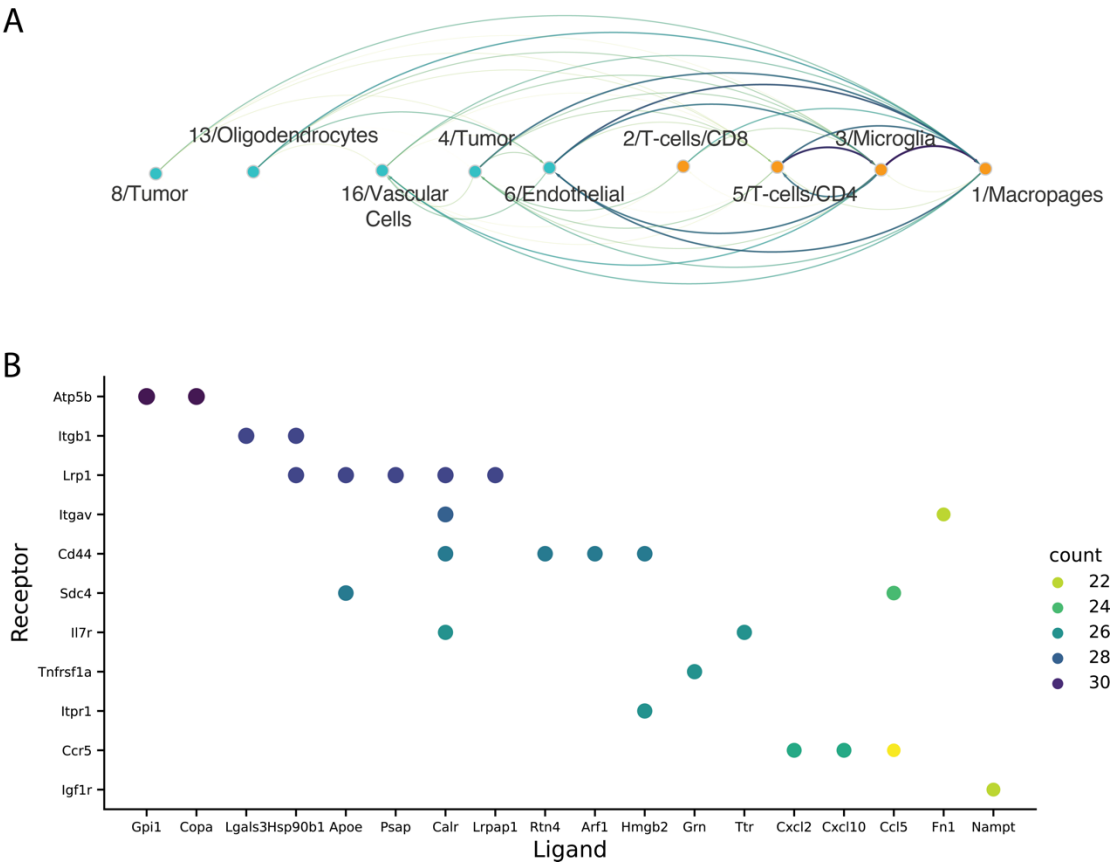


Figure 5. Global interaction map. (A) Global interaction plot, demonstrating interactions between the different clusters in the dataset. Edges radiate from the signal sender cluster to the signal receiving cluster, edge colors and width represent the strength of the interaction. (B) top 25 Ligand-receptor interactions that were identified as active between multiple cell types in the dataset.

Validation of identified receptors

To test our results, we used the recently published dataset (Yeini et al. 2021), which contains different cell populations from the microenvironment of GL261 murine brain tumor model.

As a first step, we checked if receptors that are up-regulated in a given cluster, are indeed associated with high flow values defined by our generated network. Thus, not only is the receptor itself upregulated but as is its downstream signaling pathway.

To this end, for each cluster, we define a set of receptors that are upregulated relative to all other clusters in the data set. We then defined an equally balanced list of receptors that are upregulated in the clusters and receptors which are not. For each receptor, we calculated the max flow value, and used these values to calculate ROC AUC of the classification of upregulated receptors (**Fig. 6**).

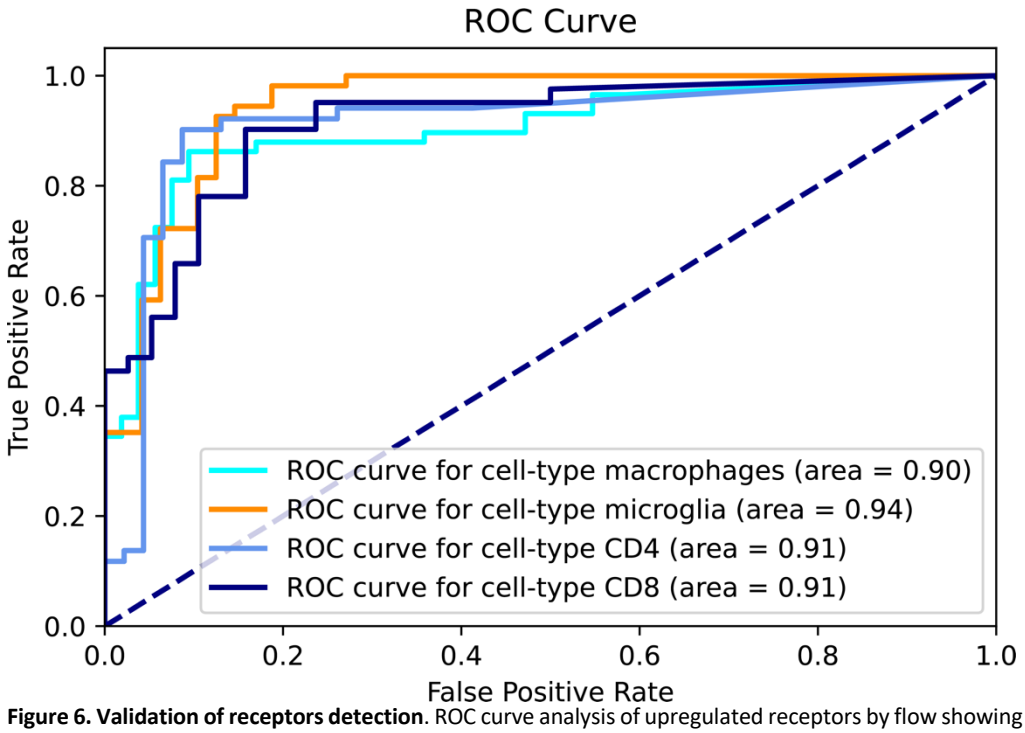


Figure 6. Validation of receptors detection. ROC curve analysis of upregulated receptors by flow showing that flow values are significantly associated with upregulated receptors.

As can be observed, the maximum flow value of a given receptor is strongly associated with the upregulation of the receptor itself, with an AUC value of 0.91 of the flow value. This further supports our hypothesis that the signal transduction pathway score can be used as a proxy for signaling via the receptor.

Validation of identified transcription factors

We hypothesized that in most cases, the interaction signal should flow toward a layer of transcription factors, which in turn will propagate the signaling into a cellular response. Therefore, we expected that transcription factors which are up-regulated in a given cluster, would also be associated with higher signaling flow value. However, without normalization, the importance value of each TF is strongly associated with the degree of centrality of the node in the PPI network. Thus, we normalized the flow value by the flow value generated in the unweighted PPI network (all capacities are one).

Indeed, we found that there is an association between up-regulated TF and the normalized flow value. Thus, to some extent the TF and its upstream signal transduction pathway are co-regulated. To further validate our result, we tested the correlation between the flow score of a transcription factor and the enrichment of its gene targets in the cluster. We assume that enrichment of the TF targeted expression is a strong indicator of TF activation. To this end, we compared the -log (Targets Enrichment P-Value) and the normalized flow score of all TFs in each cell type. We found a correlation of 0.3-0.4 for each of the immune cell populations (**Fig. 7B**). Thus, we can further conclude that our flow score contains some indication for the activation of a given transcription factor.

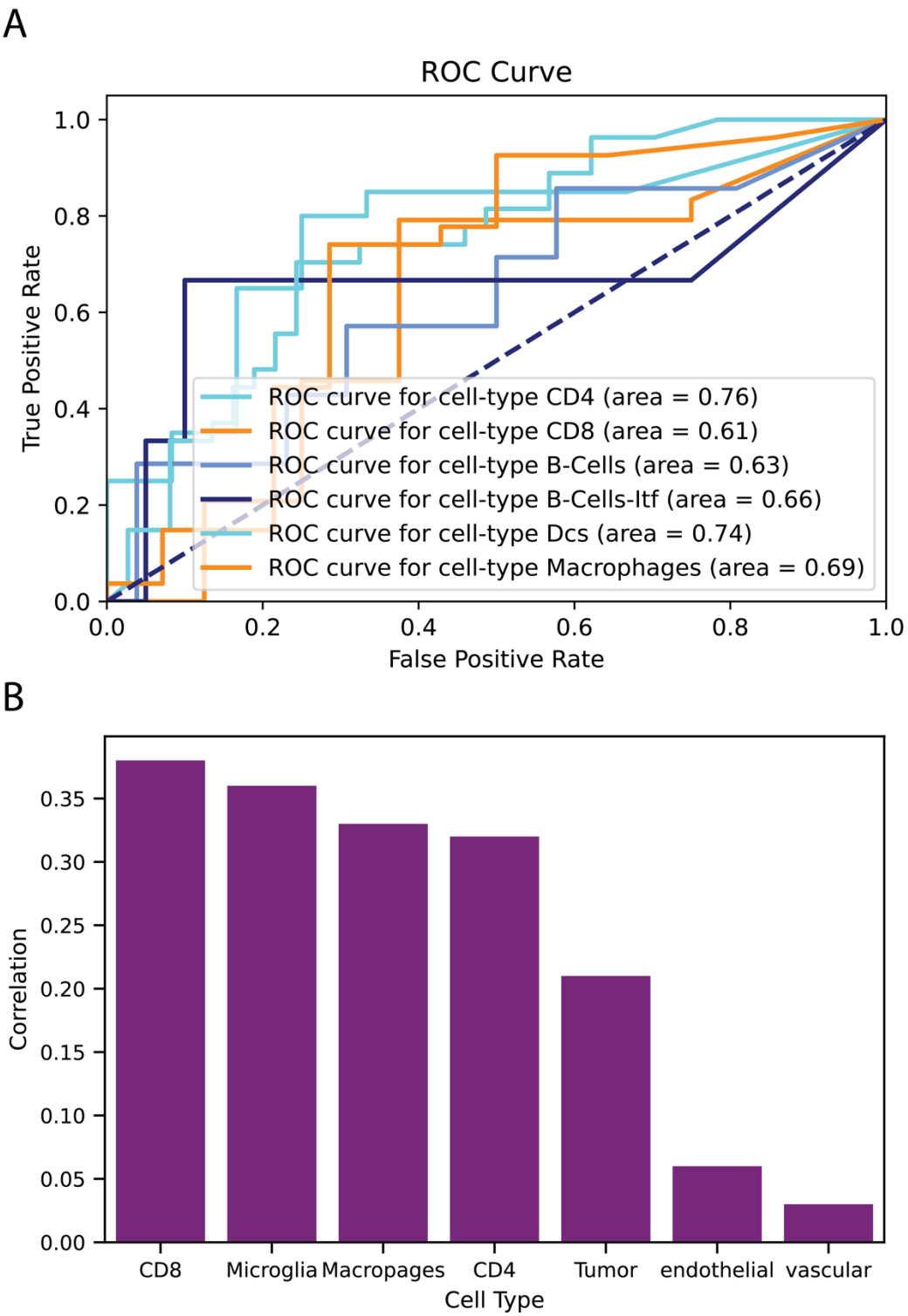


Figure 7. Validation of converging transcription factors. (A) ROC curve analysis of upregulated transcription factor by normalized flow score. (B) Correlation between TFs flow score and the target enrichment score per cell type.

As can be seen in Figure 7, there is a relatively high variance in the quality of performance for different cell populations. These could potentially be explained by the significant bias for commonly studied cell populations, especially in cell-specific TF and their targets. Thus, commonly studied cell populations are likely to have relatively larger known subset of TFs and their targets. Therefore, one can expect to observe a better performance of FLOW on those cell populations. Likewise, one can assume that not all cell populations in a biological system

are affected in the same manner from cell-to-cell interactions. Thus, those cells which are not a key interaction axis in the system will show a lower result from the FLOW analysis.

In the global analysis of the current dataset, we expect to find stronger interactions to or from immune compared to non-immune clusters. To quantify such a trend, we calculated the mean interaction score of

immune-related interactions (at least one of the clusters in the interaction represents immune cell type), to the mean score of non-immune interaction (**Fig. 8A**). As expected, the immune interaction score was indeed stronger.

To check the robustness of our framework, we ran the flow pipeline on the published Single Cell RNA seq dataset of a lymph node that was published at (R. Lopez et al. 2022), a summary of the results is as shown in Supplementary Figure 2.

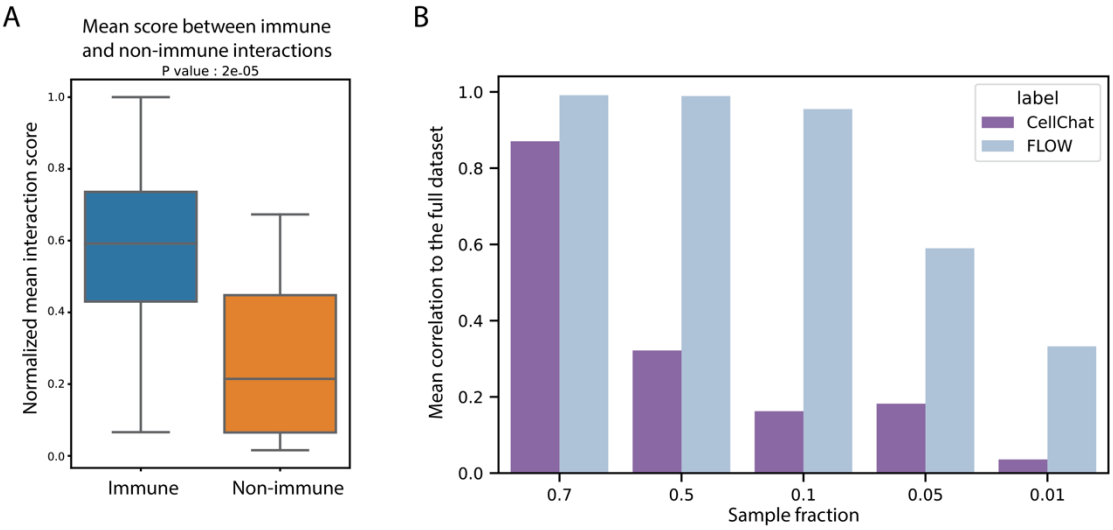


Figure 8. Validation of global immune and non-immune interactions and method robustness. (A) Normalized mean interaction score of immune cell types compared to interactions between non-immune cell types, p-value was calculated using t-test statistic. (B) Mean correlation of dataset sub-sampling to the full dataset scores.

Comparison to state-of-the-art methods

We compared our results to the recently published CellChat algorithm (Jin et al. 2021). As discussed previously, comparing the identification processes of clusters specific receptor-ligand pairs using different tools is a non-trivial task. Thus, one can only test the performance of those tools relative to a set of biological assumptions.

The CellChat provides cell-cell communication on a global scale, whereas we were aiming to see if the same biological trends can be observed using the CellChat toolkit, as we observed by FLOW. We generated all-to-all interactions using CellChat, and compared the Immune interactions to the non-immune interactions score. CellChat mean score of the immune interaction was significantly higher than the non-immune interaction (p-value = 0.03), which is thus in agreement with the FLOW method (**Fig. 8B**).

As we cannot compare the methods to a “gold standard” data set, we next compared the relationship between the methods’ scores. To this end, we compared FLOW to the score of

each receptor. We discovered a significant correlation of 0.4 between our flow scores and CellChat communication probability.

Furthermore, we compared our flow permutation score between receptors that were common to both methods and receptors that were not. We found a significant difference in the mean score value between the two groups (p-value = 0.006), and ROC AUC of 0.7.

Finally, we compared the robustness of the two methods. For that, we interactively sampled the data set in fractions of 0.7, 0.5, 0.1, 0.05, 0.01 and for each fraction, we calculated the correlations in interaction scores to the scores gained from the full dataset. We found that the FLOW was extremely robust with a drop in correlation observed only with a fraction size of 5%, while CellChat showed a significant drop in correlation at 50% sampling of the data.

Discussion

The task of quantitative modeling of cell-to-cell interaction based on transcriptomic data and single-cell transcriptomics specifically is heuristic by nature. First, gene expression data is not able to detect direct interaction between different cells, which leaves us only with the detection of the interaction potential. Second, gene expression also does not necessarily reflect gene activation. Third, the relatively low mRNA capture rate of the single-cell RNAseq technologies generates “zero-inflated” datasets, which makes it harder to detect such complex events per cell. Finally, the lack of “gold standard “datasets makes it extremely hard to compare different methods.

