NicheNet: Modeling intercellular communication by linking ligands to target genes

NicheNet is a new computational method to study intercellular communication from bulk and single-cell expression data. NicheNet predicts ligand-target links between interacting cells by combining their expression data with existing knowledge. The prior knowledge contained within NicheNet goes beyond ligand-receptor interactions and incorporates intracellular signaling and gene regulation as well. This makes that NicheNet can prioritize ligand-receptor interactions based on their gene regulatory effects on target cells. Moreover, NicheNet can be used to predict which target genes are affected by each ligand and which signaling mediators may be involved.

NicheNet was implemented as an R package available at github.com/saeyslab/nichenetr.

Context: why do we need another computational tool to study intercellular communication?

Intercellular communication is a crucial process in multicellular organisms. One way of studying intercellular communication is to profile the gene and/or protein expression in interacting cells. Single-cell and spatial -omics technologies in particular are promising because of their ability to analyze tissue composition in single-cell and/or spatial resolution.

The potential of single-cell transcriptomics to investigate intercellular communication has recently been demonstrated in several studies. In these studies, computational methods were applied to infer links between ligands (i.e., extracellular protein signals) expressed by a sender cell and their corresponding receptors expressed by a receiver cell. These computational approaches, such as CellPhoneDB, provide insights into intercellular communication processes because they elucidate which cells can communicate with each other and through which extracellular signals.

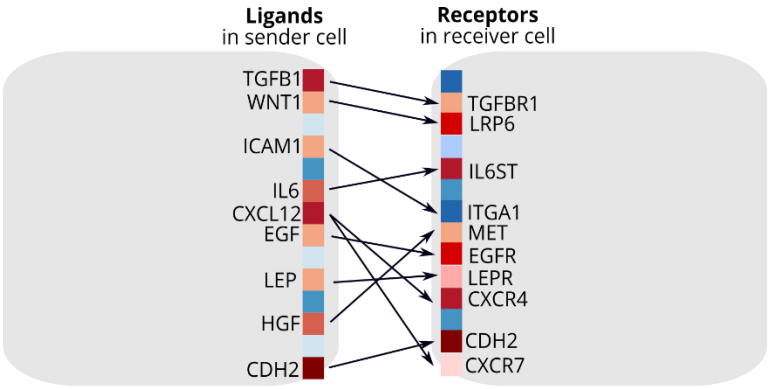

Figure 1: Overview of current ligand-receptor network inference methods

Nevertheless, functional understanding of a communication process also requires knowledge about how signals are interpreted intracellularly: the signaling pathways that get activated and the genes of which the expression changes as a consequence. Hence, there is a need for computational

methodologies that use expression data of interacting cells to infer the effects of ligands expressed by one cell on gene expression in another cell. To overcome this lack, we have developed NicheNet.

Differences between NicheNet and current tools that predict ligand-receptor interactions

Contrary to existing approaches, NicheNet looks at gene regulatory effects of ligands because the used prior knowledge goes beyond ligand-receptor interactions and incorporates intracellular signaling and transcriptional regulation as well. As a result, NicheNet allows to predict which ligands influence the expression in another cell, which target genes are affected by each ligand and which signaling mediators may be involved. By generating these novel types of hypotheses, NicheNet can drive an improved functional understanding of a cell-cell communication process of interest. The figure below summarizes the conceptual differences between most current ligand-receptor network inference approaches (top panel) and NicheNet (bottom panel) and visualizes the power of NicheNet in prioritizing ligand-receptor interactions based on gene expression effects.

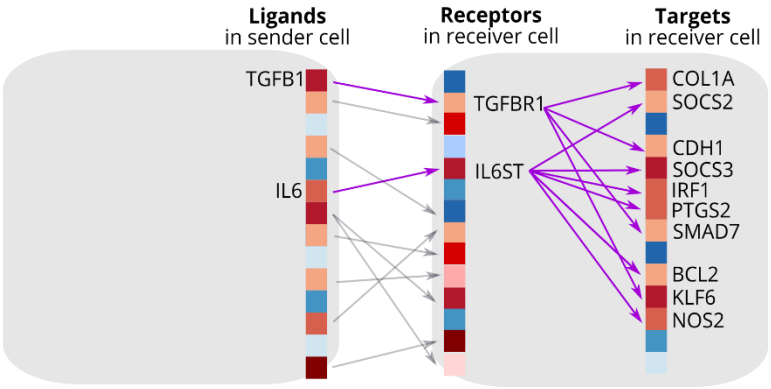

Figure 2: Overview of NicheNet, which prioritizes ligand-receptor interactions based on their affected target genes

NicheNet links ligands to target genes via network-based data-integration

NicheNet requires human or mouse gene expression data of interacting cells as input and combines this with a prior model that integrates existing knowledge on ligand-to-target signaling paths. Hereby, NicheNet can predict active regulatory links between ligands in sender cells and their corresponding affected target genes in receiver cells. A visualization of the workflow of NicheNet is shown below:

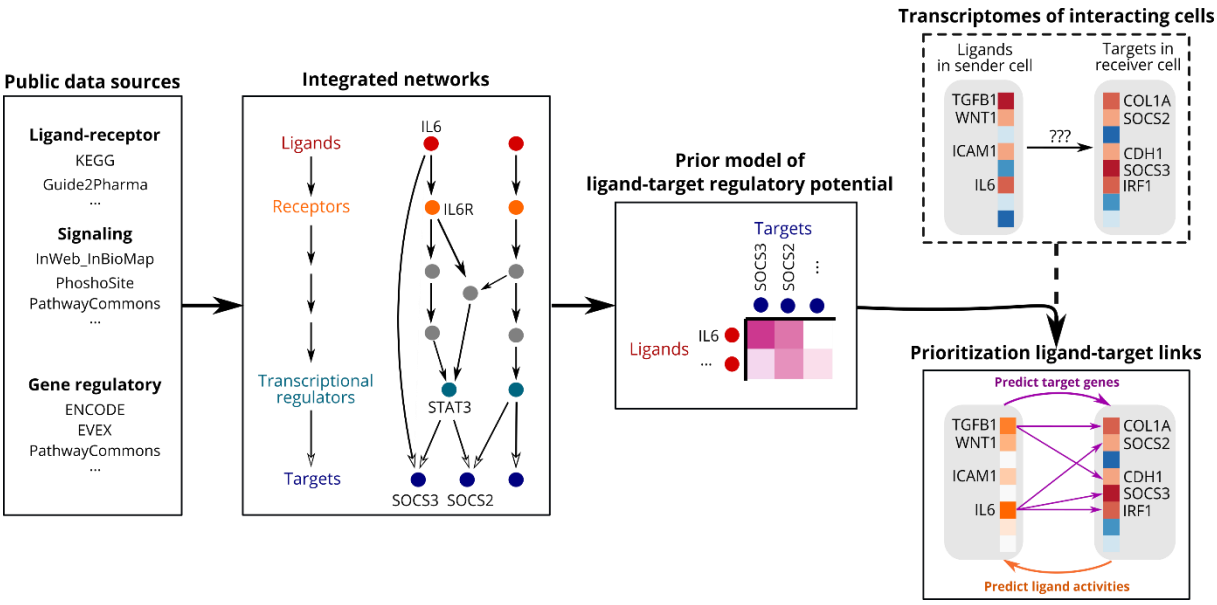

Figure 3: Workflow of NicheNet.

The prior model at the basis of NicheNet denotes how strongly existing knowledge supports that a ligand may regulate the expression of a target gene. To calculate this ligand-target regulatory potential, we integrated biological knowledge about ligand-to-target signaling paths as follows.

First, we collected multiple complementary data sources covering ligand-receptor, signal transduction (e.g., protein-protein and kinase-substrate interactions) and gene regulatory interactions (e.g., inferred from ChIP-seq and motifs.

Secondly, we integrated these individual data sources into two weighted networks: 1) a ligand-signaling network, which contains protein-protein interactions covering the signaling paths from ligands to downstream transcriptional regulators; and 2) a gene regulatory network, which contains gene regulatory interactions between transcriptional regulators and target genes. To let informative data sources contribute more to the final model, we weighted each data source during integration. These data source weights were automatically determined via model-based parameter optimization to improve the accuracy of ligand-target predictions.

Finally, we combined the ligand-signaling and gene regulatory network to calculate a regulatory potential score between all pairs of ligands and target genes. A ligand-target pair receives a high regulatory potential if the regulators of the target gene are lying downstream of the signaling network of the ligand. To calculate this, we used network propagation methods on the integrated networks to propagate the signal starting from a ligand, flowing through receptors, signaling proteins, transcriptional regulators, and ultimately ending at target genes.

This process is visualized below:

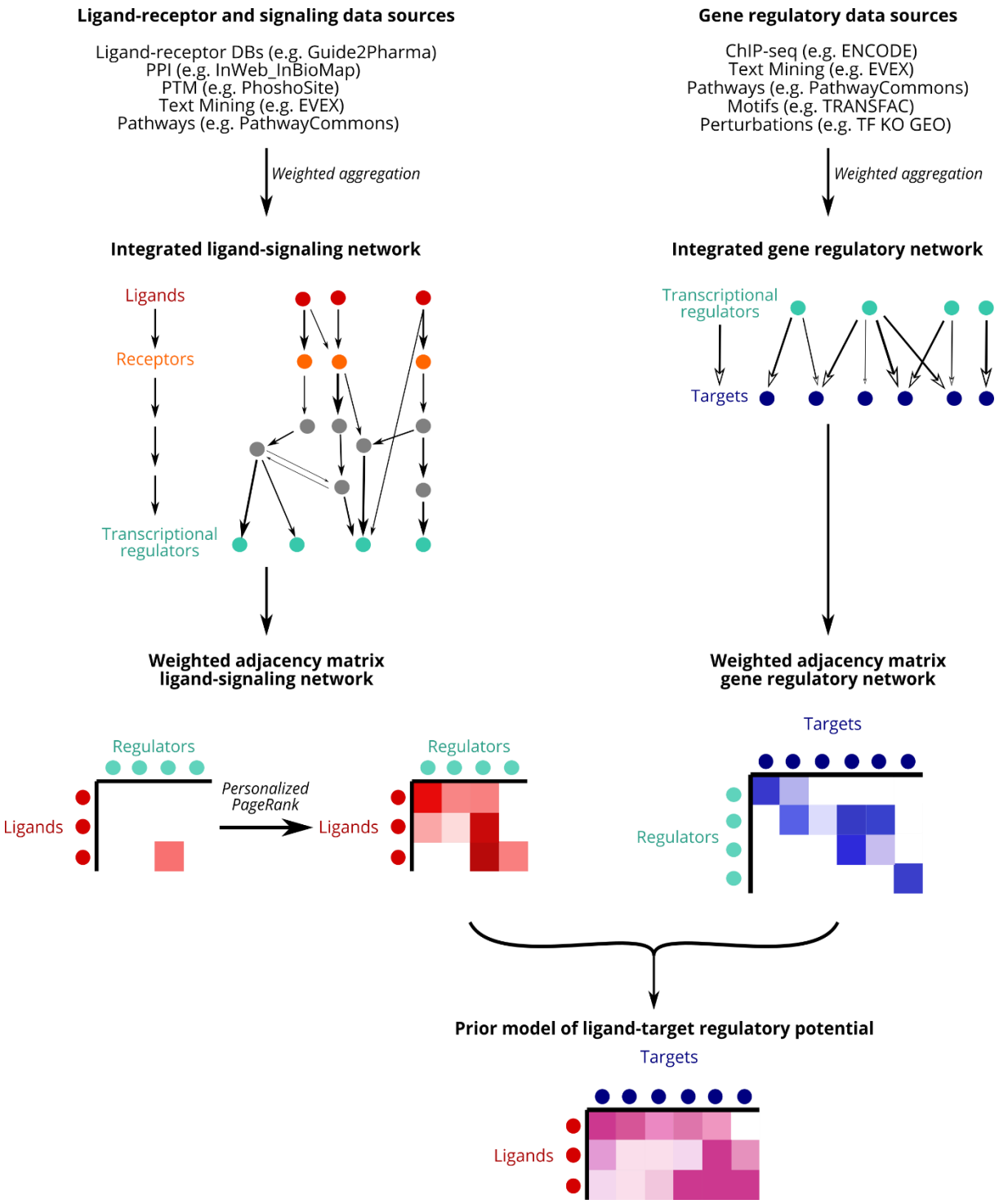

Figure 4: Overview of how the prior model of ligand-target regulatory potential is calculated

When applying NicheNet to investigate communication between interacting cells of interest, this general prior model of ligand-target regulatory potential is combined with the gene expression data of these cells. A first question NicheNet can address is predicting which ligands produced by a sender cell are the most active in affecting gene expression in the receiver cell. For this, NicheNet assesses how well these ligands predict the observed changes in gene expression and ranks them according to this. After ligand prioritization, NicheNet infers active ligand-target links by looking for genes that are affected in the receiver cell and have a high potential to be regulated by the prioritized ligands. Finally, users can visualize possible signaling paths between ligands and target genes of interest to analyze

why the model infers specific ligand-target links. This allows to prioritize signaling mediators and to check how the predictions are supported by all the collected data sources.

Evaluation of NicheNet's prior model

To validate the prior model of ligand-target regulatory potential, optimize the model parameters and compare different models, we established an evaluation procedure. For this procedure, we gathered public transcriptome data of several different cell types before and after they were treated by one or two ligands in culture. Using these ligand treatment datasets for validation has the advantage that observed gene expression changes can be directly attributed to the addition of the ligand(s). Hence, differentially expressed genes can be considered as a gold standard of target genes of a particular ligand. Note that this gold standard has some limitations, such as the popularity bias in ligand selection and the possible presence of secondary target genes. Moreover, the cell type and other context-specific factors have a strong influence on how a ligand affects the expression within a cell. As a consequence, not all possible target genes of a ligand are affected in a specific condition. The imperfect ligand treatment gold standard is therefore more useful to optimize parameters and compare different models than to indicate the exact predictive ability of NicheNet.

The ligand treatment datasets were used to define two complementary evaluation measures: target gene prediction performance and ligand activity prediction performance. To evaluate target gene prediction, we determined how well the model predicts which genes are differentially expressed after treatment with a ligand. To determine ligand activity prediction accuracy, we assessed how well the model predicts whether a ligand was added to cells or not. In other words, we evaluated how well NicheNet prioritizes ligands according to their potential to regulate a set of affected genes. This procedure is based on the following assumption: the better a ligand predicts the transcriptional response compared to other ligands, the more likely it is that this ligand is active.

We compared target gene and ligand activity prediction performances of the final optimized model to an unoptimized model and to models constructed from random networks. Parameter optimization was performed to maximize both target gene and ligand activity prediction accuracy. To assess the need for integrating multiple data sources, we also estimated performances of models that include fewer data sources, such as models built from only one comprehensive ligand-receptor, signaling and gene regulatory database. In conclusion, extensive data-integration and parameter optimization improved both target gene and ligand activity prediction performance, although even unoptimized models already performed considerably better than random network models (Fig. 5a-b).

Furthermore, we also benchmarked NicheNet's ligand activity prediction against Upstream Regulator Analysis of Ingenuity Pathway Analysis (IPA®) and CCCExplorer. IPA's Upstream Regulator Analysis is a state-of-the-art proprietary software tool that is built on a manually curated knowledgebase and allows to prioritize possible factors (including ligands) regulating a specific gene set. CCCExplorer combines signaling pathway information and bulk transcriptomics data to predict ligand-to-target crosstalk pathways between interacting cell types. We found that NicheNet and IPA outperformed CCCExplorer (Fig. 5b). Notably, NicheNet was much less biased than both IPA and CCCExplorer towards those ligands that are most frequently mentioned in the literature, with IPA consistently ranking well-studied ligands such as TNF, TGFB1 and IL1B higher than the actual ligand (> 40% of datasets) (Fig. 5c-e). This points to the added value of complementing validated but biased ligand-target links from the literature by more unbiased links inferred from integration of data sources covering interactions in ligand-to-target signaling paths.

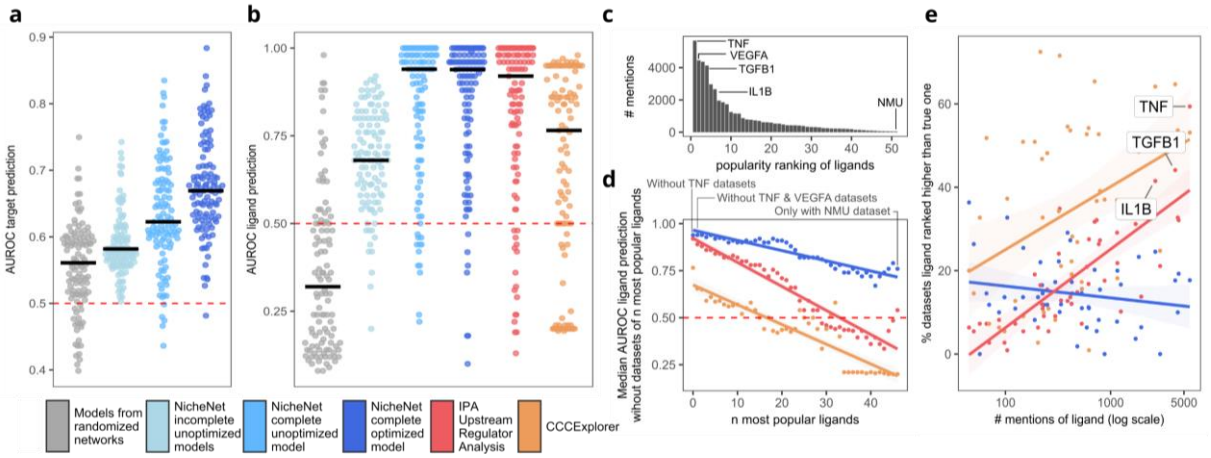

Figure 5: Evaluation of NicheNet's prior model. (b) Evaluation of target gene prediction performance: for 111 datasets profiling the transcriptional response of cells to a ligand (51 different ligands), we assessed performance in predicting which genes are differentially expressed upon treatment with a particular ligand. We compared performance between: 1) models constructed from 100 randomized networks; 2) 280 one-vs-one-vs-one models calculated from only one ligand-receptor, one signaling, and one gene regulatory database; 3) NicheNet without parameter optimization; 4) optimized NicheNet. For 1) and 2), we calculated the median performance for each dataset over respectively all 100 and 280 models. Each dot indicates the performance for one dataset, and the black line indicates the median performance over all datasets. (c) Evaluation of ligand activity prediction performance: for the same datasets, we assessed the performance in predicting the added ligand based on the differentially expressed genes. We compared performance between the models mentioned in (b), CCCExplorer, and IPA® Upstream Regulator Analysis. Visual representation of the performance per dataset is the same as in (b). (d) Ranking of all 51 profiled ligands in function of popularity (i.e., the number of studies in PubMed in which a ligand is described). (e) Analysis of popularity bias in ligand activity prediction performance by iteratively leaving out datasets of the n most popular ligands. (f) Analysis of popularity bias in the ligand rankings from the ligand activity prediction procedure.

Because the transcriptional response to ligands is cell-type dependent, we assessed whether the target gene and ligand activity prediction accuracy of NicheNet could be biased towards certain cell types. While we did observe substantial variation in performance between different cell types treated with the same ligand, there was no apparent bias towards particular cell types across different ligands (Fig. 6).

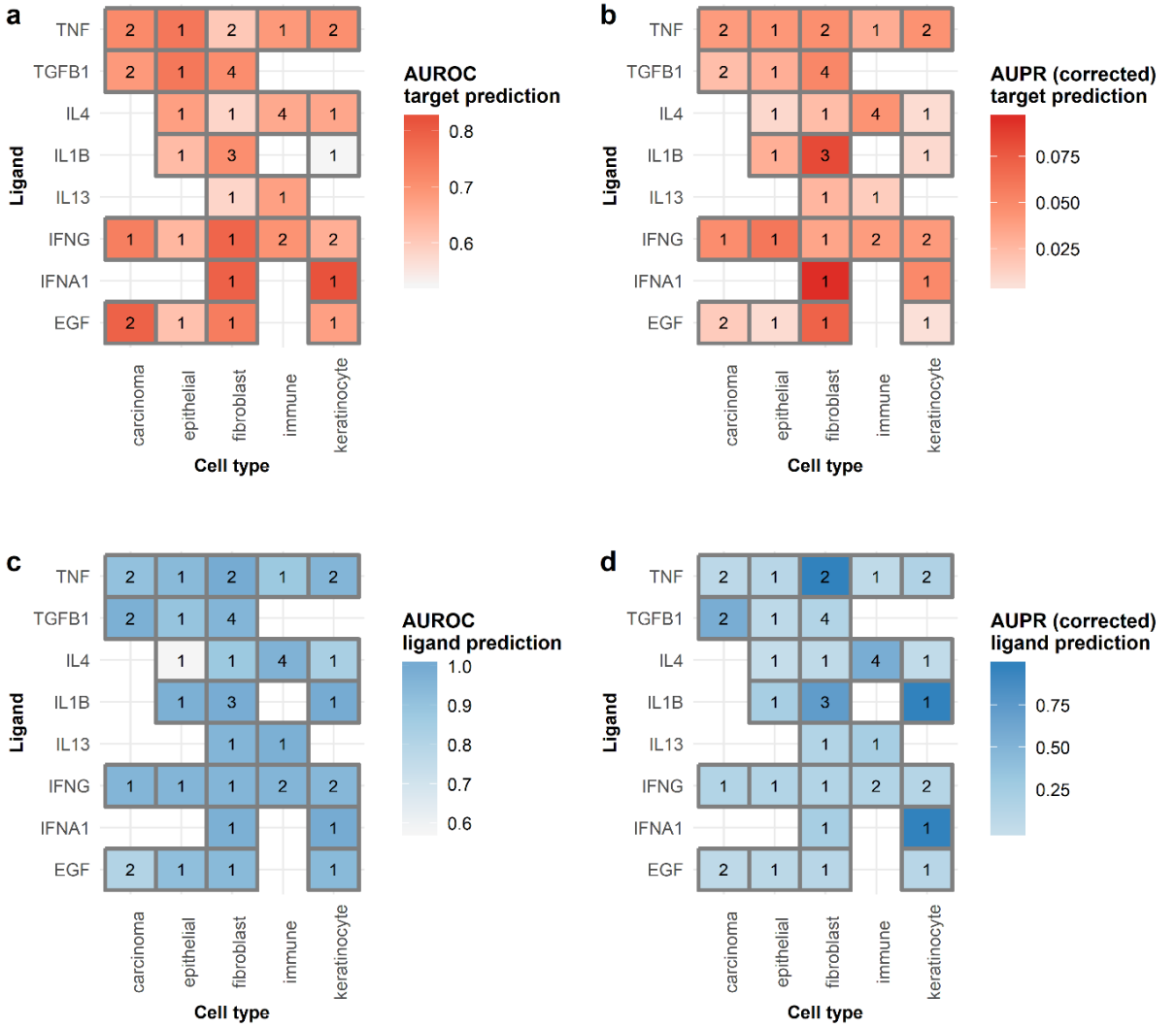

Figure 6: Evaluation of cell type bias in target gene and ligand activity prediction performance of NicheNet. To evaluate the possible presence of a cell type bias, we assessed the predictive performance of NicheNet on ligand treatment datasets that were only selected out of the total set of 111 datasets if they measure the response to a ligand for which the response was profiled in at least two different cell types that were stimulated by at least two different ligands. For the 43 selected ligand treatment expression datasets, we show the target gene (a-b) and ligand activity (c-d) prediction performances in function of the ligand added to the cells and the cell type that was stimulated. For each ligand-cell type combination, we also indicate the number of analyzed datasets. When multiple datasets were analyzed for a specific ligand-cell type combination, the average performance is shown on the heatmap.

Application of NicheNet on datasets of interacting cells

In the following sections, we show how NicheNet can be used to prioritize putative ligand-target links between interacting cells in tumor and immune cell microenvironments.

How do fibroblast-ligands regulate a p-EMT program in neighboring malignant cells in head and neck squamous cell carcinoma?

In the first case study, we applied NicheNet to human single-cell RNA-seq data that was generated by Puram et al. to study the malignant and non-malignant cell composition in head and neck squamous cell carcinoma (HNSCC) tumors. Puram et al. observed that several specific gene programs in malignant cells were heterogeneously expressed within and between tumors. One such gene program was the partial epithelial-to-mesenchymal transition (p-EMT) program, which could be linked to metastasis. Interestingly, the subset of malignant cells that express a p-EMT program were found to be located at the edge of tumors close to cancer-associated fibroblasts (CAFs), pointing to possible interactions between both cell types. Complementary to this observation, Puram et al. were able to computationally infer many potential ligand-receptor interactions between CAFs and malignant cells. Notably, some of these CAF-ligands, such as TGFB3 and CXCL12, are known to regulate classic EMT. Because of all these observations, they hypothesized that the expression of p-EMT genes could possibly be regulated by CAF-ligands. To support this hypothesis, Puram et al. experimentally assessed and validated the role of the CAF-ligand TGF-β in promoting p-EMT in vitro. They decided to validate the role of TGF-β because of its well-studied role in classic EMT and the presence of its well-known target gene TGFBI in the p-EMT program.

We applied NicheNet to extend on this ligand-receptor analysis and further investigate the hypothesis that CAF-ligands may regulate the p-EMT program in neighboring malignant cells. First, we performed ligand activity prediction to do a data-driven prioritization of the most probable p-EMT-regulating ligands based on how well they predict the entire p-EMT gene set (Fig. 7c). Noteworthy, the by Puram et al. experimentally validated factor TGFB3 was among the top-ranked ligands. Moreover, a role in classic EMT has been described previously in other studies for 18 of the 20 top-ranked ligands.

Subsequently, we assessed to what extent the p-EMT program may be regulated by CAF-ligands. For this analysis, we determined how well the 20 top-ranked CAF-ligands predict which genes belong to the p-EMT program. We found that 25% of the p-EMT genes were among the 5% most strongly predicted target genes, compared to 4.7% of non-p-EMT genes (enrichment p-value from Fisher's Exact Test: 2.5 x 10-10). This suggests that top-ranked CAF-ligands might indeed regulate a substantial set of p-EMT genes. Therefore, we then inferred specific regulatory interactions that are likely active between CAF-ligands and p-EMT target genes in malignant cells (Fig. 7c). This inference of active ligand-target links allows to generate more functional hypotheses concerning the specific p-EMT-regulatory role of CAF-ligands than limiting the analysis to ligand-receptor network inference.

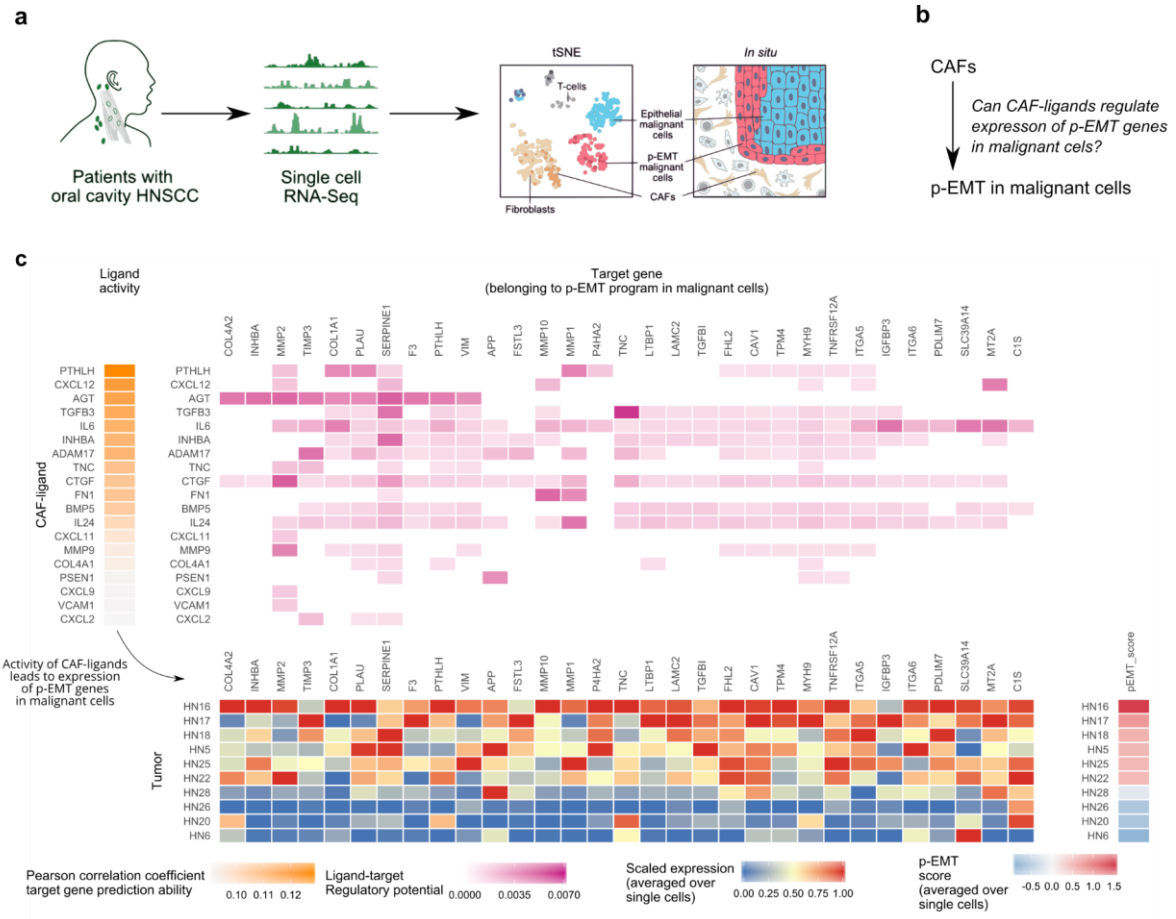

Figure 7: NicheNet prioritizes cancer-associated fibroblast ligands based on their potential to regulate p-EMT genes in adjacent malignant cells in head and neck squamous cell carcinoma. (a) Puram et al. performed single-cell RNA-seq of head and neck squamous cell carcinoma (HNSCC) tumors from different patients. Malignant cells expressing a partial epithelial-to-mesenchymal transition (p-EMT) gene program are located at the leading edge of primary tumors and are likely interacting with cancer-associated fibroblasts (CAFs). Figure adapted from ref. 9. (b) Here, we applied NicheNet to single-cell RNA-seq data of malignant cells and CAFs from 10 tumors to investigate how CAF-ligands might promote the p-EMT program in neighboring malignant cells. (c) Top left: outcome of NicheNet's ligand activity prediction on the p-EMT gene set: results are shown for the 20 (out of 131) CAF-ligands best predicting the p-EMT gene set. As the ligand activity ranking metric, we used the Pearson correlation coefficient between prior regulatory potential scores and p-EMT gene set assignments (n = 6072 genes, of which 96 belong to the p-EMT program). This Pearson correlation indicates the ability of each ligand to predict the p-EMT target genes, and better predictive ligands are thus ranked higher. Top center: NicheNet's ligand-target matrix denoting the regulatory potential between CAF-ligands and target genes from the p-EMT program. Bottom center: scaled expression of p-EMT genes in malignant cells, averaged per tumor. Bottom right: each cell was scored for expression of genes belonging to the p-EMT program, and this score is averaged per tumor.

Importantly as well, NicheNet enables the user to check the validity of predicted ligand-target links by providing the underlying signaling interactions and the data sources that support these interactions. We demonstrated this functionality by inferring signaling paths between TGFB3 and some of its top-predicted p-EMT target genes TGFBI, LAMC2 and TNC (Fig. 8). Interestingly, an important mediator in this predicted signaling network was SMAD3, a transcription factor involved in classic EMT.

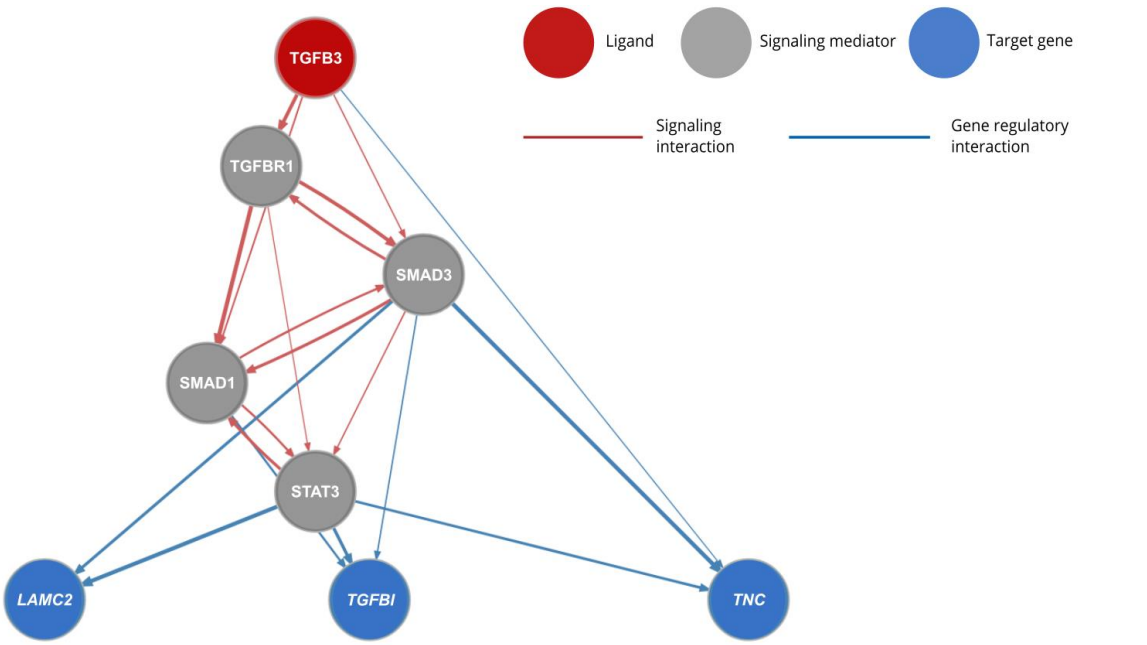

Figure 8: Network showing high-confidence interactions forming putative signaling paths between the ligand TGFB3 and its predicted target genes TGFBI, LAMC2 and TNC. From NicheNet's weighted integrated signaling and gene regulatory networks, we inferred the most important signaling mediators and transcriptional regulators involved in the signaling from TGFB3 to TGFBI, LAMC2, and TNC. Shown here is the network of interactions (as documented by NicheNet's data sources) between ligand, signaling mediators, transcriptional regulators, and target genes. The ligand node is indicated in red, the target gene nodes in blue and nodes of signaling and transcriptional regulators in grey. Edge line thickness is proportional to the weight of the represented interaction in the weighted integrated networks. Edges representing signaling interactions are colored red, gene regulatory interactions blue.

The table below demonstrates that you can check the individual data sources underlying each interaction in this network:

from	to	source
SMAD1	TGFBI	regnetwork_source
SMAD1	TGFBI	Remap_5
SMAD3	LAMC2	harmonizome_CHEA
SMAD3	LAMC2	harmonizome_TRANSFAC_CUR
SMAD3	LAMC2	pathwaycommons_controls_expression_of
SMAD3	TGFBI	harmonizome_CHEA
SMAD3	TNC	harmonizome_CHEA
SMAD3	TNC	regnetwork_source
SMAD3	TNC	trrust
SMAD3	TNC	Remap_5
SMAD3	TNC	pathwaycommons_controls_expression_of
STAT3	LAMC2	harmonizome_CHEA
STAT3	LAMC2	harmonizome_ENCODE
…	…	…

Which signals from hepatic stellate cells, liver sinusoidal endothelial cells and hepatocytes imprint a Kupffer cell identity on engrafted monocytes?

Check following publication from Bonnardel, T'Jonck et al. to see how we used NicheNet to predict upstream niche signals driving Kupffer cell differentiation Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche.

When is it possible to apply NicheNet at your data and when not?

NicheNet studies how ligands affect gene expression in neighboring cells. This makes that you need to have data about this effect in gene expression you want to study. So, you need to have a clear set of genes that are putatively affected by ligands from one of more interacting cells. For example, this were the genes from the p-EMT program in the first discussed example, and genes that are differentially expressed during niche-guided monocyte-to-Kupffer cell differentiation in the second example.

So to perform a NicheNet analysis, you need to have:

- A set of genes that are putatively affected by interacting cells
- Expression data of the interacting cells to define the list of possible ligands
    - All ligands expressed in sender cells, for which a corresponding receptor is expressed in the receiver cell as well

If you would only have "steady-state" data from which you cannot deduce which genes are regulated due to intercellular communication processes, it is not possible to apply NicheNet.

Conclusion

NicheNet allows to prioritize ligands based on their predicted effects on genome-wide expression, to infer their target genes that are affected in the interacting cells, and to generate hypotheses about potential underlying signaling mechanisms between them and their target genes. Whether prioritized ligand-target links are important in the process of interest should be experimentally validated, certainly because the wiring of cellular signaling can be condition and cell-type specific. Including cell-type-specific signaling and gene regulatory networks (e.g., inferred from ATAC-seq data) within the modular NicheNet framework could presumably improve the accuracy of context-specific ligand-target predictions. This is an exciting topic for future research and the provided software includes the ability to personalize the data sources within the NicheNet framework. Better predictions could also result from using expression data that provides information about the spatial context or dynamics of intercellular interactions. The former type of data can give a better definition of truly interacting cells, whereas the latter allows taking into account the time delay between the expression of ligands in sender cells and their transcriptional response in receiver cells. Extending the NicheNet framework in the future to fully exploit these types of data could even further advance computational ways to study the functional effects of cell-cell communication networks in health and disease states.

In conclusion, we expect that NicheNet will be a useful tool to better study the functional effects of cell-cell communication networks in health and disease states.