

# NicheNet: modeling intercellular communication by linking ligands to target genes

Robin Browaeys  <sup>1,2</sup>, Wouter Saelens  <sup>1,2,3</sup> and Yvan Saeys  <sup>1,2,3\*</sup>

**Computational methods that model how gene expression of a cell is influenced by interacting cells are lacking. We present NicheNet (<https://github.com/saeyslab/nichener>), a method that predicts ligand-target links between interacting cells by combining their expression data with prior knowledge on signaling and gene regulatory networks. We applied NicheNet to tumor and immune cell microenvironment data and demonstrate that NicheNet can infer active ligands and their gene regulatory effects on interacting cells.**

Cells are influenced by extracellular signals produced by cells in their microenvironment. One way of studying intercellular communication is to profile gene expression in interacting cells. Especially single-cell-omics technologies are promising because of their ability to analyze tissue composition in high resolution<sup>1,2</sup>. Furthermore, spatial and sophisticated single-cell transcriptomics technologies have recently been developed to assay the spatial context of a cell<sup>3–6</sup>, characterize cellular niches<sup>7</sup> or construct physical cell–cell interaction maps<sup>8</sup>.

Several studies have recently demonstrated the potential of single-cell transcriptomics to investigate intercellular communication<sup>9–14</sup>. There, computational methods, such as CellPhoneDB<sup>14</sup>, were applied to infer links between ligands (that is, extracellular protein signals) expressed by sender cells and receptors expressed by receiver cells. Thus, these computational approaches elucidate by which extracellular signals cells can communicate. Nevertheless, a functional understanding of cell–cell communication requires knowledge about the influence of these ligand–receptor interactions on target gene expression. Hence, there is a need for computational methodologies that use expression data of interacting cells to infer the effects of sender-cell ligands on receiver-cell expression. Although a few basic methods were developed for bulk data (for example, CCCExplorer)<sup>15,16</sup>, comprehensive and predictive methods that are generally applicable to both bulk and single-cell expression data are lacking.

To address this problem, we have developed a computational method called NicheNet (<https://github.com/saeyslab/nichener>). NicheNet requires human or mouse gene expression data of interacting cells as user input and combines this with a prior model built by integrating prior knowledge on ligand-to-target signaling paths (Fig. 1a). Contrary to existing approaches, NicheNet's prior model goes beyond ligand–receptor interactions and incorporates intracellular signaling as well. As a result, NicheNet can predict which ligands influence the expression in another cell, which target genes are affected by each ligand and which signaling mediators may be involved.

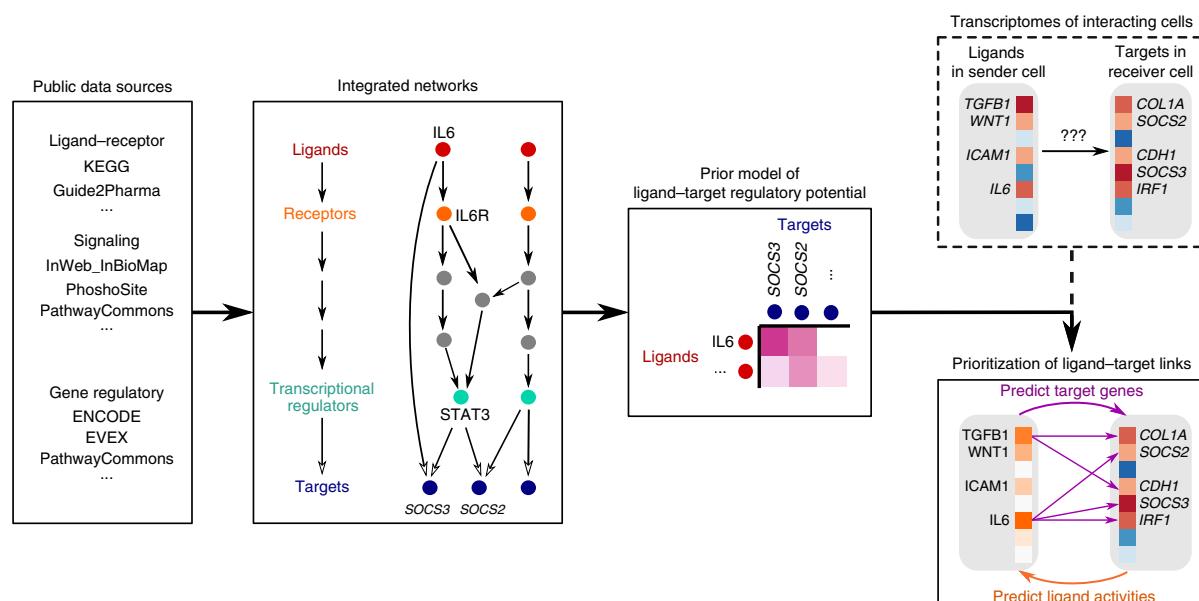
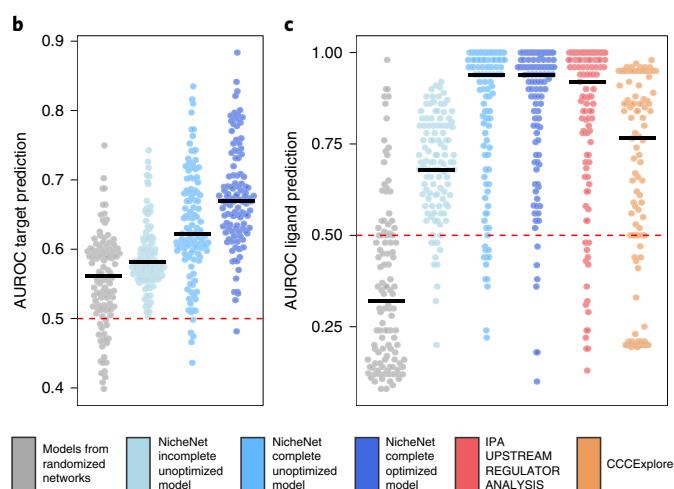
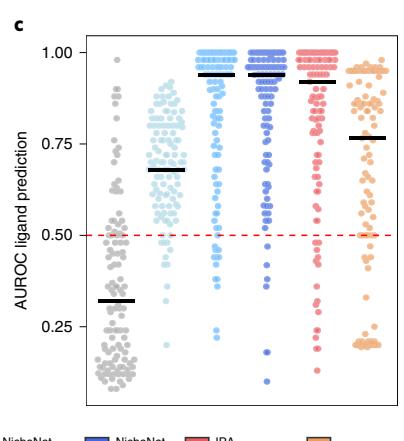
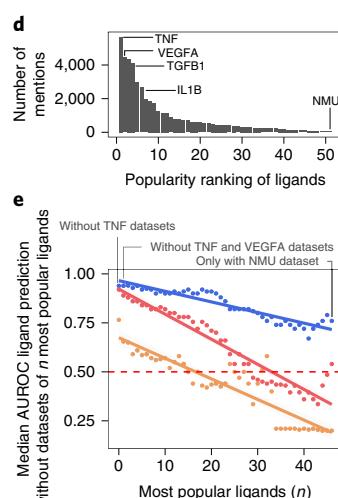
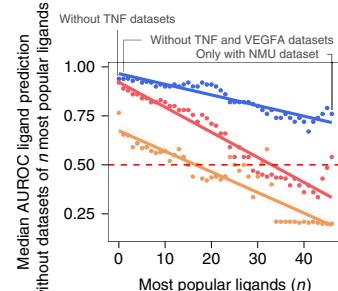
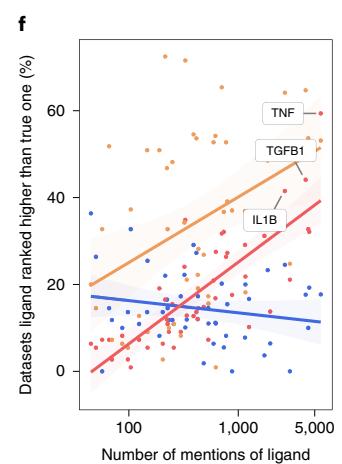
The prior model at the basis of NicheNet denotes how strongly existing knowledge supports that a ligand may regulate the expression of a target gene. To calculate this ligand–target regulatory

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 and gene regulatory interactions (Supplementary Fig. 1, Supplementary Table 1a–c and Supplementary Note 1). Secondly, we integrated these individual data sources into weighted networks (Supplementary Note 2). These contain protein–protein interactions covering the signaling paths from ligands to downstream transcriptional regulators, and gene regulatory interactions between transcriptional regulators and target genes. To let informative data sources contribute more to the final model, we used model-based parameter optimization to weight each data source to maximize the accuracy of ligand–target predictions (Supplementary Note 2). Finally, we calculated a regulatory potential score between all pairs of ligands and target genes (Supplementary Note 3). A ligand–target pair receives a high regulatory potential if the regulators of the target gene are downstream of the signaling network of the ligand. To calculate this, we used network propagation methods on the integrated networks to propagate the signal from a ligand, over receptors, signaling proteins and transcriptional regulators, to end at target genes.

When applying NicheNet to investigate communication between cells, this general prior model of ligand–target regulatory potential is combined with the gene expression data resulting from these cells. First, users can apply NicheNet to prioritize which sender cell ligands are most likely to have affected the gene expression in interacting receiver cells. This procedure, called ligand activity prediction, ranks ligands according to how well their prior target gene predictions correspond to the observed gene expression changes resulting from communication with sender cells. To then predict active ligand–target links, NicheNet searches for genes that are affected in receiver cells and are possibly regulated by these prioritized ligands as indicated with a high regulatory potential score. 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 users to prioritize signaling mediators and to understand how all of the collected data sources support the predictions.

To validate the prior model of ligand–target regulatory potential, optimize the model parameters and compare different models, we established an evaluation procedure for which we gathered public transcriptome data of several cell types before and after they were stimulated by one or two ligands in culture (Supplementary Table 2a,b and Supplementary Fig. 2). The NicheNet prior model can then be evaluated by comparing observed differentially expressed genes to the top target genes of the model. Note that this gold standard has some limitations and is therefore more useful to optimize parameters and compare different models than to indicate the exact predictive ability of NicheNet (Supplementary Note 4, Supplementary Fig. 3 and Supplementary Table 2c,d).

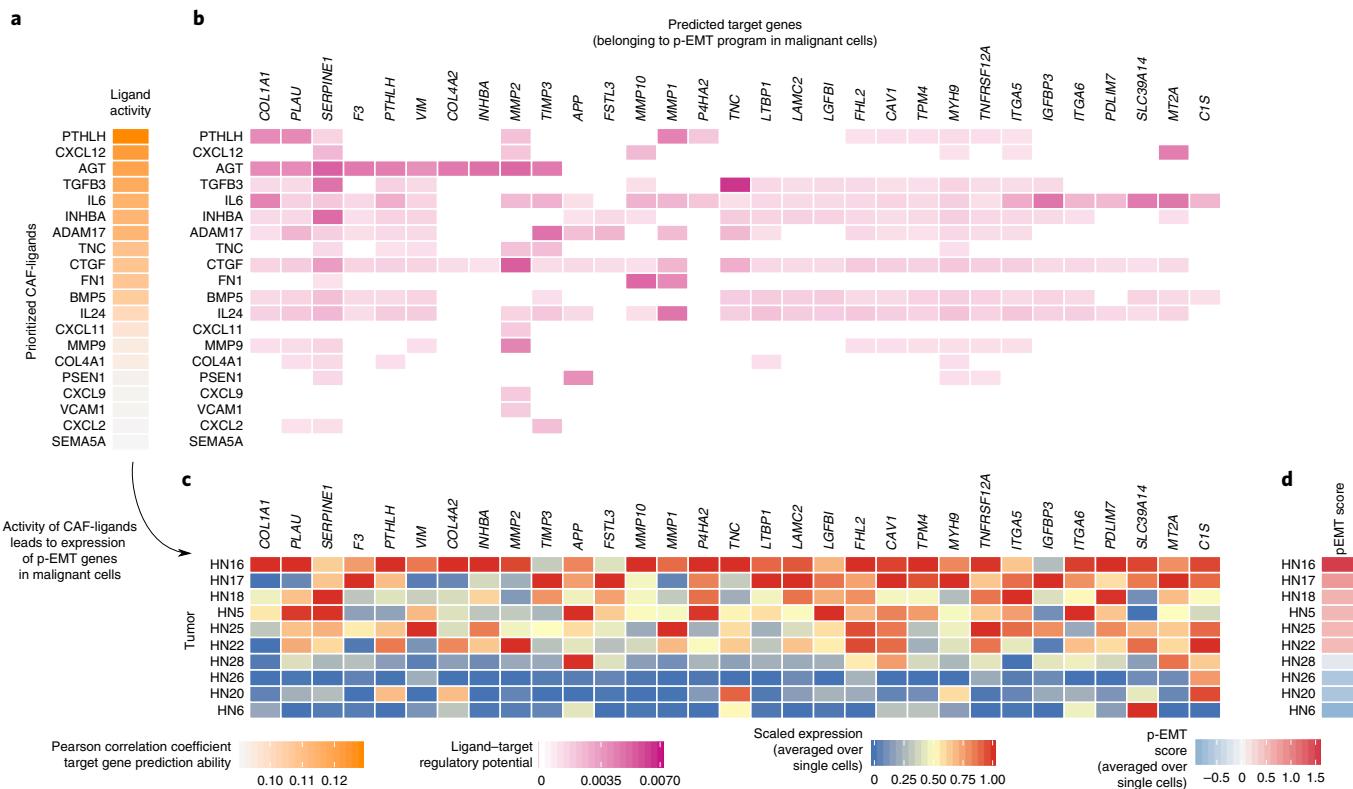
<sup>1</sup>Data Mining and Modeling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium. <sup>2</sup>Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium. <sup>3</sup>These authors jointly supervised this work: Wouter Saelens, Yvan Saeys. \*e-mail: [yvan.saeys@ugent.be](mailto:yvan.saeys@ugent.be)

**a****b****c****d****e****f**

**Fig. 1 | General workflow and validation of NicheNet.** **a**, NicheNet infers active ligand–target links between interacting cells by combining their expression data with a prior knowledge model on ligand–target links. **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 on treatment with a particular ligand. We compared performance among: (1) models constructed from 100 randomized networks; (2) 280 one-versus-one-versus-one models calculated from only one ligand–receptor, one signaling and one gene regulatory database; (3) NicheNet without parameter optimization; and (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 among 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 (that is, 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. The smoothing lines shown in **e** and **f** are the result of fitting a linear regression model ( $n=51$  different ligands). The accompanying shaded region indicates the 95% confidence interval. Dashed line in **b**, **c** and **e** indicates the performance of random guessing.

The ligand treatment datasets were used to define two complementary evaluation measures: target gene prediction 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 evaluated how well NicheNet predicts by which ligand the cells are treated (Supplementary Fig. 4). This measure indicates how well NicheNet prioritizes ligands according to their potential to regulate a set of affected genes.

We compared target gene and ligand activity prediction performances of the final optimized model with an unoptimized model and models constructed from random networks (Fig. 1b,c, Supplementary Figs. 5–7, Supplementary Table 3 and Supplementary Note 5). Parameter optimization was performed to maximize both target gene and ligand activity prediction accuracy (Supplementary Table 4a,b and Supplementary Note 6). To assess the need for integrating multiple data sources, we also estimated the performance of models built from fewer data sources (Fig. 1b,c, Supplementary



**Fig. 2 | NicheNet prioritizes CAF-ligands based on their potential to regulate p-EMT genes in adjacent malignant cells in HNSCC.** **a**, Outcome of NicheNet's ligand activity prediction on the p-EMT gene set: results are shown for the 20 (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 ( $n=6,072$  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. **b**, NicheNet's ligand-target matrix denoting the regulatory potential between CAF-ligands and target genes from the p-EMT program. **c**, Scaled expression of p-EMT genes in malignant cells, averaged per tumor. **d**, Each cell was scored for expression of genes belonging to the p-EMT program, and this score is averaged per tumor.

Table 4c and Supplementary Figs. 8–10). In conclusion, extensive data-integration and parameter optimization improved both target gene and ligand activity prediction performance, although even unoptimized models already performed better than random network models.

Next, we evaluated the possible presence of cell type bias in the model (Supplementary Fig. 3). Performance varied between different cell types treated with the same ligand, but there was no apparent bias toward particular cell types across ligands (Supplementary Fig. 11).

Furthermore, we also benchmarked NicheNet's ligand activity prediction against Upstream Regulator Analysis of Ingenuity Pathway Analysis (IPA)<sup>17</sup> and CCCExplorer<sup>15</sup> (Supplementary Note 7). NicheNet and IPA outperformed CCCExplorer, and NicheNet was much less biased than both IPA and CCCExplorer toward those ligands that are most frequently mentioned in the literature (Fig. 1c–f and Supplementary Figs. 6, 12 and 13). This points to the complementary value of combining well-studied but biased ligand–target links from the literature by more unbiased links inferred from integrating molecular interactions in ligand-to-target signaling paths.

Altogether, the validation study supports the proposed methodology and demonstrates that the final prior model of ligand–target links can be generally applicable to a broad set of different biological systems.

As a first case study, we applied NicheNet to human single-cell RNA sequencing data that were generated by Puram et al. to study the malignant and nonmalignant cell composition in head and

neck squamous cell carcinoma (HNSCC) tumors<sup>9</sup>. There, Puram et al. described a subset of malignant cells expressing a partial epithelial-to-mesenchymal transition (p-EMT) program. Interestingly, these malignant cells were located at the edge of tumors close to cancer-associated fibroblasts (CAFs). Moreover, Puram et al. were able to predict many ligand–receptor interactions between CAFs and malignant cells. Therefore, they hypothesized that CAF-ligands might regulate the expression of p-EMT genes in neighboring malignant cells.

We applied NicheNet to extend on this ligand–receptor analysis and further investigate this hypothesis. 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. 2a, Supplementary Table 5a,b, Supplementary Fig. 14 and Supplementary Note 8). Noteworthily, one of the top-ranked ligands was TGFB3, which was experimentally validated in vitro by Puram et al. Moreover, a role in classic epithelial-to-mesenchymal transition (EMT) has been described previously in other studies for 18 of the 20 top-ranked ligands (Supplementary Table 6).

Subsequently, we assessed to what extent CAF-ligands may regulate the p-EMT program. 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 with 4.7% of non-p-EMT genes ( $P$  value Fisher's exact test:  $2.5 \times 10^{-10}$ ) (Supplementary Table 5c–g). This suggests that top-ranked CAF-ligands might indeed regulate a substantial set of p-EMT genes.

We then inferred specific regulatory interactions that could be active between CAF-ligands and p-EMT target genes in malignant cells (Fig. 2b). The inference of active ligand-target links allows generating functional hypotheses concerning the specific p-EMT regulatory role of prioritized CAF-ligands.

Importantly, users can also check the validity of predicted ligand–target links because NicheNet provides 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* (Supplementary Table 7 and Supplementary Fig. 15). Interestingly, an important mediator in this predicted signaling network was SMAD3, a transcription factor involved in classic EMT<sup>18</sup>.

NicheNet can also be applied to single-cell datasets to specifically analyze the heterogeneity in ligand activity between individual cells. This complementary type of analysis suggested intertumoral heterogeneity in crosstalk between CAFs and malignant cells (Supplementary Note 9 and Supplementary Table 8).

As the second case study, we applied NicheNet to explore immune cell crosstalk in response to viral infection. In particular, we used single-cell data of mouse immune cells in the T-cell area in the inguinal lymph node before and after lymphocytic choriomeningitis virus (LCMV) infection<sup>7</sup>. NicheNet prioritized immune cell ligands based on how well they could induce the differential expression in T cells on viral infection (Supplementary Note 10 and Supplementary Table 9). Ligands known to play a role in antiviral responses, such as Il27, Ifng and Il12a, were enriched at the top of the ranking compared with ligands involved in other processes.

In both case studies, we have illustrated the type of hypotheses that NicheNet can provide. These hypotheses are more functional than those provided by current analysis methods that infer ligand–receptor interactions because NicheNet explores how these interactions affect gene expression (Supplementary Note 11 reviews ligand–receptor network inference methods, CellPhoneDB and NicheNet). Whether prioritized ligand–target links are essential in the process of interest should be experimentally validated<sup>19</sup>, certainly because the wiring of cellular signaling can be condition and cell type specific. Including cell type-specific signaling and gene regulatory networks 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 (Supplementary Note 12).

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.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-019-0667-5>.

Received: 5 March 2019; Accepted: 29 October 2019;  
Published online: 9 December 2019

## References

- Yosef, N. & Regev, A. Writ large: genomic dissection of the effect of cellular environment on immune response. *Science* **354**, 64–68 (2016).
- Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
- Stähli, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Rodrigues, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
- Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235 (2019).
- Vickovic, S. et al. High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
- Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
- Boisset, J.-C. et al. Mapping the physical network of cellular interactions. *Nat. Methods* **15**, 547–553 (2018).
- Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
- Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533 (2017).
- Pavlicev, M. et al. Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal–fetal interface. *Genome Res.* **27**, 349–361 (2017).
- Zepp, J. A. et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* **170**, 1134–1148.e10 (2017).
- Cohen, M. et al. Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting. *Cell* **175**, 1031–1044 (2018).
- Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347 (2018).
- Choi, H. et al. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep.* **10**, 1187–1201 (2015).
- Komurov, K. Modeling community-wide molecular networks of multicellular systems. *Bioinformatics* **28**, 694–700 (2012).
- Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
- Valcourt, U., Kowanetz, M., Niimi, H., Heldin, C.-H. & Moustakas, A. TGF-β and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Mol. Biol. Cell* **16**, 1987–2002 (2005).
- Bonnardel, J. et al. Stellate cells, hepatocytes, and endothelial cells imprint the kupffer cell identity on monocytes colonizing the liver macrophage niche. *Immunity* **51**, 638–354 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

## Methods

**Collection of ligand–receptor, intracellular signaling and gene regulatory interactions.** Ligand–receptor interactions were collected from KEGG<sup>20</sup>, Ramilowski et al.<sup>21</sup> and IUPHAR Guide to Pharmacology<sup>22</sup> (via Harmonizome<sup>23</sup>). In addition to this, we predicted ligand–receptor interactions by searching in protein–protein interaction databases for interactions between genes annotated as ligands and receptors.

Signaling and protein–protein interactions were collected from Omnipath<sup>24</sup>, PathwayCommons<sup>25</sup>, InWeb\_InBioMap<sup>26</sup>, ConsensusPathDB<sup>27</sup>, Vinayagam et al.<sup>28</sup>, EVEX<sup>29</sup>, KEA<sup>30</sup>, PhosphoSite<sup>31</sup>, DEPOD<sup>32</sup> and Harmonizome kinase–substrate predictions<sup>23</sup> (the last four retrieved via Harmonizome<sup>23</sup>). Networks containing binary protein–protein interactions and interactions between members of the same protein complex are intrinsically nondirectional. However, for linking ligands to target genes in this study, the signaling direction is important, and directional networks are required. Therefore, we converted the nondirectional networks to directional networks by considering the interaction  $A \leftrightarrow B$  as  $A \rightarrow B$  and  $B \rightarrow A$  and applying the following adaptations: (1) both  $A \rightarrow B$  and  $B \rightarrow A$  were kept if neither interaction was documented in another data source implying a directional interaction; (2) only one interaction of  $A \rightarrow B$  or  $B \rightarrow A$  was kept if a directional data source covers either interactions. Furthermore, we also adapted the signaling networks by removing self-links and by removing interactions going to ligands, because in this study only signaling from ligands to downstream genes is relevant.

Gene regulatory interactions were collected from RegNetwork<sup>33</sup>, TRRUST<sup>34</sup>, HTRIDB<sup>35</sup>, ReMap<sup>36</sup>, EVEX<sup>29</sup>, PathwayCommons<sup>25</sup>, Ontogenet<sup>37,38</sup>, CHEA<sup>39</sup>, ENCODE<sup>40</sup>, JASPAR<sup>41</sup>, TRANSFAC<sup>42</sup>, MOTIFMAP<sup>43</sup> and the Gene Expression Omnibus (GEO)<sup>44</sup> and MSigDB<sup>45</sup>. Signatures of Differentially Expressed Genes for Gene Perturbations (the last seven retrieved via Harmonizome<sup>23</sup>).

Almost all data sources provided their gene–gene interactions in a readily available network format. A minority of data sources required some additional parsing or conversion to network format. Necessary information regarding the collected data sources and their additional processing can be found in Supplementary Notes 1 and 13 and Supplementary Table 1.