In this article, we propose a computational framework view of the maximum flow problem. Assuming that a significant proportion of the signals received by the cell via its receptors will propagate downstream towards a transcription factor. Provided the right ques, this transcription factor will in turn affect a vast number of genes. Thus, an analog to maximum flow in a network enables us not only to uncover important receptor-ligand pairs participating in cell-cell crosstalk, but also, to detect likely pathways by which the signaling occurs and the transcription factors that aggregate the signals from different receptors. Importantly, by providing clear start and end points for a signaling pathway, the suggested framework avoids making prior assumptions on the pathway length or limiting the analysis to a single pathway. This approach enables us to model complex interaction systems, such as the immune response in tumor microenvironment, across a relatively large subset of channels in each cell type.

Cell-cell interaction involves the activation of multiple receptors followed by complex intracellular signal transduction. These signals must be aggregated for the cell to reach a reaction decision thus highlighting the role of key transcription factors as the layer that aggregates the different signals into a cellular response is unique to our approach. Further characterization of transcription factors identified in a specific biological setting is, however, needed. Nevertheless, this approach opens new therapeutic avenues for the perturbation of a cellular response following cell-cell interactions.

In this article have shown that the amount of flow of a given receptor is associated with its upregulation in different cell populations. The signal flow was also found to be associated with the upregulation of the transcription factor predicted to further converge the signals. Thus, our interaction potential between two clusters includes not only the expression of the receptor and ligand but its entire signal transduction pathway.

On a global scale, our framework allows us to inspect the interaction map of an entire single-cell data set. Indeed, various microenvironments are characterized by multiple cell-type populations that constantly interact with each other. Such complex and dynamic interaction as well as a comparison between different states of the system could greatly benefit from such global analysis.

It is important to note, that the activation of a signaling pathway in biological systems usually includes also post-translational modification, which are not measured by any form of RNA seq data. Thus, any activation score that is based solely on RNA expression will never capture the entire pathway activation process.

Using data sets taken from a tumor microenvironment, we have shown that our method can reflect known dynamics in the data that agreed with known biological assumptions. To reduce the effect of the data sparsity associated with single-cell RNAseq most of our calculations are done at the cluster level. However, we believe that our suggested framework, with the proper normalization, may enable us to assign an interaction potential score to every single cell in the dataset, and by that reveal more complex biological dynamics inside the clusters. Combining such cell-specific scores with other single-cell analysis approaches will allow us to ask more complex questions, such as how the interactions can affect cell differentiation route, and understanding how cell-to-cell interactions change in different biological conditions.

Funding

A.M. was supported by The Alon fellowship for outstanding young scientists, Israel Council for Higher Education, from the Israel Science Foundation (1700/21), from the Israel Cancer Association (ICA, 01028753) and from the Israel Cancer Research Fund (ICRF) Research Career Development Awards. RS was supported by a research grant from the Israel Science Foundation (grant no. 715/18). Ron Sheinin was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University.

Conflict of Interest: none declared.

References

1. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. Cell. https://doi.org/10.1016/j.cell.2021.04.048

2. Dinitz, Y. (2006). Dinitz' algorithm: The original version and even's version. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/11685654_10.

3. Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports. https://doi.org/10.1038/s41598-019-41695-z.

4. Yeini, E., Ofek, P., Pozzi, S., Albeck, N., Ben-Shushan, D., Tiram, G., … Satchi-Fainaro, R. (2021). P-selectin axis plays a key role in microglia immunophenotype and glioblastoma progression. Nature Communications. https://doi.org/10.1038/s41467-021-22186-0:(Yeini et al. 2021)

5. Kurtulus, S., Madi, A., Escobar, G., Klapholz, M., Nyman, J., Christian, E., … Anderson, A. C. (2019). Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1 − CD8 + Tumor-Infiltrating T Cells. Immunity. https://doi.org/10.1016/j.immuni.2018.11.014

6. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., & Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Research. https://doi.org/10.1101/gr.240663.118

7. Ma, F., Zhang, S., Song, L., Wang, B., Wei, L., & Zhang, F. (2021). Applications and analytical tools of cell communication based on ligand-receptor interactions at single cell level. Cell and Bioscience. https://doi.org/10.1186/s13578-021-00635-z

8. Browaeys, R., Saelens, W., & Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nature Methods. https://doi.org/10.1038/s41592-019-0667-5

9. Shao, X., Liao, J., Li, C., Lu, X., Cheng, J., & Fan, X. (2021). CellTalkDB: A manually curated database of ligand-receptor interactions in humans and mice. Briefings in Bioinformatics. https://doi.org/10.1093/bib/bbaa269

10. Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., … Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nature Communications. https://doi.org/10.1038/s41467-021-21246-9

11. Van Der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research.

12. Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008

13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research.

14. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. https://doi.org/10.1038/s41592-019-0686-2

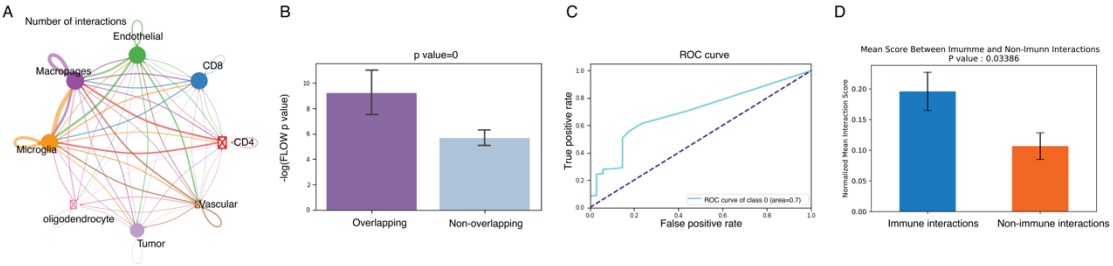
15. Teschendorff, A. E., & Feinberg, A. P. (2021). Statistical mechanics meets single-cell biology. Nature Reviews Genetics. https://doi.org/10.1038/s41576-021-00341-z

16. Biran, H., Kupiec, M., & Sharan, R. (2019). Comparative analysis of normalization methods for network propagation. Frontiers in Genetics. https://doi.org/10.3389/fgene.2019.00004

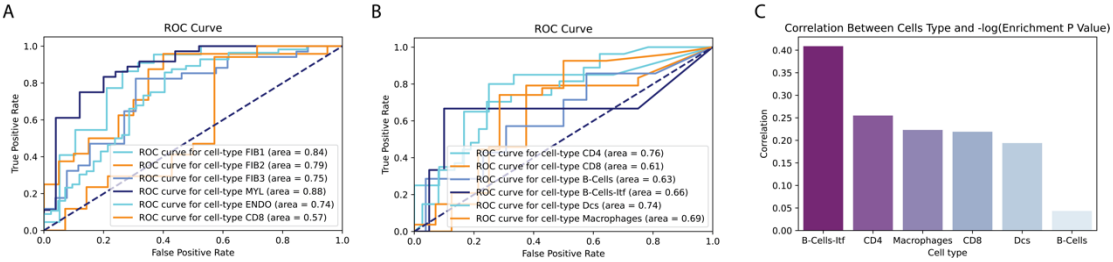
17. Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., … Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nature Communications. https://doi.org/10.1038/s41467-021-21246-9

18. R. Lopez*, B. Li*, H. Keren-Shaul*, P. Boyeau, M. Kedmi, D. Pilzer, A.Jelinski, E. David, A. Wagner, Y. Addadi, M.I. Jordan, I. Amit†, N. Yosef† Nature Biotechnology (accepted) , 2022. Multi-resolution deconvolution of spatial transcriptomics data reveals continuous patterns of inflammation

Supplementary Figures:



Supplementary Figure 1. Comparison to the CellChat method. (A) CellChat global interaction map. (B) Mean score of overlapping and non-overlapping interactions between the methods. (C) ROC AUC results for the overlapping interactions by the FLOW score. (D) Normalized mean interactions score of immune cell types compared to interactions between non immune cell types, p value was calculated using t test statistic.



Supplementary Figure 2. Validation on Lymph node dataset. (A) ROC curve analysis of upregulated receptors by FLOW. (B) ROC curve analysis of upregulated transcription factors by normalized FLOW score. (C) Correlation between TFs FLOW score and the target enrichment score per cell type.