nature research

Corresponding author(s): Yvan Saeys

Last updated by author(s): Sep 26, 2019

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/nichenetr. 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

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging