nature biotechnology

Article

https://doi.org/10.1038/s41587-023-01782-z

Comparative analysis of cell–cell communication at single-cell resolution

In the format provided by the authors and unedited

Table of contents

This PDF file includes:

SUPPLEMENTARY INFORMATION

Supplementary Text

Selection of optimal parameters for single-cell resolution CCC analysis with Scriabin

Implementing Scriabin requires users to select several parameters, particularly during ligand activity ranking as well as in the interaction program discovery workflow. Here, we discuss these parameters, their impact on obtained results, and suggestions for adjusting them based on biological and experimental considerations.

User-defined parameters for ligand activity ranking include:

1) the minimum percentage of cells expressing a ligand for that ligand to be considered expressed (% expressed ligands) – this parameter defines which ligands will be considered in Scriabin's implementation of NicheNet to identify biologically-active ligands,
2) the number of variable genes to generate the MCA embedding for calculating a cell's gene signature – this parameter reflects what genes define variation in cell state that will be captured in ligand activity ranking,
3) the distance quantile within the MCA biplot to define a cell's gene signature – this parameter defines how many genes will be used as target genes to predict ligand activities,
4) the Pearson cutoff to define an active ligand – this parameter represents the threshold of ligand activity above which the ligand will be considered biologically-active, and
5) the method for weighting the CCIM for ligand activities – this parameter determines if predicted ligand activities are considered as equally valid evidence of an interaction as expression of a cognate receptor.

To quantify the degree to which these parameters impact Scriabin's results, we tested 217 combinations of these parameters for three different datasets using Scriabin's CCIM workflow. Specifically, we first used each combination of parameters to predict ligand activity and weight the CCIM by predicted ligand activity. Next, for each of the resulting weighted CCIMs we calculated the differentially-expressed ligand-receptor pairs defining communication between each combination of cell types in each dataset. Finally, we calculated the Jaccard overlap index of DE testing results for each CCIM, reflecting the similarity of results obtained for each parameter combination.

For all three datasets, we found generally stable results across parameter combinations (mean Jaccard index 0.47-0.71; **Extended Data Fig. 8a**). We found highly stable results when the weight method "product" is used (where in the weighted CCIM the final interaction vectors are the product of ligand/receptor expression and the predicted ligand activity). In all datasets, we saw discordant results when three parameter trends were present: 1) weight method "sum", 2) low Pearson cutoff, and 3) high distance quantile. These three characteristics correspond to the least stringent parameters for CCIM weighting: in weight method "sum" predicted ligand activity

is considered equally valid evidence of an interaction as expression of a cognate receptor, ligands with relatively lower predicted activities are considered active, and genes that are less specific to an individual cell are used when generating that cell's gene signature. The percentage of expressed ligands, as well as the number of variable genes to use for gene signature calculation, did not have a substantial impact on result stability (**Extended Data Fig. 8a**). These results lead us to conclude that highly stable results can be achieved when at least one stringent ligand activity ranking parameter, including the "product" weight method, high Pearson cutoff, or low distance quantile, is used.

However, stable results are not necessarily equivalent to the most correct results. Thus, to determine how analysis parameters impacted result accuracy, we leveraged data from our in vitro NK-B cell co-culture system as a heuristic (see **Fig. 3**). In this experiment, we introduced CD40L-CD40 as an additional ligand-receptor pair edge between NK cells and B cells–thus, we evaluated which combination of parameters resulted in the most specific prediction of the CD40L-CD40 edge as differential between transfected vs. untransfected cell-cell pairs. Specifically, we applied the CCIM workflow with ligand activity ranking followed by differential expression testing using an ROC test between cells from the CD40L-CD40 transfected condition and cells from the GFP-GFP transfected condition. We define the predictive power for each ligand-receptor pair as $p = |AUC - 0.5| * 2$, and the relative predictive power for CD40L-CD40 as the difference between $p_{CD40L-CD40}$ and the average p for all other ligand-receptor pairs.

We found that ligand activity ranking with every combination of parameters improved the relative log(fold-change) of predicting CD40L-CD40 as a differential ligand-receptor edge (**Extended Data Fig. 8b-d**). Using multiple regression to identify the relative impacts of each parameter on relative predictive power, we found that a lower Pearson cutoff, a higher % expressed ligands, and the use of the "sum" weighting method resulted in a higher relative prediction power. The distance quantile and number of variable genes for cell signature calculation had little impact on the power to specifically identify CD40L-CD40 as a differential edge. We hypothesize these trends are due to the relatively high expression of *CD40LG* in this dataset–a higher % expressed ligands prevents more lowly expressed ligands from being considered, and the "sum" weighting method allows cells with low *CD40* expression to be included in the weighting. We previously showed that the "sum" weighting method, when used in combination with less stringent activity ranking parameters, can decrease result stability (**Extended Data Fig. 8a**); however, here we find that the "sum" weighting method can improve result accuracy by treating predicted ligand activities as equal evidence of interaction as expression of the ligand itself.

These results lead us to make the following recommendations for appropriate selection of these parameters:

1) The minimum percentage of cells expressing a ligand for that ligand to be considered expressed (default: 2.5%). We encourage users to select this parameter based on biological intuition as well as which sequencing platform was used. When ranking ligands across multiple cell types, users should set this value lower than the proportion

of the rarest cell type that could be contributing to communication. This value can be further lowered when analyzing data from low coverage platforms (like droplet-based methods). We encourage users to examine expression levels of potential ligands of interest before setting this value. Setting this parameter too high may result in the exclusion of ligands that are biologically important but either lowly expressed or expressed by a small subset of cells. The main issue in setting this parameter too low is an increase in computational requirements to analyze ligands that are unlikely to be biologically important. Although this parameter has only a small impact on result stability (**Extended Data Fig. 8a**) we nonetheless encourage users to test multiple thresholds to observe the robustness and impact of this parameter choice, and to err on the side of setting this parameter low.

2) The number of variable genes to generate the MCA embedding for calculating a cell's gene signature (default: 500). This value should be concordant with the total number of genes in a dataset that are expected to be upregulated by biologically active ligands. This parameter appears to have only a small impact on the results of the ligand activity ranking workflow.

3) The distance quantile within the MCA biplot to define a cell's gene signature (default: 0.05). This default value indicates that the top 5% of genes nearest a cell in the MCA biplot will be considered part of that cell's gene signature. This parameter has only a small impact on ligand activity ranking results when weight method "product" is used; however, this parameter should be set to 0.05 or lower when using weight method "sum", in order to avoid distant genes from being considered in calculation of a cell's gene signature.

4) The Pearson cutoff to define an active ligand (default: 0.075). Based on benchmarking and recommendations from the author's of NicheNet[20], reasonable thresholds for defining an active ligand may lie between 0.05-0.2. We encourage users to examine the distribution of all Pearson coefficients and the median Pearson coefficient per cell to determine if a change from the default threshold is necessary. Generally, we have found these distributions to be right-skewed, with the right tail representing putative biological activity. **Extended Data Fig. 8e** depicts an example where a Pearson threshold of 0.075 would likely capture biologically-active ligands in a system where many unknown ligands are influencing communication. In situations where a user knows a priori that only a subset of cells are responding to very few ligands, the Pearson threshold can be raised to exclude activities the user believes may be less important. We recommend setting this parameter to 0.15 or above when using weight method "sum."

5) The method for weighting the CCIM for ligand activities. Scriabin implements two different methods for weighting the CCIM based on predicted ligand activities. Method "product" (default) multiplies individual elements of the CCIM by scaled ligand activities that exceed the defined Pearson threshold. Method "sum" treats an active ligand prediction as equally valid evidence for an interaction as the expression of a corresponding receptor and sums receptor expression values and scaled ligand activities when calculating CCIM elements (see **Methods**). In our direct experimental validation (**Extended Data Fig. 8b-d**), which uses a low coverage microwell-based

scRNA-seq platform, we show that the "sum" weighting method can improve result accuracy, presumably by rescuing receiver cells with low receptor expression. However, as we have found that weight method "sum" can be associated with less stable results, particularly when low pearson cutoffs and high distance quantiles are used, we recommend using method "product" if either of the following is true: 1) a platform with high capture/coverage (eg. Smart-seq) is used, or 2) there is a factor that decreases confidence in NicheNet's ligand-target gene linkages (for example, analysis of a non-human dataset). If neither of these considerations apply, method "sum" represents one strategy to decrease CCIM sparsity and rescue bona fide interactions that involve lowly expressed receptors.

Another important user-defined parameter in Scriabin's workflow is the soft thresholding power (softPower) used in generating the adjacency matrix for interaction program discovery. In traditional gene correlation network analyses, soft thresholding increases the degree of gene-gene connectivity required to form a module, thereby lowering the influence of spurious correlations within the similarity matrix. The authors of WGCNA have recommended utilizing the lowest softPower that results in a scale-free topology fitting index (R^2) of greater than 0.8[22].

Because a single ligand can interact with multiple receptors, and vice versa, the variables of a CCIM that are used to calculate the similarity matrix for WGCNA are not independent. We hypothesized that using the same soft thresholding guidelines as recommended for WGCNA may result in highly connected interaction programs where only a single ligand or receptor is represented. We thus evaluated the impact of the R^2 threshold on interaction program size, composition, and statistical significance.

We found that, at the standard R^2 = 0.8 threshold, the mean recommended softPower was 3, and this decreased as the R^2 threshold decreased (**Extended Data Fig. 8f**). Higher R^2 thresholds were associated with smaller program sizes, particularly when R^2 > 0.5, and there was a moderate increase in the percentage of programs composed of only 1 ligand or receptor with increasing R^2 thresholds. Additionally, we observed that decreasing the R^2 threshold led to a moderate increase in the percentage of non-statistically significant programs, indicative of spurious correlations. These data indicate that an optimal R^2 threshold to avoid both spurious programs as well as programs composed of only a single ligand or receptor may lie between 0.5 and 0.75. The default R^2 for interaction program discovery is now set at 0.6.

<u>Conceptual requirements for dataset alignment for comparative analyses of summarized interaction graphs</u>

Comparing summarized interaction graphs from multiple samples requires that cells from different samples share a set of labels or annotations denoting what cells represent the same identity. Each identity class to be compared then requires representation from each of the samples to be compared. This annotation typically comes in the form of coarse, low-resolution

labels like cluster or cell type calls. We sought to minimize the degree of agglomeration required for comparative CCC analysis by maximizing the resolution of cell type identity labels.

We hypothesized that high-resolution clustering or sub-clustering is an inadequate solution because greater transcriptional perturbation between samples necessitates lower clustering resolutions to capture representation from each sample. To illustrate this observation, we analyzed a toy dataset of peripheral blood monocytes from a longitudinal experiment (**Extended Data Fig. 9**). Cells from the week 4 timepoint show a high degree of transcriptional perturbation from the other samples, visually evidenced by little overlap in low dimensional manifold embeddings. At default clustering resolution, the cluster that constitutes this sample does not contain cells from two of the samples we wish to compare. Decreasing the cluster resolution (ie. increasing the degree of agglomeration) to 0.05 improves representation of other samples but still fails to capture cells from all timepoints in each cluster. An optimal strategy for comparing summarized interaction graphs thus involves manifold alignment rather than clustering.

Scriabin's binning workflow

Identifying mutual nearest neighbors (MNNs)[25] between datasets has been implemented in several methods for dataset integration[21,36,93,94]. We reasoned that because MNNs represent pairs of cells with a shared molecular state, MNNs themselves encode single-cell resolution inter-dataset correspondences that could be generalized to high-resolution identity labels for all cells in a dataset. We refer to this process as "binning" to distinguish it from graph-based clustering. Binning with Scriabin begins by identifying MNNs between all datasets to be compared, as implemented by Seurat v3[21]. MNNs are then filtered, as we have observed that cross-cell type anchor pairs tend to have lower scores (**Extended Data Fig. 10**). Cells are initially binned with the set of MNNs with which they share the highest connectivity in the shared nearest neighbor (SNN) graph, and these bin assignments are further optimized for SNN connectivity and representation of all datasets to be compared. At the end of the binning process, each cell will have a high-resolution bin identity linking it to at least one cell from all other datasets to be compared. These identities can be used as the basis for comparative analysis of CCC that maintain near single-cell resolution.

We illustrate the utility of Scriabin's binning workflow by analyzing a dataset that contains PBMCs from both a human and a mouse donor. These samples contain the same broad cell types, but there are strong species-specific phenotypic differences that prevent the same cell types from clustering together (**Extended Data Fig. 10**). While we can annotate these broad cell types for the individual samples, we can only perform comparative analyses of CCC at the level of granularity that we can manually annotate. Scriabin's binning workflow enables a comprehensive alignment of the gene expression manifolds between species (**Extended Data Fig. 10**).

We also evaluated the performance of Scriabin's binning strategy for comparative CCC analyses when more than two samples need to be aligned. In a toy dataset of ~14,000 cells with nine sub-datasets, Scriabin identified a total of 456 bins with a median bin size of 25 cells,

maintaining near single-cell resolution (**Extended Data Fig. 10**). Additionally, cells from each bin generally shared the same orthogonal reference-based cell type annotation (**Extended Data Fig. 10**). For example, all plasmacytoid dendritic cells (pDCs) fell into a single bin that was composed of only pDCs. Bins whose cells did not share the same cell type annotation generally shared related annotations; for instance, intermediate and naive B cells frequently occupied the same bin (**Extended Data Fig. 10**).

Heterogeneity of interaction program structure and expression

The discovery of interaction programs in the intestinal development dataset[49] revealed several programs that were significantly up- or down-regulated between anatomical sampling locations, reflecting either changes in the magnitude of program expression or a shift in the proportions of cell types expressing the program (**Extended Data Fig. 6**). We also hypothesized that interaction program structure may vary between samples even if the expression magnitude of the program does not change. Thus, we calculated intramodular connectivity scores for each gene in each sampling location analyzed. We identified ligand-receptor pairs in non-differentially expressed interaction programs that had differential intramodular connectivity between sampling locations (**Extended Data Fig. 6**). For example, the *EFNB2* - *EPHB4* ligand-receptor pair has higher intramodular connectivity in hindgut than other anatomical locations, and the hindgut is the location that expresses the highest level of both of these genes (**Extended Data Fig. 6**). This indicates that, in the hindgut, when *EFNB2* and *EPHB4* are expressed, they are co-expressed along with other ligand receptor pairs in this interaction program. Collectively, this analysis provides an example of the heterogeneity, specificity, and nuanced co-expression patterns that can be revealed through the scalable discovery of interaction programs in atlas-scale datasets.

Validation of Scriabin through transfection of exogenous CCC edges

To provide additional evidence that Scriabin recovers differentially-expressed and biologically active CCC edges, we designed an experiment where we would force overexpression of a specific ligand-receptor pair, co-culture the cells, and then profile the co-cultured cells by scRNA-seq. Thus, we transfected isolated human NK cells with CD40L-encoding mRNA and isolated human B cells with CD40-encoding mRNA, followed by 12 hour co-culture and single-cell transcriptional profiling by Seq-Well[79]. We chose to profile interactions between NK cells and B cells because at baseline these cell types do not express a large number of cognate ligand-receptor pairs. We selected the CD40L-CD40 interaction because the transcriptional response downstream of CD40 in B cells is well-characterized. Further, we decided to co-culture the cells for 12 hours because this is the time at which the transcriptional response to CD40 signaling peaks[95]. Finally, we transfected the isolated NK and B cells with mRNA complexed with Charge-Altering Releasable Transporters, a class of cationic di-block oligomers that degrade into neutral small molecule metabolites after delivering anionic cargo to a cell, both facilitating cargo release and avoiding toxicity associated with persistent cations[96,97]. We chose CARTs because we have recently shown that, unlike lipofection and electroporation, CARTs transfect resting primary immune cells with high efficacy while causing minimal reconfiguration

of cell surface proteome[78], making them suitable delivery vehicles for specific manipulation of cell phenotype with few off-target effects. We also elected to transfect cells with a CART initiated with a difluoroboron-β-diketonate (BDK) fluorophore so that transfected cells could be easily distinguished by flow cytometry.

Flow cytometric analysis at the start of co-culture indicated successful transfection, with 21% of B cells expressing CD40 (with a higher mean fluorescent intensity and compared to 8% at baseline), and 13.5% of NK cells expressing CD40L (compared to 1% at baseline; **Extended Data Fig. 3**). After co-culture, B cells had upregulated CD40, consistent with known outcomes of CD40L/CD40 signaling[20], and NK cells had lost expression of CD40L, suggesting internalization of CD40L post-ligation, a well-established process for ligands and receptors expressed by NK cells (**Extended Data Fig. 3**)[98,99]. scRNA-seq profiling by Seq-Well indicated that cells expressed the exogenous mRNAs expected from the individual co-cultures (**Extended Data Fig. 3**).

Supplementary Tables

Sample	T cells	Myeloid cells	Total
LLA	1377	201	2996
LLC	380	56	1052
LLD	732	328	2671
LLT	320	56	902
LLW	337	150	983
RR1	1153	1014	5062
RR5	690	52	3721
RRJ	213	54	1574
RRP	317	35	2216

Supplementary Table 1. Sample sizes for analysis of leprosy granuloma dataset published by Ma, et al.

Marker	Fluorophore	Titration	Clone	Source	Catalog #
CD3	BV421	1:200	OKT3	BioLegend	317343
CD7	PE-Dazzle594	1:100	CD7-6B7	BioLegend	343119
CD14	PE-Cy5	1:400	61D3	ThermoFisher	15-0149-42
CD16	AlexaFluor700	1:400	3G8	BioLegend	302025
CD19	PerCP-Cy5.5	1:200	HIB19	BD Biosciences	561295
CD56	PE-Cy7	1:50	HDC56	BioLegend	318317
CD40	BV510	1:400	5C3	BioLegend	334330
CD40L	BV605	1:50	24-31	BioLegend	310826

Supplementary Table 2. Antibodies used for flow cytometric analysis.