bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Context-aware deconvolution of cell-cell communication with Tensor-cell2cell

Erick Armingol[1,2,7], Hratch M. Baghdassarian[1,2,7], Cameron Martino[1,2,3], Araceli Perez-Lopez[4], Caitlin Aamodt[2], Rob Knight[2,3,5,6], Nathan E. Lewis[2,6*]

[1] Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA
[2] Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA
[3] Center for Microbiome Innovation, University of California San Diego, La Jolla, CA 92093, USA
[4] Biomedicine Research Unit, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla, State of México 54090, México
[5] Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA
[6] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
[7] These authors equally contributed to this work: Erick Armingol, Hratch Baghdassarian
*Correspondence: nlewisres@ucsd.edu

Abstract

Cell interactions determine phenotypes, and intercellular communication is shaped by cellular contexts such as disease state, organismal life stage, and tissue microenvironment. Single-cell technologies measure the molecules mediating cell-cell communication, and emerging computational tools can exploit these data to decipher intercellular communication. However, current methods either disregard cellular context or rely on simple pairwise comparisons between samples, thus limiting the ability to decipher complex cell-cell communication across multiple time points, levels of disease severity, or spatial contexts. Here we present Tensor-cell2cell, an unsupervised method using tensor decomposition, which is the first strategy to decipher context-driven intercellular communication by simultaneously accounting for multiple stages, states, or locations of the cells. To do so, Tensor-cell2cell uncovers context-driven patterns of communication associated with different phenotypic states and determined by unique combinations of cell types and ligand-receptor pairs. As such, Tensor-cell2cell robustly improves upon and extends the analytical capabilities of existing tools. We show Tensor-cell2cell can identify multiple modules associated with distinct communication processes (e.g., participating cell-cell and ligand receptor pairs) linked to COVID-19 severities and Autism Spectrum Disorder. Thus, we introduce an effective and easy-to-use strategy for understanding complex communication patterns across diverse conditions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Organismal phenotypes arise as cells adapt and coordinate their functions through cell-cell interactions within their microenvironments[1]. Variations in these interactions and the resulting phenotypes can occur because of genotypic differences (e.g. different subjects) or the transition from one biological state or condition to another[2] (e.g. from one life stage into another, migration from one location into another, and transition from health to disease states). These interactions are mediated by changes in the production of signals and receptors by the cells, causing changes in cell-cell communication (CCC). Thus, CCC is dependent on temporal, spatial and condition-specific contexts[3], which we refer to here as cellular contexts. “Cellular contexts” refer to variation in genotype, biological state or condition that can shape the microenvironment of a cell and therefore its CCC. Thus, CCC can be seen as a function of a context variable that is not necessarily binary and can encompass multiple levels (e.g. multiple time points, gradient of disease severities, different subjects, distinct tissues, etc.). Consequently, varying contexts trigger distinct strength and/or signaling activity[1,4–6] of communication, leading to complex dynamics (e.g. increasing, decreasing, pulsatile and oscillatory communication activities across contexts). Importantly, unique combinations of cell-cell and ligand-receptor (LR) pairs can follow different context-dependent dynamics, making CCC hard to decipher across multiple contexts.

Single-cell omics assays provide the necessary resolution to measure these cell-cell interactions and the ligand-receptor pairs mediating CCC. While computational methods for inferring CCC have been invaluable for discovering the cellular and molecular interactions underlying many biological processes, including organismal development and disease pathogenesis[5], current approaches cannot account for high variability in contexts (e.g., multiple time points or phenotypic states) simultaneously. Existing methods lose the correlation structure across contexts since they involve repeating analysis for each context separately, disregarding informative variation in CCC across such factors as disease severities, time points, subjects, or cellular locations[7]. Additional analysis steps are required to compare and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compile results from pairwise comparisons[8–11], reducing the statistical power and hindering efforts to link phenotypes to CCC. Moreover, this roundabout process is computationally expensive, making analysis of large sample cohorts intractable. Thus, new methods are needed that analyze CCC while accounting for the correlation structure across multiple contexts simultaneously.

Tensor-based approaches such as Tensor Component Analysis[12] (TCA) can deconvolve patterns associated with the biological context of the system of interest. While matrix-based dimensionality reduction methods such as Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF), Uniform Manifold Approximation and Projection (UMAP) and t-distributed Stochastic Neighbor Embedding (t-SNE) can extract low-dimensional structures from the data and reflect important molecular signals[13,14], TCA is better suited to analyze multidimensional datasets obtained from multiple biological contexts or conditions[7] (e.g. time points, study subjects and body sites). Indeed, TCA outperforms matrix-based dimensionality reduction methods when recovering ground truth patterns associated with, for example, dynamic changes in microbial composition across multiple patients[15] and neuronal firing dynamics across multiple experimental trials[12]. TCA exhibits superior performance because it does not require the aggregation of datasets across varying contexts into a single matrix. It instead organizes the data as a tensor, the higher order generalization of matrices, which better preserves the underlying context-driven correlation structure by retaining mathematical features that matrices lack[16,17]. Thus, with the correlation structure retained, the use of TCA with expression data across many contexts allows one to gain a detailed understanding of how context shapes communication, as well as the specific molecules and cells mediating these processes.

Here, we introduce Tensor-cell2cell, a TCA-based strategy that deconvolves intercellular communication across multiple contexts and uncovers modules, or latent context-dependent patterns, of CCC. These data-driven patterns reveal underlying communication changes given the simultaneous interaction between contexts, ligand-receptor pairs, and cells. We demonstrate that Tensor-cell2cell is

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

broadly applicable, enabling the study of diverse biological questions about CCC in multiple organisms and contexts. We first show that Tensor-cell2cell successfully extracts complex temporal patterns from a simulated dataset. While our approach can simultaneously analyze more than two samples, we show that Tensor-cell2cell is faster, demands less memory and can achieve better accuracy in separating context-specific information in simpler analyses accessible to other tools. We further demonstrate that Tensor-cell2cell can leverage existing CCC tools by using their output communication scores to analyze multiple contexts.

Finally, we apply Tensor-cell2cell to two datasets, identifying communication patterns associated with COVID-19 severity and Autism Spectrum Disorder (ASD). In the case of COVID-19, we demonstrate that from just the single analysis of one dataset recapitulates findings across tens of published literature resources. For example, we identify stronger interactions between epithelial and immune cells with increased severity[18] and M1- and M2-like macrophage communication distinguishing moderate from severe cases. In the case of ASD, we focus on a data-driven analysis, exemplifying the possible downstream analyses that can be run on the outputs of Tensor-cell2cell, including pathway enrichment, clustering, and the generation of factor-specific communication networks. From these analyses we also show that combinations of CCC dysregulation across multiple context-dependent patterns distinguish ASD patients from controls. Thus, Tensor-cell2cell's easily interpretable output enables quick identification of key mediators of cell-cell communication across contexts, both reproducing known results and identifying novel interactors.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Deciphering context-driven communication patterns with Tensor-cell2cell

Organizing biological data through a tensor preserves the underlying correlation structure of the biological conditions of interest[12,15,17]. Extending this approach to infer cell-cell communication enables analysis of important ligand-receptor pairs and cell-cell interactions in a context-aware manner. Accordingly, we developed Tensor-cell2cell, a method based on tensor decomposition[17] that extracts context-driven latent patterns of intercellular communication in an unsupervised manner. Briefly, Tensor-cell2cell first generates a 4D-communication tensor that contains non-negative scores to represent cell-cell communication across different conditions (Figures 1a-c). Then, a non-negative TCA[19] is applied to deconvolve the latent CCC structure of this tensor into low-dimensional components or factors (Figures 1d-e). Thus, each of these factors can be interpreted as a module or pattern of communication whose dynamics across contexts is indicated by the loadings in the context dimension (Figure 1e).

To demonstrate how Tensor-cell2cell recovers latent patterns of communication, we simulated a system of 3 cell types interacting through 300 LR pairs across 12 contexts (represented in our simulation as time points) (Figure 2a). We built a 4D-communication tensor that incorporates a set of embedded patterns of communication that were assigned to certain LR pairs used by specific pairs of interacting cells, and represented through oscillatory, pulsatile, exponential, and linear changes in communication scores (Figures 2a-f; see Supplementary Notes for further details of simulating and decomposing this tensor). Using Tensor-cell2cell, we found that four factors led to the decomposition that best minimized error (Supplementary Figure S1a), consistent with the number of introduced patterns (Figure 2f). This was robustly observed in multiple independent simulations (Supplementary Figure S2a).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our simulation-based analysis further demonstrates that Tensor-cell2cell accurately detects context-dependent changes of communication, and identifies which LR pairs, sender cells, and receiver cells are important (Figure 2g). In particular, the context loadings of the TCA on the simulated tensor accurately recapitulate the introduced patterns (Figures 2f-g), while ligand-receptor and cell loadings properly capture the ligand-receptor pairs, sender cells and receiver cells assigned as participants of the cognate pattern (Figure 2g). Indeed, we observed a concordance between the “ground truth” LR pairs assigned to a pattern and their respective factor loadings through Jaccard index and Pearson correlation metrics (Supplementary Tables S1-S2**)**. Moreover, Tensor-cell2cell robustly recovered communication patterns when we added noise to the simulated tensor (Supplementary Figure S2 and Supplementary Notes).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

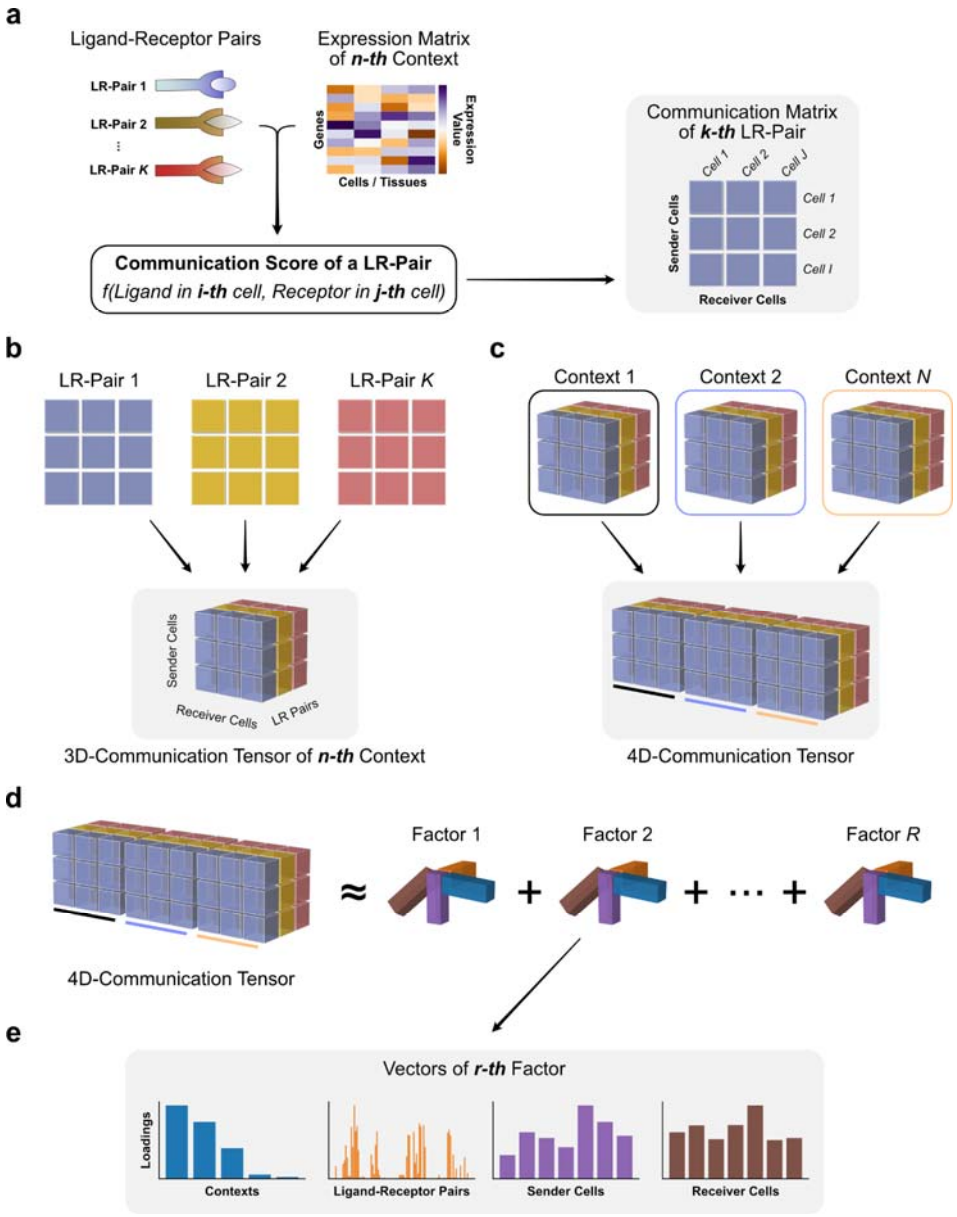


Figure 1. Tensor representation and factorization of cell-cell communication. In a given context (n-th context among N total contexts), cell-cell communication scores (see available scoring functions in REF[5]) are computed from the expression of the ligand and the receptor in a LR pair (k-th pair among K pairs) for a specific sender-receiver cell pair (i-th and j-th cells among I and J cells, respectively). This results in a communication matrix containing all pairs of sender-receiver cells for that LR pair (**a**). The same process is repeated for every single LR pair in the input list of ligand-receptor interactions, resulting in a set of communication matrices that generate a 3D-communication tensor (**b**). 3D-communication tensors are built for all contexts and are used to generate a 4D-communication tensor wherein each dimension represents the contexts (colored lines), ligand-receptor pairs, sender cells and receiver cells (**c**). A non-negative TCA model approximates this tensor by a lower-rank tensor equivalent to the sum of multiple factors of rank-one (R factors in total) (**d**). Each component or factor (r-th factor) is built by the outer product of interconnected descriptors (vectors) that contain the loadings for describing the relative contribution that contexts, ligand-receptor pairs, sender cells and receiver cells have in the factor (**e**). For interpretability, the behavior that context loadings follow represent a communication pattern across contexts. Hence, the communication captured by a factor is more relevant or more likely to be occurring in contexts with higher loadings. Similarly, ligand-receptor pairs with higher loadings are the main mediators of that communication pattern. By constructing the tensor to account for directional interactions (panels a-b), ligands and receptors in LR pairs with high loadings are mainly produced by sender and receiver cells with high loadings, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

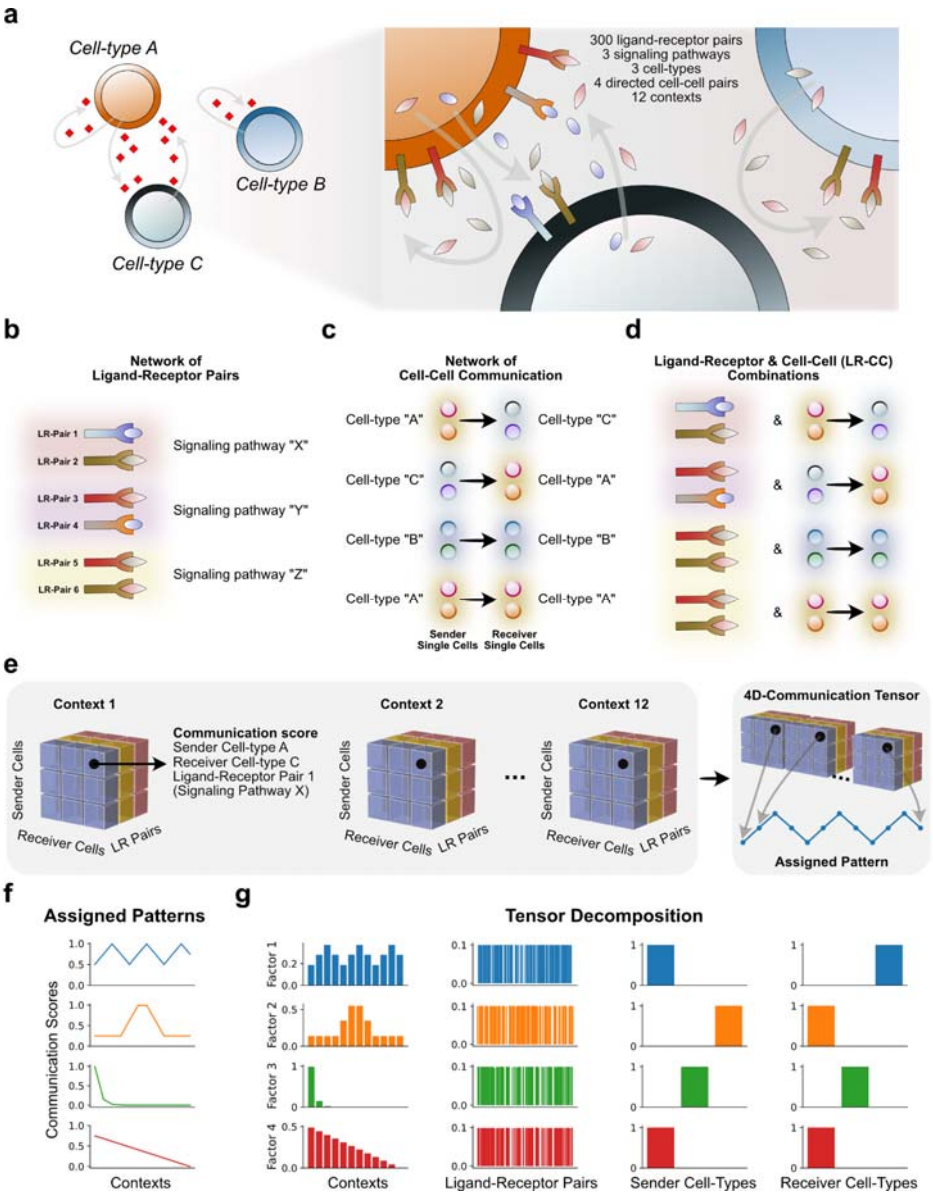


Figure 2. Tensor-cell2cell recovers simulated communication patterns. (**a**) Cell-cell communication scenario used for simulating patterns of communication across different contexts (here each a different time point). (**b**) Examples of specific ligand-receptor (LR) and (**c**) cell-cell pairs that participate in the simulated interactions. Individual LR pairs and cell pairs were categorized into groups of signaling pathways and cell types, respectively. In this simulation, signaling pathways did not overlap in their LR pairs, and each pathway was assigned 100 different LR pairs. (**d**) Distinct combinations of signaling pathways with sender-receiver cell type pairs were generated (LR-CC combinations). LR-CC combinations that were assigned the same signaling pathway overlap in the LR pairs but not in the interacting cell types. (**e**) A simulated 4D-communication tensor was built from each time point's 3D-communication tensor. Here, a communication score was assigned to each ligand-receptor and cell-cell member of a LR-CC combination. Each communication score varied across time points according to a specific pattern. (**f**) Four different patterns of communication scores were introduced to the simulated tensor by assigning a unique pattern to a specific LR-CC combination. From top to bottom, these patterns were an oscillation, a pulse, an exponential decay and a linear decrease. The average communication score (y-axis) is shown across time points (x-axis). This average was computed from the scores assigned to every ligand-receptor and cell-cell pair in the same LR-CC combination. (**g**) Results of running Tensor-cell2cell on the simulated tensor. Each row represents a factor, and each column a tensor dimension, wherein each bar represents an element of that dimension (e.g. a time point, a ligand-receptor pair, a sender cell or a receiver cell). Factor loadings (y-axis) are displayed for each element of a given dimension. Here, the factors were visually matched to the corresponding latent pattern in the tensor, and their loadings were normalized to unit Euclidean length.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tensor-cell2cell adds new capabilities to cell-cell communication analysis

To demonstrate the power of accounting for multiple contexts simultaneously, we compared the computational efficiency and accuracy of our method with respect to CellChat[10], the only tool that summarizes multiple pairwise comparisons in an automated manner (Table 1). Since CellChat cannot extract patterns of CCC across multiple contexts, we instead use the output of its joint manifold learning on pairwise-based changes in signaling pathways as a comparable proxy to the output of Tensor-cell2cell. Despite the use of these proxy comparisons, we emphasize that the conceptual outputs reported by Tensor-cell2cell are unique. Briefly, we found that Tensor-cell2cell is faster, uses less memory, and achieves higher accuracy when analyzing CCC of multiple samples (Supplementary Figure S3); using a GPU further increases computational speed of Tensor-cell2cell. See more details regarding this comparison in the Methods and *Tensor-cell2cell is fast and accurate* section of the Supplementary Notes.

A major advantage of Tensor-cell2cell is that it acts as a robust dimensionality reduction method for any communication scores arranged as a tensor. To illustrate this, we set out to harness the sample-wise communication scoring outputs of other tools. Tensor-cell2cell can restructure these outputs into a 4D-communication tensor (Figure 1), extending their capabilities to recover context-dependent patterns of communication. This generalizability enables users to employ any scoring method. Thus, we ran Tensor-cell2cell on communication scores generated by sample-specific analysis with CellPhoneDB[20], CellChat[10], NATMI[9], and SingleCellSignalR[21], as well as the built-in scoring of Tensor-cell2cell. Specifically, we analyzed twelve bronchoalveolar lavage fluid (BALF) samples from patients with different severities of COVID-19 (healthy, moderate and severe) with each method listed above. We assessed the consistency of decomposition between all five scoring methods by using the CorrIndex[22]. The CorrIndex value lies between 0 and 1, with a higher score indicating more dissimilar decomposition outputs; we thus report our similarity results as (1-CorrIndex). Our results indicate that Tensor-cell2cell

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

194 can consistently identify context-dependent communication patterns independent of the initial
195 communication scoring method (Figure 3a, Supplementary Figure S4), with a mean similarity score of
196 0.82. Furthermore, differences in decomposition results are driven at the ligand-receptor resolution, yet
197 tend not to propagate to the cell- or context-resolution (Supplementary Notes and Supplementary
198 Figures S5-6). While these results agree with previous reports regarding the inconsistency of scoring
199 methods for ligand-receptor interactions[23], they also show the power of tensor decomposition to resolve
200 these inconsistencies and identify biologically and conceptually consistent communication patterns.

201 Since Tensor-cell2cell requires the use of multiple conditions or samples, we also assessed biases that
202 may have been introduced by batch effects during gene expression count transformation (e.g.,
203 normalization, batch correction, etc). Specifically, we assessed the impact of applying the log(CPM+1)
204 and the fraction of non-zero cells as preprocessing methods[24], and ComBat[25] and Scanorama[26] as
205 batch-effect correction. Here, we also used the BALF COVID-19 samples and built the 4D-tensors
206 using the gene expression values obtained in each case. After running the tensor decomposition, these
207 strategies generated results that seem biologically comparable, as measured with a mean similarity
208 score of 0.86 (Figure 3b). As expected, using the raw counts leads to the most biased and different
209 results in comparison to the other preprocessing methods; the mean similarity score between raw
210 counts and all other approaches is 0.77. In contrast, the highest similarity was between the log(CPM+1)
211 and the non-zero fraction of cells. This result is also expected since the non-zero fraction of cells is
212 comparable to the log(CPM+1). However, the non-zero fraction performs better in comparisons of lowly
213 expressed genes[24](e.g. receptors on the cell surface[27]), so we included this fraction as part of the
214 Tensor-cell2cell built-in workflow. Thus, Tensor-cell2cell can detect consistent CCC signatures
215 independent of the method by which gene expression is corrected, with the exception of raw counts, as
216 indicated by the high similarities observed (Figure 3b).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Methodological strategy and context-based analysis in available tools

Tool	Communication Score[a]	Context Evaluation	Simultaneous Contexts	Multimeric LR pairs	Data Resolution	Platform	Refs.
Tensor-cell2cell	Expression Mean, Expression Product and Geometric Mean	Builds a tensor with all contexts simultaneously and runs a tensor decomposition, accounting for the correlation structure across contexts	Unlimited[b]	Yes	Bulk, Single Cell	Python	This work
CellChat	Mass-action-based probability	Runs separate analyses of each context, does pairwise comparisons and harmonizes them through a joint manifold learning	2	Yes	Single Cell	R	[10]
CellPhoneDB	Expression Mean	None	1	Yes	Single Cell	Python	[20]
CellTalker	Differential Combinations	Differential analysis between two contexts	2	No	Single Cell	R	[8]
Connectome	Modified Expression Product	Differential analysis between two contexts. An overall analysis of cell-type importance can be done for more contexts	2	No	Single Cell	R	[11]
ICELLNET	Expression Product	None	1	Yes	Bulk, Single Cell	R	[28]
iTalk	Differential Combinations	Differential analysis between two contexts	2	No	Single Cell	R	[29]
NATMI	Expression Product and Normalized Expression Product	None	1	No	Bulk, Single Cell	Python	[9]
NicheNet	Personalized-PageRank-based score	None	1	No	Bulk, Single Cell	R	[30]
scAgeCom	Geometric Mean	Differential analysis between two contexts	2	Yes	Single Cell	R	[31]
scTensor	Expression Product	None	1	No	Single Cell	R	[32]
SingleCellSignalR	Regularized Expression Product	None	1	No	Single Cell	R	[21]

[a] For further details about distinct communication scores, see REF[5] and/or respective references for each tool.
[b] Dependent on computational resources (e.g. memory availability)
LR, ligand-receptor

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

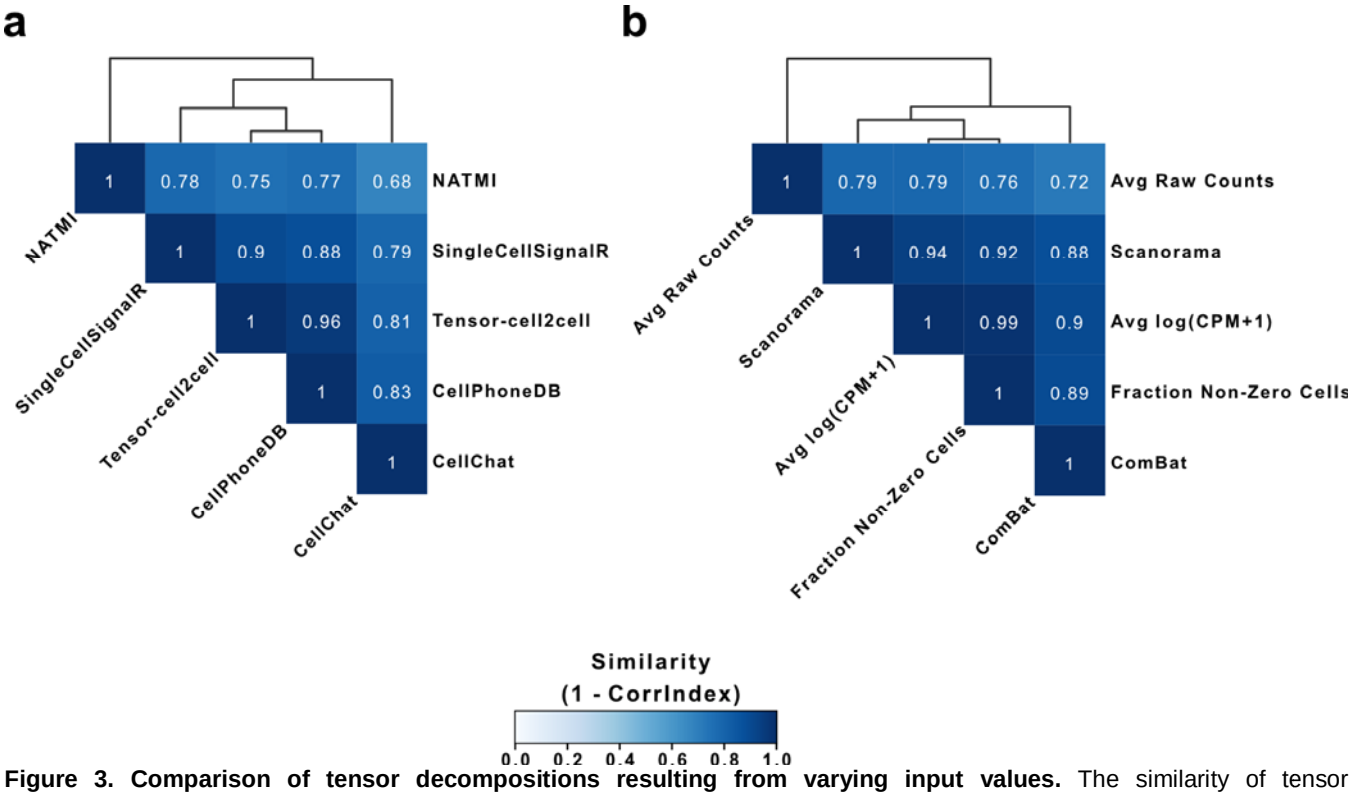


Figure 3. Comparison of tensor decompositions resulting from varying input values. The similarity of tensor decompositions performed on 4D-communication tensors constructed from the single-cell dataset of BALF in patients with varying severities. For a given comparison, constructed tensors have the same elements in each dimension. (**a**) Similarity between tensor decompositions performed on 4D-communication tensors, each corresponding to communication scores computed from different tools for inferring cell-cell communication. The scoring functions correspond to those of CellChat[10], CellPhoneDB[20], NATMI[9], SingleCellSignalR[21] and the built-in methods in Tensor-cell2cell. (**b**) Similarity between tensor decompositions performed on 4D-communication tensors, each modifying the gene expression values by different preprocessing methods (log(CPM+1) and the fraction of non-zero cells[24]) or batch-effect correction methods (Combat[25] and Scanorama[26]), as well as using the raw counts. The communication scores in (**b**) were calculated as the mean expression between the partners in each LR pair, previously aggregating gene expression at the single-cell level into the cell-type level. In (**a**) and (**b**) similarity was measured as (1-CorrIndex), where the CorrIndex[22] is a distance metric for comparing different decompositions on tensors containing the same indices and its values range from 0 to 1 (more similar to more dissimilar). Assessed methods were hierarchically clustered by the similarities of their tensor decompositions.

Tensor-cell2cell links intercellular communication with varying severities of COVID-19

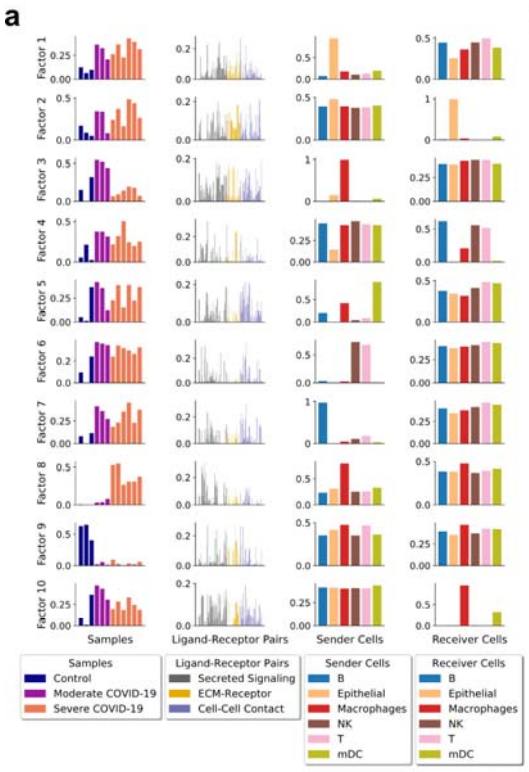
Great strides have been made to unravel molecular and cellular mechanisms associated with SARS-CoV-2 infection and COVID-19 pathogenesis. Thus, we tested our method on a single-cell dataset of BALF samples from COVID-19 patients[33] to see how many cell-cell and LR pair relationships in COVID-19 could be revealed by Tensor-cell2cell. By decomposing the tensor associated with this dataset into 10 factors (Figure 4a and Supplementary Figure S1b), Tensor-cell2cell found factors representing communication patterns that are highly correlated with COVID-19 severity (Figure 4c) and other factors

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that distinguish features of the different disease stages (Supplementary Figure S7), consistent with the high performance that the classifier achieved for this dataset (Supplementary Figure S3f,h). Furthermore, these factors involve signaling molecules previously linked with severity in separate works (Supplementary Table S3).

The first two factors capture CCC involving autocrine and paracrine interactions of epithelial cells with immune cells in BALF (Figure 4a). The sample loadings of these factors reveal a communication pattern wherein the involved LR and cell-cell interactions become stronger as severity increases (Spearman correlation of 0.72 and 0.61, Figure 4c and Supplementary Figure S7). Although this observation was not reported in the original study, it is consistent with a previous observation of a correlation between COVID-19 severity and the airway epithelium-immune cell interactions[18]. Specifically, epithelial cells are highlighted by Tensor-cell2cell as the main sender cells in factor 1 (Figure 4a), and we further provide new details of the molecular mechanisms involving top ranked signals such as APP, MDK, MIF and CD99 (Figure 4b). These molecules have been reported to be produced by epithelial cells[34–40] and participate in immune cell recruiting[36–38,41], in response to mechanical stress in lungs[39] and regeneration of the alveolar barrier during viral infection[40]. In addition, epithelial cells act as the main receiver in factor 2 (Figure 4a), involving proteins such as PLXNB2, SDC4 and F11R (Figure 4b), which were previously determined important for tissue repair and inflammation during lung injury[42–44]. Remarkably, a new technology for experimentally tracing CCC revealed that SEMA4D-PLXNB2 interaction promotes inflammation in a diseased central nervous system[45]; our approach suggests a similar role promoting inflammation in severe COVID-19, specifically mediating the communication between immune and epithelial cells, as reflected in factor 2 (Figure 4b).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



b

Top-5 Ligand-Receptor Pairs

Factor 1		Factor 2	
APP - CD74	0.268	SEMA4D - PLXNB2	0.212
MDK - NCL	0.246	SEMA4A - PLXNB2	0.203
MIF - CD74 & CD44	0.231	MDK - SDC4	0.202
MDK - ITGA4 & ITGB1	0.227	COL9A2 - SDC4	0.186
CD99 - CD99	0.220	F11R - F11R	0.186

Factor 3		Factor 4	
SIGLEC1 - SPN	0.194	MIF - CD74 & CD44	0.321
RETN - CAP1	0.182	LGALS9 - CD44	0.289
MDK - NCL	0.177	COL9A2 - CD44	0.245
FN1 - ITGA4 & ITGB1	0.176	LAMB3 - CD44	0.242
FN1 - ITGA4 & ITGB7	0.172	LAMB2 - CD44	0.234

Factor 5		Factor 6	
CD99 - CD99	0.213	CD99 - CD99	0.333
ITGB2 - CD226	0.211	CCL5 - CCR5	0.305
CD86 - CTLA4	0.210	CCL5 - CCR1	0.293
ITGB2 - ICAM2	0.201	GZMA - F2R	0.241
ITGB2 - ICAM1	0.192	PTPRC - MRC1	0.222

Factor 7		Factor 8	
CD99 - CD99	0.307	CCL2 - CCR2	0.297
MIF - CD74 & CXCR4	0.241	CCL3 - CCR1	0.275
CD22 - PTPRC	0.231	CCL8 - CCR1	0.261
MIF - CD74 & CD44	0.229	CCL3 - CCR5	0.238
SELL - MADCAM1	0.191	CCL3L1 - CCR1	0.238

Factor 9		Factor 10	
FN1 - CD44	0.274	CD99 - PILRA	0.191
MIF - CD74 & CD44	0.269	LGALS9 - HAVCR2	0.170
APP - CD74	0.266	ANXA1 - FPR1	0.163
PTPRC - MRC1	0.263	MDK - LRP1	0.154
RETN - CAP1	0.243	PTPRC - MRC1	0.153

c

Factor	Spearman Coefficient	Gini Coefficient
Factor 1	0.72	0.50
Factor 2	0.61	0.76
Factor 3	-0.26	0.75
Factor 4	0.39	0.48
Factor 5	0.40	0.59
Factor 6	0.25	0.65
Factor 7	0.51	0.68
Factor 8	0.92	0.24
Factor 9	-0.51	0.09
Factor 10	-0.02	0.74

Figure 4. Deconvolution of intercellular communication in patients with varying severity of COVID-19. (**a**) Factors obtained after decomposing the 4D-communication tensor from a single-cell dataset of BALF in patients with varying severities of COVID-19. 10 factors were selected for the analysis, as indicated in Supplementary Figure S1b. Here, the context corresponds to samples coming from distinct patients (12 in total, with three healthy controls, three moderate infections, and six severe COVID-19 cases). Each row represents a factor and each column represents the loadings for the given tensor dimension (samples, LR pairs, sender cells and receiver cells), normalized to unit Euclidean length. Bars are colored by categories assigned to each element in each tensor dimension, as indicated in the legend. (**b**) List of the top 5 ligand-receptor pairs ranked by loading for each factor. The corresponding ligands and receptors in these top-ranked pairs are mainly produced by sender and receiver cells with high loadings, respectively. Ligand-receptor pairs with supporting evidence (Supplementary Table S3) for a relevant role in general immune response (black bold) or in COVID-19-associated immune response (red bold) are highlighted. (**c**) Coefficients associated with loadings of each factor: Spearman coefficient quantifying correlation between sample loadings and COVID-19 severity, and Gini coefficient quantifying the dispersion of the edge weights in each factor-specific cell-cell communication network (to measure the imbalance of communication). Important values are highlighted in red (higher absolute Spearman coefficients represent stronger correlations; while smaller Gini coefficients represent distributions with similar edge weights).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our strategy also elucidates communication patterns attributable to specific groups of patients according to disease severity (Figure 4a). For example, we found interactions that are characteristic of severe (factor 8) and moderate COVID-19 (factors 3 and 10), and healthy patients (factor 9) (adj. P-value < 0.05, Supplementary Figure S7). Factor 8 was the most correlated with severity of the disease (Spearman coefficient 0.92, Figure 4c) and highlights macrophages playing a major role as pro-inflammatory sender cells. Their main signals include CCL2, CCL3 and CCL8, which are received by cells expressing the receptors CCR1, CCR2 and CCR5 (Figure 4b). Consistent with our result, another study of BALF samples[18] revealed that critical COVID-19 cases involve stronger interactions of cells in the respiratory tract through ligands such as CCL2 and CCL3, expressed by inflammatory macrophages[18]. Moreover, the inhibition of CCR1 and/or CCR5 (receptors of CCL2 and CCL3) has been proposed as a potential therapeutic target for treating COVID-19[18,46]. Tensor-cell2cell also deconvolved patterns attributable to moderate rather than severe COVID-19, also highlighting interactions driven by macrophages (factors 3 and 10; Figure 4a). However, top-ranked molecules (Figure 4b) and gene expression patterns (Supplementary Figure S8) suggest that the intercellular communication is led by macrophages with an anti-inflammatory M2-like phenotype, in contrast to factor 8 (pro-inflammatory phenotype). Multiple top-ranked signals in factors 3 and 10 have been associated with an M2 macrophage phenotype acting in the immune response to SARS-CoV-2[47–52].

In contrast to severe and moderate COVID-19 patients, communication patterns associated with healthy subjects involve all sender-receiver cell pairs with a similar importance. In particular, factor 9 (Figure 4a) demonstrated the smallest Gini coefficient (0.09; Figure 4c), which measures the extent to which edge weights between sender and receiver cells are evenly distributed in the factor-specific cell-cell communication network. Smaller Gini coefficients show more even distributions, i.e., more equally weighted potential of communication across sender and receiver cell pairs (see ***Methods***). This indicates that the intercellular communication represented by factor 9 is ubiquitous across cell types.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Thus, this conservation across cells may be an indicator of communication during homeostasis, since the context loadings for this factor are not associated with disease (Supplementary Figure S7). Interestingly, a top-ranked LR pair in factor 9 is MIF-CD74/CD44 (Figure 4b), which is consistent with ubiquitous expression of MIF across tissues and its protective role in normal conditions[40,53]. Thus, Tensor-cell2cell extracts communication patterns distinguishing one group of patients from another and detects known mechanisms of immune response during disease progression (Supplementary Notes), which is important for therapeutic applications.

Tensor-cell2cell elucidates communication mechanisms associated with Autism Spectrum Disorders

Dysregulation of neurodevelopment in Autism Spectrum Disorders (ASD) is associated with perturbed signaling pathways and CCC in complex ways[54]. To understand these cellular and molecular mechanisms, we analyzed single-nucleus RNA-seq (snRNA-seq) data from postmortem prefrontal brain cortex (PFC) from 13 ASD patients and 10 controls[55]. We built a 4D-communication tensor containing 16 cell types present in all samples, including neurons and non-neuronal cells, and 749 LR pairs; then we used Tensor-cell2cell to deconvolve their associated CCC into 6 context-driven patterns (Figure 5a and Supplementary Figure S1c). In these factors, we observe communication between all neurons (factor 1), as well as communication of specific neurons in the cortical layers I-VI (factors 2 and 3), interneurons (factor 4), astrocytes and oligodendrocytes (factor 5), and endothelial cells (factor 6).

Tensor-cell2cell's outputs can be further dissected using downstream analyses with common approaches. To illustrate this, we ranked the LR pairs by their loadings in a factor-specific fashion, and ran Gene Set Enrichment Analysis[56] (GSEA) using LR pathway sets built from KEGG pathways[57] (see ***Methods***). We observed that each factor was associated with different biological functions including axon guidance, cell adhesion, extracellular-matrix-receptor interaction, ERBB signaling, MAPK

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

signaling, among others (Figure 5b). Dysregulation of axon guidance, synaptic processes and MAPK pathway have been previously linked to ASD from differential analysis[55,58], supporting our observations. Moreover, our results extend to other roles associated with extracellular matrix, focal adhesion of cells, regulation of actin cytoskeleton, and signaling through ErbB receptors, which involves Akt, PI3K, and mTOR pathways, as well as regulation of cell proliferation, migration, motility, differentiation, and apoptosis[59]. Thus, Tensor-cell2cell outputs can be used to assign macro-scale biological functions to each of the factors, extending the interpretation of factor-specific CCC.

After identifying main pathways involved in each factor, one can further use sample loadings to identify how these functions are associated with each sample group. By doing so, we found that factors 3 and 4 significantly distinguish ASD from typically-developing controls (Figure 5c). Neurons in cortical layers are the main sender cells in factor 3, while interneurons are key receiver cell types in factor 4 (Figure 5a and Supplementary Figure S9), with parvalbumin interneurons (IN-PV), and SV2C-expressing interneurons (IN-SV2C) as the top ranked cells, consistent with the previously reported cell types that are more affected in ASD condition[55] (i.e., with a greater number of dysregulated genes), and that correspond to neurons in the cortical layers I-VI, IN-SV2C and IN-PV. Thus, considering the overall decreased sample loadings in the ASD group, the GSEA results, and the factor-specific CCC networks built from the cell loadings (Supplementary Figure S9), our analysis suggests that there is a downregulation of axon guidance, cell adhesion, and ERBB signaling involving neurons in cortical layers I-VI and interneurons in ASD patients. See Supplementary Notes for further discussion.

Clustering methods can be applied for grouping samples in an unsupervised manner. Thus, we can assess the overall similarity between samples across all factors; considering combinations of factors can offer additional insights to the analysis as compared to considering one factor at a time. We use hierarchical clustering to group samples into four main clusters (Figure 5d). Cluster 1 mainly groups controls, cluster 2 is not associated with any category, cluster 3 mostly represents ASD patients, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cluster 4 is completely related to ASD condition. These clusters also reveal that combinations of factors separate samples by ASD and control groups. For example, samples in cluster 1 seem to have smaller loadings in factors 1 and 5, and higher loadings in factors 3 and 4. Interestingly, the only ASD sample present in this cluster had the smallest ASD clinical score, suggesting that CCC mechanisms are more similar to controls when the phenotype is mild. In contrast, cluster 3 shows an opposite CCC behavior to cluster 1. Cluster 4 also reveals that the combination of factor 6 with low sample loadings and factors 1 and 5 with high values is a strong marker of several ASD patients, even though factors 1, 5, and 6 did not show significant differences between sample groups (Figure 5c). Based on this, patients in cluster 4 had increased CCC through the NRXNs-NRLGs, CTNs-NRCAMs, and NCAMs-NCAMs interactions (synapse and cell adhesion) in neurons as senders and receivers, and astrocytes and oligodendrocytes as receivers, as well as a decreased CCC through VEGFs-FLT1, PTPRM-PTPRM, and PTN-NCL interactions (angiogenesis, neural migration and neuroprotection) related to endothelial cells as the main receivers (Supplementary Table S4). Finally, both ASD-clusters seem to be slightly distinct in terms of phenotype, considering their mean clinical scores of 25.0 and 22.8, respectively for clusters 3 and 4, but without significant differences. Thus, downstream analyses reveal that multiple dysregulations of CCC patterns captured by Tensor-cell2cell occur simultaneously in ASD condition (Figure 5d), even though these patterns could not explain phenotypic differences when considered in isolation (Figures 5c).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

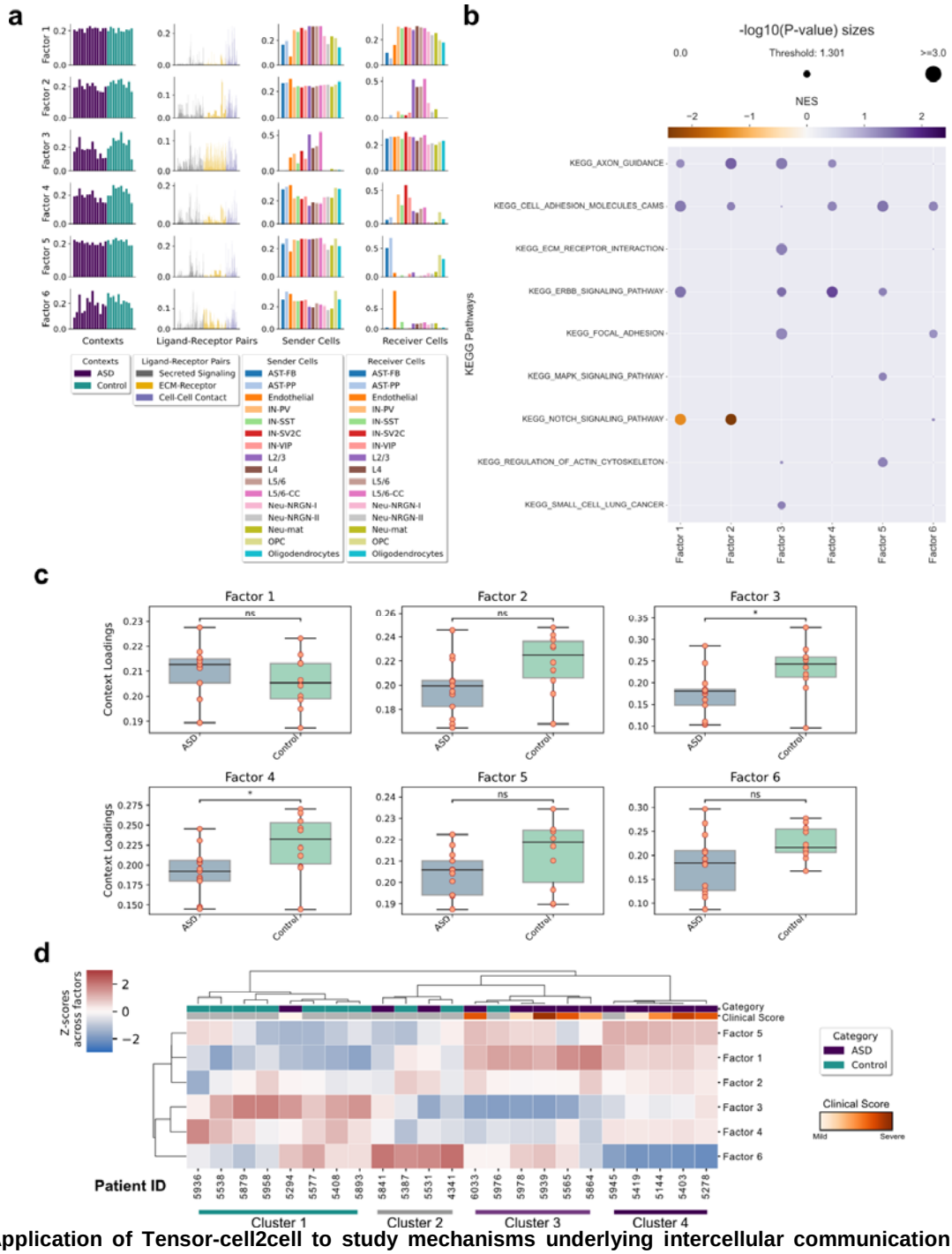


Figure 5. Application of Tensor-cell2cell to study mechanisms underlying intercellular communication in patients with ASD. (**a**) Factors obtained after decomposing the 4D-communication tensor from a single-nucleus dataset of prefrontal brain cortex samples from patients with or without ASD. Six factors were selected for the analysis, as indicated in Supplementary Figure S1c. Here, the context corresponds to samples coming from distinct patients (23 in total, with thirteen ASD patients and ten controls). Each row represents a factor and each column represents the loadings for the given tensor dimension (samples, LR pairs, sender cells and receiver cells), normalized to unit Euclidean length. Bars are colored by categories assigned to each element in each tensor dimension, as indicated in the legend. Cell-type annotations are those used in REF[55]. (**b**) GSEA performed on the pre-ranked LR pairs by their respective loadings in each factor, and using KEGG pathways. Dot sizes are proportional to the negative logarithmic of the P-values, as indicated at the top of the panel. The threshold value indicates the size of a P-value=0.05. The dot colors represent the normalized enrichment score (NES) after

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the permutations performed by the GSEA, as indicated by the colorbar. (**c**) Boxplot representation for ASD and control groups of patients. Each panel represents the sample loadings, grouped by disease condition, in each of the factors. Groups were compared by a two-sided independent t-test, followed by a Bonferroni correction. Annotations of significance are ns (P-value >= 0.05) and * (0.01 < P-value < 0.05). (**d**) Heatmap of the standardized sample loadings across factors (z-scores) for each of the samples. An agglomerative hierarchical clustering was performed on the z-scores to group both samples and factors. Major clusters of the samples are indicated at the bottom. The category of each sample is colored on the top, according to the legend. A clinical score of each patient is also shown, according to the colorbar. Controls, and ASD samples without an assigned score, were colored gray. This clinical score summarizes the social interactions, communication, repetitive behaviors, and abnormal development of the patients, as indicated in REF[55].

Discussion

Here we present Tensor-cell2cell, a computational approach that identifies modules of cell-cell communication and their changes across contexts (e.g., across subjects with different disease severity, multiple time points, different tissues, etc.). Our approach can rank LR pairs based on their contribution to each communication module and connect these signals to specific cell types and phenotypes. Tensor-cell2cell's ability to consider multiple contexts simultaneously to identify context-dependent communication patterns is novel, in contrast to state-of-the-art tools that are either unaware of the context driving CCC[5,20,30,60] or require analysis of each context separately to perform pairwise comparisons in posterior steps[10,61]. Tensor-cell2cell is therefore a flexible method that can integrate multiple datasets and readily identify patterns of intercellular communication in a context-aware manner, reporting them through interconnected and easily interpretable scores.

Tensor-cell2cell robustly detects communication patterns using many other scoring methods[13]. Thus, our method is not only an improvement over other tools, but also greatly extends these tools, enabling new analyses with existing methods. One can choose any tool of interest, run it on each context separately, and use the resulting communication scores to build and deconvolve a 4D-communication tensor. Other tools, such as CellChat, allow the generation of scores at the signaling pathway level instead of LR pairs. This, combined with Tensor-cell2cell, could provide additional information about changes in signaling pathways. Thus, Tensor-cell2cell can also be used for analyzing any other score linking gene expression from cell pairs, beyond just scores based on protein-protein interactions. In this regard, our tool outputs consistent results regardless of the preprocessing and batch correction method

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

413 we evaluated (Figure 3b). Nevertheless, it is best practice to employ integration/batch-correction
414 methods to correct gene expression and annotate cell types before running Tensor-cell2cell to ensure
415 this source of variation is controlled[62].

416 Tensor-cell2cell is faster for analyzing multiple samples than pairwise comparisons, providing a
417 considerable improvement in running time and reduced memory requirements (Supplementary Notes).
418 Tensor-cell2cell's runtime can be further accelerated when a GPU is available (Supplementary Figure
419 S3a). It is also more accurate, resulting in 10-20% higher classification accuracy of subjects with
420 COVID-19 when compared to CellChat (Supplementary Figures S3e-h). However, we note that
421 benchmarking CCC prediction tools is challenging due to the lack of a ground truth[5], and it is hard to
422 compare and evaluate tools because of the qualitative differences in their outputs[23] (Supplementary
423 Notes). While pairwise comparisons can be informative about differential cellular and molecular
424 mediators of communication, the results are less interpretable (Supplementary Figures S10-13), do not
425 provide the multi-scale resolution available in Tensor-cell2cell (Figures 4a and 5a), and do not identify
426 context-dependent patterns.

427 Meaningful biology can be easily identified from Tensor-cell2cell. For example, a manual interpretation
428 of the BALF COVID-19 decomposition found communication results not previously observed in the
429 original study[33] and recapitulated findings spanning tens of peer-reviewed articles (Supplementary
430 Table S3). This included a correlation between the lung epithelium-immune cell interactions and
431 COVID-19 severity[18] and molecular mediators that distinguished moderate and severe COVID-19 (see
432 *Tensor-cell2cell elucidates molecular mechanisms distinguishing moderate from severe COVID-19* in
433 the Supplementary Notes). Additionally, Tensor-cell2cell results can be coupled with downstream
434 analysis methods to facilitate interpretation and provide further insights of underlying mechanisms. In
435 our ASD case-study (Figure 5), such analyses included GSEA, clustering, visualization and statistical
436 comparison of factors, and factor-specific analysis of sender-receiver communication networks

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Supplementary Figure S9). In the ASD case-study, we found dysregulated CCC directly distinguished ASD patients from controls and was linked with a downregulation of axon guidance, cell adhesion, synaptic processes, and ERBB signaling in cortical neurons and interneurons (Figures 5a,b), consistent with previous evidence[55,58,63,64]. Moveover, accounting for the combinatorial relationship of samples across factors demonstrated additional complex relationships of CCC dysregulation (Figure 5d).

A limitation to consider is the potential of missing communication scores in the tensor (e.g., when a rare cell type appears in only one sample). Although Tensor-cell2cell can handle cell types that are missing in some conditions, the implemented tensor decomposition algorithm can be further optimized for missing values. Since the implemented algorithm is not optimized for this purpose, we built a 4D-communication tensor that contains only the cell types that are shared across all samples in our COVID-19 and ASD study cases. Thus, further developments will facilitate analyses with missing values to include all possible members of communication (i.e., LR pairs and cell types that may be missing in certain contexts).

In addition to single cell data analyzed here, Tensor-cell2cell also accepts bulk transcriptomics data (an example of a time series bulk dataset of *C. elegans* is included in a Code Ocean capsule, see ***Methods***), and it could further be used to analyze proteomic data. We demonstrated the application of Tensor-cell2cell in cases where samples correspond to distinct patients, but it can be applied to many other contexts. For instance, our strategy can be readily applied to time series data by considering time points as the contexts, and to spatial transcriptomic datasets, by previously defining cellular niches or neighborhoods as the contexts. We have included Tensor-cell2cell as a part of our previously developed tool cell2cell[65], enabling previous functionalities such as employing any list of LR pairs (including protein complexes), multiple visualization options, and personalizing the communication scores to account for other signaling effects such as the (in)activation of downstream genes in a signaling pathway[30,66]. Thus, these attributes make Tensor-cell2cell valuable for identifying key cell-cell

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and LR pairs mediating complex patterns of cellular communication within a single analysis for a wide range of studies.

Methods

RNA-seq data processing

RNA-seq datasets were obtained from publicly available resources. Datasets correspond to a large-scale single-cell atlas of COVID-19 in humans[67], a COVID-19 dataset of single-cell transcriptomes for BALF samples[33]. COVID-19 datasets were collected as raw count matrices from the NCBI's Gene Expression Omnibus[68] (GEO accession numbers GSE158055 and GSE145926, respectively), while the ASD dataset was obtained from https://cells.ucsc.edu/autism/downloads.html. In total, the first dataset contains 1,462,702 single cells, the second 65,813 and the last one 104,559 single nuclei. The first dataset contains samples of patients with varying severities of COVID-19 (control, mild/moderate and severe/critical) and we selected just 60 PBMC samples among all different sample sources (20 per severity type). In the second dataset, we considered the 12 BALF samples of patients with varying severities of COVID-19 (3 control, 3 moderate and 6 severe) and preprocessed them by removing genes expressed in fewer than 3 cells, which left a total of 11,688 genes in common across samples. In the ASD dataset, PFC samples from 23 patients with and without ASD condition (13 ASD patients and 10 controls) were considered, and preprocessed similarly to the BALF dataset, resulting in a total of 24,298 genes in common across samples. In all datasets, we used the cell type labels included in their respective metadata. We aggregated the gene expression from single cells/nuclei into cell types by calculating the fraction of cells in the respective label with non-zero counts, as previously recommended for properly representing genes with low expression levels[24], as usually happens with genes encoding surface proteins[27].

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ligand-receptor pairs

A human list of 2,005 ligand-receptor pairs, 48% of which include heteromeric-protein complexes, was obtained from CellChat[10]. We filtered this list by considering the genes expressed in the PBMC and BALF expression datasets and that match the IDs in the list of LR pairs, resulting in a final list of 1639 and 189 LR pairs, respectively. While in the ASD dataset, 749 LR pairs that matched the gene IDs were considered.

Building the context-aware communication tensor

For building a context-aware communication tensor, three main steps are followed: 1) A communication matrix is built for each ligand-receptor pair contained in the interaction list from the gene expression matrix of a given sample. To build this communication matrix, a communication score[5] is assigned to a given LR pair for each pair of sender-receiver cells. The communication score is based on the expression of the ligand and the receptor in the respective sender and receiver cells (Figure 1a). 2) After computing the communication matrices for all LR pairs, they are joined into a 3D-communication tensor for the given sample (Figure 1b). Steps 1 and 2 are repeated for all the samples (or contexts) in the dataset. 3) Finally, the 3D-communication tensors for each sample are combined, each of them representing a coordinate in the 4th-dimension of the 4D-communication tensor (or context-aware communication tensor; Figure 1c).

To build the tensor for all datasets, we computed the communication scores as the mean expression between the ligand in a sender cell type and cognate receptor in a receiver cell type, as previously described[20]. For the LR pairs wherein either the ligand or the receptor is a multimeric protein, we used the minimum value of expression among all subunits of the respective protein to compute the communication score. In all cases we further considered cell types that were present across all samples. Thus, the 4D-communication tensor for the PBMC, BALF and ASD datasets resulted in a size

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of 60 x 1639 x 6 x 6; 12 x 189 x 6 x 6, and 23 x 749 x 16 x 16 respectively (that is, samples x ligand-receptor pairs x sender cell types x receiver cell types).

Non-negative tensor component analysis

Briefly, non-negative TCA is a generalization of NMF to higher-order tensors (matrices are tensors of order two). To detail this approach, let χ represent a *C* x *P* x *S* x *T* tensor, where *C*, *P*, *S* and *T* correspond to the number of contexts/samples, ligand-receptor pairs, sender cells and receiver cells contained in the tensor, respectively. Similarly, let χ_{ijkl} denote the representative interactions of context *i*, using the LR pair *j,* between the sender cell *k* and receiver cell *l*. Thus, the TCA method underlying Tensor-cell2cell corresponds to CANDECOMP/PARAFAC[69,70], which yields the decomposition, factorization or approximation of χ through a sum of *R* tensors of rank-1 (Figure 1d):

$$\chi \approx \sum_{r=1}^{R} c^r \otimes p^r \otimes s^r \otimes t^r \quad (1)$$

Where the notation $\otimes$ represents the outer product and $c^r, p^r, s^r and\ t^r$ are vectors of the factor *r* that contain the loadings of the respective elements in each dimension of the tensor (Figure 1e). These vectors have values greater than or equal to zero. Similar to NMF, the factors are permutable and the elements with greater loadings represent an important component of a biological pattern captured by the corresponding factor. Values of individual elements in this approximation are represented by:

$$\chi_{ijkl} \approx \sum_{r=1}^{R} c_i^r \otimes p_j^r \otimes s_k^r \otimes t_l^r \quad (2)$$

The tensor factorization is performed by iterating the following objective function until convergence through an alternating least squares minimization[17,71]:

$$min_{\{c,p,s,t\}} \left|\left|\chi - \sum_{r=1}^{R} c^r \otimes p^r \otimes s^r \otimes t^r\right|\right|_F^2 \quad (3)$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Where $||\ .\ ||_F^2$ represent the squared Frobenius norm of a tensor, calculated as the sum of element-wise squares in the tensor:

$$||\ \chi\ ||_F^2 = \sum_{i=1}^{C}\sum_{j=1}^{P}\sum_{k=1}^{S}\sum_{l=1}^{T} \chi_{ijkl}{}^2 \quad (4)$$

All the described calculations were implemented in Tensor-cell2cell through functions available in Tensorly[72], a Python library for tensors.

Measuring the error of the tensor decomposition

Depending on the number of factors used for approximating the 4D-communication tensor, the reconstruction error calculated in the objective function can vary. To quantify the error with an interpretable value, we used a normalized reconstruction error as previously described[12]. This normalized error is on a scale of zero to one and is analogous to the fraction of unexplained variance used in PCA:

$$\frac{\left|\left|\chi - \sum_{r=1}^{R} c^r \otimes p^r \otimes s^r \otimes t^r\right|\right|_F^2}{||\ \chi\ ||_F^2} \quad (5)$$

Running Tensor-cell2cell with communication scores from external tools

We assessed the similarity of tensor decomposition on the BALF dataset using different communication scoring methods (CellChat[10], CellPhoneDB[20], NATMI[9], SingleCellSignalR[21], and Tensor-cell2cell's built-in scoring). To enable consistency between methods, we used the same ligand-receptor PPI database (CellChat – see "Ligand-receptor pairs") and ran each method via LIANA[23]. LIANA provides a number of advantages over running each tool separately, including consistent thresholding and parameters, interoperability between methods and LR databases, and modifications to allow methods that could not originally account for protein complexes to do so. We adjusted parameters to match those of Tensor-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cell2cell's built-in scoring by not filtering for minimal proportions of expression by cell type or thresholding for differentially expressed genes.

As input to LIANA, we constructed a Seurat object with log(CPM+1) normalized counts for each sample. For each tool and sample, LIANA outputs an edge-list of communication scores for a given combination of sender and receiver cells, as well as ligand-receptor pairs. We extended Tensor-cell2cell's functionalities to restructure a set of these edge-lists, each associated with a sample, into a 4D-communication tensor (Figure 1). This functionality enables users to either provide input expression matrices and use Tensor-cell2cell's built-in scoring, or to run their communication scoring method of choice on each sample and provide the resultant edge-lists as input. To further ensure consistency, we subsetted each resultant tensor to the intersection of ligand-receptor pairs scored across all 5 methods. For each method, this resulted in a tensor consisting of 12 samples, 172 ligand-receptor pairs, and 6 sender- and receiver- cells.

Evaluating the effect of gene expression preprocessing and batch-effect correction on Tensor-cell2cell

To evaluate how gene expression preprocessing and batch-effect correction impact the results of Tensor-cell2cell, we assessed the similarity of tensor decomposition on the BALF dataset. To compute the communication scores for building the tensors (Figure 1a), we used different gene expression values, including the raw UMI counts, the preprocessed values with log(CPM+1) and the fraction of non-zero cells[24], and the batch-corrected values with ComBat[25] and Scanorama[26]. Except by the fraction of non-zero cells, which already aggregated single-cells into cell-types, other values were aggregated into the cell-type level by computing their average value for each gene across single cells with the same cell-type label. As the communication score, we used the expression mean of the interacting partners in each LR pair. Thus, we built 4D-communication tensors as mentioned for the BALF data in the Methods subsection ***Building the context-aware communication tensor***. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tensor decomposition resulting with the fraction of non-zero cells in this case corresponds to the same in Figure 4.

Measuring the similarity between distinct tensor decomposition runs

To assess decomposition consistency between different scoring methods or preprocessing pipelines, we employed the CorrIndex[22]. The CorrIndex is a permutation- and scaling-invariant distance metric that enables consistent comparison of decompositions between tensors containing the same elements, without need to align the factors obtained in each case (separate tensor decompositions can output similar factors but in different order). The CorrIndex value lies between 0 and 1, with a higher score indicating more dissimilar decomposition outputs. To score tensor decompositions, the output factor matrices must first be vertically stacked. We implemented a modification that instead assesses each tensor dimension separately (see Supplementary Note for more details). While taking the minimal score between all dimensions tends to be more stringent, it disregards the combinatorial effects of all dimensions together. These combinatorial effects are important because they better reflect the goal of tensor decomposition and because similarity in those dimensions that are not the minimal one may be artificially inflated. To facilitate the use of the CorrIndex and its modified version, we wrote a Python implementation that is available on the Tensorly package[72].

Downstream analyses using the loadings from the tensor decomposition

We incorporate several downstream analyses of Tensor-cell2cell's decomposition outputs to further elucidate the underlying cell- and molecular- mediators of cell-cell communication. Each of these analyses are associated with a specific tensor dimension, and thus, a specific biological resolution. This includes 1) statistical, correlative, and clustering analyses to understand context associations for each factor, 2) gene set enrichment analysis of ligand-receptor loadings to identify granular signaling pathways associated with factors, 3) the generation of factor-specific cell-cell communication networks to represent the overall communication state of cells in that factor.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We can understand the context associations for a factor by comparing the loadings of samples associated with distinct contexts. For statistical significance, we conduct an independent t-test pairwise between each context group associated with the samples and use Bonferonni's correction to account for multiple comparisons. We use this for both the COVID-19 BALF dataset (Supplementary Figures S7 and S8) and the ASD dataset (Figure 5c). We also conduct correlative analyses – assuming ordinal contexts (i.e., healthy control < moderate COVID-19 < severe COVID-19), we take the Spearman correlation between the sample loadings and sample severity (Figure 4c). Finally, we also hierarchically cluster the samples using their loadings across all factors (Figure 5d). For this purpose, we use the normalized loadings resulting from the tensor decomposition, and standardize them across all factors. Then, we apply an agglomerative hierarchical clustering by using Ward's method and the Euclidean distance as a metric. Note that this type of clustering analysis can be applied to the other tensor dimensions.

We can use the LR-pair loadings of a factor to identify the signaling pathways associated with it, by using the Gene Set Enrichment Analysis[56] (GSEA). Before running the analysis, pathways of interest have to be assigned to a list of associated LR pairs. We do that by considering the KEGG gene sets available at http://www.gsea-msigdb.org/. We annotate a LR pair available in CellChat with the gene sets that contain all genes participating in that LR interaction. Then, by filtering LR pathway sets to those containing at least 15 LR pairs, we end up with 22 LR pathway sets. To run GSEA, we rank the LR pairs in each factor by their loadings, and use the PreRanked GSEA function in the package gseapy, by including the 22 LR pathway sets as input. As parameters of the "gseapy.prerank" function, we consider 999 permutations, gene sets (LR pathway sets here) with at least 15 elements, and a score weight of 1 for computing the enrichment scores[56].

Finally, we generate factor-specific cell-cell communication networks. To do so, for a factor *r*, we take the outer product between the sender-cell loadings vector, $\boldsymbol{s}^r$, and the receiver-cell loadings vector, $\boldsymbol{t}^r$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conceptually, this outer product represents an adjacency matrix of a factor-specific cell-cell communication network, where each value is an edge weight representing the overall communication between a pair of sender-receiver cells (Supplementary Figure S9). We can further use this network to understand the communication distribution inequality between sender- and receiver- cells. We compute a Gini coefficient[73] ranging between 0 and 1 on the distribution of edge weights in the adjacency matrix (Figure 4c). A value of 1 represents maximal inequality of overall communication between cell pairs (i.e. one cell pair has a high overall communication value while the others have a value of 0) and 0 indicates minimal inequality (i.e. all cell pairs have the same overall communication values). More generally, the outer product between any two tensor dimension loadings for a given factor conceptually represents the joint distribution of the elements in those two dimensions and can be informative of how the specific elements are related.

Benchmarking of computational efficiency of tools

We measured the running time and memory demanded by Tensor-cell2cell and CellChat to analyze the COVID-19 dataset containing PBMC samples. Each tool was evaluated in two scenarios: either using each sample individually, or by first combining samples by severity (control, mild/moderate, and severe/critical) by aggregating the expression matrices. The latter was intended to favor CellChat by diminishing the number of pairwise comparisons to always be between three contexts; thus, increases in running time or memory demand in this case are not due to an exponentiation of comparisons (*n* samples choose 2). CellChat was run by following the procedures outlined in the "Comparison_analysis_of_multiple_datasets" vignette (https://github.com/sqjin/CellChat/tree/master/tutorial). Briefly, signaling pathway communication probabilities were first individually calculated for each sample or context. Next, pairwise comparisons between each sample or context were obtained by computing either a "functional" or a "structural" similarity. The functional approach computes a Jaccard index to compare the signaling pathways that

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

642 are active in two cellular communication networks, while the structural approach computes a network
643 dissimilarity[74] to compare the topology of two signaling networks (see REF[10] for further details). Finally,
644 CellChat performs a manifold learning approach on sample similarities and returns UMAP embeddings
645 for each signaling pathway in each different context (e.g. if CellChat evaluates 10 signaling pathways in
646 3 different contexts, it will return embeddings for 30 points) which can be used to rank the similarity of
647 shared signaling pathways between contexts in a pairwise manner.

648 The analyses of computational efficiency were run on a compute cluster of 2.8GHz x2 Intel(R) Xeon(R)
649 Gold 6242 CPUs with 1.5 TB of RAM (Micron 72ASS8G72LZ-2G6D2) across 32 cores. Each timing
650 task was limited to 128 GB of RAM on one isolated core and one thread independently where no other
651 processes were being performed. To limit channel delay, data was stored on the node where the job
652 was performed, where the within socket latency and bandwidth are 78.9 ns and 46,102 MB/s
653 respectively. For all timing jobs, the same ligand-receptor pairs and cell types were used. Furthermore,
654 to make the timing comparable, all samples in the dataset were subsampled to have 2,000 single cells.
655 In the case of Tensor-cell2cell, the analysis was also repeated by using a GPU, which corresponded to
656 a Nvidia Tesla V100.

657 ***Training and evaluation of a classification model***

658 A Random Forest[75] (RF) model was trained to predict disease status based on both COVID-19 status
659 (healthy-control vs. patient with COVID-19) and severity (healthy-control, moderate symptoms, and
660 severe symptoms). The RF model was trained using a Stratified K-Folds cross-validation (CV) with 3-
661 Fold CV splits. On each CV split a RF model with 500 estimators was trained and RF probability-
662 predictions were compared to the test set using the Receiver Operating Characteristic (ROC). The
663 mean and standard deviation from the mean were calculated for the area under the Area Under the
664 Curve (AUC) across the CV splits. This classification was performed on the context loadings of Tensor-
665 cell2cell, and the two UMAP dimensions of the structural and functional joint manifold learning of

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CellChat, for both the BALF and PBMC COVID-19 datasets. All classification was performed through Scikit-learn (v. 0.23.2)[76].

Code and data availability

Tensor-cell2cell is implemented in our cell2cell suite[65], which is available in a GitHub repository (https://github.com/earmingol/cell2cell). All the code and input data used for the analyses are available online in a Code Ocean capsule for reproducible runs (https://doi.org/10.24433/CO.0051950.v2). While the code for benchmarking the computational efficiency in a local computer is available in a GitHub repository (https://github.com/LewisLabUCSD/CCC-Benchmark).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Hwang, S., Kim, S., Shin, H. & Lee, D. Context-dependent transcriptional regulations between signal transduction pathways. *BMC Bioinformatics* **12**, 19 (2011).
2. Shakiba, N., Jones, R. D., Weiss, R. & Del Vecchio, D. Context-aware synthetic biology by controller design: Engineering the mammalian cell. *Cell Syst* **12**, 561–592 (2021).
3. Rachlin, J., Cohen, D. D., Cantor, C. & Kasif, S. Biological context networks: a mosaic view of the interactome. *Mol. Syst. Biol.* **2**, 66 (2006).
4. Schubert, M. *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20 (2018).
5. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* 1–18 (2020).
6. Griffiths, J. I. *et al.* Circulating immune cell phenotype dynamics reflect the strength of tumor–immune cell interactions in patients during immunotherapy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 16072–16082 (2020).
7. Omberg, L., Golub, G. H. & Alter, O. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18371–18376 (2007).
8. Cillo, A. R. *et al.* Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. *Immunity* vol. 52 183–199.e9 (2020).
9. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 1–11 (2020).
10. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
11. Raredon, M. S. B. *et al.* Computation and visualization of cell-cell signaling topologies in single-cell systems data using Connectome. *Sci. Rep.* **12**, 4187 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Williams, A. H. *et al.* Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron* **98**, 1099–1115.e8 (2018).
13. Stein-O'Brien, G. L. *et al.* Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet.* **34**, 790–805 (2018).
14. Sun, S., Zhu, J., Ma, Y. & Zhou, X. Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis. *Genome Biol.* **20**, 269 (2019).
15. Martino, C. *et al.* Context-aware dimensionality reduction deconvolutes gut microbial community dynamics. *Nat. Biotechnol.* **39**, 165–168 (2021).
16. Anandkumar, A., Jain, P., Shi, Y. & Niranjan, U. N. Tensor vs. Matrix Methods: Robust Tensor Decomposition under Block Sparse Perturbations. in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* (eds. Gretton, A. & Robert, C. C.) vol. 51 268–276 (PMLR, 2016).
17. Rabanser, S., Shchur, O. & Günnemann, S. Introduction to Tensor Decompositions and their Applications in Machine Learning. *arXiv [stat.ML]* (2017).
18. Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0602-4.
19. Friedlander, M. P. & Hatz, K. Computing non-negative tensor factorizations. *Optim. Methods Softw.* **23**, 631–647 (2008).
20. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* (2020) doi:10.1038/s41596-020-0292-x.
21. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
22. Sobhani, E., Comon, P., Jutten, C. & Babaie-Zadeh, M. CorrIndex: A permutation invariant

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performance index. *Signal Processing* **195**, 108457 (2022).

23. Dimitrov, D. *et al.* Comparison of Resources and Methods to infer Cell-Cell Communication from Single-cell RNA Data. *bioRxiv* 2021.05.21.445160 (2021) doi:10.1101/2021.05.21.445160.
24. Booeshaghi, A. S. & Pachter, L. Normalization of single-cell RNA-seq counts by log(x + 1)* or log(1 + x). *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab085.
25. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
26. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
27. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
28. Noël, F. *et al.* Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089 (2021).
29. Wang, Y. *et al.* iTALK: an R Package to Characterize and Illustrate Intercellular Communication. *Cancer Biology* (2019) doi:10.1101/507871.
30. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* (2019) doi:10.1038/s41592-019-0667-5.
31. Lagger, C. *et al.* scAgeCom: a murine atlas of age-related changes in intercellular communication inferred with the package scDiffCom. *bioRxiv* 2021.08.13.456238 (2021) doi:10.1101/2021.08.13.456238.
32. Tsuyuzaki, K., Ishii, M. & Nikaido, I. Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. *bioRxiv* 566182 (2019) doi:10.1101/566182.
33. Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
34. Schmitt, T. L., Steiner, E., Klingler, P., Lassmann, H. & Grubeck-Loebenstein, B. Thyroid epithelial

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cells produce large amounts of the Alzheimer beta-amyloid precursor protein (APP) and generate potentially amyloidogenic APP fragments. *J. Clin. Endocrinol. Metab.* **80**, 3513–3519 (1995).

35. Puig, K. L., Manocha, G. D. & Combs, C. K. Amyloid precursor protein mediated changes in intestinal epithelial phenotype in vitro. *PLoS One* **10**, e0119534 (2015).
36. Zemans, R. L., Colgan, S. P. & Downey, G. P. Transepithelial migration of neutrophils: mechanisms and implications for acute lung injury. *Am. J. Respir. Cell Mol. Biol.* **40**, 519–535 (2009).
37. Schenkel, A. R., Mamdouh, Z., Chen, X., Liebman, R. M. & Muller, W. A. CD99 plays a major role in the migration of monocytes through endothelial junctions. *Nat. Immunol.* **3**, 143–150 (2002).
38. Pasello, M., Manara, M. C. & Scotlandi, K. CD99 at the crossroads of physiology and pathology. *J. Cell Commun. Signal.* **12**, 55–68 (2018).
39. Sanino, G., Bosco, M. & Terrazzano, G. Physiology of Midkine and Its Potential Pathophysiological Role in COVID-19. *Front. Physiol.* **11**, 616552 (2020).
40. Farr, L., Ghosh, S. & Moonah, S. Role of MIF Cytokine/CD74 Receptor Pathway in Protecting Against Injury and Promoting Repair. *Front. Immunol.* **11**, 1273 (2020).
41. Weckbach, L. T., Muramatsu, T. & Walzog, B. Midkine in inflammation. *ScientificWorldJournal* **11**, 2491–2505 (2011).
42. Xia, J. *et al.* Semaphorin-Plexin Signaling Controls Mitotic Spindle Orientation during Epithelial Morphogenesis and Repair. *Dev. Cell* **33**, 299–313 (2015).
43. Nikaido, T. *et al.* Serum Syndecan-4 as a Possible Biomarker in Patients With Acute Pneumonia. *J. Infect. Dis.* **212**, 1500–1508 (2015).
44. Azari, B. M. *et al.* Transcription and translation of human F11R gene are required for an initial step of atherogenesis induced by inflammatory cytokines. *J. Transl. Med.* **9**, 98 (2011).
45. Clark, I. C. *et al.* Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science* **372**, (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

775 46. Zhang, F. *et al.* IFN- γ and TNF- α drive a CXCL10 + CCL2 + macrophage phenotype expanded in
776 severe COVID-19 and other diseases with tissue inflammation. *bioRxiv* (2020)
777 doi:10.1101/2020.08.05.238360.
778 47. Kohyama, M. *et al.* Monocyte infiltration into obese and fibrilized tissues is regulated by PILRα.
779 *Eur. J. Immunol.* **46**, 1214–1223 (2016).
780 48. Saheb Sharif-Askari, N. *et al.* Enhanced expression of immune checkpoint receptors during SARS-
781 CoV-2 viral infection. *Mol Ther Methods Clin Dev* **20**, 109–121 (2021).
782 49. Martinez, F. O., Combes, T. W., Orsenigo, F. & Gordon, S. Monocyte activation in systemic Covid-
783 19 infection: Assay and rationale. *EBioMedicine* **59**, 102964 (2020).
784 50. Ocaña-Guzman, R., Torre-Bouscoulet, L. & Sada-Ovalle, I. TIM-3 Regulates Distinct Functions in
785 Macrophages. *Front. Immunol.* **7**, 229 (2016).
786 51. Grant, R. A. *et al.* Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia.
787 *Nature* **590**, 635–641 (2021).
788 52. Matsuyama, T., Kubli, S. P., Yoshinaga, S. K., Pfeffer, K. & Mak, T. W. An aberrant STAT pathway
789 is central to COVID-19. *Cell Death Differ.* **27**, 3209–3225 (2020).
790 53. Florez-Sampedro, L., Soto-Gamez, A., Poelarends, G. J. & Melgert, B. N. The role of MIF in
791 chronic lung diseases: looking beyond inflammation. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **318**,
792 L1183–L1197 (2020).
793 54. de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of
794 autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).
795 55. Velmeshev, D. *et al.* Single-cell genomics identifies cell type-specific molecular changes in autism.
796 *Science* **364**, 685–689 (2019).
797 56. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting
798 genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
799 57. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

28, 27–30 (2000).

58. Astorkia, M., Lachman, H. M. & Zheng, D. Characterization of Cell-cell Communication in Autistic Brains with Single Cell Transcriptomes. *bioRxiv* 2021.10.15.464577 (2021) doi:10.1101/2021.10.15.464577.
59. Avraham, R. & Yarden, Y. Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nat. Rev. Mol. Cell Biol.* **12**, 104–117 (2011).
60. Almet, A. A., Cang, Z., Jin, S. & Nie, Q. The landscape of cell-cell communication through single-cell transcriptomics. *Current Opinion in Systems Biology* (2021) doi:10.1016/j.coisb.2021.03.007.
61. Raredon, M. S. B. *et al.* Connectome: computation and visualization of cell-cell signaling topologies in single-cell systems data. *bioRxiv* 2021.01.21.427529 (2021) doi:10.1101/2021.01.21.427529.
62. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
63. Abbasy, S. *et al.* Neuregulin1 types mRNA level changes in autism spectrum disorder, and is associated with deficit in executive functions. *EBioMedicine* **37**, 483–488 (2018).
64. Gazestani, V. H. *et al.* A perturbed gene network containing PI3K-AKT, RAS-ERK and WNT-β-catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nat. Neurosci.* **22**, 1624–1634 (2019).
65. Armingol, E. *et al.* Inferring a spatial code of cell-cell interactions across a whole animal body. *bioRxiv* 2020.11.22.392217 (2020) doi:10.1101/2020.11.22.392217.
66. Wang, S., Karikomi, M., MacLean, A. L. & Nie, Q. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* **47**, e66 (2019).
67. Ren, X. *et al.* COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 1895–1913.e19 (2021).
68. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

69. Carroll, J. D. & Chang, J.-J. Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika* **35**, 283–319 (1970).
70. Harshman, R. A. & Others. Foundations of the PARAFAC procedure: Models and conditions for an' explanatory' multimodal factor analysis. (1970).
71. Anandkumar, A., Ge, R. & Janzamin, M. Guaranteed Non-Orthogonal Tensor Decomposition via Alternating Rank-1 Updates. *arXiv [cs.LG]* (2014).
72. Kossaifi, J., Panagakis, Y., Anandkumar, A. & Pantic, M. TensorLy: Tensor Learning in Python. *arXiv [cs.LG]* (2016).
73. Farris, F. A. The Gini Index and Measures of Inequality. *Am. Math. Mon.* **117**, 851–864 (2010).
74. Schieber, T. A. *et al.* Quantification of network structural dissimilarities. *Nat. Commun.* **8**, 13928 (2017).
75. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
76. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).

Acknowledgements

EA is supported by the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through its scholarship program DOCTORADO BECAS CHILE/2018 - 72190270 and by the Fulbright Commission Chile. HMB is supported by NIMH T32GM008806. APL is supported by the InnovaUNAM of the National Autonomous University of Mexico (UNAM) and Alianza UCMX of the University of California. This work was further supported by NIGMS (R35 GM119850) and the Novo Nordisk Foundation (NNF20SA0066621) to NEL. The authors also thank Daniel McDonald for providing useful guidance about the timing analysis of the tools, the Code Ocean team for providing extra computational time for developing the capsule associated with this work, Aaron Meyer for giving practical insights about tensor decomposition methods, and Daniel Dimitrov for providing helpful guidance about running LIANA.

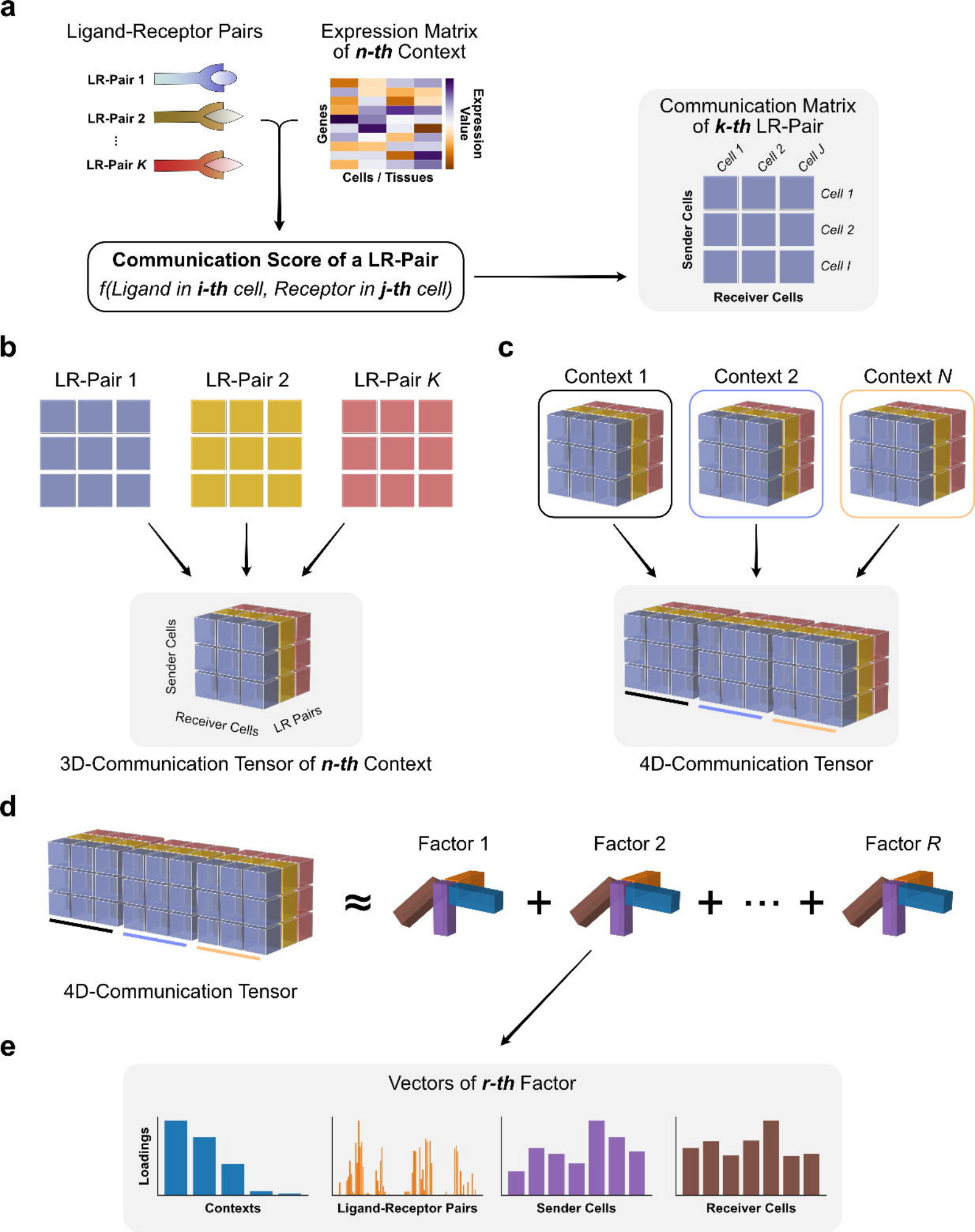
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted March 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

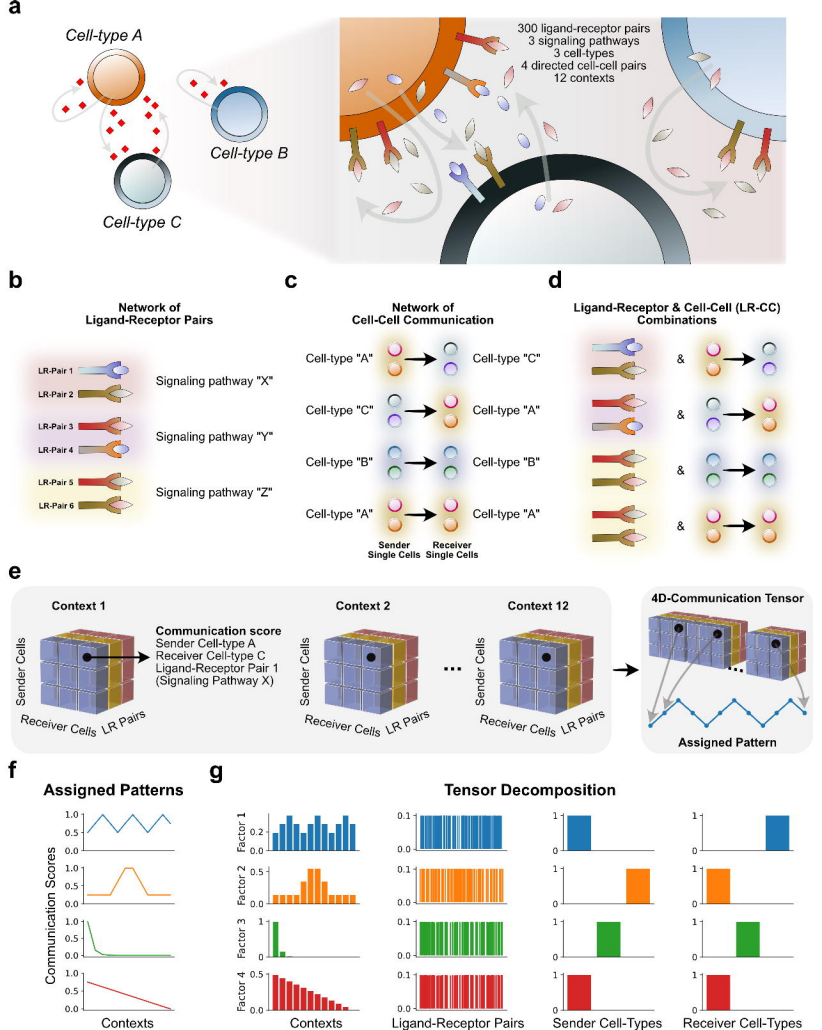
Author contributions

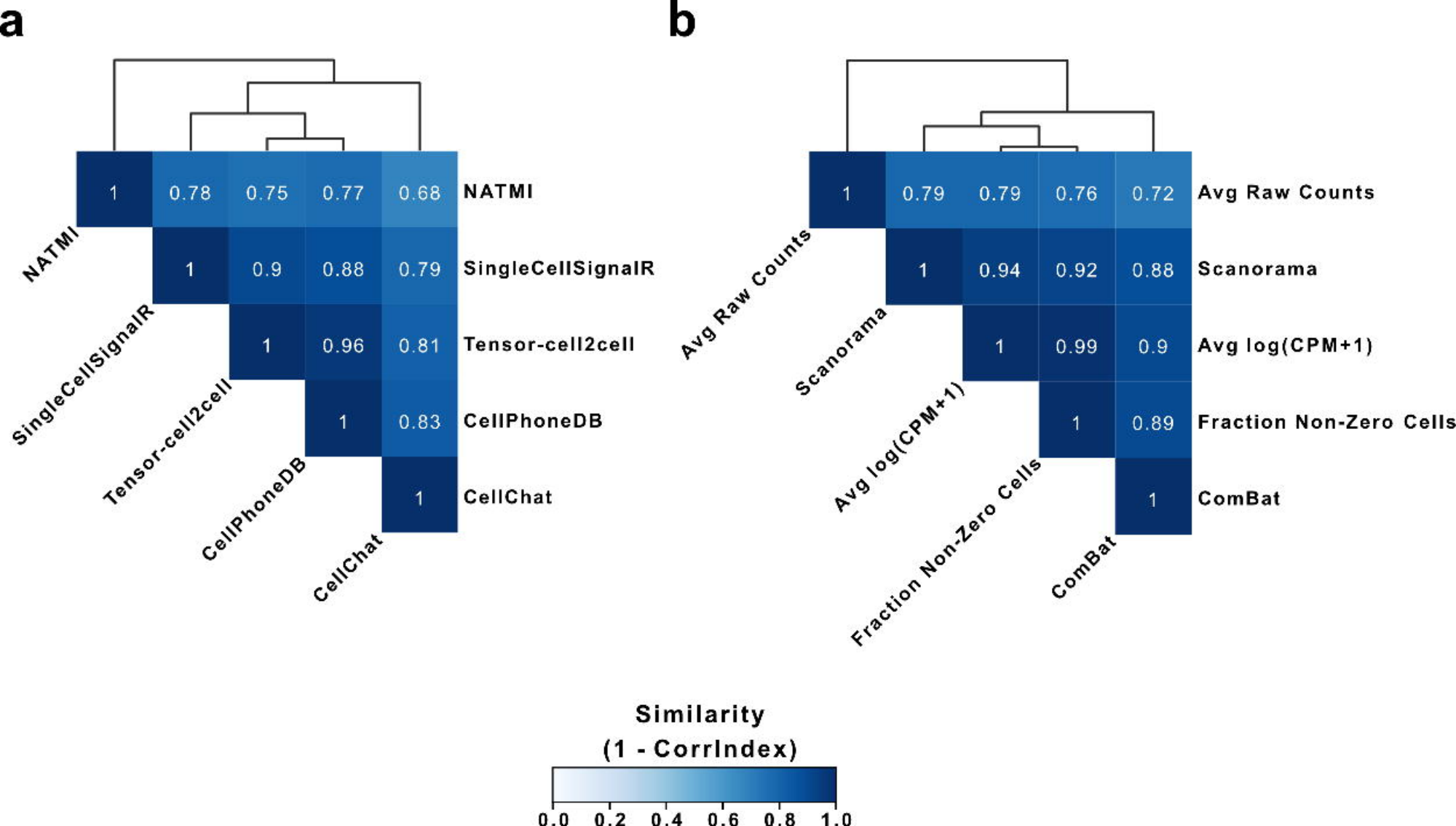
EA, HB and NEL conceived the work. CM contributed important insights for creating Tensor-cell2cell. EA implemented Tensor-cell2cell and performed the analyses on the datasets of COVID-19 and ASD. HB designed and created the simulated 4D-communication tensor and performed the analyses on the simulated data. EA, HB and CM performed benchmarking and statistical analyses. CM trained classifiers and compared Tensor-cell2cell to CellChat. HB performed benchmarking analyses using different external CCC tools. EA performed benchmarking analyses using different pre-processing and batch-correction methods. EA and HB developed downstream analyses. APL helped to interpret the COVID-19 results and researched literature. CA helped to interpret the ASD study case and researched literature. RK contributed to the benchmarking analyses. EA and HB wrote the paper and all authors carefully reviewed, discussed and edited the paper.

Competing interests

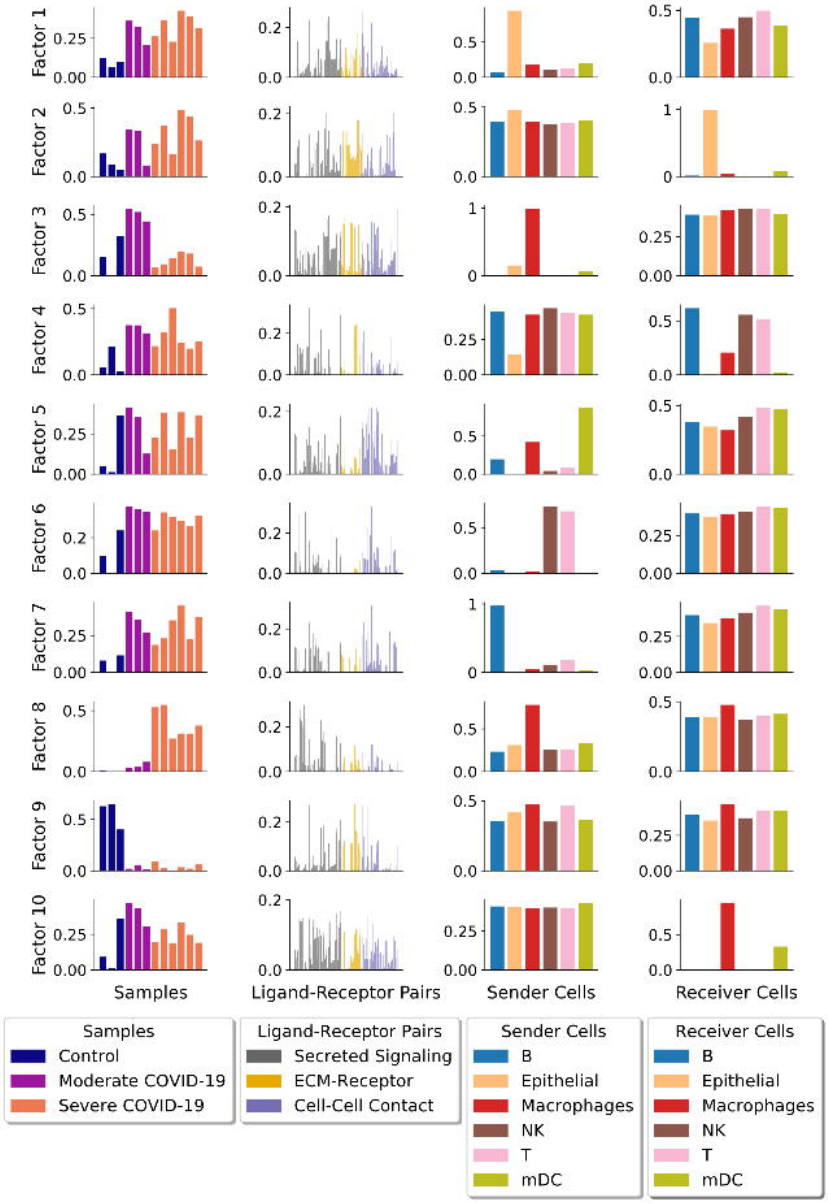
The authors declare no competing interests.


a
Ligand-Receptor Pairs
LR-Pair 1
LR-Pair 2
LR-Pair K
Expression Matrix of n-th Context
Genes
Expression Value
Cells / Tissues
Communication Score of a LR-Pair
f(Ligand in i-th cell, Receptor in j-th cell)
Communication Matrix of k-th LR-Pair
Cell 1
Cell 2
Cell J
Sender Cells
Cell 1
Cell 2
Cell I
Receiver Cells
b
LR-Pair 1
LR-Pair 2
LR-Pair K
Sender Cells
Receiver Cells
LR Pairs
3D-Communication Tensor of n-th Context
c
Context 1
Context 2
Context N
4D-Communication Tensor
d
4D-Communication Tensor
≈
Factor 1
Factor 2
Factor R
e
Vectors of r-th Factor
Loadings
Contexts
Ligand-Receptor Pairs
Sender Cells
Receiver Cells


a
Cell-type A
Cell-type B
Cell-type C
300 ligand-receptor pairs
3 signaling pathways
3 cell-types
4 directed cell-cell pairs
12 contexts
b
Network of
Ligand-Receptor Pairs
LR-Pair 1
LR-Pair 2
LR-Pair 3
LR-Pair 4
LR-Pair 5
LR-Pair 6
Signaling pathway "X"
Signaling pathway "Y"
Signaling pathway "Z"
c
Network of
Cell-Cell Communication
Cell-type "A"
Cell-type "C"
Cell-type "C"
Cell-type "A"
Cell-type "B"
Cell-type "B"
Cell-type "A"
Cell-type "A"
Sender
Single Cells
Receiver
Single Cells
d
Ligand-Receptor & Cell-Cell (LR-CC)
Combinations
&
&
&
&
e
Context 1
Sender Cells
Receiver Cells
LR Pairs
Communication score
Sender Cell-type A
Receiver Cell-type C
Ligand-Receptor Pair 1
(Signaling Pathway X)
Context 2
Context 12
4D-Communication Tensor
Assigned Pattern
f
Assigned Patterns
Communication Scores
Contexts
g
Tensor Decomposition
Factor 1
Factor 2
Factor 3
Factor 4
Contexts
Ligand-Receptor Pairs
Sender Cell-Types
Receiver Cell-Types


a
1
0.78
0.75
0.77
0.68
NATMI
1
0.9
0.88
0.79
SingleCellSignalR
1
0.96
0.81
Tensor-cell2cell
1
0.83
CellPhoneDB
1
CellChat
NATMI
SingleCellSignalR
Tensor-cell2cell
CellPhoneDB
CellChat
b
1
0.79
0.79
0.76
0.72
Avg Raw Counts
1
0.94
0.92
0.88
Scanorama
1
0.99
0.9
Avg log(CPM+1)
1
0.89
Fraction Non-Zero Cells
1
ComBat
Avg Raw Counts
Scanorama
Avg log(CPM+1)
Fraction Non-Zero Cells
ComBat
Similarity
(1 - CorrIndex)
0.0
0.2
0.4
0.6
0.8
1.0

a



b

Top-5 Ligand-Receptor Pairs

Factor 1		Factor 2	
APP - CD74	0.268	SEMA4D - PLXNB2	0.212
MDK - NCL	0.246	SEMA4A - PLXNB2	0.203
MIF - CD74 & CD44	0.231	MDK - SDC4	0.202
MDK - ITGA4 & ITGB1	0.227	COL9A2 - SDC4	0.186
CD99 - CD99	0.220	F11R - F11R	0.186

Factor 3		Factor 4	
SIGLEC1 - SPN	0.194	MIF - CD74 & CD44	0.321
RETN - CAP1	0.182	LGALS9 - CD44	0.289
MDK - NCL	0.177	COL9A2 - CD44	0.245
FN1 - ITGA4 & ITGB1	0.176	LAMB3 - CD44	0.242
FN1 - ITGA4 & ITGB7	0.172	LAMB2 - CD44	0.234

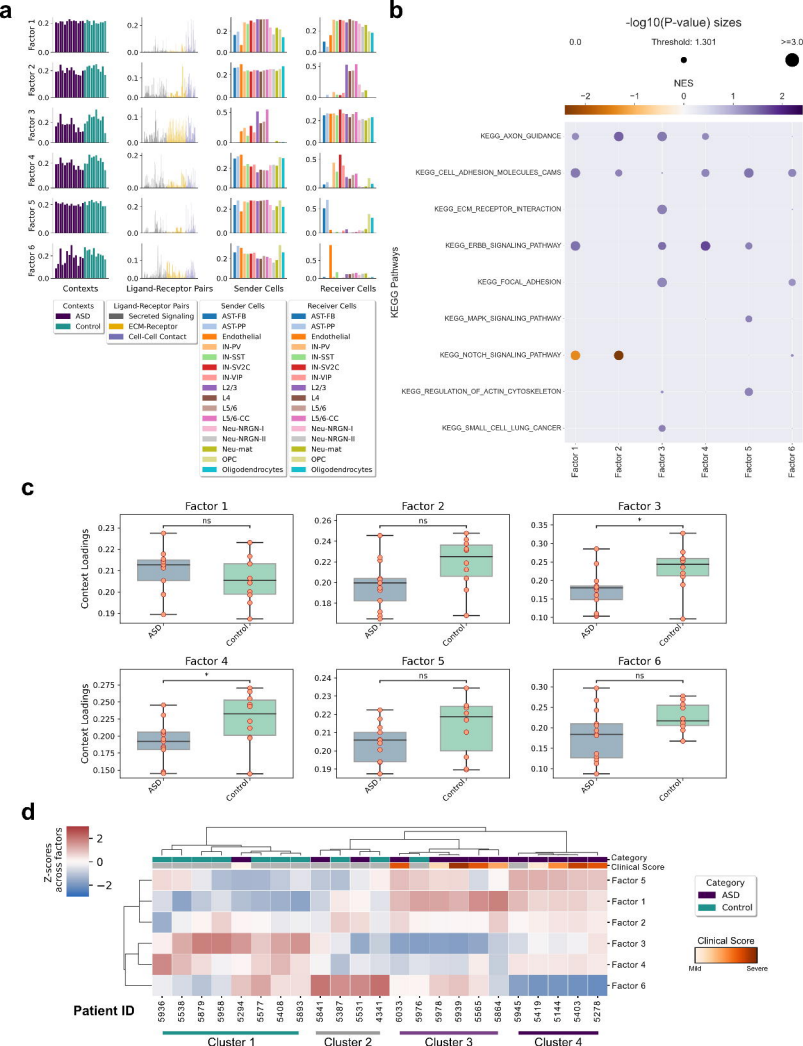
Factor 5		Factor 6	
CD99 - CD99	0.213	CD99 - CD99	0.333
ITGB2 - CD226	0.211	CCL5 - CCR5	0.305
CD86 - CTLA4	0.210	CCL5 - CCR1	0.293
ITGB2 - ICAM2	0.201	GZMA - F2R	0.241
ITGB2 - ICAM1	0.192	PTPRC - MRC1	0.222

Factor 7		Factor 8	
CD99 - CD99	0.307	CCL2 - CCR2	0.297
MIF - CD74 & CXCR4	0.241	CCL3 - CCR1	0.275
CD22 - PTPRC	0.231	CCL8 - CCR1	0.261
MIF - CD74 & CD44	0.229	CCL3 - CCR5	0.238
SELL - MADCAM1	0.191	CCL3L1 - CCR1	0.238

Factor 9		Factor 10	
FN1 - CD44	0.274	CD99 - PILRA	0.191
MIF - CD74 & CD44	0.269	LGALS9 - HAVCR2	0.170
APP - CD74	0.266	ANXA1 - FPR1	0.163
PTPRC - MRC1	0.263	MDK - LRP1	0.154
RETN - CAP1	0.243	PTPRC - MRC1	0.153

c

Factor	Spearman Coefficient	Gini Coefficient
Factor 1	0.72	0.50
Factor 2	0.61	0.76
Factor 3	-0.26	0.75
Factor 4	0.39	0.48
Factor 5	0.40	0.59
Factor 6	0.25	0.65
Factor 7	0.51	0.68
Factor 8	0.92	0.24
Factor 9	-0.51	0.09
Factor 10	-0.02	0.74


a
Factor 1
Factor 2
Factor 3
Factor 4
Factor 5
Factor 6
Contexts
Ligand-Receptor Pairs
Sender Cells
Receiver Cells
Contexts
ASD
Control
Ligand-Receptor Pairs
Secreted Signaling
ECM-Receptor
Cell-Cell Contact
Sender Cells
AST-FB
AST-PP
Endothelial
IN-PV
IN-SST
IN-SV2C
IN-VIP
L2/3
L4
L5/6
L5/6-CC
Neu-NRGN-I
Neu-NRGN-II
Neu-mat
OPC
Oligodendrocytes
Receiver Cells
AST-FB
AST-PP
Endothelial
IN-PV
IN-SST
IN-SV2C
IN-VIP
L2/3
L4
L5/6
L5/6-CC
Neu-NRGN-I
Neu-NRGN-II
Neu-mat
OPC
Oligodendrocytes
b
-log10(P-value) sizes
0.0
Threshold: 1.301
>=3.0
NES
−2
−1
0
1
2
KEGG Pathways
KEGG_AXON_GUIDANCE
KEGG_CELL_ADHESION_MOLECULES_CAMS
KEGG_ECM_RECEPTOR_INTERACTION
KEGG_ERBB_SIGNALING_PATHWAY
KEGG_FOCAL_ADHESION
KEGG_MAPK_SIGNALING_PATHWAY
KEGG_NOTCH_SIGNALING_PATHWAY
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON
KEGG_SMALL_CELL_LUNG_CANCER
Factor 1
Factor 2
Factor 3
Factor 4
Factor 5
Factor 6
c
Factor 1
Factor 2
Factor 3
Factor 4
Factor 5
Factor 6
Context Loadings
ns
*
ASD
Control
d
Z-scores across factors
2
0
−2
Category
Clinical Score
Factor 5
Factor 1
Factor 2
Factor 3
Factor 4
Factor 6
Category
ASD
Control
Clinical Score
Mild
Severe
Patient ID
5936
5538
5879
5958
5284
5577
5408
5893
5841
5387
5531
4341
6033
5976
5978
5939
5685
5684
5945
5419
5144
5403
5278
Cluster 1
Cluster 2
Cluster 3
Cluster 4