Lignature provides a curated resource of ligand induced transcriptomic signatures for signaling inference

Ying Xin[1#], Md Amanullah[1#], Cheng Qian[1#], Chingmo Zhou[1], Jiang Qian[1,2*]

Author Affiliations

[1] Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[2] The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

# These authors contributed equally to this work.

* Corresponding author: jiang.qian@jhmi.edu

Running Title: Lignature for Signaling Inference

Abstract

Ligand-receptor interactions mediate intercellular communication, inducing transcriptional changes that regulate physiological and pathological processes. Ligand-induced transcriptomic signatures can be used to infer ligand activity; however, the absence of a comprehensive set of ligand-response signatures has limited their practical application in predicting ligand-receptor interactions. To bridge this gap, we develop Lignature, a curated database encompassing intracellular transcriptomic signatures for 362 human ligands, significantly expanding the repertoire of ligands with available intracellular response signatures such as CytoSig and ImmuneDictionary. Lignature compiles signatures from published transcriptomic datasets, generating both gene- and pathway-based signatures for each ligand. We apply Lignature to prioritize ligand-associated transcriptional activity in controlled *in vitro* experiments and real-world single-cell sequencing datasets. Across these settings, Lignature consistently improves the prioritization of experimentally supported ligands compared with existing approaches. We additionally develop a regression-based framework to model combinatorial regulation by multiple ligands. These results establish Lignature as a robust platform for ligand signaling inference, providing a powerful tool to explore ligand-receptor interactions across diverse experimental and physiological contexts.

Introduction

Cell-cell communication is essential for coordinating biological processes in multicellular organisms, with ligand-receptor (LR) interactions serving as a primary mechanism in this

process(Armingol et al. 2021; Armingol et al. 2024; Su et al. 2024). The rapid advancement of sequencing technologies has enabled the development of numerous computational methods for predicting potential LR interactions from transcriptomic data.

Current methods for predicting cell-cell communication, particularly LR interactions, fall into three major categories. The first category identifies LR interactions based on the co-expression of ligand genes in sender cells and receptor genes in receiver cells. Tools such as CellPhoneDB(Efremova et al. 2020), CellChat(Jin et al. 2021), SingleCellSignalR(Cabello-Aguilar et al. 2020), NATMI(Hou et al. 2020), and ICELLNET(Massenet-Regad et al. 2021) exemplify this approach. While these methods are effective, they often generate a high number of false positives by relying solely on ligand and receptor gene expression levels.

To enhance accuracy, a second type of approach integrates intracellular networks to model the impact of ligands on target genes in receiver cells. Tools like CCCExplorer(Yeung et al. 2019), NicheNet(Browaeys et al. 2020), SoptSC(Wang et al. 2019), and LRLoop(Xin et al. 2022; Qian et al. 2024) predict LR interactions by comparing observed transcriptional changes with predicted target genes associated with each ligand. However, the incomplete and potentially noisy nature of current signaling and regulatory networks leads to inaccuracies in target gene prediction. Additionally, these methods fail to model the direction and magnitude of target gene expression changes in response to specific ligands.

A third group of methods, including SpaOTsc(Cang and Nie 2020) and COMMOT(Cang et al. 2023), incorporates spatial information to better account for potential signaling between cells, further refining predictions of cell-cell communication. However, these

methods rely on spatial transcriptomics data and do not resolve the inherent challenges of expression-based prediction.

In this work, we propose a new approach to predict the specific ligands responsible for the transcriptomic changes within the receiver cells. Instead of merely modeling the target genes of ligands, our method will identify the signatures of ligands by measuring the transcriptomic changes derived from actual gene expression data. When cells are treated with a given ligand, and their gene expression is measured before and after treatment, we can then create a distinct signature unique to that particular ligand. It's worth noting that the concept of ligand signatures is widely accepted in the field. Many researchers have identified the “responsive genes” of the ligands(Sanda et al. 2006; Dölken et al. 2008; Qin et al. 2009; Sohn et al. 2010; Cecchi et al. 2015). Previous works such as CytoSig(Jiang et al. 2021) and ImmuneDictionary(Cui et al. 2024) indeed collect the signatures of ligands, while PROGENy leverages publicly perturbation experiments for modeling oncogenic and tumor suppressor signaling footprints(Schubert et al. 2018). However, they either focus exclusively on cytokines, or on tumor related pathways, limiting the practical applications of their datasets to the prediction of cell-cell communication. Here we provide a comprehensive ligand signature database, Lignature, which contains the signatures for 362 ligands across diverse categories. Multiple signatures are retained for individual ligands to capture potential cell type–specific transcriptional responses induced by the same ligand. Additionally, we constructed gene- and pathway-based signatures for these ligands. We validated these signatures using in vitro experiments and single-cell sequencing datasets, demonstrating their reliability. The accompanying R(Team *2023*)-based computational tool further enables the signature-

based ligand–transcriptional program association and corresponding ligand-receptor interaction predictions.

Results

Overall design

The transcriptomic changes induced by a specific ligand are referred to as the ligand's signatures. By comparing observed gene expression changes to a comprehensive library of ligand signatures, it is possible to predict which ligands are likely to induce observed gene expression changes. To achieve this, we collected transcriptomic datasets where gene expression profiles were measured before and after treatment with specific ligands. These datasets were used to construct a library of ligand signatures based on the resulting transcriptomic changes.

Recognizing that ligand-induced signatures can vary depending on cell types and experimental conditions, we included multiple signatures for a given ligand when such datasets were available. Additionally, we developed a companion software tool designed to predict and visualize transcriptional program-associated ligands by leveraging this curated collection of ligand signatures.

Generation of the Lignature database

We developed an automatic pipeline to search Gene Expression Omnibus (GEO)(Edgar et al. 2002) for datasets with an experimental design that would potentially contain transcriptomic responses to these ligands. Starting with the metadata of all 25,868

platform identifiers downloaded from GEO, we collected metadata for 72,738 data-series corresponding to Homo sapiens platforms and molecule type “total RNA”, “poly(A) RNA”, or “nuclear RNA”, where each data-series could include multiple datasets. The curated metadata of the GEO data-series was then queried with a predefined set of 859 ligands based on our previous work LRLoop(Xin et al. 2022). We searched the metadata with ligand genes’ full names, symbols, and aliases, together with a set of 105 experiment keywords such as “ligand”, “receptor”, “treatment”, “stimulate”, “overexpress”, “block”, “knockdown”, and “silence”, resulted in a list of 9,187 data-series for manual curation. Finally, after carefully examining the collected data-series in GEO, we removed those that were not relevant to our purpose or exhibited technical issues such as the use of non-genome-wide platforms, ambiguous data description, or the lack of accessible raw/processed data. This curation yielded a list of 460 data-series, containing transcriptomic response data for 213 ligands that were not included in the databases CytoSig(Jiang et al. 2021) and ImmuneDictionary(Cui et al. 2024) (Fig. 1A).

The transcriptomic datasets, including gene expression array and RNA-seq datasets, were then processed for differential expression analysis (see Methods for details), vectors of $\log_2$FC (fold change) values for expressed genes were collected as the gene-level ligand signatures. For datasets with multiple treatment conditions (e.g., durations or doses), a separate signature was created for each different condition. After combining the data of 149 ligands from CytoSig(Jiang et al. 2021) and ImmuneDictionary(Cui et al. 2024), we obtain the datasets for 362 ligands (Fig. 1B). In addition to the gene-level response signatures, we also calculated KEGG(Kanehisa and Goto 2000; Kanehisa 2019; Kanehisa et al. 2025) pathway signatures for each ligand. Specifically, we performed

Gene Set Enrichment Analysis across 302 KEGG pathways (each containing 10–500 genes) and calculated the Normalized Enrichment Score (NES) for each pathway. The resulting NES vectors were compiled as the pathway-level signatures. (Fig. 1A). For each signature, we comprehensively summarized information on cell types, organisms, sources of the cell type or tissue, and the conditions pertaining to the ligand treatment dose and duration (Supplemental Table S1).

The 362 ligands with signatures in Lignature are classified into 100 groups based on the HUGO Gene Nomenclature Committee categories(Seal et al. 2023), including a group labeled “Other” consisting of 64 ligands that either have no gene groups, or have only very generalized (such as “Receptor ligands”) or biochemical structure-focused (such as “WAP four-disulfide core domain containing”) gene groups specified in HGNC. The largest ligand groups include CD molecules, interleukins, neuropeptides, TNF superfamily, and chemokine ligands (Fig. 1C, Supplemental Table S2). This classification also highlights data source biases; CytoSig and ImmuneDictionary primarily focus on interleukins, TNF superfamily, interferons, and fibroblast growth factors, while our newly curated data cover a broader range, including most CD molecules, neuropeptides, Ig-like cell adhesion molecules, and semaphorins.

Using the Lignature database as a reference, we provided a companion computational tool that infers the ligands or ligand combinations responsible for transcriptomic changes within receiver cells, and the corresponding cell-cell interaction networks between multiple cell types or clusters. The tool provides multiple options for calculating expression-based LR interaction strength, providing an intuitive summary of the predicted LR activities across different conditions (Fig. 1D).

Overall properties of the ligand signatures

A single ligand may have multiple signatures, as it may produce different responses in various cell types with diverse treatment durations and concentrations. In fact, across the total of 4215 signatures corresponding to 362 ligands, the number of signatures per ligand ranges from 1 to 271, with TGFB1 having the largest number of signatures (Fig. 1E). We systematically compared the similarity of signatures of the same ligand with those from different ligands by calculating Pearson correlation coefficients for all signature pairs (Fig. 2A). The mean correlation coefficients between signatures of the same ligand are 0.174, 0.161, and 0.064 in Lignature, CytoSig, and ImmuneDictionary, respectively. In contrast, the mean correlation coefficients between signatures of different ligands are 0.003, 0.022, and 0.058 for Lignature, CytoSig, and ImmuneDictionary, respectively. This analysis indicates that signatures derived from the same ligand tend to exhibit higher similarity than those from different ligands. Correlation analysis of five representative ligands, TNF, TGFB1, SHH, IFNA (family), and IFNG, further supports this observation (Supplemental Fig. S1A).

Although signatures generated by the same ligand exhibit higher similarity than those from different ligands, the same ligand can still induce diverse responses across cell types, treatment conditions, and other biological contexts. To further investigate the sources of this diversity, we analyzed signature variation within individual ligands. For example, TGFB1 signatures show that profiles derived from the same cell types or cell lines (e.g., fibroblasts, A549, MCF10), whether from the same or different studies, tend to cluster together, whereas signatures from different cell types display substantial variation (Fig. 2B). Similarly, SHH has more than twenty signatures from datasets across different

organisms, cell types, treatment modalities, doses, durations, and platforms. Signatures derived from the same cell types, such as mouse neuralized embryoid bodies or medulloblastoma cells, consistently cluster together, while signatures across distinct cell types show context-dependent variation (Fig. 2B). Together, these analyses indicate that cell type is a major source of signature diversity for a given ligand.

To confirm the biological relevance of the ligand signatures, we calculated the enrichment of upregulated and downregulated genes of ligand signatures in KEGG pathways. For instance, the enriched pathways for the signature genes of ligand LEP (which encodes leptin) include allograft rejection, intestinal immune network for IgA production, graft-versus-host disease, complement and coagulation cascades. These pathways are related to LEP's involvement in inflammatory processes(Molinero et al. 2016; Cui et al. 2017; Scheja and Heeren 2019). Conversely, the TP53 pathway is enriched in downregulated genes, consistent with leptin's inhibitory effect on cell cycle processes by decreasing the key TP53 proliferation regulator(Toro et al. 2014) (Fig. 2C). Similarly, the signature genes of ligand TNFSF13 shows enrichment of upregulated genes in NF-κB signaling, apoptosis, TNF signaling pathway and other pathways that are highly consistent with the known cellular functions of TNFSF13(Baert et al. 2018). It is worth noting that TNFSF13 is also closely associated with allograft rejection in organ transplantation(Jo et al. 2019) (Fig. 2C, for more examples, see Supplemental Fig. S1B). These enriched pathways underscore the biological significance of the ligand signatures and can be used to construct pathway-based signatures and predict the active ligands.

Finally, we examined whether ligands from the same family or with similar biological functions tend to induce similar transcriptional responses. To do this, we generated a

merged signature for each ligand and performed hierarchical clustering based on the similarity of these merged signatures. This analysis revealed several coherent groups corresponding to well-established ligand families (Supplemental Fig. S1C). For example, members of the TGF-beta family, including TGFB1, TGFB2, and TGFB3, and BMP2, BMP4, and BMP6, clustered tightly, consistent with their convergent downstream signaling pathways(Miyazono et al. 2001; Elsafadi et al. 2019; Wu et al. 2024). Likewise, most interferons (IFNs) clustered together, and many interleukins (ILs) showed relatively close similarity. In addition, TNF, IL1B, and CD40LG, which all regulate NF-κB signaling, grouped together(Lawrence 2009; Hostager and Bishop 2013; Jiang et al. 2021). In contrast, CD (cluster of differentiation) molecules, whose nomenclature is based on structural or localization features rather than shared functions, were dispersed across the clustering, reflecting their broad functional diversity.

Assessment in signaling inference with bulk *in vitro* datasets

We assessed the ability of Lignature to predict active ligands using ligand signatures. To this end, we collected and processed additional transcriptomic datasets as test cases, where known ligands were used to treat cell lines, and transcriptomic changes were measured before and after treatment. The goal was to predict the treatment ligands based on these observed transcriptomic changes. A total of 90 test datasets were used, including 24 human and 66 mouse datasets, representing 27 unique treatment ligands (Fig. 3A). Note that the test datasets were not included in the initial Lignature database used to generate ligand signatures, and the same test datasets were used when comparing Lignature with other existing methods.

For each test dataset, we measured the similarity between the observed transcriptomic changes and signatures in the Lignature library by the Pearson correlation coefficients. The signatures were ranked based on these similarities, and the ranked position of the treatment ligand was examined. For example, in a dataset treated with ligand IFNG, the transcriptomic changes were compared against the signatures of the ligands in Lignature. Both the gene-level and pathway-level signatures for ligand IFNG exhibit high correlation coefficients with the test data, accurately identifying it as an acting ligand responsible for the observed transcriptomic changes (ranked number 1; Fig. 3B). Three other methods for ligand activity prediction based on receiver-cell responses were compared to Lignature. NicheNet ranked IFNG number 11, lower than the prediction from Lignature yet a reasonably high rank for identifying IFNG as a treatment ligand. Another method iTALK that predicts LR interactions based on differential expression of receptor genes alone ranked IFNG much lower at number 279 and 306, respectively, when scoring each ligand by the maximum or mean absolute $\log_2$FC values of its cognate receptors (denoted by maxRlfc and meanRlfc) (Fig. 3B).

Across all 90 test datasets, the median rank of treatment ligands was 3.5 and 8.5 for predictions from Lignature gene- and pathway-level signatures (Fig. 3C), the median AUC values of the corresponding ROC (receiver operating characteristic) curves are 0.989 and 0.951 (Fig. 3D), and the AUC values of the combined ROC curves from predictions for all 90 test datasets are 0.925 and 0.883, respectively (Fig. 3E). Applying NicheNet, iTALK(maxRlfc), and iTALK(meanRlfc) to the same datasets yielded median ranks of 33.5, 64.5, and 78 for the treatment ligands (Fig. 3C), median AUC values of the corresponding ROC curves 0.843, 0.817, and 0.792 (Fig. 3D), and AUC values of combined ROC curves

0.685, 0.749, and 0.713 (Fig. 3E). In this well-controlled *in vitro* setting, Lignature shows significantly enhanced ability to prioritize experimentally supported ligand-transcriptional program associations relative to existing methods.

We next asked whether the signatures from one species can be applied to another. Since Lignature contains mostly signatures from human datasets, we compared its performance on 24 human and 66 mouse test datasets to assess cross-species applicability (Supplemental Fig. S2A-B). In human test datasets, the median rank of the treatment ligands was 2 for both gene- and pathway-level signatures. In mouse test datasets, the corresponding median ranks were 4 and 13, respectively. These comparable performances indicate that transcriptional responses to homologous ligands are largely conserved across species, enabling effective cross-species inference of transcriptional program-associated ligands.

We next examined whether the number of available signatures for a ligand influenced prediction performance. Our analysis revealed a weak-to-moderate negative correlation (–0.39) between signature library size and prediction rank, indicating that ligands with more signatures tend to achieve better prediction performance. However, this trend was not absolute: ligands with only a few signatures often ranked highly, whereas some ligands with large signature libraries performed only moderately (Supplemental Fig. S2C).

Combinatorial regulation

We next evaluated whether ligand signatures could predict combinatorial ligand effects. Our analysis used 38 datasets in which cells were treated with pairs of ligands drawn from 13 distinct factors, across varying doses and treatment durations. When each

ligand was considered independently, the signatures frequently identified either both treatment ligands or one of them, with the latter likely reflecting cases where a single ligand dominated the transcriptional response (Supplemental Fig. S3A–B).

To more directly address the challenge of deconvoluting combinatorial signaling, we developed a dedicated feature that applies regularized linear regression (Ridge, LASSO, or Elastic Net) to model a differential expression profile as a weighted combination of multiple ligand activities. Using LASSO regression on the 38 dual-ligand datasets, this combinatorial approach improved the ranking of the true treatment ligands relative to individual signature matching at both the gene and pathway levels (Supplemental Fig. S3A–B). These results indicate that ligand signatures can be extended to predict more complex combinatorial signaling interactions.

Application and assessment of Lignature with scRNA-seq datasets

We then applied Lignature to two published scRNA-seq datasets that preserved more natural *in vivo* conditions involving multiple cell types.

Example 1: SHH-Induced Thalamic Differentiation To investigate SHH-mediated thalamic differentiation, organoids were treated with recombinant SHH, and scRNA-seq was performed on samples collected on day 70 with or without SHH treatment(Kiral et al. 2023). Major cell clusters were identified, including excitatory neurons (ExNs), inhibitory neurons (INs), astrocytes (ASs), endothelial cells (ECs), and ependymal cells (Epn) (Fig. 4A-B). Since SHH activation promotes the generation of INs, we focused on the IN cluster from the SHH day 14–22 treatment sample compared to the untreated sample.

Candidate ligands in Lignature were first filtered based on the detection rate of their receptors, retaining ligands with at least one receptor detected at a rate above 0.1. Using Lignature (gene-level signatures), SHH was ranked as one of the top ligands (number four) predicted to be transcriptional program-associated, inducing the observed transcriptomic changes (green dots in Fig. 4C, Supplemental Fig. S4A). Three ligands, VEGFA, IFNG, and IGF2 ranked higher than SHH. Among them, IFNG can be reasonably excluded, because the expression level of *IFNG* in all cells are zero (Fig. 4C). VEGFA and IGF2 are known to be the downstream regulators of SHH, which play a role in neurogenesis, neuroprotection, and the processes of synaptic formation throughout neural development(Lawson et al. 2002; Rosenstein et al. 2010; Mackenzie and Ruhrberg 2012; Alberini 2023; Pandey et al. 2023). This result suggests that Lignature captures the transcriptomic response from the signaling cascade initiated by SHH. In contrast, with the same set of candidate ligands and filtering criteria, NicheNet, iTALK (maxRlfc) and a receptor expression-based method (maximum expression of the cognate receptors) ranked SHH much lower at number 37, 114 and 126, respectively. In summary, Lignature successfully prioritized SHH, an exogenously introduced ligand, as one of the active ligands associated with transcriptomic changes between treated and untreated samples, demonstrating improved performance relative to other methods in linking ligand activity to transcriptional programs during induced IN generation.

Example 2: Microglia and Müller Glia (MG) Regeneration Microglia cells have been implicated in suppressing Müller glia (MG) regeneration capacity, as microglia ablation improved retinal neurogenesis from MG(Todd et al. 2020). Furthermore, treatment with TNF in microglia-ablated mice suppressed the enhanced neurogenesis observed in the

absence of microglia(Todd et al. 2020), suggesting that microglia-originated TNF might inhibit MG regeneration. scRNA-seq was performed on microglia and MG before and after microglia ablation(Todd et al. 2020) (Fig. 5A-B), and we aimed to predict the ligands responsible for the transcriptomic changes in MG.

Application of Lignature to the scRNA data in MG predicted TNF as the top ranked ligand (green dots in Fig. 5C, Supplemental Fig. S4B). In contrast, when we used NicheNet to assess the activities of the same ligands, TNF was positioned number 37. Several ligands highly ranked by NicheNet, such as IL1A, GIP, IL17F, C5, and LIF, were assigned lower rankings by Lignature. Upon examining the expression levels of these ligands in the sender cells (microglia) and their receptor's expression in MG before microglia ablation, we found that either the ligands or their receptors were undetectable or detected in a very limited portion of the cells, indicating they are unlikely to be the major contributors to the transcriptomic changes in MG. Moreover, iTALK and a receptor expression-based method ranked TNF number 148 and 42, failed in ranking TNF highly as a transcriptional program-associated ligand for the transcriptomic changes in MG.

We next used CellPhoneDB, NATMI, SingleCellSignalR, and CellChat, four well established gene expression-based methods for predicting ligand-receptor interactions, to identify active ligands responsible for the transcriptomic changes in MG. The ligands were ranked based on their maximum interaction strength with corresponding receptors. TNF was ranked 17th, 7th, 7th, and 6th by CellPhoneDB, NATMI, SingleCellSignalR, and CellChat, respectively (Fig. 5D and Supplemental Fig. S4C). This analysis suggests that gene expression alone might not be sufficient for the identification of transcriptional

program-associated ligands, while in practice, filtering the candidate ligands by the gene expression levels would help reduce noises (Fig. 5D).

Usage of Lignature

We demonstrate the usage of Lignature by applying a scRNA-seq dataset of immune cell populations in checkpoint blockade-associated colitis and control samples(Luoma et al. 2020). Our program takes the Seurat object, specified receiver and sender cell types, and a pair of sample conditions as input. In this case study, we focused on the transcriptomic changes between samples from checkpoint inhibitor treated melanoma patients with and without histologically colitis, labeled “colitis” and “NC”, in myeloid, with all cell types in the dataset as potential ligand sending cells (Fig. 6A-B).

Lignature scores the candidate ligands based on users’ choice of the similarity measurement metric (e.g., Pearson correlation coefficient, Spearman’s correlation coefficient, Cosine Similarity, Euclidean distance) or signature weights from combinatorial analysis by regularized linear regression (Ridge, LASSO, or Elastic Net). Here we used the Pearson correlation coefficient, which was set as default in Lignature. In parallel, Lignature calculates the LR interaction strengths using specified methods such as CellPhoneDB, SingleCellSignalR, or NATMI, which was used in this case. Transcriptional program-associated ligands can be then predicted based on both Lignature score and LR interaction strength (Fig. 6C). Lignature also provides options for filtering the list of LR pairs by gene expression, differential expression, Lignature score, confidence score, interaction strength and specificity. Finally, Lignature produces various plots displaying enriched pathways, inferred ligands, and the associated LR interactions, providing intuitive visualizations for the analysis (Fig. 6B-F and Supplemental Fig. S5). Through this

analysis, Lignature successfully identified transcriptional program-associated ligand signaling to myeloid cells such as IFNG and TNF from T cells, as well as potential therapeutic targets for checkpoint inhibitor-induced colitis such as TNF.

Discussion

Understanding cell-cell communication is fundamental in decoding complex biological processes, and ligand-receptor interactions play a pivotal role in mediating these signals. In this study, we introduced Lignature, a comprehensive database of ligand signatures derived from experimental data, coupled with a predictive framework for prioritizing transcriptional program-associated ligands based on observed transcriptomic changes. By leveraging this resource, we address several limitations of existing computational methods, such as the challenge of ligand-transcriptional program association prioritization from gene expression-based approaches and biases introduced by incomplete signaling network models.

Compared to network-based methods such as NicheNet, Lignature captures not only the target genes of the ligands, but also the direction (i.e., up- and downregulation) and the magnitude of the expression changes of the target genes. Furthermore, it also captures cell type- and condition-specific transcriptomic changes. Our findings demonstrate that Lignature consistently outperforms existing methods, including NicheNet and iTALK, in prioritizing experimentally supported ligand–transcriptional program association. For example, in both controlled *in vitro* datasets and more complex *in vivo* scRNA-seq

datasets, Lignature was able to rank the actual treatment ligands or experimentally validated biologically relevant ligands at the top.

Ligand–receptor redundancy, where different ligands can interact with the same receptor, and a single ligand can bind to multiple receptors and activate distinct pathways, is prevalent in biological systems. This redundancy enables both convergent functional synergies and divergent, context-specific responses. For example, lymphotoxin proteins in the TNF superfamily, such as LTA and LTB, can bind TNF receptors and trigger intracellular signaling cascades similar to other TNF superfamily ligands(Seleznik et al. 2014). Likewise, certain interleukins share convergent signaling pathways with specific interferons; for instance, IL2, IL3, IL6, and IFNE, IFNK all activate the JAK-STAT signaling pathway(O'Shea et al. 2015; Philips et al. 2022). Consistent with these known relationships, our clustering analysis positioned LTA/B, several other TNF superfamily members, interleukins, and interferons in close proximity (Supplemental Fig. S1C). Conversely, some ligands, such as growth factors like EGF, can activate multiple signaling pathways or pathway branches, including PI3K/Akt/mTOR, JAK-STAT, and MAPK/ERK. Such versatility allows ligands to induce divergent functions depending on tissue- and cell-type-specific contexts and supports our decision to retain multiple signatures for a single ligand.

Despite the strong performance of Lignature across our analyses, several limitations remain. First, a fundamental challenge in benchmarking ligand-receptor inference methods is the lack of definitive ground truth and the biological complexity arising from multiple ligands acting simultaneously within tissues. Accordingly, although Lignature can prioritize experimentally supported ligands, other highly ranked ligands predicted by

alternative methods cannot be simply classified as false positives, particularly in single-cell contexts where concurrent signaling is common. Second, because Lignature infers activity from downstream transcriptional programs, it may prioritize ligands with convergent signaling effects, even in the absence of detectable ligand expression. For example, in the SHH dataset, Lignature ranked IFNG among the top predicted ligands, despite *IFNG* expression being undetectable across cells. Therefore, Lignature is not intended to function as a standalone method for definitive ligand attribution. Instead, we recommend interpreting Lignature-derived activity scores in conjunction with ligand and receptor expression–based methods, which can help refine and validate candidate ligands, particularly in settings where specificity is critical. Third, the current version of Lignature depends on condition-based comparisons to extract ligand-induced transcriptional signatures and is therefore not directly applicable when matched conditions are unavailable, such as in many human tissue datasets. In these cases, users can still apply Lignature by defining biologically meaningful contrasts within a single dataset, for example, leveraging cell type labels, clinical groupings, or positions along a developmental trajectory to construct pseudo-conditions for comparison.

Several future directions could further strengthen Lignature's coverage and usability. First, although Lignature retains predictive capability even for unseen cell types, expanding the library as additional transcriptomic datasets become available across diverse tissues/cell types and pathological conditions will improve the overall predictive power of the database. Second, extending signatures to more species will enhance cross-species generalizability. Although Lignature achieved comparable performance on human and mouse datasets in benchmarking analyses, its performance was more robust on human

datasets, likely due to the predominance of human-derived signatures in the current database. Third, incorporating non-peptide ligands, such as neurotransmitters, will broaden the range of signaling activities that can be accurately inferred(Xin et al. 2025). Finally, while we provide companion software with Lignature to predict ligand-transcriptional program associations, Lignature can also function as a standalone database, both integrate easily with other analytical tools. For example, Lignature can be combined with existing spatial transcriptomics workflows. Lignature infers ligand-transcriptional program associations from transcriptional changes in receiver cells, whereas spatial analyses can evaluate the physical plausibility of inferred interactions based on co-localization patterns.

In summary, Lignature represents a significant advancement in the field of cell-cell communication, providing a robust, experimentally validated resource for inferring ligand-transcriptional program association. Its ability to address the limitations of existing computational methods while offering mechanistic insights underscores its potential as a valuable tool for both basic and translational research.

Methods

Datasets collection

We collected 859 human ligands based on our previous work LRLoop(Xin et al. 2022). We then developed an automatic pipeline to search Gene Expression Omnibus (GEO)(Edgar et al. 2002) for datasets with an experimental design that would potentially contain transcriptomic responses to each ligand.

Specifically, our collection of the relevant datasets includes the following three steps. We started by the list of all 25,868 platform accession identifiers downloaded from GEO and collected the metadata for each of them using the R package “GEOquery”(Davis and Meltzer 2007), filtered the list by restricting the organism to “Homo sapiens”, and collected the list of 97,693 data-series identifiers corresponding to the 6,278 filtered platforms, where each data-series could include multiple datasets. Then again using GEOquery, we curated metadata for the collected data-series including data “type”, “organism”, “molecule”, “platform id”, “cell line”, “cell type”, “tissue”, as well as data “title”, “abstract”, “url”, “geo accession”, “overall design”, “sample taxid”, “summary”, etc. The list of data-series was then filtered by requiring the “molecule” entry of the corresponding metadata to include the strings “total RNA”, “polyA RNA” (for poly(A) RNA) or “nuclear RNA”, resulted in a list of 72,738 data-series identifiers for further systematic searching of ligand and experiment keywords.

In order to search the curated metadata of the GEO series systematically for potential ligand signatures, we collected a set of “ligand keywords” including gene full names, symbols, and aliases for each ligand, as well as a set of 105 experiment keywords such as “ligand”, “receptor”, “treatment”, “stimulate”, “activate”, “expose”, “block”, “overexpress”, “knockout”, “knockdown”, “silence”, “inhibit” and “in the presence of”. The curated metadata of the GEO series list was then queried with the ligand keywords and experiment keywords. This automatic searching process reduced the number of candidate data-series in the list to 9,187 for manual curation.

Finally, through careful manual review of the GEO data series, we excluded entries that were irrelevant to our aims or presented technical limitations, such as the use of non–

genome-wide platforms, unclear experimental descriptions, or missing raw or processed data. This curation resulted in 460 high-quality data series, providing transcriptomic response profiles for 213 ligands not represented in CytoSig(Jiang et al. 2021) and ImmuneDictionary(Cui et al. 2024) (Fig. 1A).

In addition, to create a comprehensive ligand signaling signature database, we took advantage of CytoSig(Jiang et al. 2021) and ImmuneDictionary(Cui et al. 2024) by including the cytokine signatures from these sources into Lignature, and included the signatures of the 90 test bulk datasets after method assessment by these test datasets.

Data processing

We have two major data types: gene expression array and RNA-seq datasets. For Affymetrix array data with raw data .cel files provided, we adopted the "maEndtoEnd"(Klaus and Reisenauer 2016) workflow, which generates and preprocesses expression data using the R package "Oligo"(Carvalho and Irizarry 2010) and perform differential expression analysis using the "limma"(Ritchie et al. 2015) package. For other forms of array data, we used processed data and performed differential expression analysis by "limma". Briefly, after RMA calibration using the "oligo::rma" function, an intensity-based filtering was performed by removing transcripts without intensities higher than a threshold of two in at least the number of arrays of the smallest experimental group, followed by annotation of the transcript clusters with gene symbols and removal of multiple mappings. Subsequently, a contrast matrix was constructed using functions "model.matrix" and "makeContrasts" and used as input of a standard differential expression analysis pipeline in limma, i.e., "lmFit", "contrasts.fit", and

472 “eBayes”. Then for each gene, the most significant ProbeSet corresponding to it with the
473 lowest adjusted p.value was identified and used as the representative of that gene.

474 RNA-seq datasets were processed with a combination of “edgeR”(Robinson et al. 2010)
475 and “limma”. In brief, a DGEList object was created using the function “DGEList” with the
476 expression count matrix, genes were annotated by gene symbols. and lowly expressed
477 genes were identified for removal by the “filterByExpr” function with default settings.
478 Following gene expression normalization through “calcNormFactors” with the “TMM”
479 method, the standard limma-voom pipeline was applied for differential expression
480 analysis.

481 After the differential expression analysis, different types of gene identifiers were converted
482 to gene symbols, and the vectors of log_2FC values of expressed genes were collected as
483 the gene-level ligand signature. For datasets with multiple treatment conditions, e.g.,
484 durations or doses, a separate signature was created for each different treatment
485 condition.

486 In addition to the gene-level signatures, we also calculated KEGG pathway-level
487 signatures for each ligand. Specifically, we collected gene sets of 302 KEGG pathways
488 with 10 ~ 500 genes by the R package “graphite”(Sales et al. 2012). Then for each gene-
489 level signature, we pre-sorted the log_2FC vector in descending order and performed Gene
490 Set Enrichment Analysis using the R package “FGSEA”(Korotkevich et al. 2019),
491 calculated the Normalized Enrichment Score (NES) for the KEGG pathways. The vectors
492 of NES were collected as the pathway-level signature (Fig. 1A).

493 **Prediction of active ligands based on ligand signatures**

Using the Lignature database as a reference for ligand signaling responses, we provided a companion R (Team *2023*)-based computational tool that infers the ligands responsible for observed transcriptomic changes within receiver cells, and the corresponding cell-cell interaction networks between multiple cell types or clusters. For a dataset of interest, taking matrices resulted from differential expression analysis, or the Seurat(Hao et al. 2024) object of single cell RNA-seq datasets as input, Lignature scores each signature in the database by the similarity compared to the input data. Options of the similarity measurement metrics include Pearson correlation coefficient (default), Spearman's correlation coefficient, cosine similarity, Euclidean distance, Manhattan distance, and the Euclidean or Manhattan distance normalized by vector lengths. In addition, Lignature features the option to perform combinatorial analysis by regularized linear regression (Ridge, LASSO, or Elastic Net) using R packages "caret"(Kuhn 2008) and "glmnet"(Friedman et al. 2010; Tay et al. 2023).

While potential active ligands are ranked using the similarity scores between observed gene expression changes and ligand signatures, we also provide two measures of statistical significance. First, permutation-based *p* values are calculated, allowing users to define a threshold for ligand scoring. Second, users may generate a confidence score for each ligand, defined as the proportion of similarity scores in a null distribution, which is constructed from comparisons between signatures of different ligands, that fall below the ligand's observed score. For example, a ligand score of 0.1 with a confidence score of 0.95 indicates that the score exceeds 95% of the null distribution.

When we perform cross-species prediction, genes from mouse datasets were converted to human gene symbols using their human homologs. Furthermore, Lignature provides

multiple options for calculating expression-based LR interaction strength such as CellPhoneDB, NATMI, and SingleCellSignalR. In addition, users are provided the choices of filtering the LR-interactions by the ligand scores, confidence scores, expression or differential expression of the ligands and their receptors, interaction strength and the interaction-specificities. Lastly, Lignature features the options to visualize predicted ligand scores, expression of the ligands and their receptors, as well as inferred strengths of LR interactions leveraging R packages “ggplot2”(Wickham 2016), “pheatmap”(Kolde 2025), and “circlize”(Gu et al. 2014), providing an intuitive summary of the predicted LR activities across different conditions (Fig. 1D).

Comparison with other methods

Five other different methods of ligand activity prediction were compared to Lignature. Specifically, for each test dataset, “NicheNet” ligand activities were calculated by the R package “nichenetr” according to its recommended parameter settings, i.e., expressed genes with $|\log_2 FC| >= 1$ and adjusted p.value < 0.1 were used as differentially expressed genes; The method “iTALK [maxRlfc]” scored each ligand by the maximum absolute $\log_2 FC$ value of the cognate receptors; “iTALK [meanRlfc]” scored each ligand by the arithmetic mean of its receptors’ absolute $\log_2 FC$ values; “Receptor expression [max]” scored each ligand by the maximum expression value of the cognate receptors; and “Receptor expression [mean]” scored each ligand by the mean expression value of the cognate receptors. On the other hand, with Lignature, “Lignature [gene $\log_2 FC$]” scored each ligand by the signature similarity scores with the gene-level signatures, while “Lignature [pathway NES]” scored each ligand by the signature similarity scores using the

pathway-level signatures, with the similarities measured by the Pearson correlation coefficients.

All methods were evaluated using the same test dataset, which consists of 90 ligand treatment datasets. These datasets were originally collected and used in NicheNet and were not included in the initial Lignature database. Known treatment ligands were labeled "TRUE" and other candidate ligands were labeled "FALSE". Ranks of the true treatment ligands were plotted in Fig. 3C. In Fig. 3D, we present the AUC values calculated from ROC curves of the methods under comparison for each test dataset separately. ROC curves of the methods under comparison from combined test results of all 90 test datasets are presented in Fig. 3E.

Software availability

Lignature database is at GitHub (https://github.com/yingxinac/LignatureData), and the companion R package is available at GitHub (https://github.com/yingxinac/Lignature) and as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This study was supported by National Eye Institute (NEI) grant R01EY031779 (JQ).

Author contributions: Conceptualization, Y.X. and J.Q.; Methodology, Y.X. and J.Q.; Data collection, Y.X.; Data processing, Y.X., M.A., and C.Q.; Software development and bioinformatics analysis, Y.X., with the help of C.Z.; Data interpretation, C.Q.; Writing - Original Draft, Y.X., C.Q., and J.Q; Writing - review & editing, Y.X., M.A., C.Q., C.Z., and J.Q. All authors read and approved the final manuscript.

References

Alberini CM. 2023. IGF2 in memory, neurodevelopmental disorders, and neurodegenerative diseases. *Trends Neurosci* **46**: 488-502.

Armingol E, Baghdassarian HM, Lewis NE. 2024. The diversification of methods for studying cell-cell interactions and communication. *Nat Rev Genet* **25**: 381-400.

Armingol E, Officer A, Harismendy O, Lewis NE. 2021. Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* **22**: 71-88.

Baert L, Manfroi B, Casez O, Sturm N, Huard B. 2018. The role of APRIL - A proliferation inducing ligand - In autoimmune diseases and expectations from its targeting. *J Autoimmun* **95**: 179-190.

Browaeys R, Saelens W, Saeys Y. 2020. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**: 159-162.

Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. 2020. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res* **48**: e55.

Cang Z, Nie Q. 2020. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun* **11**: 2084.

Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, Atwood SX, Nie Q. 2023. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* **20**: 218-228.

Carvalho BS, Irizarry RA. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**: 2363-2367.

Cecchi F, Lih CJ, Lee YH, Walsh W, Rabe DC, Williams PM, Bottaro DP. 2015. Expression array analysis of the hepatocyte growth factor invasive program. *Clin Exp Metastasis* **32**: 659-676.

Cui A, Huang T, Li S, Ma A, Pérez JL, Sander C, Keskin DB, Wu CJ, Fraenkel E, Hacohen N. 2024. Dictionary of immune responses to cytokines at single-cell resolution. *Nature* **625**: 377-384.

Cui H, López M, Rahmouni K. 2017. The cellular and molecular bases of leptin and ghrelin resistance in obesity. *Nat Rev Endocrinol* **13**: 338-351.

Davis S, Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**: 1846-1847.

Dölken L, Ruzsics Z, Rädle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P et al. 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *Rna* **14**: 1959-1972.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**: 207-210.

Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. 2020. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**: 1484-1506.

Elsafadi M, Shinwari T, Al-Malki S, Manikandan M, Mahmood A, Aldahmash A, Alfayez M, Kassem M, Alajez NM. 2019. Convergence of TGFβ and BMP signaling in regulating human bone marrow stromal cell differentiation. *Sci Rep* **9**: 4977.

Friedman JH, Hastie T, Tibshirani R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**: 1 - 22.

Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**: 2811-2812.

Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C et al. 2024. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**: 293-304.

Hostager BS, Bishop GA. 2013. CD40-Mediated Activation of the NF-κB2 Pathway. *Front Immunol* **4**: 376.

Hou R, Denisenko E, Ong HT, Ramilowski JA, Forrest ARR. 2020. Predicting cell-to-cell communication networks using NATMI. *Nat Commun* **11**: 5011.

Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R, Mirza A, Altan-Bonnet G, Liu L, Ruppin E et al. 2021. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat Methods* **18**: 1181-1191.

Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. 2021. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**: 1088.

Jo HA, Han SS, Lee S, Kim JY, Yang SH, Lee H, Yang JS, Lee JP, Joo KW, Lim CS et al. 2019. The association of tumor necrosis factor superfamily 13 with recurrence of immunoglobulin A nephropathy in living related kidney transplantation. *BMC Nephrol* **20**: 33.

Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* **28**: 1947-1951.

Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. 2025. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res* **53**: D672-d677.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27-30.

Kiral FR, Cakir B, Tanaka Y, Kim J, Yang WS, Wehbe F, Kang YJ, Zhong M, Sancer G, Lee SH et al. 2023. Generation of ventralized human thalamic organoids with thalamic reticular nucleus. *Cell Stem Cell* **30**: 677-688.e675.

Klaus B, Reisenauer S. 2016. An end to end workflow for differential gene expression using Affymetrix microarrays. *F1000Res* **5**: 1384.

Kolde R. 2025. pheatmap: Pretty Heatmaps.

Korotkevich G, Sukhov V, Sergushichev A. 2019. Fast gene set enrichment analysis. *bioRxiv* doi:10.1101/060012: 060012.

Kuhn M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**: 1 - 26.

Lawrence T. 2009. The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb Perspect Biol* **1**: a001651.

Lawson ND, Vogel AM, Weinstein BM. 2002. sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. *Dev Cell* **3**: 127-136.

Luoma AM, Suo S, Williams HL, Sharova T, Sullivan K, Manos M, Bowling P, Hodi FS, Rahma O, Sullivan RJ et al. 2020. Molecular Pathways of Colon Inflammation Induced by Cancer Immunotherapy. *Cell* **182**: 655-671.e622.

Mackenzie F, Ruhrberg C. 2012. Diverse roles for VEGF-A in the nervous system. *Development* **139**: 1371-1380.

Massenet-Regad L, Noël F, Soumelis V. 2021. [In silico study of cell to cell communication with ICELLNET]. *Med Sci (Paris)* **37**: 981-983.

Miyazono K, Kusanagi K, Inoue H. 2001. Divergence and convergence of TGF-beta/BMP signaling. *J Cell Physiol* **187**: 265-276.

Molinero LL, Yin D, Lei YM, Chen L, Wang Y, Chong AS, Alegre ML. 2016. High-Fat Diet-Induced Obesity Enhances Allograft Rejection. *Transplantation* **100**: 1015-1021.

O'Shea JJ, Schwartz DM, Villarino AV, Gadina M, McInnes IB, Laurence A. 2015. The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annu Rev Med* **66**: 311-328.

Pandey K, Bessières B, Sheng SL, Taranda J, Osten P, Sandovici I, Constancia M, Alberini CM. 2023. Neuronal activity drives IGF2 expression from pericytes to form long-term memory. *Neuron* **111**: 3819-3836.e3818.

Philips RL, Wang Y, Cheon H, Kanno Y, Gadina M, Sartorelli V, Horvath CM, Darnell JE, Jr., Stark GR, O'Shea JJ. 2022. The JAK-STAT pathway at 30: Much learned, much more to do. *Cell* **185**: 3857-3876.

Qian C, Xin Y, Qi C, Wang H, Dong BC, Zack DJ, Blackshaw S, Hattar S, Zhou FQ, Qian J. 2024. Intercellular communication atlas reveals Oprm1 as a neuroprotective factor for retinal ganglion cells. *Nat Commun* **15**: 2206.

Qin H, Chan MW, Liyanarachchi S, Balch C, Potter D, Souriraj IJ, Cheng AS, Agosto-Perez FJ, Nikonova EV, Yan PS et al. 2009. An integrative ChIP-chip and gene expression profiling to model SMAD regulatory modules. *BMC Syst Biol* **3**: 73.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
Rosenstein JM, Krum JM, Ruhrberg C. 2010. VEGF in the nervous system. *Organogenesis* **6**: 107-114.
Sales G, Calura E, Cavalieri D, Romualdi C. 2012. graphite - a Bioconductor package to convert pathway topology to gene network. *BMC Bioinformatics* **13**: 20.
Sanda C, Weitzel P, Tsukahara T, Schaley J, Edenberg HJ, Stephens MA, McClintick JN, Blatt LM, Li L, Brodsky L et al. 2006. Differential gene induction by type I and type II interferons and their combination. *J Interferon Cytokine Res* **26**: 462-472.
Scheja L, Heeren J. 2019. The endocrine function of adipose tissues in health and cardiometabolic disease. *Nat Rev Endocrinol* **15**: 507-524.
Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, Garnett MJ, Blüthgen N, Saez-Rodriguez J. 2018. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications* **9**: 20.
Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, Bruford EA. 2023. Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res* **51**: D1003-d1009.
Seleznik GM, Zoller J, O'Connor T, Graf R, Heikenwalder M. 2014. The role of lymphotoxin signaling in the development of autoimmune pancreatitis and associated secondary extra-pancreatic pathologies. *Cytokine Growth Factor Rev* **25**: 125-137.
Sohn P, Cox M, Chen D, Serra R. 2010. Molecular profiling of the developing mouse axial skeleton: a role for Tgfbr2 in the development of the intervertebral disc. *BMC Dev Biol* **10**: 29.
Su J, Song Y, Zhu Z, Huang X, Fan J, Qiao J, Mao F. 2024. Cell-cell communication: new insights and clinical implications. *Signal Transduct Target Ther* **9**: 196.
Tay JK, Narasimhan B, Hastie T. 2023. Elastic Net Regularization Paths for All Generalized Linear Models. *Journal of Statistical Software* **106**: 1 - 31.
Team RC. *2023*. R: A Language and Environment for Statistical Computing. (ed. RFfS Computing).
Todd L, Finkbeiner C, Wong CK, Hooper MJ, Reh TA. 2020. Microglia Suppress Ascl1-Induced Retinal Regeneration in Mice. *Cell Rep* **33**: 108507.
Toro AR, Maymó JL, Ibarbalz FM, Pérez-Pérez A, Maskin B, Faletti AG, Sánchez-Margalet V, Varone CL. 2014. Leptin is an anti-apoptotic effector in placental cells involving p53 downregulation. *PLoS One* **9**: e99187.
Wang S, Karikomi M, MacLean AL, Nie Q. 2019. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res* **47**: e66.
Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
Wu M, Wu S, Chen W, Li YP. 2024. The roles and regulatory mechanisms of TGF-β and BMP signaling in bone and cartilage development, homeostasis and disease. *Cell Res* **34**: 101-123.
Xin Y, Jin Y, Qian C, Blackshaw S, Qian J. 2025. MetaLigand provides a prior-knowledge-guided framework for predicting non-peptide ligand mediated cell-cell communication. *Cell Rep Methods* **5**: 101217.

Xin Y, Lyu P, Jiang J, Zhou F, Wang J, Blackshaw S, Qian J. 2022. LRLoop: a method to predict feedback loops in cell-cell communication. *Bioinformatics* **38**: 4117-4126.
Yeung TL, Sheng J, Leung CS, Li F, Kim J, Ho SY, Matzuk MM, Lu KH, Wong STC, Mok SC. 2019. Systematic Identification of Druggable Epithelial-Stromal Crosstalk Signaling Networks in Ovarian Cancer. *J Natl Cancer Inst* **111**: 272-282.

Figure 1. Curation and overview of Lignature. (A) Flowchart of generating the Lignature database. (B) Number of the ligands collected in the Lignature database from different data sources. (C) HUGO Gene Nomenclature Committee categories of the ligands curated in the Lignature database. (D) Schematic diagram of application of Lignature in predicting and visualizing LR interactions based on ligand signatures and LR expression. (E) Distribution of the number of signatures of each ligand in Lignature.

Figure 2. Properties of ligand signatures in Lignature. (A) Distributions of Pearson correlation coefficients of gene-level signatures of the same ligand and signatures from different ligands. P-values were calculated using KS-test. (B) Heatmaps of sample gene-level signatures of TGFB1 and SHH (x-axis represents genes in the signatures) exhibit diversity in signatures for individual ligands. (C) Examples of enriched upregulated and downregulated KEGG pathways in sample ligand signatures (LEP and TNFSF13).

Figure 3. Comparison of Lignature to other methods in signaling inference with bulk in vitro datasets. (A) A total of 90 transcriptomic datasets of ligand-treated cell lines were used as test datasets, including 24 human and 66 mouse datasets representing 27 unique treatment ligands. (B) Top ranked ligands (among all ligands included in Lignature) and the corresponding ligand scores of an example test dataset treated with IFNG, predicted from gene-level signatures, pathway-level signatures, and three other methods. (C) Ranks of the actual treatment ligands among all ligands included in Lignature of 90 test datasets predicted by each method. (D) AUC values of the ROC curves for predictions of the 90 test datasets from each method. (E) ROC curves and the corresponding AUC values for combined predictions of all the 90 test datasets from each method. In this *in vitro* setting, true positives are defined as the experimentally added ligands, whereas false positives correspond to all other ligands.

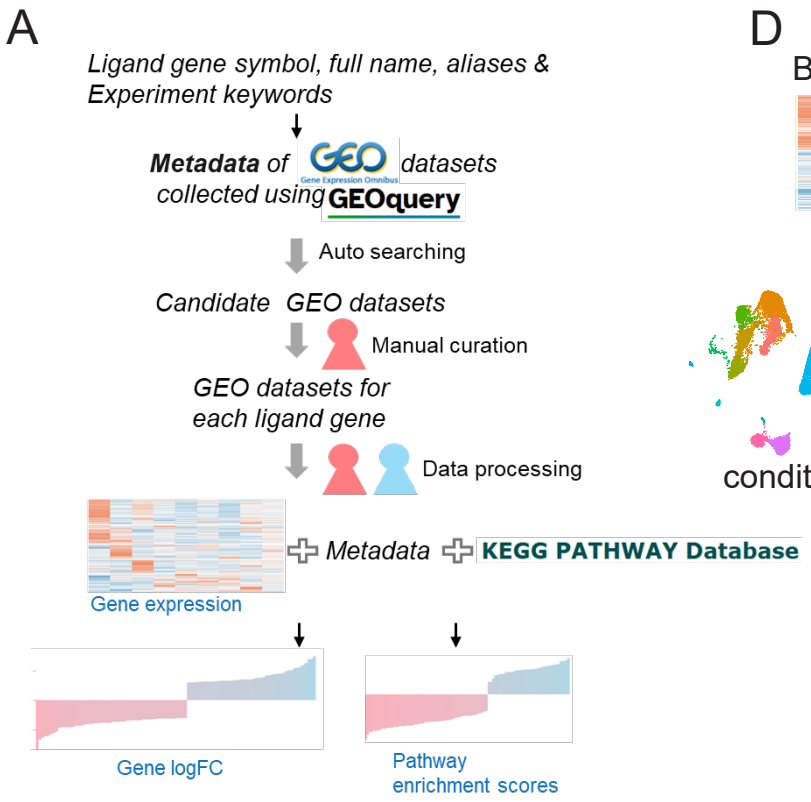
Figure 4. Application and assessment of Lignature on SHH-induced thalamic differentiation scRNA-seq dataset. (A-B) UMAP dimensionality reductions, example feature plots, and DotPlot of selected marker

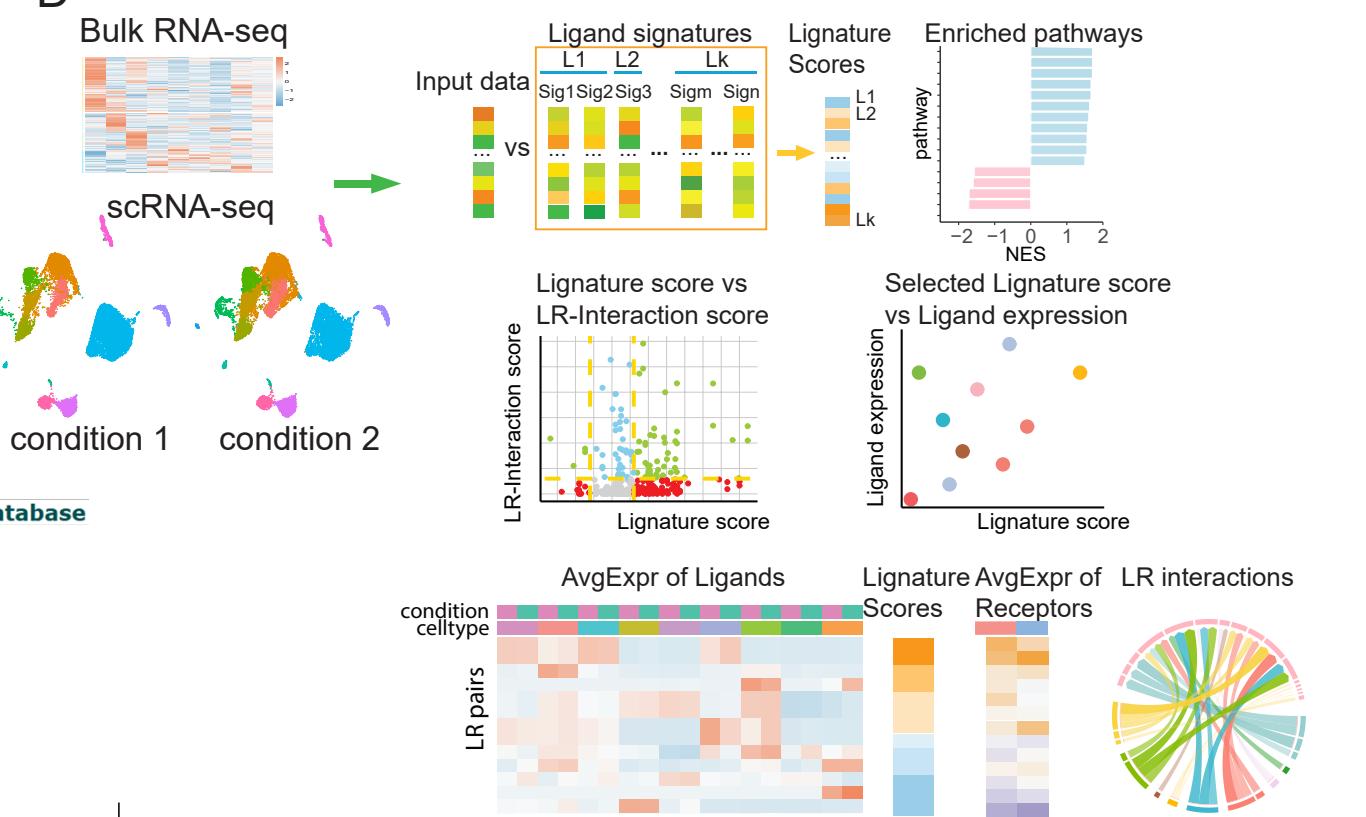
genes of major cell clusters of the SHH-induced thalamic differentiation scRNA-seq dataset. (C) Top predicted ligand scores obtained for the IN cells. The transcriptomes from the SHH day 14–22 treatment samples were compared with untreated samples, and transcriptional program-associated ligands were predicted using Lignature alongside three other methods. The first column shows the detection rates of the ligands across all cells, and the second column shows the maximum detection rate of their cognate receptors in IN cells.

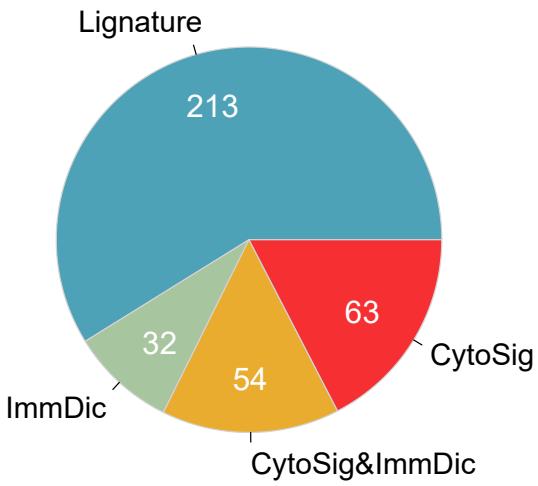
Figure 5. Application and assessment of Lignature on microglia and Müller glia regeneration scRNA-seq dataset. (A-B) UMAP dimensionality reductions of the scRNA-seq data on microglia and MG before and after microglia ablation. (C) Heatmaps of top predicted ligand scores from Lignature and three other methods, together with detection rate of the ligands in microglia and the maximum detection rate of their cognate receptors in MG before microglia ablation. (D) LR interaction strengths were calculated by CellPhoneDB and NATMI based on the expression of ligand genes in microglia and the expression of the receptors in MG, and each ligand was scored by the maximum interaction strength among the LR pairs between the ligand and its receptors. CellPhoneDB and NATMI-induced ligand scores, as well as gene-level Lignature predicted ligand scores, zeroed out by detection rates of the ligands and their receptors at 0.3, were plotted in scatter plots against the detection rates of the ligands in microglia.

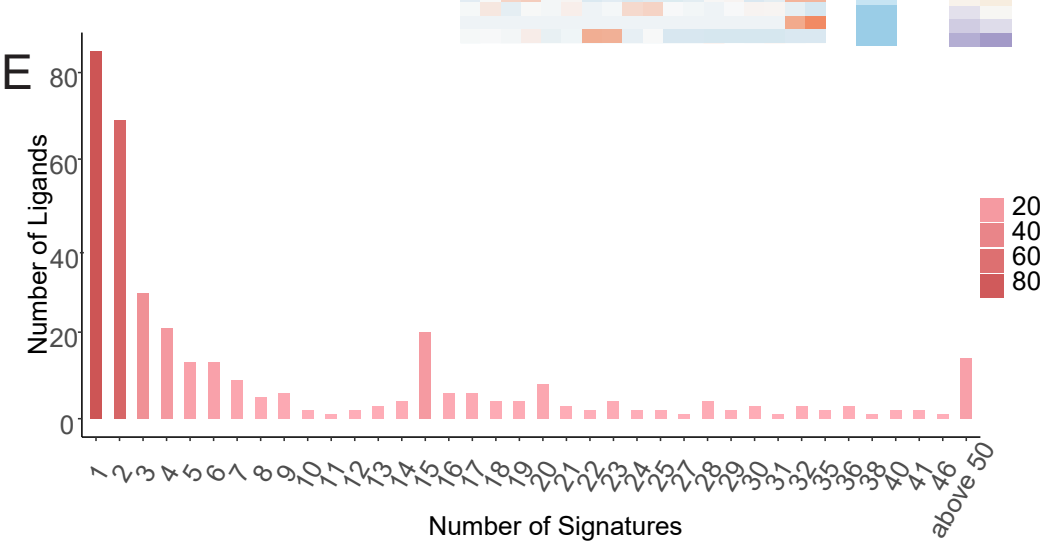
Figure 6. Application of Lignature on checkpoint blockade-associated colitis scRNA-seq dataset. (A) UMAP dimensionality reductions show cell types identified in the scRNA-seq data of immune cell populations in checkpoint blockade-associated colitis and control samples. (B) Enriched upregulated and downregulated KEGG pathways in myeloid cells between colitis and NC (normal colonic mucosa endoscopically and histologically) samples. (C) Scatter plot of ligand signature similarity scores predicted by Lignature (x-axis) and the maximum LR interaction strengths calculated by the method NATMI across all sender cells and conditions (y-axis) for all LR pairs. (D) Scatter plot of similarity scores predicted by Lignature (x-axis) and the maximum average expression across all sender cells and conditions (y-axis) of the ligands, for identified LR interactions (detection rate of the ligand genes above 0.25 in any cell type, detection rate of the receptor genes above 0.25 in myeloid, and the ligand signature similarity scores above 0.1). (E) Heatmaps of ligand signature similarity scores predicted by Lignature (middle), the average expression of the ligands in all cell

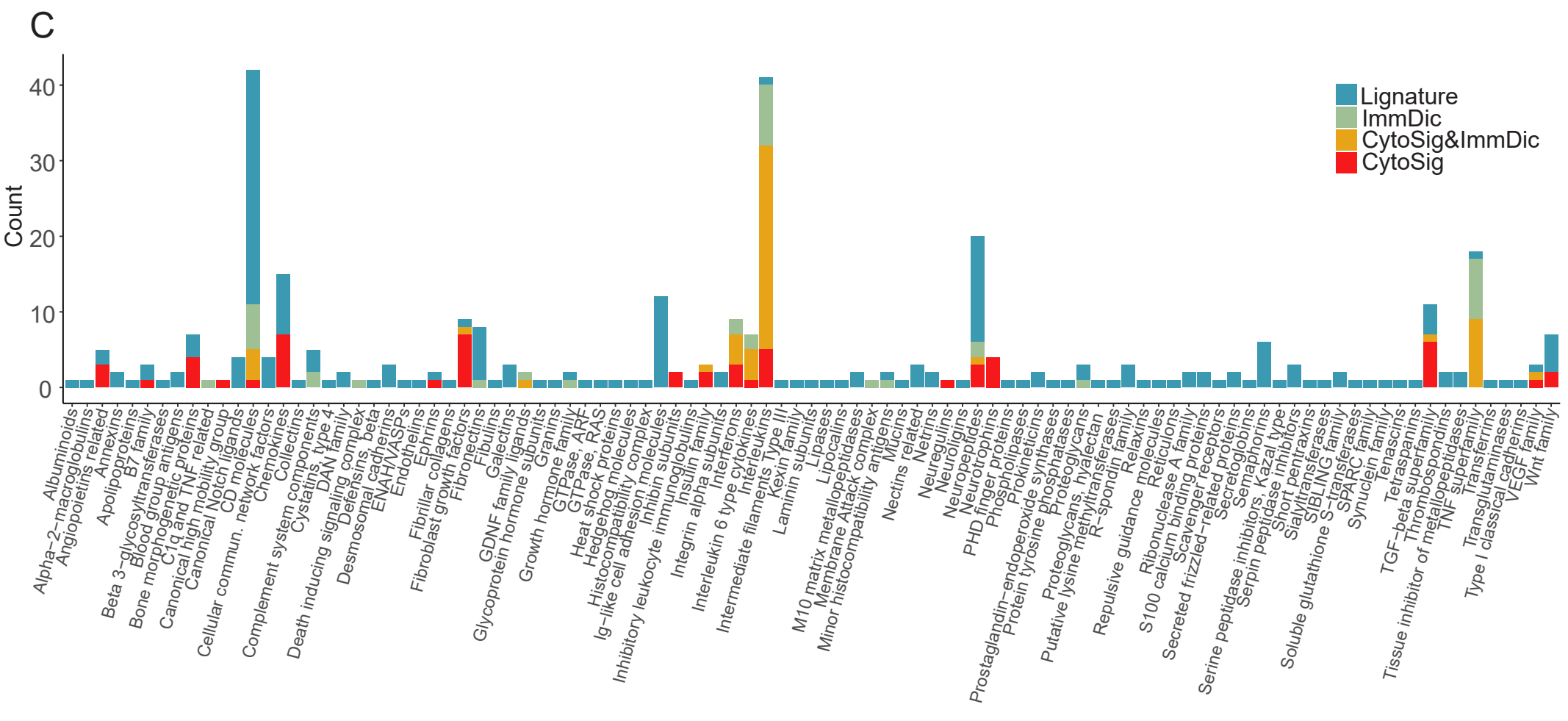
types identified in the colitis and NC samples (left), and the average expression of their receptors in myeloid cells in the colitis and NC samples (right), for LR interactions identified by the same criteria as in (D). (F) Circos plot (generated using R package “circlize”(Gu et al. 2014)) of maximum interaction strengths calculated by NATMI across the colitis and NC samples, of the LR pairs from all cell types to myeloid, with detection rates of both the ligand and receptor genes above 0.25, ligand signature similarity scores above 0.1, and ligand expression in sender cells above 75% of all cell types.

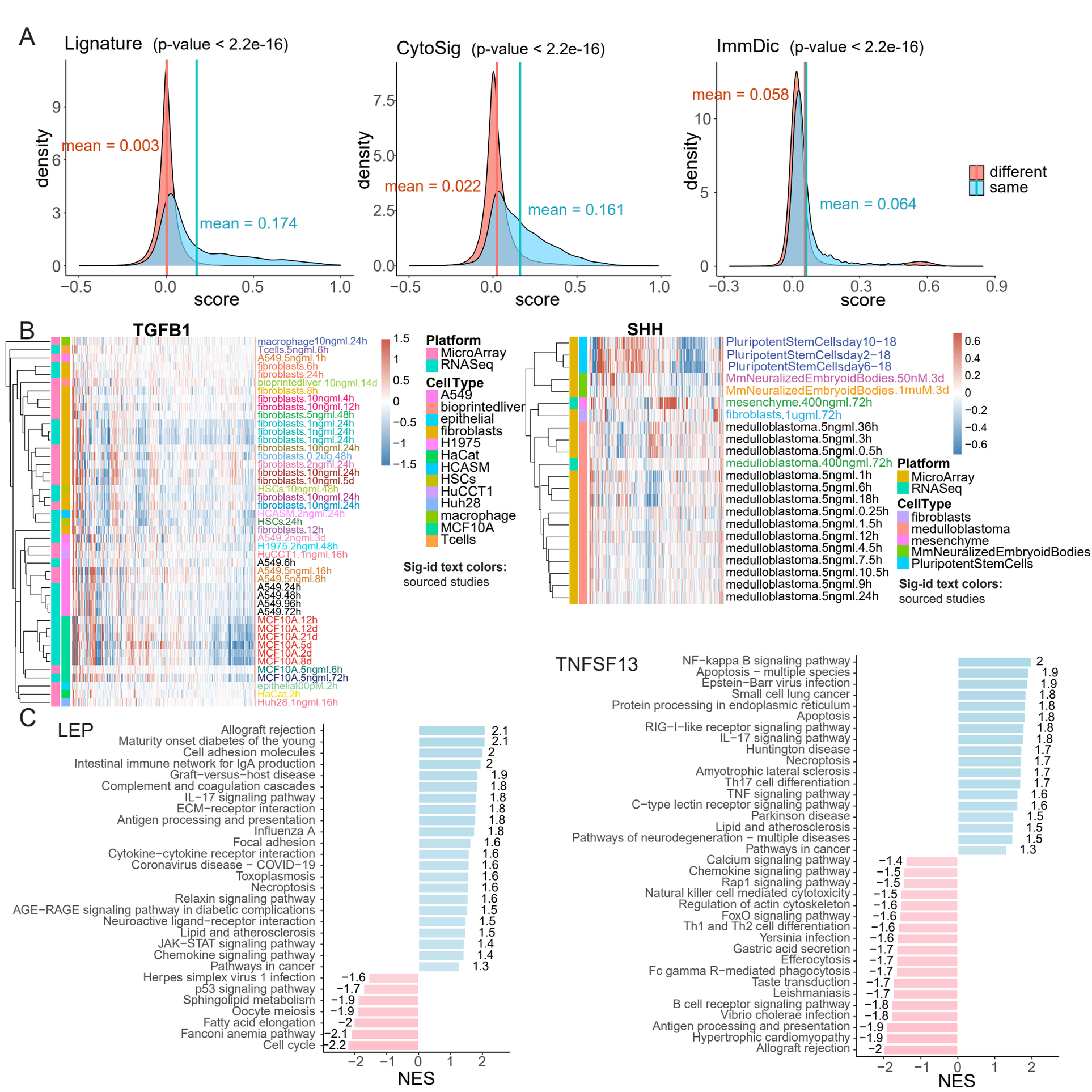

A
Ligand gene symbol, full name, aliases & Experiment keywords
Metadata of GEO datasets collected using GEOquery
Auto searching
Candidate GEO datasets
Manual curation
GEO datasets for each ligand gene
Data processing
Metadata
KEGG PATHWAY Database
Gene expression
Gene logFC
Pathway enrichment scores

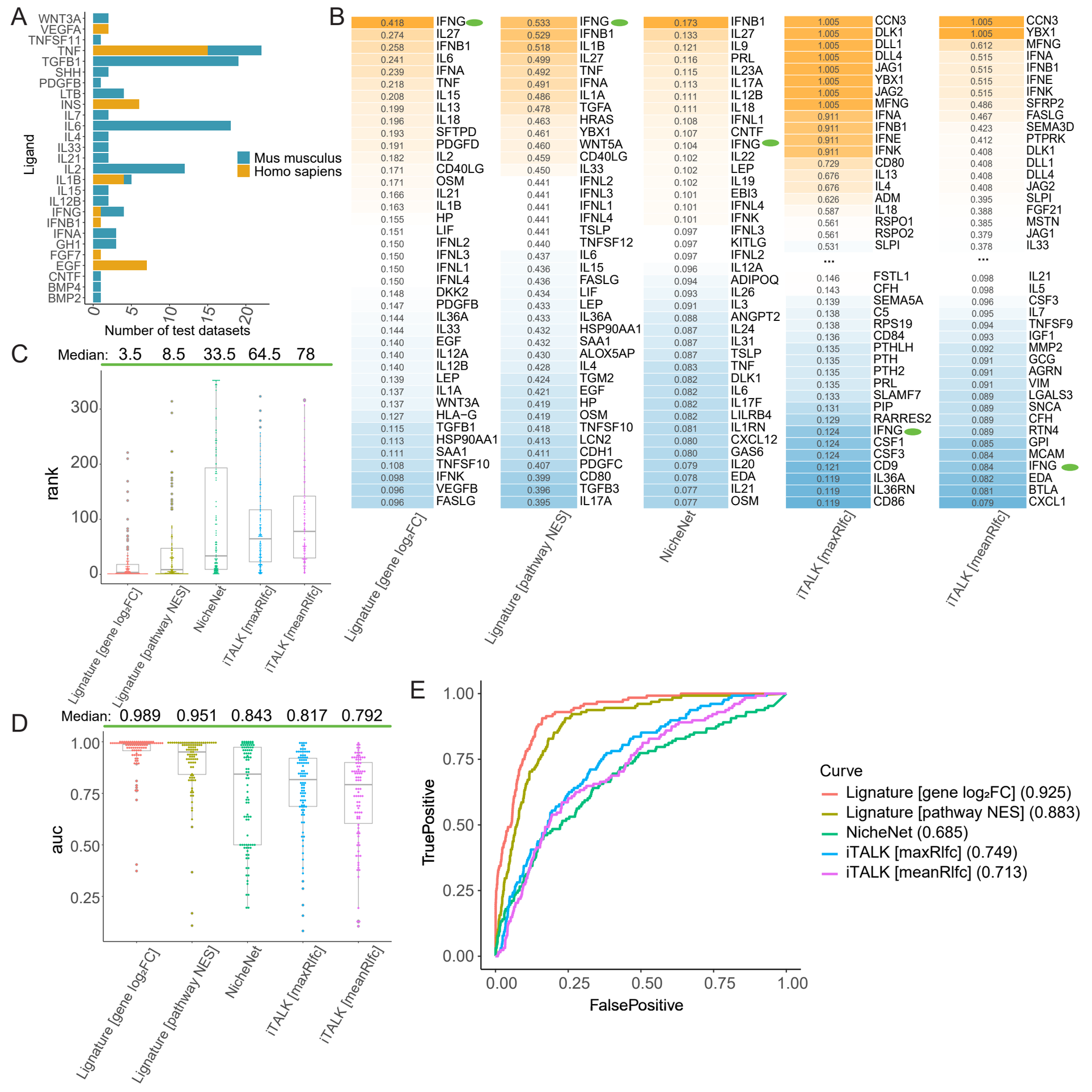

D
Bulk RNA-seq
scRNA-seq
condition 1
condition 2
Input data
vs
Ligand signatures
L1
L2
Lk
Sig1
Sig2
Sig3
Sigm
Sign
Lignature Scores
Enriched pathways
pathway
NES
Lignature score vs LR-Interaction score
LR-Interaction score
Lignature score
Selected Lignature score vs Ligand expression
Ligand expression
AvgExpr of Ligands
condition
celltype
LR pairs
Lignature Scores
AvgExpr of Receptors
LR interactions

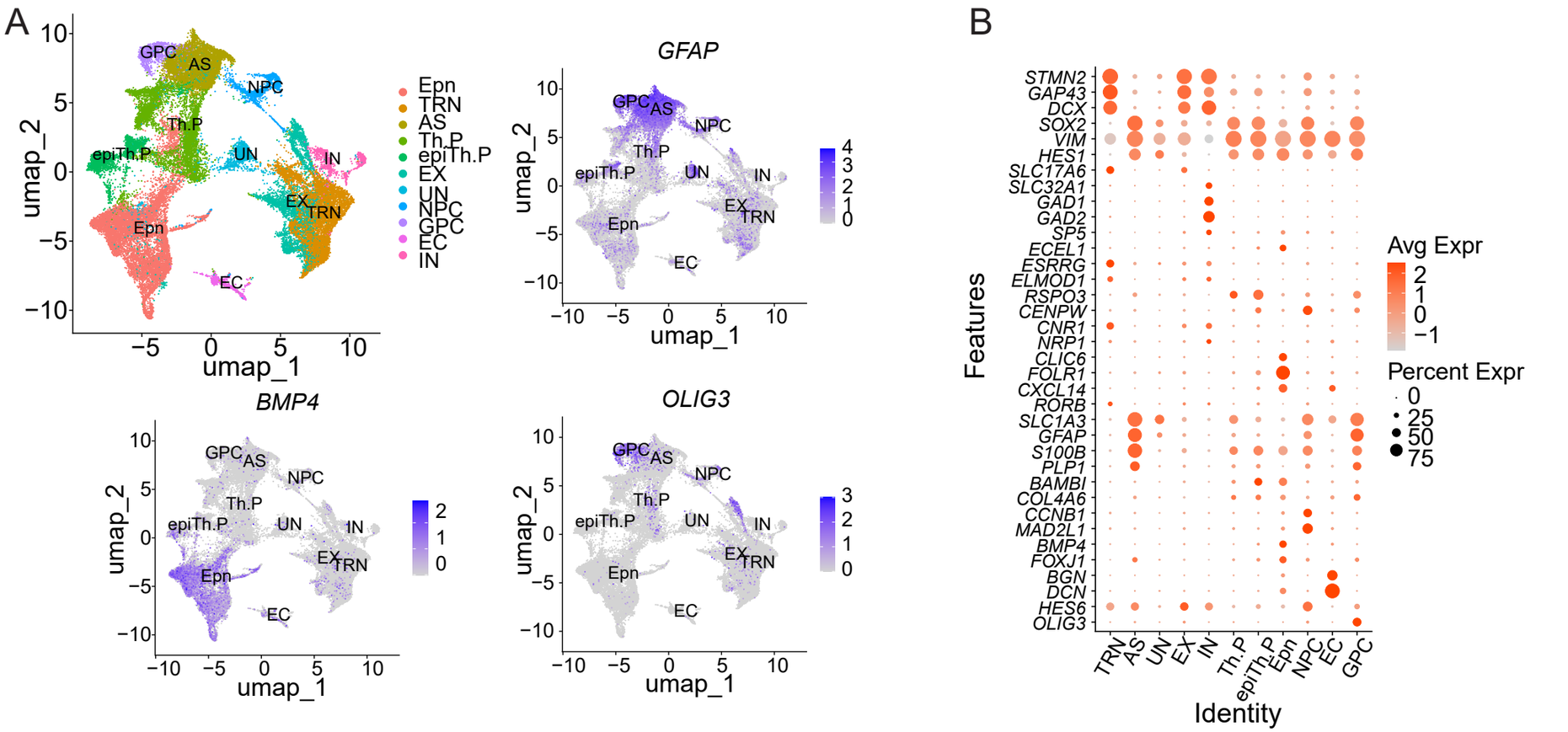

B
Lignature
213
63
32
54
CytoSig
ImmDic
CytoSig&ImmDic


E
Number of Ligands
Number of Signatures
20
40
60
80
above 50


C
Count
Lignature
ImmDic
CytoSig&ImmDic
CytoSig
Albuminoids
Alpha−2−macroglobulins
Angiopoietins related
Annexins
Apolipoproteins
B7 family
Beta 3−glycosyltransferases
Blood group antigens
Bone morphogenetic proteins
C1q and TNF related
Canonical high mobility group
Canonical Notch ligands
CD molecules
Cellular commun. network factors
Chemokines
Collectins
Complement system components
Cystatins, type 4
DAN family
Death inducing signaling complex
Defensins, beta
Desmosomal cadherins
ENAH/VASPs
Endothelins
Ephrins
Fibrillar collagens
Fibroblast growth factors
Fibronectins
Fibulins
Galectins
GDNF family ligands
Glycoprotein hormone subunits
Granins
Growth hormone family
GTPase, ARF
GTPase, RAS
Heat shock proteins
Hedgehog molecules
Histocompatibility complex
Ig−like cell adhesion molecules
Inhibitory leukocyte immunoglobulins
Inhibin subunits
Insulin family
Integrin alpha subunits
Interferons
Interleukin 6 type cytokines
Interleukins
Intermediate filaments Type III
Kexin family
Laminin subunits
Lipases
Lipocalins
M10 matrix metallopeptidases
Membrane Attack complex
Minor histocompatibility antigens
Mucins
Nectins related
Netrins
Neuregulins
Neuroligins
Neuropeptides
Neurotrophins
PHD finger proteins
Phospholipases
Prokineticins
Prostaglandin−endoperoxide synthases
Protein tyrosine phosphatases
Proteoglycans
Putative lysine methyltransferases
Proteoglycans, hyalectan
R−spondin family
Relaxins
Repulsive guidance molecules
Reticulons
Ribonuclease A family
S100 calcium binding proteins
Secreted frizzled−related proteins
Scavenger receptors
Secretoglobins
Semaphorins
Serine peptidase inhibitors, Kazal type
Serpin peptidase inhibitors
Short pentraxins
Sialyltransferases
Soluble glutathione S−transferases
SIBLING family
SPARC family
Synuclein family
Tenascins
Tetraspanins
TGF−beta superfamily
Thrombospondins
Tissue inhibitor of metallopeptidases
TNF superfamily
Transferrins
Transglutaminases
Type I classical cadherins
VEGF family
Wnt family


A
Lignature (p-value < 2.2e-16)
mean = 0.003
mean = 0.174
density
score
CytoSig (p-value < 2.2e-16)
mean = 0.022
mean = 0.161
ImmDic (p-value < 2.2e-16)
mean = 0.058
mean = 0.064
different
same
B
TGFB1
macrophage10ngml.24h
Tcells.5ngml.6h
A549.5ngml.1h
fibroblasts.6h
fibroblasts.24h
bioprintedliver.10ngml.14d
fibroblasts.8h
fibroblasts.10ngml.4h
fibroblasts.10ngml.12h
fibroblasts.5ngml.48h
fibroblasts.1ngml.24h
fibroblasts.1ngml.24h
fibroblasts.1ngml.24h
fibroblasts.10ngml.24h
fibroblasts.0.2ug.48h
fibroblasts.2ngml.24h
fibroblasts.10ngml.24h
fibroblasts.10ngml.5d
HSCs.10ngml.48h
fibroblasts.10ngml.24h
fibroblasts.10ngml.24h
HCASM.2ngml.24h
HSCs.24h
fibroblasts.12h
A549.2ngml.3d
H1975.2ngml.48h
HuCCT1.1ngml.16h
A549.6h
A549.5ngml.16h
A549.5ngml.8h
A549.24h
A549.48h
A549.96h
A549.72h
MCF10A.12h
MCF10A.12d
MCF10A.21d
MCF10A.5d
MCF10A.2d
MCF10A.8d
MCF10A.5ngml.6h
MCF10A.5ngml.72h
epithelial100pM.2h
HaCat.2h
Huh28.1ngml.16h
Platform
MicroArray
RNASeq
CellType
A549
bioprintedliver
epithelial
fibroblasts
H1975
HaCat
HCASM
HSCs
HuCCT1
Huh28
macrophage
MCF10A
Tcells
Sig-id text colors:
sourced studies
SHH
PluripotentStemCellsday10−18
PluripotentStemCellsday2−18
PluripotentStemCellsday6−18
MmNeuralizedEmbryoidBodies.50nM.3d
MmNeuralizedEmbryoidBodies.1muM.3d
mesenchyme.400ngml.72h
fibroblasts.1ugml.72h
medulloblastoma.5ngml.36h
medulloblastoma.5ngml.3h
medulloblastoma.5ngml.0.5h
medulloblastoma.400ngml.72h
medulloblastoma.5ngml.1h
medulloblastoma.5ngml.6h
medulloblastoma.5ngml.18h
medulloblastoma.5ngml.0.25h
medulloblastoma.5ngml.1.5h
medulloblastoma.5ngml.12h
medulloblastoma.5ngml.4.5h
medulloblastoma.5ngml.7.5h
medulloblastoma.5ngml.10.5h
medulloblastoma.5ngml.9h
medulloblastoma.5ngml.24h
Platform
MicroArray
RNASeq
CellType
fibroblasts
medulloblastoma
mesenchyme
MmNeuralizedEmbryoidBodies
PluripotentStemCells
Sig-id text colors:
sourced studies
C
LEP
Allograft rejection
Maturity onset diabetes of the young
Cell adhesion molecules
Intestinal immune network for IgA production
Graft−versus−host disease
Complement and coagulation cascades
IL−17 signaling pathway
ECM−receptor interaction
Antigen processing and presentation
Influenza A
Focal adhesion
Cytokine−cytokine receptor interaction
Coronavirus disease − COVID−19
Toxoplasmosis
Necroptosis
Relaxin signaling pathway
AGE−RAGE signaling pathway in diabetic complications
Neuroactive ligand−receptor interaction
Lipid and atherosclerosis
JAK−STAT signaling pathway
Chemokine signaling pathway
Pathways in cancer
Herpes simplex virus 1 infection
p53 signaling pathway
Sphingolipid metabolism
Oocyte meiosis
Fatty acid elongation
Fanconi anemia pathway
Cell cycle
NES
TNFSF13
NF−kappa B signaling pathway
Apoptosis − multiple species
Epstein−Barr virus infection
Small cell lung cancer
Protein processing in endoplasmic reticulum
Apoptosis
RIG−I−like receptor signaling pathway
IL−17 signaling pathway
Huntington disease
Necroptosis
Amyotrophic lateral sclerosis
Th17 cell differentiation
TNF signaling pathway
C−type lectin receptor signaling pathway
Parkinson disease
Lipid and atherosclerosis
Pathways of neurodegeneration − multiple diseases
Pathways in cancer
Calcium signaling pathway
Chemokine signaling pathway
Rap1 signaling pathway
Natural killer cell mediated cytotoxicity
Regulation of actin cytoskeleton
FoxO signaling pathway
Th1 and Th2 cell differentiation
Yersinia infection
Gastric acid secretion
Efferocytosis
Fc gamma R−mediated phagocytosis
Taste transduction
Leishmaniasis
B cell receptor signaling pathway
Vibrio cholerae infection
Antigen processing and presentation
Hypertrophic cardiomyopathy
Allograft rejection
NES


A
Ligand
WNT3A
VEGFA
TNFSF11
TNF
TGFB1
SHH
PDGFB
LTB
INS
IL7
IL6
IL4
IL33
IL21
IL2
IL1B
IL15
IL12B
IFNG
IFNB1
IFNA
GH1
FGF7
EGF
CNTF
BMP4
BMP2
0
5
10
15
20
Number of test datasets
Mus musculus
Homo sapiens
B
0.418 IFNG
0.274 IL27
0.258 IFNB1
0.241 IL6
0.239 IFNA
0.218 TNF
0.208 IL15
0.199 IL13
0.196 IL18
0.193 SFTPD
0.191 PDGFD
0.182 IL2
0.171 CD40LG
0.171 OSM
0.166 IL21
0.163 IL1B
0.155 HP
0.151 LIF
0.150 IFNL2
0.150 IFNL3
0.150 IFNL1
0.150 IFNL4
0.148 DKK2
0.147 PDGFB
0.144 IL36A
0.144 IL33
0.140 EGF
0.140 IL12A
0.140 IL12B
0.139 LEP
0.137 IL1A
0.137 WNT3A
0.127 HLA−G
0.115 TGFB1
0.113 HSP90AA1
0.111 SAA1
0.108 TNFSF10
0.098 IFNK
0.096 VEGFB
0.096 FASLG
Lignature [gene log2FC]
0.533 IFNG
0.529 IFNB1
0.518 IL1B
0.499 IL27
0.492 TNF
0.491 IFNA
0.486 IL1A
0.478 TGFA
0.463 HRAS
0.461 YBX1
0.460 WNT5A
0.459 CD40LG
0.450 IL33
0.441 IFNL2
0.441 IFNL3
0.441 IFNL1
0.441 IFNL4
0.441 TSLP
0.440 TNFSF12
0.437 IL6
0.436 IL15
0.436 FASLG
0.434 LIF
0.433 LEP
0.433 IL36A
0.432 HSP90AA1
0.432 SAA1
0.430 ALOX5AP
0.428 IL4
0.424 TGM2
0.421 EGF
0.419 HP
0.419 OSM
0.418 TNFSF10
0.413 LCN2
0.411 CDH1
0.407 PDGFC
0.399 CD80
0.396 TGFB3
0.395 IL17A
Lignature [pathway NES]
0.173 IFNB1
0.133 IL27
0.121 IL9
0.116 PRL
0.115 IL23A
0.113 IL17A
0.111 IL12B
0.111 IL18
0.108 IFNL1
0.107 CNTF
0.104 IFNG
0.102 IL22
0.102 LEP
0.102 IL19
0.101 EBI3
0.101 IFNL4
0.101 IFNK
0.097 IFNL3
0.097 KITLG
0.097 IFNL2
0.096 IL12A
0.094 ADIPOQ
0.093 IL26
0.091 IL3
0.088 ANGPT2
0.087 IL24
0.087 IL31
0.087 TSLP
0.083 TNF
0.082 DLK1
0.082 IL6
0.082 IL17F
0.082 LILRB4
0.081 IL1RN
0.080 CXCL12
0.080 GAS6
0.079 IL20
0.078 EDA
0.077 IL21
0.077 OSM
NicheNet
1.005 CCN3
1.005 DLK1
1.005 DLL1
1.005 DLL4
1.005 JAG1
1.005 YBX1
1.005 JAG2
1.005 MFNG
0.911 IFNA
0.911 IFNB1
0.911 IFNE
0.911 IFNK
0.729 CD80
0.676 IL13
0.676 IL4
0.626 ADM
0.587 IL18
0.561 RSPO1
0.561 RSPO2
0.531 SLPI
...
0.146 FSTL1
0.143 CFH
0.139 SEMA5A
0.138 C5
0.138 RPS19
0.136 CD84
0.135 PTHLH
0.135 PTH
0.135 PTH2
0.135 PRL
0.133 SLAMF7
0.131 PIP
0.129 RARRES2
0.124 IFNG
0.124 CSF1
0.124 CSF3
0.121 CD9
0.119 IL36A
0.119 IL36RN
0.119 CD86
iTALK [maxRlfc]
1.005 CCN3
1.005 YBX1
0.612 MFNG
0.515 IFNA
0.515 IFNB1
0.515 IFNE
0.515 IFNK
0.486 SFRP2
0.467 FASLG
0.423 SEMA3D
0.412 PTPRK
0.408 DLK1
0.408 DLL1
0.408 DLL4
0.408 JAG2
0.395 SLPI
0.388 FGF21
0.385 MSTN
0.379 JAG1
0.378 IL33
...
0.098 IL21
0.098 IL5
0.096 CSF3
0.095 IL7
0.094 TNFSF9
0.093 IGF1
0.092 MMP2
0.091 GCG
0.091 AGRN
0.091 VIM
0.089 LGALS3
0.089 SNCA
0.089 CFH
0.089 RTN4
0.085 GPI
0.084 MCAM
0.084 IFNG
0.082 EDA
0.081 BTLA
0.079 CXCL1
iTALK [meanRlfc]
C
Median: 3.5 8.5 33.5 64.5 78
rank
0
100
200
300
Lignature [gene log2FC]
Lignature [pathway NES]
NicheNet
iTALK [maxRlfc]
iTALK [meanRlfc]
D
Median: 0.989 0.951 0.843 0.817 0.792
auc
0.25
0.50
0.75
1.00
Lignature [gene log2FC]
Lignature [pathway NES]
NicheNet
iTALK [maxRlfc]
iTALK [meanRlfc]
E
TruePositive
0.00
0.25
0.50
0.75
1.00
FalsePositive
Curve
Lignature [gene log2FC] (0.925)
Lignature [pathway NES] (0.883)
NicheNet (0.685)
iTALK [maxRlfc] (0.749)
iTALK [meanRlfc] (0.713)



C

Lignature [gene log₂FC]

L Expr%	R Expr%	Score	
0.441	0.338	0.086	VEGFA
0.000	0.221	0.059	IFNG
0.133	0.464	0.050	IGF2
1.000	0.211	0.047	SHH
0.164	0.446	0.043	BMP2
0.925	0.281	0.042	APP
0.001	0.148	0.041	TNF
0.442	0.221	0.040	GRN
0.044	0.164	0.038	TGFB1
0.208	0.338	0.037	FGF2
0.544	0.760	0.036	CADM1
0.001	0.473	0.036	C3
0.023	0.174	0.035	CALCA
0.000	0.281	0.030	IFNE
0.008	0.303	0.029	EGF
0.015	0.391	0.028	IL34
0.214	0.446	0.028	BMP7
0.036	0.281	0.026	BDNF
0.220	0.779	0.026	WNT5A
0.000	0.281	0.025	IFNB1
0.066	0.372	0.025	FN1
0.012	0.435	0.022	INHBA
0.037	0.435	0.022	INHBB
0.000	0.281	0.022	IFNK
0.488	0.372	0.021	VCAN
0.065	0.281	0.021	LPL
0.036	0.164	0.021	TGFB3
0.132	0.281	0.020	TFPI
0.218	0.446	0.019	BMP5
0.221	0.268	0.019	ITGAV
0.045	0.174	0.019	POSTN
0.441	0.372	0.019	LGALS1
0.074	0.464	0.018	IGF1
0.145	0.189	0.018	VEGFC
0.119	0.174	0.018	ADM
0.011	0.356	0.018	DKK1
0.162	0.325	0.018	KITLG
0.081	0.338	0.017	FGF1
0.003	0.779	0.017	MYOC
0.002	0.331	0.017	FGF19

NicheNet

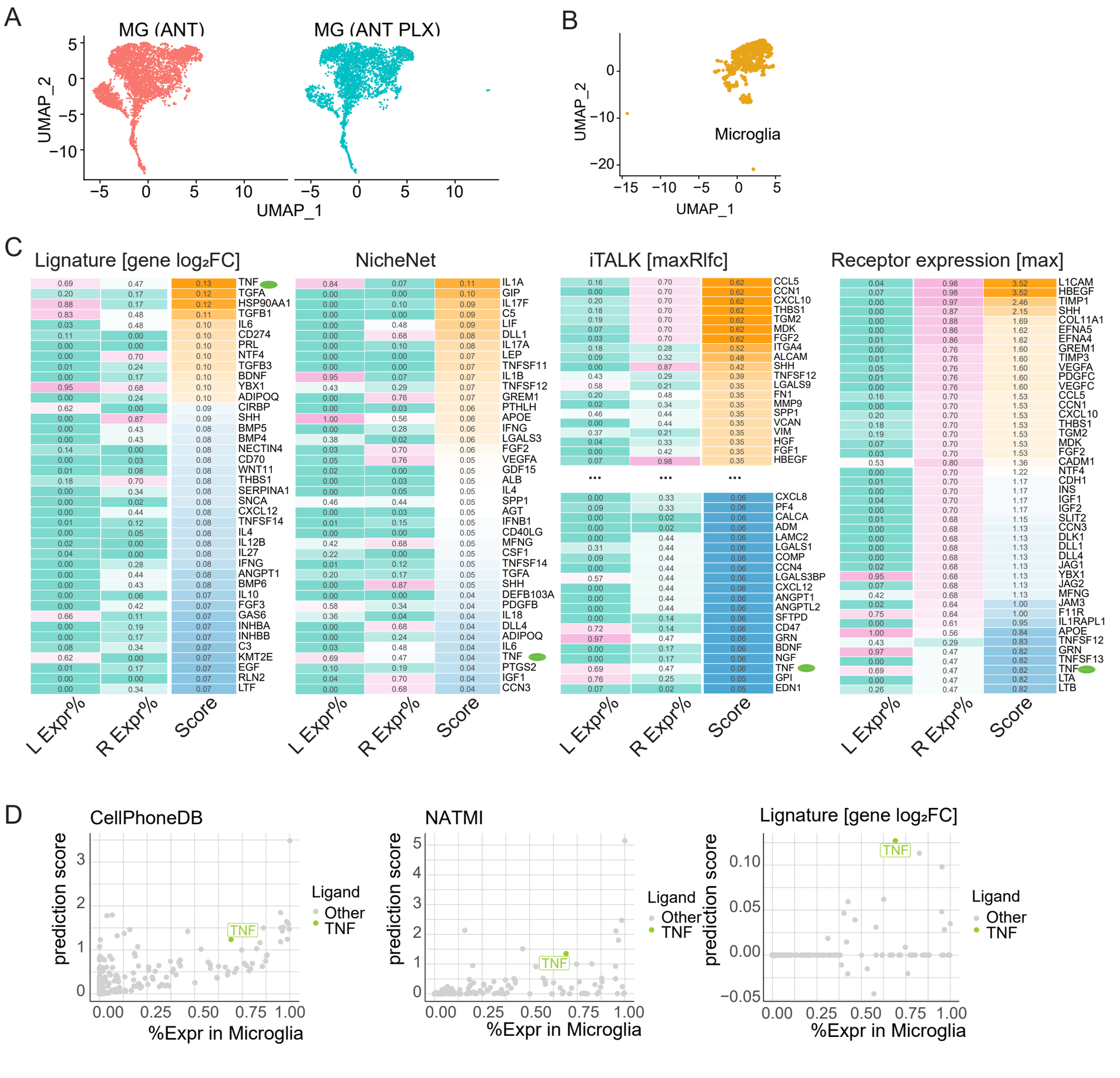
L Expr%	R Expr%	Score	
0.001	0.779	0.081	WNT1
0.218	0.446	0.057	BMP5
0.214	0.446	0.057	BMP7
0.023	0.174	0.056	CALCA
0.154	0.372	0.053	ANGPTL2
0.000	0.446	0.046	GDF2
0.000	0.338	0.042	SEMA3B
0.350	0.315	0.038	JAM3
0.116	0.446	0.037	BMP4
0.074	0.464	0.036	IGF1
0.140	0.372	0.036	CXCL12
0.003	0.303	0.035	TGFA
0.012	0.435	0.033	INHBA
0.121	0.473	0.033	JAG1
0.044	0.164	0.033	TGFB1
0.037	0.435	0.032	INHBB
0.111	0.356	0.032	DKK2
0.011	0.356	0.031	DKK1
0.836	0.281	0.031	CALR
0.002	0.394	0.031	COMP
0.000	0.221	0.031	IFNG
0.000	0.446	0.030	BMP10
0.544	0.760	0.030	CADM1
0.002	0.331	0.029	FGF19
0.491	0.331	0.029	HRAS
0.011	0.366	0.029	MSTN
0.008	0.303	0.028	EGF
0.164	0.446	0.028	BMP2
0.208	0.338	0.027	FGF2
0.076	0.445	0.026	SEMA3A
0.101	0.180	0.026	EPCAM
0.164	0.435	0.026	GDF11
0.356	0.448	0.025	TNC
0.005	0.331	0.024	FGF23
0.011	0.107	0.024	EDN1
0.162	0.413	0.023	COL11A1
1.000	0.211	0.023	SHH
0.084	0.122	0.022	PCSK9
0.066	0.372	0.022	FN1
0.005	0.446	0.022	GDF6

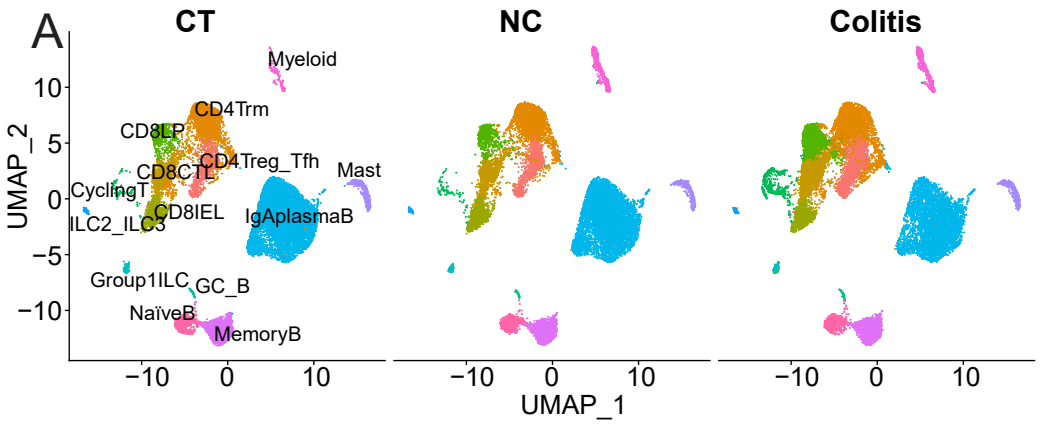
iTALK [maxRlfc]

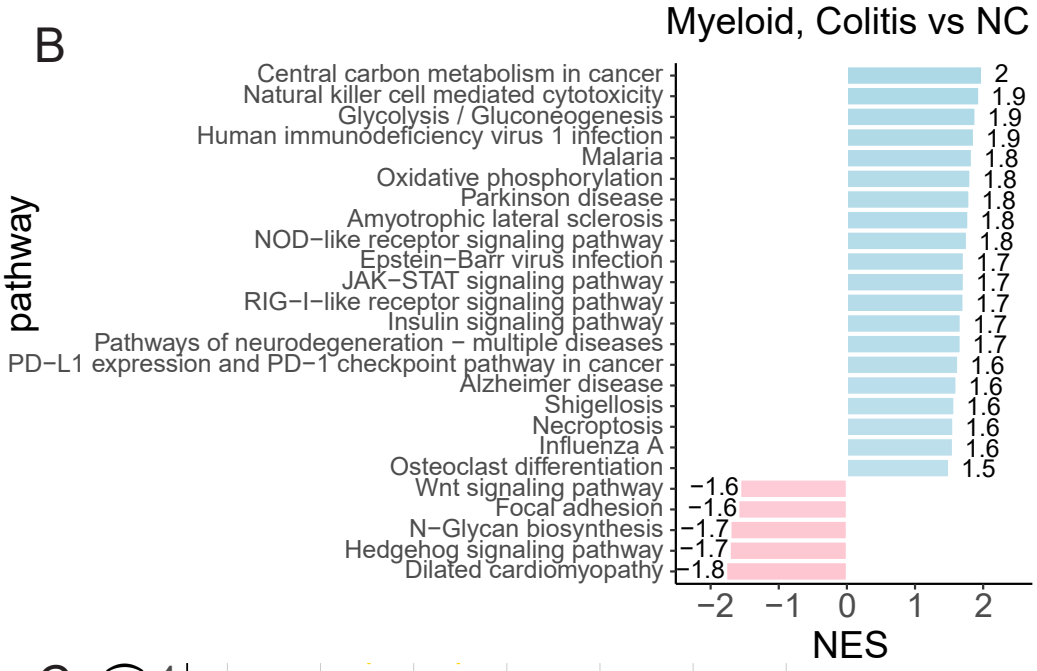
L Expr%	R Expr%	Score	
0.181	0.265	4.335	EFNA5
0.106	0.265	4.335	EFNA4
0.145	0.189	3.302	VEGFC
0.076	0.445	3.269	SEMA3A
0.000	0.338	3.269	SEMA3B
0.039	0.338	3.269	SEMA3F
0.441	0.338	3.269	VEGFA
0.101	0.180	2.711	EPCAM
0.298	0.962	2.643	SLIT2
0.544	0.760	2.532	CADM1
0.925	0.281	2.067	APP
0.001	0.148	2.067	TNF
0.422	0.165	1.669	GAS6
0.342	0.281	1.657	APOE
0.084	0.122	1.657	PCSK9
0.132	0.281	1.657	TFPI
0.065	0.281	1.657	LPL
0.131	0.994	1.514	NLGN3
0.011	0.107	1.290	EDN1
0.005	0.165	1.210	KISS1
...	...	...	
0.018	0.464	0.212	CDH1
0.000	0.464	0.212	INS
0.074	0.464	0.212	IGF1
0.011	0.356	0.207	DKK1
0.031	0.356	0.207	RSPO1
0.003	0.356	0.207	SOST
0.111	0.356	0.207	DKK2
0.000	0.158	0.205	EPO
0.302	0.331	0.179	NAMPT
0.491	0.331	0.179	HRAS
0.811	0.331	0.179	ARF1
0.162	0.413	0.146	COL11A1
0.000	0.221	0.122	IFNG
1.000	0.211	0.098	SHH
0.221	0.268	0.092	ITGAV
0.324	0.315	0.090	FSTL1
0.009	0.372	0.089	LAMC2
0.441	0.372	0.089	LGALS1
0.006	0.372	0.089	TGM2
0.050	0.372	0.089	SPP1

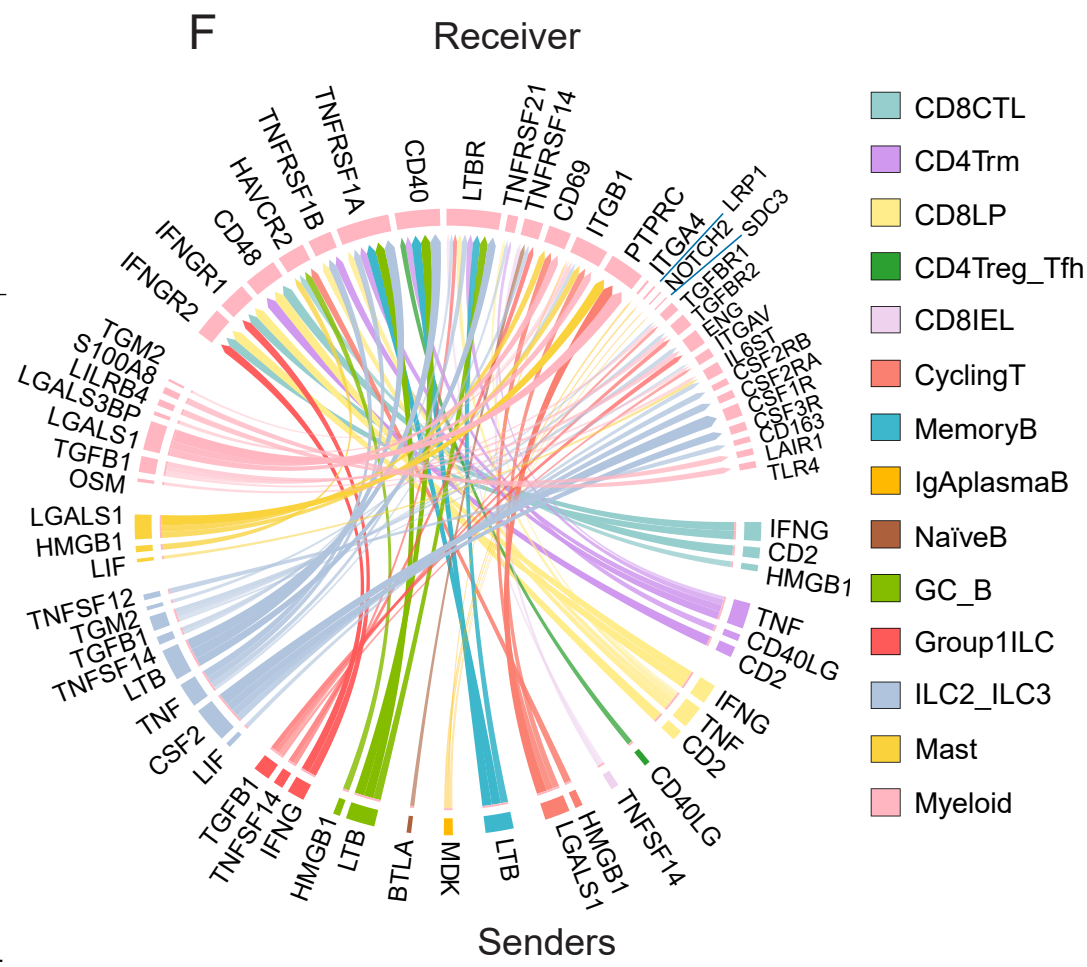
Receptor expression [max]

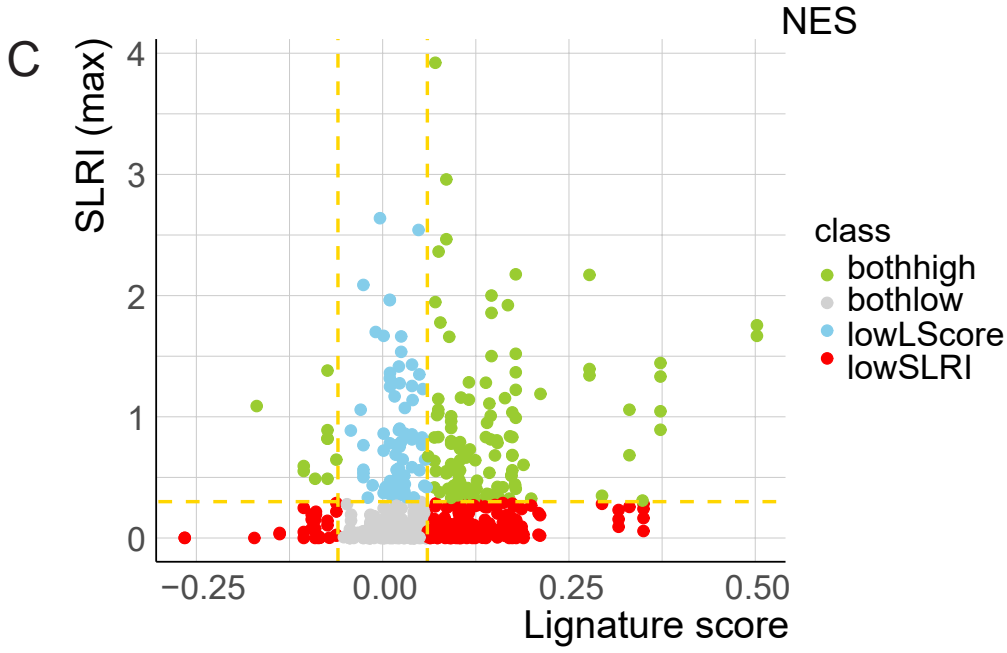
L Expr%	R Expr%	Score	
0.131	0.994	2.759	NLGN3
0.298	0.962	2.118	SLIT2
0.302	0.842	1.735	L1CAM
0.333	0.729	1.400	ALCAM
0.544	0.760	1.367	CADM1
0.001	0.779	1.286	WNT1
0.003	0.779	1.286	MYOC
0.220	0.779	1.286	WNT5A
0.322	0.666	1.158	TIMP1
0.949	0.694	1.035	RTN4
0.148	0.647	1.002	IL1RAPL1
0.193	0.599	0.918	AGRN
0.025	0.338	0.877	SEMA3D
0.076	0.445	0.877	SEMA3A
0.000	0.338	0.877	SEMA3B
0.039	0.338	0.877	SEMA3F
0.649	0.338	0.877	VEGFB
0.441	0.338	0.877	VEGFA
0.060	0.338	0.877	FGF7
0.208	0.338	0.877	FGF2
...	...	...	
0.000	0.268	0.342	ADIPOQ
0.023	0.174	0.318	CALCA
0.119	0.174	0.318	ADM
0.644	0.237	0.317	CD99
0.145	0.189	0.315	VEGFC
0.005	0.165	0.304	KISS1
0.038	0.196	0.302	NTN4
0.084	0.122	0.302	PCSK9
0.101	0.180	0.298	EPCAM
0.422	0.165	0.293	GAS6
0.442	0.221	0.283	GRN
0.053	0.221	0.283	NTS
0.014	0.221	0.283	NGF
0.000	0.221	0.259	IFNG
0.045	0.174	0.248	POSTN
1.000	0.211	0.241	SHH
0.000	0.158	0.228	EPO
0.545	0.148	0.209	ENHO
0.044	0.164	0.202	TGFB1
0.036	0.164	0.201	TGFB3

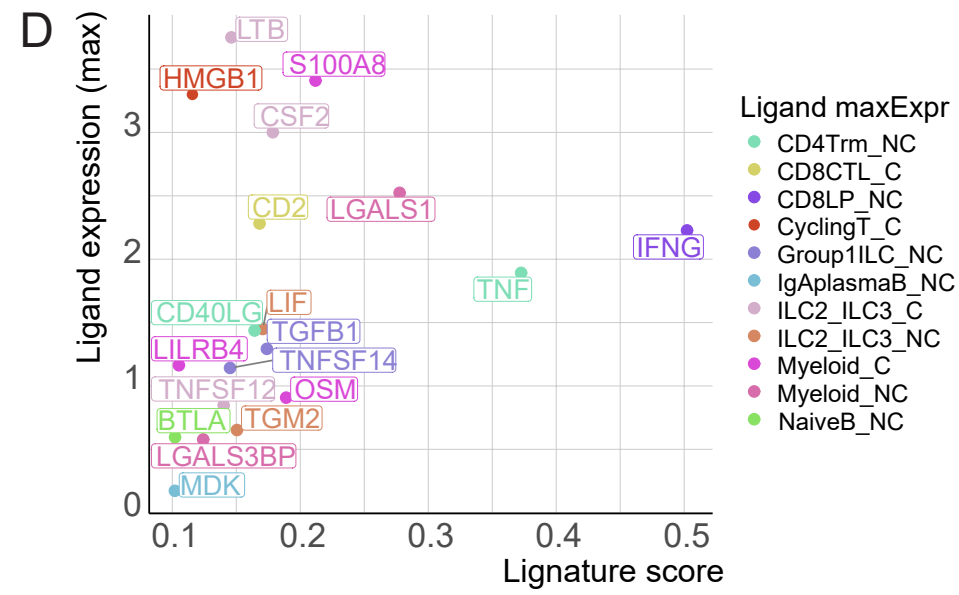

A
MG (ANT)
MG (ANT PLX)
UMAP_2
UMAP_1
B
Microglia
C
Lignature [gene log2FC]
NicheNet
iTALK [maxRlfc]
Receptor expression [max]
L Expr%
R Expr%
Score
D
CellPhoneDB
NATMI
Lignature [gene log2FC]
prediction score
%Expr in Microglia
Ligand
Other
TNF

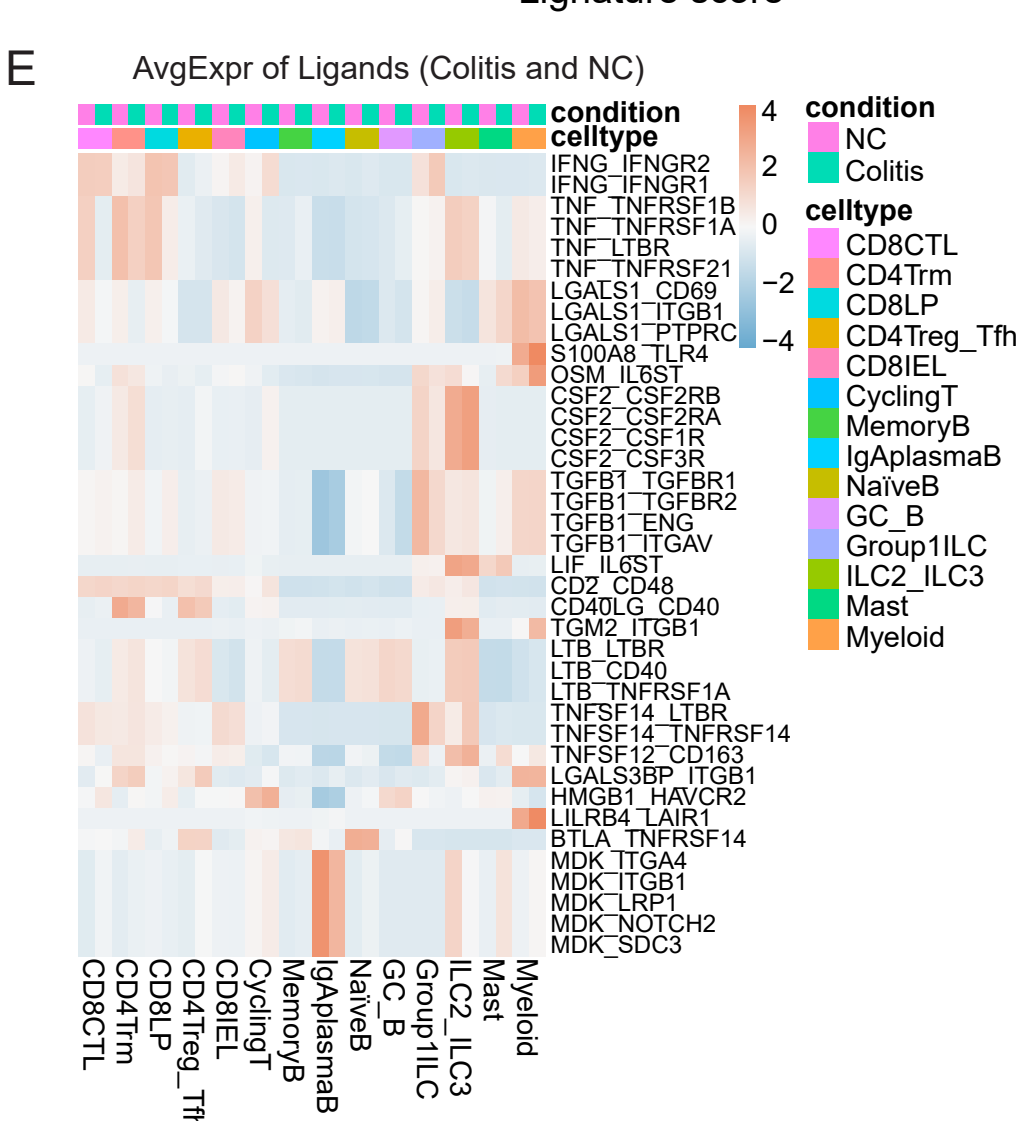

A
CT
NC
Colitis
Myeloid
CD4Trm
CD8LP
CD4Treg_Tfh
CD8CTL
Mast
CyclingT
CD8IEL
IgAplasmaB
ILC2_ILC3
Group1ILC
GC_B
NaïveB
MemoryB
UMAP_2
UMAP_1

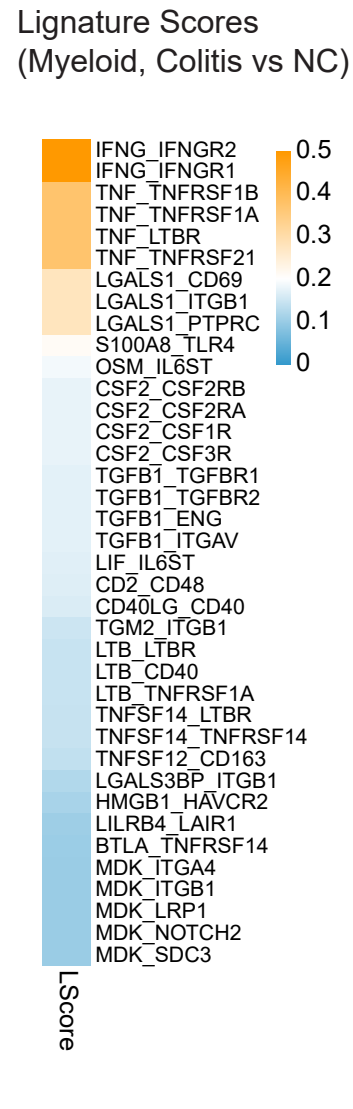

B
Myeloid, Colitis vs NC
pathway
Central carbon metabolism in cancer
Natural killer cell mediated cytotoxicity
Glycolysis / Gluconeogenesis
Human immunodeficiency virus 1 infection
Malaria
Oxidative phosphorylation
Parkinson disease
Amyotrophic lateral sclerosis
NOD−like receptor signaling pathway
Epstein−Barr virus infection
JAK−STAT signaling pathway
RIG−I−like receptor signaling pathway
Insulin signaling pathway
Pathways of neurodegeneration − multiple diseases
PD−L1 expression and PD−1 checkpoint pathway in cancer
Alzheimer disease
Shigellosis
Necroptosis
Influenza A
Osteoclast differentiation
Wnt signaling pathway
Focal adhesion
N−Glycan biosynthesis
Hedgehog signaling pathway
Dilated cardiomyopathy
NES

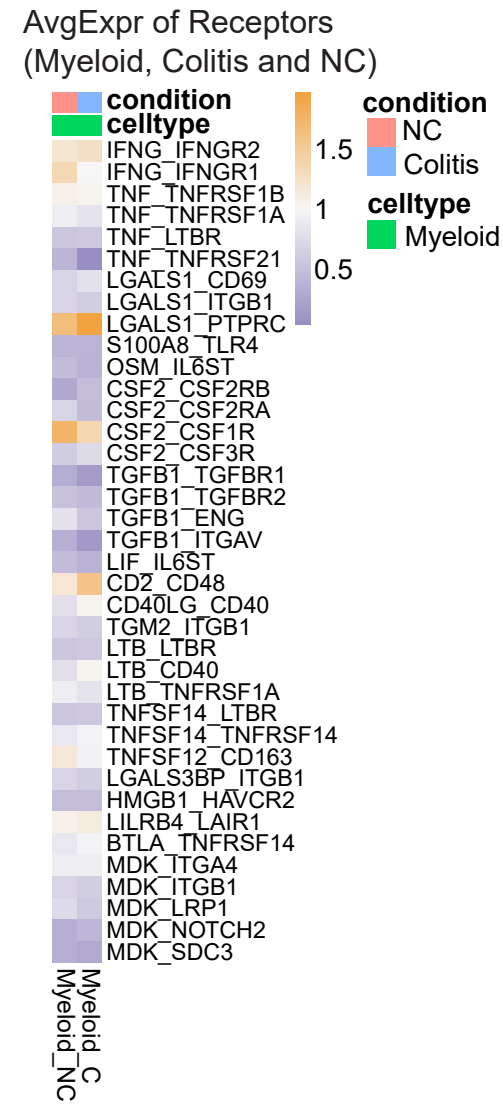

F
Receiver
Senders
CD8CTL
CD4Trm
CD8LP
CD4Treg_Tfh
CD8IEL
CyclingT
MemoryB
IgAplasmaB
NaïveB
GC_B
Group1ILC
ILC2_ILC3
Mast
Myeloid


C
SLRI (max)
Lignature score
class
bothhigh
bothlow
lowLScore
lowSLRI


D
Ligand and expression (max)
Lignature score
Ligand maxExpr
CD4Trm_NC
CD8CTL_C
CD8LP_NC
CyclingT_C
Group1ILC_NC
IgAplasmaB_NC
ILC2_ILC3_C
ILC2_ILC3_NC
Myeloid_C
Myeloid_NC
NaiveB_NC


E
AvgExpr of Ligands (Colitis and NC)
condition
celltype
NC
Colitis
CD8CTL
CD4Trm
CD8LP
CD4Treg_Tfh
CD8IEL
CyclingT
MemoryB
IgAplasmaB
NaïveB
GC_B
Group1ILC
ILC2_ILC3
Mast
Myeloid


Lignature Scores
(Myeloid, Colitis vs NC)
LScore


AvgExpr of Receptors
(Myeloid, Colitis and NC)
condition
celltype
NC
Colitis
Myeloid
Myeloid_C
Myeloid_NC