**Weighted aggregation of data sources into integrated networks.** In the end, every data source was converted into a source-specific directed graph  $G = \langle V, E \rangle$ , where  $V$  is the set of nodes representing genes and  $E$  is the set of edges representing a possible signaling interaction between nodes for ligand–receptor and signaling data sources and an expression regulation interaction for gene regulatory data sources.  $A$  is the corresponding adjacency matrix with elements  $A_{ij}$  representing an interaction from node  $v_i$  to  $v_j$ .

Ligand–receptor and signaling data source networks were aggregated into an integrated weighted ligand-signaling network; gene regulatory data source networks were aggregated into a integrated weighted gene regulatory network. We did not directly integrate all data sources in one common integrated network because the edges of ligand–receptor and signaling data source networks represent a different type of interaction than edges from gene regulatory networks. Namely, a protein–protein interaction for ligand–receptor and signaling data sources, and a gene regulatory interaction in the gene regulatory network.

Instead of simply summing the adjacency matrices of the separate data source networks, we created a weighted sum of adjacency matrices to construct the adjacency matrix  $W$  of the integrated networks as follows:

$$W = \sum_{i=1}^n (w_i A_i)$$

where  $n$  is the number of data sources in a specific layer and the weight  $0 \leq w_i \leq 1$  assigned to a source-specific network  $i$  represents the contribution of network  $i$  to the performance of the final model. The weighting coefficients were determined by data source optimization, as explained in the section Parameter optimization via mlrMBO. In an unoptimized setting, the weighting coefficients for every data source-specific network would be equal.

To mitigate the potential negative influence of over-dominant hubs on the final model, we defined two parameters: a ligand-signaling network hub correction factor and a gene regulatory network hub correction factor. The hub correction factor decreases the weight of edges directed to hubs proportional to their indegree:

$$W_{\text{corrected}} = WD^{-h}$$

with  $D$  the diagonal matrix denoting the indegree of each node and  $h$  a hub correction factor ( $0 \leq h \leq 1$ ). Hub correction factor values were automatically determined via parameter optimization as explained in the section Parameter optimization via mlrMBO.

**Randomization of weighted networks.** Randomization was done for the ligand-signaling and gene regulatory integrated network separately after all nonzero interaction weights received the same weight. To preserve node degree, we randomized the network via the *sample\_degseq* function of the *igraph* R package. Preservation of node degree is vital to assess whether the predictive performance of

NicheNet is due to the biologically meaningful nature of links and not merely due to the degree of the nodes.

**Calculation of ligand–target regulatory potential scores.** Firstly, for a particular ligand, a signaling importance score for every gene was calculated by applying Personalized PageRank (PPR) on the ligand-signaling network with the ligand of interest as seed node<sup>12</sup>. The PPR vector of a ligand is the steady-state probability distribution of a random walk with restart that returns to the ligand node with probability  $1 - d$  (with  $0 \leq d < 1$  known as the damping factor, automatically determined via parameter optimization as explained in the section Parameter optimization via mlrMBO). The PPR equation can be written as<sup>46,47</sup>:

$$\mathbf{v} = (1 - d) \times \mathbf{u} + d \times W \times \mathbf{v}$$

where  $\mathbf{v}$  is the vector of PPR importance scores for every node in the ligand-signaling network based on the steady-state distribution of the random walker;  $\mathbf{u}$  is the ‘preference’ vector in which the personal preference for every node is denoted (normalized such that the sum of the elements of  $\mathbf{u}$  is 1): here, the ligand of interest gets value 1, the other genes 0; and  $W$  is the normalized adjacency matrix of the weighted ligand-signaling network in which an entry  $w_{ij}$  represents the edge weight of the link from  $j$  to  $i$  divided by the outdegree of  $j$ .

Secondly, we applied a cutoff on the PPR vector of a ligand such that only genes strongly ‘enriched’ in the graph neighborhood of the ligand compared with the complete graph get nonzero importance scores. The PPR score of a gene should be larger than or equal to a specific quantile, which belongs to the interval [0.9, 0.999] and was determined via automatic parameter optimization (see Supplementary Note 3).

By applying these two procedures for every ligand of interest, an  $n \times m$  matrix of ligand–gene signaling importance scores was calculated with  $n$  the number of ligands of interest and  $m$  the total number of genes over all networks. In case a PPR vector needed to be calculated for a ‘ligand combination’ instead of a single ligand, all ligands of the specific ligand combination were chosen as seed nodes. The latter was the case when we evaluated target gene prediction when cells were treated with more than one ligand (see the sections Processing ligand treatment expression datasets and Evaluation of target gene prediction).

Finally, we multiplied the  $n \times m$  ligand–gene signaling importance matrix with the  $m \times m$  weighted adjacency matrix of the integrated gene regulatory network to obtain the  $n \times m$  ligand–target matrix  $L$ , with  $n$  the number of considered ligands and  $m$  the number of possible target genes, in which an entry  $l_{ij}$  is a regulatory potential score corresponding to the confidence that a particular ligand  $i$  can regulate the expression of a particular target gene  $j$  (given the integrated data sources):

$$l_{ij} = \sum_{k=1}^m (\text{PPR}_{ik} \times \text{GRN}_{kj})$$

with  $\text{PPR}_{ik}$  the importance of gene  $k$  in the signaling of ligand  $i$  as determined via PPR and the gene regulatory network ( $\text{GRN}_{kj}$ ) the edge weight corresponding to the evidence that gene  $k$  regulates the expression of target gene  $j$ .

**Construction of leave-one-in and one-versus-one-versus-one models.** To investigate the importance of the integration of multiple complementary data sources and to characterize the information content of individual data sources, we constructed ‘leave-one-in’ models. Ligand-signaling leave-one-in models were constructed from one ligand–receptor or signaling data source and all gene regulatory data sources. Gene regulatory leave-one-in models were constructed from all ligand–receptor or signaling data sources and one gene regulatory data source. Hyperparameters and data source weights were the same as for the optimized version of NicheNet.

To investigate the importance of integrating multiple databases per layer instead of one database per layer, we constructed ‘one-versus-one-versus-one’ models. In one-versus-one-versus-one models, ligand–target regulatory potential scores were calculated after using only one comprehensive ligand–receptor, one signaling and one gene regulatory database. Note that we included all interactions documented in a particular database and not data source, meaning that we included all data sources belonging to a specific database (see Supplementary Table 1 and Supplementary Note 1). We did this to construct models that would be a better representation of what users typically would do, namely: selecting one comprehensive ligand–receptor, signaling and gene regulatory database and not dividing comprehensive databases in data sources and using only one data source per layer. Because one-versus-one-versus-one models are incomplete, some models will not contain all ligands for which we collected ligand treatment datasets. To be able to validate these models on all datasets, we assigned all target genes of undocumented ligands a random score between 0 and the minimum nonzero ligand–target regulatory potential score of documented ligands. Hyperparameters were the same as for the optimized version of NicheNet (except for the models in Fig. 1b,c, where unoptimized parameter values were used—see the section Parameters of models evaluated) and every database received a weight of 1.

**Processing ligand treatment expression datasets.** For the validation of NicheNet, we used data of 111 microarray studies in which the transcriptional response of

human or mouse cells before and after treatment with one or two protein ligand(s) was analyzed (two-group design) or in which the transcriptome was analyzed at several time points after treatment (time-course design) (see Supplementary Table 2a for more information, including accession codes). Data were collected from the National Center for Biotechnology Information (NCBI) GEO public repository<sup>44</sup> or ArrayExpress<sup>48</sup>. If available, preprocessed data were collected and used for further analysis. When no preprocessed data were available, preprocessing for one-color assays was performed by applying the Robust Multi-array Average procedure<sup>49,50</sup> and for two-color assays via the procedure implemented in the limma package<sup>51</sup>. To determine the genes that are differentially expressed after ligand treatment, two-group design studies and time-course design studies with a small number of replicated time points were each analyzed as recommended in the user manual of limma. In the end, for both types of study designs, a log fold change value and *P* value are obtained for every gene: for two-group design studies, the  $\log_2$  fold change in expression level between the control and treated group is taken as the final log fold change value, whereas for the time-course design studies the maximum of log<sub>2</sub> fold changes in expression between consecutive time points is chosen. *P* values were corrected for multiple testing by controlling the false discovery rate using the Benjamini–Hochberg procedure<sup>52</sup>. Genes with an absolute  $\log_2$  fold change  $\geq 1$  and adjusted  $P \leq 0.1$  were considered differentially expressed. This relatively tolerant *P* value was chosen so that we retained as many true positives as possible in every dataset (considering the class imbalance problem). When searching for expression datasets, only datasets with  $>20$  and  $<1,500$  differentially expressed genes were used for evaluation in this study. Because the final ligand–target model is formalized in human gene symbols, microarray gene identifiers were mapped to these (including one-to-one orthology conversion if microarray data were of mouse origin) (see Supplementary Note 13: conversion of gene identifiers).

**Evaluation of target gene prediction.** Under the assumption that true downstream target genes of a ligand will be differentially expressed after treatment of cells with this ligand, we evaluated how well the ligand–target regulatory potential scores of NicheNet predict the observed ligand-induced transcriptional response. These ligand–target predictions were evaluated as a binary classification task in which genes differentially expressed on ligand treatment were considered as positive targets of that ligand and all other genes as negatives. Predictions for target genes of which the expression was not profiled in the specific dataset were ignored for evaluation. The following evaluation metrics were calculated to assess predictive performance: AUROC (area under receiver operating characteristic curve), AUPR (area under the precision–recall curve), Pearson correlation, Spearman’s rank correlation, negative natural logarithm of the mean-rank gene-set enrichment *P* value<sup>51,53</sup> (from limma) and the area under the cumulative recovery curve (considering the top 3% of the ranking) (from iRegulon<sup>54</sup>). Because the AUPR is higher when the fraction of positives is higher, comparison of the AUPR over different datasets requires to subtract the fraction of positives from the AUPR (since the AUPR for classifiers not better than random predictions equals the fraction of positives). The iRegulon area under the cumulative recovery curve depends as well on the fraction of positives. Therefore, this metric was also calculated for a randomly sampled set of  $n$  genes instead of the true set of  $n$  differentially expressed genes, such that this metric could be corrected for random prediction.

**Evaluation of ligand activity prediction.** As the second type of validation, we evaluated how well NicheNet predicts the ligands that were added to a cell culture, given the expression data. We tackled this problem as a binary classification task by predicting for each ligand, whether it is active or not (class label: active or not). The central assumption underlying this validation procedure is that for a good model of ligand–target regulatory potential scores, the ligand that best predicts the differential expression observed in a dataset is probably the true ligand with which the cells were treated. Supplementary Fig. 4 shows a graphical overview of this procedure, which consists of two steps: (1) calculation of feature importance scores for each ligand for every dataset and (2) using these feature importance scores for prediction of the ligand activity state.

For this validation procedure, we considered as input space the  $n \times m$  ligand–target matrix with  $m$  the number of profiled ligands (that is, all ligands of which at least one expression dataset was analyzed) and  $n$  the total number of target genes. We used this matrix to predict for each dataset whether a gene was differentially expressed after ligand treatment. Ideally, the true ligand with which the cells were treated should predict the differential expression observed in a dataset better than the other ligands. In other words, informative feature importance scores should be higher for this ligand than for others (conceptually we can thus consider feature importance scores as measures for ligand activity). For one expression dataset, we calculated multiple feature importance measures for all ligands, and then we repeated this procedure for every dataset. As feature importance measures were calculated for each ligand separately, we calculated the target gene evaluation metrics described above.

Next, we evaluated how well the feature importance scores of each dataset–ligand combination can predict the correct ligand activity state: class label ‘active’ if, for a specific ligand–dataset combination, the dataset contains indeed expression

data of cells treated with that ligand; and ‘nonactive’ if the dataset contains expression data of cells treated with another ligand. By calculating classification evaluation metrics (AUROC and AUPR) of each feature importance measure in predicting the ligand activity state, we obtained an indication of the predictive ability of each feature importance measure. As performance for NicheNet in predicting ligand activity, we took the performance of the feature importance measure resulting in the highest geometric average between the AUROC and AUPR (the geometric average was preferred over the arithmetic because of the differences in scale between the AUROC and AUPR). This was the Pearson correlation coefficient (Supplementary Fig. 7).

**Evaluation of cell type bias.** To evaluate the possible presence of a cell type bias, we assessed the predictive performance of NicheNet on ligand treatment datasets profiling the response of the same set of ligands over different cell types. For this analysis, we grouped all collected 111 ligand treatments in different cell type groups. The specific grouping of datasets in cell type groups is shown in Supplementary Table 2. Note that we divided datasets into broad cell type categories to have at least six datasets per cell type group (for example, we grouped ‘T cells’ and ‘peripheral blood mononuclear cells’ under the broad category ‘immune cells’).

Next, we selected only those datasets that measured 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. This was done to have an as good as possible balanced design in ligand–cell type combinations. For those selected datasets (43 of 111), target gene and ligand activity prediction performances were calculated.

**IPA Upstream Regulator Analysis on ligand treatment datasets.** For each ligand treatment expression dataset, IPA (Spring 2019 release; QIAGEN, <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>) core analysis was performed. Genes with an absolute  $\log_2$  fold change  $\geq 1$  and adjusted  $P \leq 0.1$  were considered differentially expressed. Upstream Regulator Analysis<sup>57</sup> results were extracted only for the ligands of which at least one ligand treatment dataset was analyzed. The negative decimal logarithm of the enrichment *P* value was used as IPA’s feature importance score. Next, we assessed how well IPA’s feature importance scores predict ligand activity (in the same way as for NicheNet, as explained above). Important to note is that IPA does not always give the official gene symbols for possible upstream regulators. To make the performance comparable to NicheNet and include these unofficial names in the rankings, we made the following conversions: ‘Interferon alpha’ → ‘IFNA1’; ‘IL12 (complex)’ → ‘IL12B’; ‘PDGF BB’ → ‘PDGFB’; ‘Growth hormone’ → ‘GH1’; and ‘Vegf’ → ‘VEGFA’.

**CCCExplorer analysis on ligand treatment datasets for ligand activity prediction.** CCCExplorer integrates expression data of interacting cells with prior information on ligand–receptor, signaling and gene regulatory interactions to predict active crosstalk-signaling pathways between cell populations in the tumor microenvironment. These crosstalk-signaling pathways are represented as a network from ligand–receptor interactions to transcription factors and their target genes. For each pathway branch activated in the crosstalk, enrichment is assessed and a *P* value given (Fisher’s test). CCCExplorer differs from NicheNet in how the prior information is used to link ligands to target genes. Moreover, CCCExplorer uses only a limited number of data sources compared with NicheNet (with KEGG as the main data source).

To apply CCCExplorer on ligand treatment datasets for ligand activity prediction, we needed to make the following changes to the in- and output of the method. As potentially active ligands, the original algorithm selects ligands that are (1) upregulated in a cell type of interest compared with a reference cell type, and (2) for which corresponding receptors are expressed in interacting cells. Because there is no ‘sender’ cell type when evaluating ligand activity prediction on ligand treatment datasets, all 51 unique ligands for which we analyzed at least one dataset were considered potentially active ligands. Note that this adaptation is similar to how we used NicheNet on ligand treatment datasets compared with data of interacting cells. Other changes include: (1) considering all transcription factors as possible regulators and not only expressed ones, and (2) considering all differentially expressed genes (absolute  $\log_2$  fold change  $\geq 1$  and adjusted  $P \leq 0.1$ ) and not only upregulated genes as the target gene set for which we wanted to predict active ligands.

Finally, the goal of the ligand activity procedure is to give an importance score to each ligand related to how likely it is that this ligand was responsible for the induction of the observed transcriptional changes. As the score for each ligand, we used the negative decimal logarithm of the *P* value given to the most significant ligand–receptor–regulator pathway to which the ligand belongs. Ligands that are part of pathways that were not retrieved were given a score of 0. We performed this procedure for each ligand treatment dataset, and we then assessed how well CCCExplorer’s feature importance scores predict ligand activity (in the same way as for NicheNet, as explained in the evaluation of ligand activity prediction section).

**Comparison of popularity bias among NicheNet, CCCExplorer and IPA Upstream Regulator Analysis.** To analyze popularity bias in ligand activity prediction performance, we assessed ligand activity prediction performance on

all ligand treatment datasets and after iteratively leaving out the datasets profiling the transcriptional response to the top  $n$  most popular ligands. The popularity of ligands is determined by the number of studies in PubMed in which a ligand is described (NCBI Entrez Gene2Pubmed database). As feature importance score of NicheNet, the Pearson correlation coefficient was used. For every analysis, we considered all ligands for which we collected one or more ligand treatment datasets as potential active ligands, so even ligands for which the corresponding ligand treatment datasets were left out.

To analyze popularity bias in the ligand rankings provided by NicheNet, CCCExplorer and IPA Upstream Regulator Analysis, we assessed for every ligand the relation between the popularity of the ligand and the percentage of datasets in which the ligand is not active but ranked higher than the actual active ligand.

**Parameter optimization via mlRMBO.** As described in the sections Weighted aggregation of data sources into integrated networks and Calculation of ligand-target regulatory potential scores, several parameters need to be defined: the weights of the 57 data sources, the ligand-signaling and gene regulatory network hub correction factors, the PPR vector cutoff and the damping factor for PPR. For each data source weight, ligand-signaling network hub correction factor and gene regulatory network hub correction factor, the lower bound was 0 and upper bound 1. For the damping factor, the lower and upper bound were respectively 0.01 and 0.99, whereas the lower and upper bounds for the PPR vector cutoff quantile were respectively 0.9 and 0.999.

These parameter values were automatically optimized via model-based, also called Bayesian, optimization, via the mlRMBO toolbox<sup>5,56</sup>. This optimization procedure was chosen because it is a state-of-the-art method that allows multi-objective optimization of expensive black-box functions. Briefly, in model-based optimization, a surrogate machine learning model is sequentially used to model the relation between parameter values and performance output. New parameter design points are proposed by optimizing a specific infill criterion. These proposed points are subsequently evaluated, the model is updated and the procedure iterates until a defined termination criterion is met. In addition to model-based optimization, we also tested a multi-objective genetic algorithm, namely NSGA-II<sup>57</sup>, which gave similar results (Supplementary Note 14).

We performed multi-objective optimization to optimize both target gene and ligand activity prediction performance on ligand treatment datasets. For some ligands, multiple datasets were collected, whereas for other ligands, only one dataset was collected (Supplementary Table 2). Because we did not want to bias the optimization toward ligands for which more datasets were collected, we decided to aggregate performances on all datasets of the same ligand by taking the median. Over these aggregated performances per ligand, we then calculated the average AUROC and AUPR (after correction for random prediction) of target gene prediction and the median AUROC and AUPR (after correction for random prediction) of ligand prediction (median was chosen here because of the strongly skewed distribution of ligand activity prediction performances, see Fig. 1c). These four aggregated measures were chosen as the optimization objectives.

As the surrogate regression model, we selected Kriging because this is recommended for parameters with numerical domains. As the specific, multi-objective, multi-point proposal method, we used a direct indicator-based method (SMS-EGO, which fits an individual Kriging model for each objective function) with the confidence bound as infill criterion and the random lambda method to generate multiple proposals per iteration (lambda parameter = 1). Optimization started with 240 parameter designs generated via maximin Latin hypercube sampling, and, during each subsequent iteration round, 24 new designs were proposed and evaluated in parallel on a grid engine cluster. Optimization stopped after 50 iterations. From the Pareto-optimal solutions, we selected the solution with the highest geometric average of the four objectives. We chose this criterion over the arithmetic mean because the four optimization criteria have different scales (for example, the AUROC for target gene prediction is between 0.6 and 0.7, while the AUROC for ligand prediction ranges between 0.75 and 1), and we wanted to weight them equally.

To define the hyperparameter values and data source weights for NicheNet's final ligand–target matrix used for application purposes, we optimized the above-mentioned criteria calculated on all ligand treatment datasets (we refer to this as optimization without cross-validation in Supplementary Note 6). However, for evaluation of the target gene and ligand prediction performance of NicheNet on ligand treatment datasets, we decided to perform a different procedure to avoid a too optimistic estimation of the performance due to possible overfitting (we refer to this procedure as optimization with cross-validation in Supplementary Note 6). The considered ligand treatment expression datasets were randomly divided into three groups, and threefold cross-validation was performed (Supplementary Table 4a,b): for one fold, two of the three groups of expression datasets were selected as the training set for optimization (that is, the optimization criteria were calculated only from these datasets) and the datasets of the left-out group for evaluation. To evaluate the prediction performance of optimized NicheNet, we calculated test set performances for datasets from the left-out group for every fold and combined these performances over all folds. All validation results shown for models using optimized parameters are thus the result of the above-described procedure and not of averaging the optimized parameter values of the models of each of the three folds.

**Collection and processing of the human single-cell transcriptomics data from Puram et al.** Processed single-cell expression data ( $\log_2$ , TPM values) of malignant and nonmalignant cells from HNSCC tumors were downloaded from the GEO with accession code GSE103322 (ref. <sup>9</sup>). As in the original study conducted by Puram et al., analyses were performed on relative expression data (calculated as described in ref. <sup>9</sup>) unless stated otherwise. For the analyses described in this paper, we only use data of cells from the following primary tumors: HN5, HN6, HN16, HN17, HN18, HN20, HN22, HN25, HN26 and HN28. To determine expressed genes, we applied the same procedure as described in the original study. However, we applied this for every cell type of interest separately (malignant cells and CAFs) across all tumors.

**Collection and processing of the mouse single-cell transcriptomics data from Medaglia et al.** Unprocessed massively parallel single-cell RNA sequencing (MARS-seq) single-cell expression data (gene counts) of cells from the T-cell area in the inguinal lymph node were downloaded from GEO with accession code GSE104054 (ref. <sup>7</sup>). Genes expressed in fewer than five cells, External RNA Controls Consortium (ERCC) spike-ins and cells with fewer than 500 unique molecular identifiers (UMIs) were discarded from the analysis. Medaglia et al.<sup>7</sup> sequenced cells from steady-state lymph nodes and from lymph nodes 72 h after they injected LCMV into the footpad of mice. Because sequencing was performed in different batches and we wanted to identify cell populations present both in steady-state and after viral infection, preprocessing was performed according to the Seurat pipeline including the alignment procedure to identify shared cell populations across both datasets<sup>58</sup>. As genes for use in the canonical correlation analysis, we selected the union of the top 1,000 genes with the highest dispersion from both datasets. After canonical correlation analysis, we selected the first 15 canonical correlation vectors for alignment and further downstream analysis (that is, t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction and Seurat clustering). Cell clusters were defined by Seurat clustering (resolution: 0.6) and annotated after conserved marker identification as B cells, CD4 T cells, CD4 regulatory T cells (Tregs), CD8 T cells, natural killer (NK) cells, monocytes or dendritic cells. The following predefined list of marker genes was used to define cell types: B cells: Ms4a1, Ly6d, Cd79b; CD4 T cells: Cd4, Il7r; CD8 T cells: Cd8b1, Cd8a; NK cells: Nkg7, Gzma, Gzmb; dendritic cells: Cst3, Fscn1; monocytes: Lyz2, Ly6c2, S100a4; Treg: Il2ra, Ikkzf2, Tnfrsf4, Foxp3.

**Definition of possible active ligands and target gene set for ligand activity prediction: application on data from Puram et al.** As potentially active ligands, we considered all ligands that were expressed in CAFs and for which at least one putative receptor was expressed in malignant cells. Putative receptors for ligands were determined by querying all collected ligand–receptor data sources. To determine which of the cells annotated as fibroblasts (provided by Puram et al.<sup>9</sup>) were CAFs, we applied the Seurat pipeline: expression data of fibroblasts were processed, fibroblasts were clustered into subtypes (resolution parameter of the Seurat clustering: 0.05) and marker genes of each of the retrieved clusters were determined<sup>58</sup>. The cluster of cells that strongly express CAF-marker genes (FAP, TGFB3, THY1, PDPN, MMP2, MMP11, PDGFRA; as provided by Puram et al.<sup>9</sup>) was considered CAFs.

As the target gene set, we considered the list of genes belonging to the p-EMT program, as given by Puram et al.<sup>9</sup>. As background, we considered all genes expressed in malignant cells.

**Definition of possible active ligands and target gene set for ligand activity prediction: application on data from Medaglia et al.** As potentially active ligands, we considered all ligands that were expressed in NK cells, monocytes, B cells and dendritic cells and for which at least one putative receptor was expressed in CD8 T cells (genes were considered to be expressed if their maximum expression value over all cells in a cell type was larger than 0). Putative receptors for ligands were determined by querying all collected ligand–receptor data sources.

As the target gene set, we considered the list of genes that are differentially expressed between CD8 T cells in steady-state and after LCMV infection (Wilcoxon rank-sum test; adjusted  $P \leq 0.05$  and  $|\text{average log fold change}| \geq 0.5$ ; 1,252 cells after LCMV infection (from two different mice) and 393 cells in steady-state (from three different mice)). As background, we considered all other genes statically expressed in CD8 T cells.

**Ligand activity prediction on predefined gene sets for application purposes.** To calculate ligand activity scores based on expression data of interacting cells, we defined two gene sets: potentially active ligands in sender cells and a set of affected genes of interest in the receiving cells. How these gene sets were defined for both case studies was described in the definition of possible active ligand sections. Ligand activity scores were then calculated as the Pearson correlation coefficient between the ligand–target regulatory potential scores of each selected ligand and the target indicator vector, which indicates whether a gene belongs to the gene set of interest or not. Pearson correlation was used because the validation study on ligand treatment datasets showed that this metric ranks ligands according to their possible activity better than other metrics (Supplementary Fig. 7).

In addition to defining ligand activities by considering each ligand individually, we also determined ligand activities as variable importance scores derived from a random forest classifier that uses the NicheNet target predictions of all ligands as features to predict the gene set of interest (R package `randomForest`). Used random forest hyperparameters: number of trees = 1,000; number of features sampled at each split = cubic root of the total number of features; calculated variable importance scores: mean decrease in Gini index and mean decrease in classification accuracy.

**Target gene prediction by multi-ligand classification models.** To assess how well ligands can predict which genes belong to a gene set of interest, we did not only look at prediction performances for individual ligands, but we also trained classification models that use the NicheNet ligand–target regulatory potential scores of all potentially active ligands as features to predict the gene set of interest. We applied random forest via the R package `randomForest`, and parameters: number of trees = 1,000; number of features sampled at each split = square root of the total number of features. We used random forest because this classifier most often performed best compared with other classifiers that were tested (Supplementary Note 15).

To avoid overfitting, five rounds of fivefold stratified cross-validation were performed. We performed stratified cross-validation such that in every fold, the same fraction of positive genes are present and class imbalance and performances on different test sets are comparable. In each fold, the positive class probabilities for the genes in the test set were calculated as a measure of how likely it is that a gene belongs to the gene set of interest. These positive class probabilities were considered target gene probabilities, and performance was evaluated by using the same metrics as discussed in the section Evaluation of target gene prediction. In addition, we performed a one-sided Fisher's exact test to test the significance of the association between the 'gene set of interest assignment' of genes and whether they are part of the 5% most strongly predicted targets. Performances described in this paper are averages over the five cross-validation rounds. Exact  $n$  values and outputs of the tests (including confidence intervals, odds ratios and  $P$  values) for every round of cross-validation are included in the corresponding Supplementary Tables 5e,g and 9f,h.

**NicheNet's single-cell ligand activity prediction on HNSCC malignant cells.** Aside from performing ligand activity prediction on the p-EMT gene set, we also performed ligand activity prediction at the single-cell level. First, we scaled the expression values ( $\log_2$  TPM values) of every expressed gene across malignant cells to a [0, 1] range after cutting off outer quantiles (cutoff: 0.05). Then we determined for every cell the genes for which the scaled expression values were larger than or equal to the 0.975 quantile. For every malignant cell, NicheNet ligand activity prediction was performed on this set of most strongly expressed genes per cell (background gene set: all other genes expressed in malignant cells). The obtained ligand activities (that is, the Pearson correlation coefficient when considering ligands individually and the mean decrease in Gini index when using random forest) were made comparable between different cells by modified  $z$ -score normalization.

**Analysis of correlation between cellular p-EMT scores and ligand activities in malignant cells.** As described in ref.<sup>9</sup>, malignant cells were scored on how strongly they express genes of the p-EMT program. Next, we assessed to what extent the p-EMT score of a cell can be predicted by the ligand activity of a ligand of interest. Hence, for every CAF-ligand separately, we performed linear regression and correlation analysis between p-EMT scores of cells and ligand activity scores. To prioritize ligands according to possible p-EMT regulating function, we ranked ligands based on the Pearson correlation coefficient.

In addition to this analysis, we performed lasso-regularized linear regression with all ligands as possible model parameters to find a limited set of ligands for which the ligand activities together explain a substantial part of the deviance in p-EMT scores. To find a tradeoff between the number of model parameters and predictive performance of the model, we selected the lasso model that explains half of the maximally explainable deviance.

We performed separate analyses for the two types of considered single-cell ligand activities: the normalized Pearson correlation coefficient when considering ligands individually and the normalized mean decrease in the Gini index when using random forest.

**Inference of putative signaling paths between a ligand and target gene of interest.** Firstly, the  $k$  most important transcriptional regulators involved in the regulation of the target gene of interest by the ligand of interest were retrieved: for the ligand and target of interest, we searched for  $k$  ligand–regulator–target combinations for which the multiplication of the ligand–regulator PPR scores and regulator–target score (that is, the edge weight in the integrated gene regulatory network) was the highest. Secondly, we determined important signaling mediators involved in the signaling from the ligand of interest to the inferred transcriptional regulators. As important signaling mediators, we considered all genes involved in the shortest path between the ligand of interest and the inferred regulators. Dijkstra's algorithm was applied on the weighted integrated ligand-signaling network (after inverting the edge weights) to infer the shortest path (that is, the path for which the summed weight of the edges is minimal). Subsequently, we

retrieved all interactions between ligand, signaling mediators, regulators and target genes that are documented in NicheNet's data sources. Hereby, we can find which of all of NicheNet's data sources contain the interactions in this signaling network and thus provide evidence for the possible ligand–target link.

To infer signaling paths between *TGFB3* and its target genes *TGFBI*, *LAMC2* and *TNC* in this study, we used  $k = 2$ , and genes that were not expressed in malignant cells were excluded from the weighted ligand-signaling and gene regulatory networks.

**Ligand–target heatmap visualization.** In the ligand–target heatmaps, we showed regulatory potential scores for interactions between the 20 top-ranked ligands and the following target genes: genes that belong to the gene set of interest and to the 250 most strongly predicted targets of at least one of the 20 top-ranked ligands. For visualization purposes, we adapted the ligand–target regulatory potential matrix as follows. Regulatory potential scores were set to 0 if their score was below a predefined threshold, which was here the 0.25 quantile of scores of interactions between the 20 top-ranked ligands and their respective top targets. As a result of this cutoff, it is possible that for some ligands no target genes will have a nonzero score in the final heatmap (as was the case in Fig. 2b for SEMA5A). Note that the specific values of the three above-mentioned cutoffs can be adapted by the user. Because the choice of these cutoffs for visualization is quite arbitrary, we recommend users to test several cutoff values.

**Parameters of models evaluated.** For the NicheNet one-versus-one-versus-one models and NicheNet with all data sources but without parameter optimization (evaluated in Fig. 1b,c), the following parameter settings were used: the weights for respectively all databases and data sources were 1; the ligand-signaling network hub correction factor and gene regulatory network hub correction factor 0; the cutoff on the PPR vector 0.9; and damping factor 0.5. For the randomized networks, hyperparameters were the same as for the final optimized NicheNet model (optimization via cross-validation).

**Statistics.** We performed a Wilcoxon rank-sum test to define differentially expressed genes in the T-cell case study as described in the section Definition of possible active ligands and target gene set for ligand activity prediction: application on data from Medaglia et al. We also performed one-sided Fisher's exact tests to test whether genes belonging to the gene set of interest are more likely to be part of the 5% most strongly predicted targets than background genes (see the paragraph Target gene prediction by multi-ligand classification models for more details).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

No new data were generated for this study. All data used in this study are publicly available: accession codes for the public ligand treatment expression datasets are given in Supplementary Table 2, accession codes for the public case study datasets are given in the Methods sections Collection and processing of the human single-cell transcriptomics data from Puram et al. and Collection and processing of the mouse single-cell transcriptomics data from Medaglia et al. The data source networks, NicheNet's ligand–target matrix and the processed ligand treatment and case study expression datasets are available on Zenodo (<https://doi.org/10.5281/zenodo.3260758>)<sup>39</sup>. Databases used to create the NicheNet model are mentioned and referred to in the Methods section Collection of ligand–receptor, intracellular signaling and gene regulatory interactions, in Supplementary Table 1 and in the Supplementary Notes. Source data underlying Figs. 1 and 2 are accessible from the respective figure legends.

## Code availability

An open-source R implementation of NicheNet is available at GitHub (<https://github.com/saeyslab/nichenepr>). The release includes tutorials and example vignettes for the following analyses: ligand activity analysis on a gene set of interest, single-cell ligand activity analysis; ligand-to-target signaling path visualization and assessment of how well prioritized ligands explain changes in gene expression in the receiver cell. In addition, we provide tutorials on model validation (evaluation of target gene and ligand activity), model construction and parameter optimization such that users can benchmark their methods against NicheNet and construct their own models (for example, with their own data sources to include more cell type specificity). The final networks, ligand–target matrix and processed expression data used in the vignettes are available at Zenodo (<https://doi.org/10.5281/zenodo.3260758>)<sup>39</sup>. The scripts used for processing of data sources and to carry out all analyses from this study are also available at Zenodo (<https://doi.org/10.5281/zenodo.3462199>).

## References

20. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2016).

21. Ramiłowski, J. A. et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
22. Pawson, A. J. et al. The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res.* **42**, D1098–D1106 (2014).
23. Rouillard, A. D. et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).
24. Türel, D., Korcsmáros, T. & Saez-Rodríguez, J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* **13**, 966–967 (2016).
25. Cerami, E. G. et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–D690 (2011).
26. Li, T. et al. A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2017).
27. Kamburov, A., Stelzl, U., Lehrach, H. & Herwig, R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* **41**, D793–D800 (2013).
28. Vinayagam, A. et al. A directed protein interaction network for investigating intracellular signal transduction. *Sci. Signal.* **4**, rs8–rs8 (2011).
29. Van Landeghem, S. et al. Exploring biomolecular literature with EVEX: connecting genes through events, homology, and indirect associations. *Adv. Bioinforma.* **2012**, e582765 (2012).
30. Lachmann, A. & Maayan, A. KEA: kinase enrichment analysis. *Bioinformatics* **25**, 684–686 (2009).
31. Linding, R. et al. NetworkKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699 (2008).
32. Duan, G., Li, X. & Köhn, M. The human DEPhOsporylation database DEPOD: a 2015 update. *Nucleic Acids Res.* **43**, D531–D535 (2015).
33. Liu, Z.-P., Wu, C., Miao, H. & Wu, H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* **2015**, bav095 (2015).
34. Han, H. et al. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* **5**, 11432 (2015).
35. Bovolenta, L. A., Acencio, M. L. & Lemke, N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* **13**, 405 (2012).
36. Griffon, A. et al. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res.* **43**, e27–e27 (2015).
37. Jojic, V. et al. Identification of transcriptional regulators in the mouse immune system. *Nat. Immunol.* **14**, 633–643 (2013).
38. Heng, T. S. P. et al. The immunological genome project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
39. Lachmann, A. et al. CheA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
40. Consortium, T. E. P. The ENCODE (ENCYclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
41. Mathelier, A. et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42**, D142–D147 (2014).
42. Matys, V. et al. TRANSFAC® and its module TRANSCompel™: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
43. Xie, X., Rigor, P. & Baldi, P. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics* **25**, 167–174 (2009).
44. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
45. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
46. Page, L., Brin, S., Motwani, R. & Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web* (Stanford InfoLab, 1999).
47. Jeh, G. & Widom, J. Scaling personalized web search, in *Proc. 12th International Conference on World Wide Web* (eds., G. Hencsey & B. White) 271–279 (ACM, 2003).
48. Kolesnikov, N. et al. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* **43**, D1113–D1116 (2015).
49. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostat. Oxf. Engl.* **4**, 249–264 (2003).
50. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
51. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
52. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
53. Michaud, J. et al. Integrative analysis of RUNX1 downstream pathways and target genes. *BMC Genomics* **9**, 363 (2008).
54. Janky, R. et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput. Biol.* **10**, e1003731 (2014).
55. Bischl, B. et al. mlrMBO: a modular framework for model-based optimization of expensive black-box functions. Preprint at <https://arxiv.org/abs/1703.03373> (2017).
56. Horn, D. & Bischl, B. Multi-objective parameter configuration of machine learning algorithms using model-based optimization. in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (eds., Y. Jin & S. Kollias) 1–8 (2016).
57. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002).
58. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
59. Browaeys, R., Saelens, W. & Saeys, Y. *Development, Evaluation and Application of NicheNet: Datasets Version 2* (Zenodo, 2019); <https://doi.org/10.5281/zenodo.3260758>

## Acknowledgements

R.B. and W.S. are supported by a PhD fellowship from The Research Foundation—Flanders (grant nos 1181318N to R.B. and 11Z4518N to W.S.). Y.S. is an ISAC Marylou Ingram scholar. This research received funding from the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” program. The authors would also like to thank G. Bader (University of Toronto) for his insightful comments on the manuscript, P. De Bleser (Ghent University) for providing us with the processed ChIP-seq data from ReMap, and M. Guilliams, W. T’Jonck, J. Bonnardel and C. Scott (Ghent University) for helpful discussions about this project.

## Author Contributions

R.B., W.S. and Y.S. conceived the method. R.B. and W.S. designed the experiments. R.B. performed the experiments and implemented the method. W.S. and Y.S. supervised the work. R.B., W.S. and Y.S. wrote and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41592-019-0667-5>.

**Correspondence and requests for materials** should be addressed to Y.S.

**Peer review information** Nicole Rusk and Lei Tang were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

## Software and code

Policy information about [availability of computer code](#)

Data collection

Information on data source collection and the scripts used for processing of data sources are available at Zenodo (10.5281/zenodo.3462199).

Data analysis

The data analysis was conducted using a custom software package available at GitHub: <https://github.com/saeyslab/nichenepr>. Scripts to carry out all analyses described in this study are available at Zenodo (10.5281/zenodo.3462199).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

No new data was generated for this study. All data used in this study is publicly available: accession codes for the public ligand treatment expression datasets are given in Supplementary Table 2, accession codes for the public case study datasets in the Online Methods sections “Collection and processing of the human single-cell transcriptomics data from Puram et al.” and “Collection and processing of the mouse single-cell transcriptomics data from Medaglia et al.”.

The data source networks, NicheNet’s ligand-target matrix, and the processed ligand treatment and case study expression datasets are available on Zenodo (<https://doi.org/10.5281/zenodo.3260758>). Databases used to create the NicheNet model are mentioned and referred to in the Online Methods section “Collection of ligand-receptor, intracellular signaling, and gene regulatory interactions”, Supplementary Table 1, and in the Supplementary Notes.

Source data underlying Figure 1 and 2 are accessible from the respective figure legends.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

In our case the number of datasets used to evaluate the model performance corresponds to the sample size. We tried to find ligand treatment datasets for as many different ligands as possible. When searching for expression datasets, only datasets with > 20 and < 1500 differentially expressed genes were used for evaluation in this study. In total, we collected 111 ligand treatment datasets for evaluation. We did not perform a power analysis in advance, but think this number of datasets is sufficient because: a large number of different ligands could be evaluated (51), for 10 ligands we have 3 or more datasets, and for 8 of these ligands, we have datasets profiling the response in 2 or more different cell types. Moreover, in a previous version of the manuscript, we analyzed only 54 datasets and main results and conclusions were identical.

Data exclusions

No data was excluded from the study.

Replication

To ensure robustness and reproducibility of the findings, we (1) made sure that the comparison of NicheNet to randomized networks is robust by generating 100 different randomized networks; (2) demonstrate the usability of NicheNet on different types of data: we applied NicheNet on: in vitro ligand treatment datasets (bulk transcriptomics), human single-cell RNA-seq data of the tumor microenvironment and mouse single-cell RNA-seq data on lymph nodes as generated by the NICHE-SEQ method; (3) we performed multiple rounds of cross-validation when assessing performance of classifiers trained via cross-validation.

Randomization

This is not relevant to our study because we do not include separate experimental groups.

Blinding

This is not relevant to our study because we do not include separate experimental groups.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involved in the study                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data               |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |