SCISPACE
formerly Typeset

Open access • Posted Content • DOI:10.1101/2021.09.20.461129

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

— Source link

Erick Armingol, Hratch M. Baghdassarian, Cameron Martino, Cameron Martino ...+3 more authors

Institutions: University of California, Berkeley, University of California, San Diego, National Autonomous University of Mexico

Published on: 23 Sep 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Context (language use)

Related papers:

- Inferring the spatial code of cell-cell interactions and communication across a whole animal body
- Inferring a spatial code of cell-cell interactions across a whole animal bodys
- Learning cell communication from spatial graphs of cells
- Machine learning of stem cell identities from single-cell expression data via regulatory network archetypes
- De novo identification of cell type hierarchy with application to compound marker detection

Share this paper:

View more about this paper here: https://typeset.io/papers/context-aware-deconvolution-of-cell-cell-communication-with-4pam0yyecl

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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], 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. 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. 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 September 23, 2021. 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 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. That is, “cellular contexts” refer to variation in genotype, biological state or condition that can shape the microenvironment of a cell and therefore its CCC. Under this idea, 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 into complex dynamics of CCC (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 the 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 usually involve repeating analysis for each context separately, disregarding informative variation across such factors as disease

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

severities, time points, subjects or cellular locations[7]. Additional analysis steps are required to compare and compile results from pairwise comparisons[8–11], reducing the statistical power and hindering efforts to link phenotypes to the interacting cells and their associated 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's superior performance resides in that 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 patterns of CCC. These

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

data-driven patterns reveal underlying communication changes given the simultaneous interaction between contexts, ligand-receptor pairs, and cells. We demonstrate that Tensor-cell2cell is 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. Next, while Tensor-cell2cell introduces the ability to simultaneously analyze any number of contexts, it can also be used for more simple analyses accessible to other existing tools, so we compare the computational efficiency and accuracy of existing tools to our method on these more basic analyses. Through this we show that Tensor-cell2cell is faster, demands less memory and can achieve better accuracy in separating context-specific information. We further apply it to study the differences of CCC associated with varying severities of COVID-19 across patients. Notably, our strategy recapitulates multiple biological features associated with COVID-19 severity, including, for example, stronger interactions between epithelial and immune cells in more severe cases[18]. Moreover, it suggests novel features distinguishing moderate from severe cases, especially in the macrophage communication with other cells in lungs, predominating signatures of M2-like and M1-like phenotypes, respectively. Thus, Tensor-cell2cell can be widely applied to any system that involves cell-cell communication that is expected to change due to context. Its easily interpretable output enables quick identification of key mediators of cell-cell communication in these contexts, both reproducing known results and identifying novel interactors.

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 inference of 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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-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 in 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).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

communication patterns when we added noise to the simulated tensor (Supplementary Figure S2 and Supplementary Notes).

Tensor-cell2cell is fast and accurate

Tensor-cell2cell is the first method for interpreting CCC across multiple conditions or contexts simultaneously, enabling a wealth of analyses and conclusions inaccessible to any other tools. Most available tools for inferring intercellular communication are designed to analyze one context at the time, while a few also allow pairwise comparisons between contexts (Table 1). However, pairwise comparisons will likely miss insights about changes of intercellular communication across multiple contexts.

To quantify computational efficiency and accuracy of Tensor-cell2cell, we compared our method to CellChat[10], the only tool that summarizes multiple pairwise comparisons in an automated manner (Table 1). While 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. Crucially, Tensor-cell2cell reports behaviors of communication modules across all contexts simultaneously and associates these modules with specific molecular and cellular mechanisms, information that CellChat is limited to reporting when comparing just two contexts. Thus, our comparison of tools is intended to assess running time and memory demand, and each tool's ability to identify signals that could separate contexts by measuring the accuracy of a context classifier trained from each tool output.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

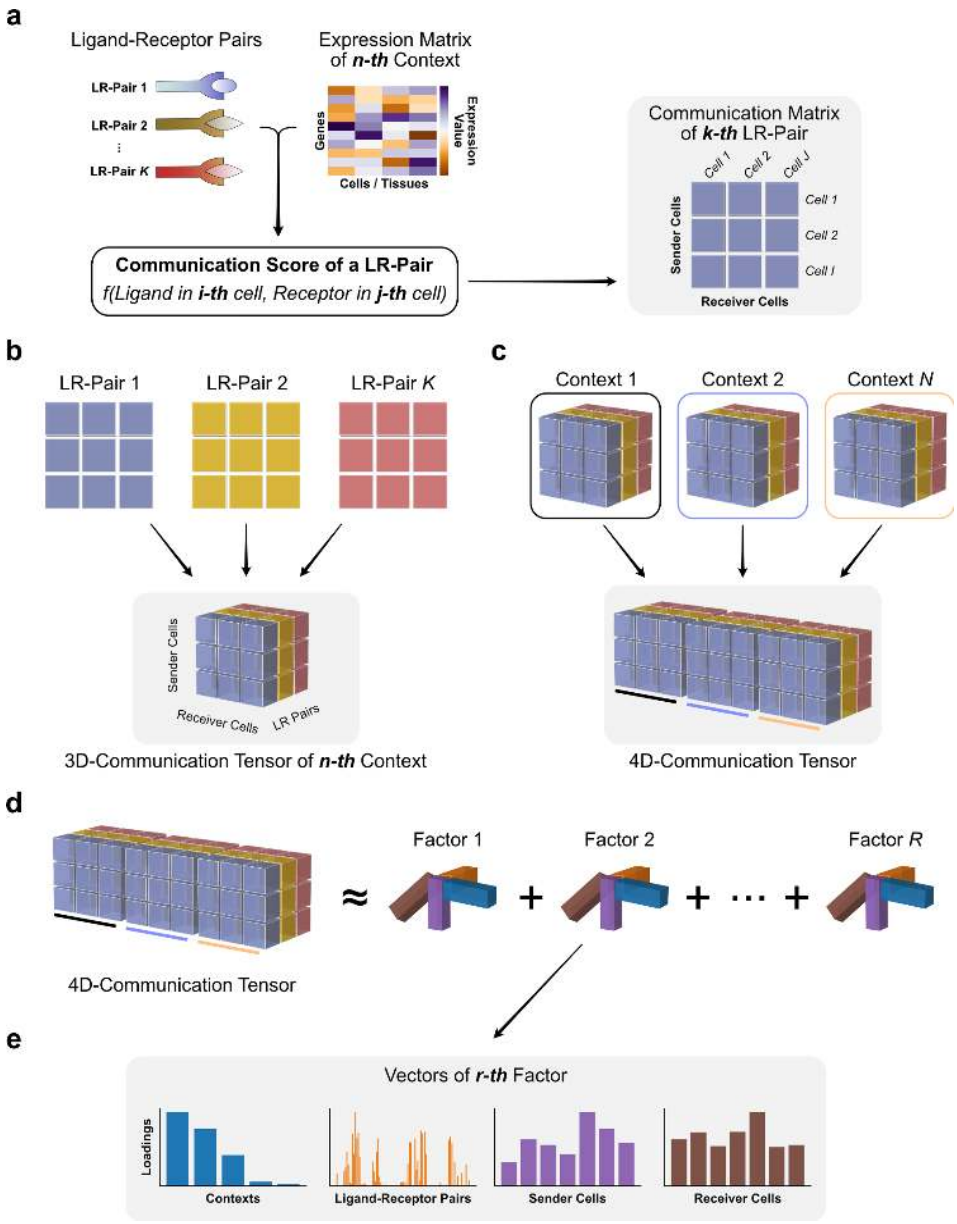


Figure 1. Tensor representation and factorization of cell-cell communication. In a given context (n-th context among N total contexts), cell-cell communication scores (see available scoring functions in REF[5]) are computed from the expression of the ligand and the receptor in a LR pair (k-th pair among 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 September 23, 2021. 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.

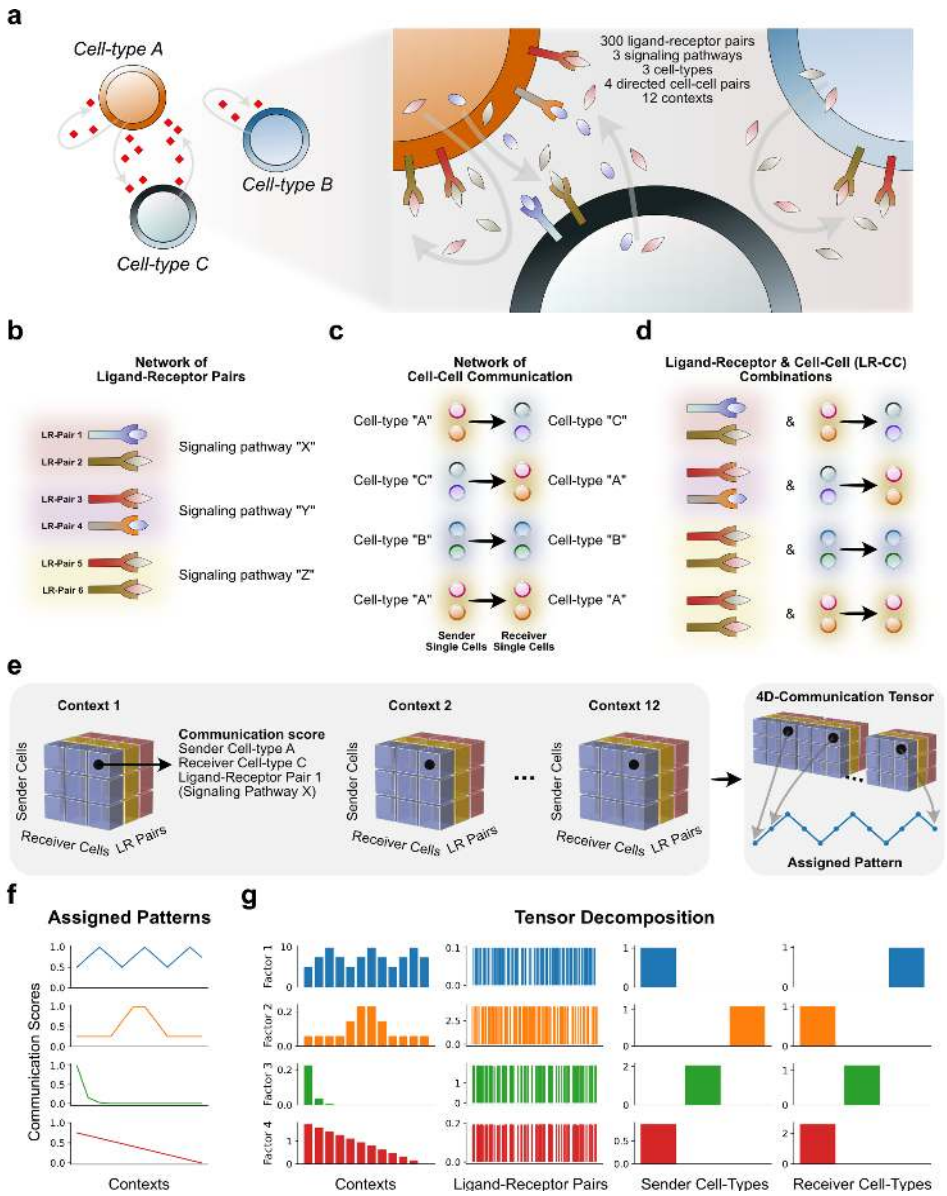


Figure 2. Tensor-cell2cell recovers simulated communication patterns. (**a**) Cell-cell communication scenario used for simulating patterns of communication across different contexts (here each a different time point). (**b**) Examples of specific ligand-receptor (LR) and (**c**) cell-cell pairs that participate in the simulated interactions. Individual LR pairs and cell pairs were categorized into groups of signalling 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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 and Expression Product	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	[21]
iTalk	Differential Combinations	Differential analysis between two contexts	2	No	Single Cell	R	[22]
NATMI	Expression Product and Normalized Expression Product	None	1	No	Bulk, Single Cell	Python	[9]
NicheNet	Personalized-Page Rank-based score	None	1	No	Bulk, Single Cell	R	[23]
scAgeCom	Geometric Mean	Differential analysis between two contexts	2	Yes	Single Cell	R	[24]
scTensor	Expression Product	None	1	No	Single Cell	R	[25]
SingleCellSignalR	Regularized Expression Product	None	1	No	Single Cell	R	[26]

[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 September 23, 2021. 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 ran Tensor-cell2cell and CellChat on a single-cell transcriptome atlas of peripheral blood mononuclear cells (PBMCs) from COVID-19 patients with varying severity[27] to measure the time and memory demands of each tool when performing the context-driven CCC analysis (Figures 3a-d). Considering the number of samples in this dataset, processing time of CellChat scales more rapidly with the high number of pairwise comparisons. To control this, we varied the number of samples and performed this benchmarking in two scenarios: 1) by considering every sample individually as a context, wherein one can obtain sample-specific signatures that may coincide with others of the same severity (Figures 3a-b), and 2) by considering every severity (control, mild/moderate and severe/critical) as contexts by aggregating cognate samples (Figures 3c-d), which keeps the number of pairwise comparisons constant at three comparisons but at the expense of losing sample-specific information.

Tensor-cell2cell performed better in all cases we tested (Figures 3a-d). The most favorable scenario led to an ~89-fold improvement in running time, which occured when 60 samples were analyzed as individual contexts and CellChat comparisons were run under the “structural” method (computes a network topology dissimilarity[10]**).** Cases when CellChat performed closer to Tensor-cell2cell were either when using the “functional” comparison (Figure 3a), which is based on a Jaccard similarity, or when samples were aggregated into major contexts (Figure 3c); however, the latter demanded substantial increases in memory usage (Figure 3d). Importantly, this improvement is obtained even though Tensor-cell2cell runs a brute-force elbow analysis (that is, by computing the error for every rank in a range of values). In this regard, this step can be either omitted (for example, when a desired number of factors is used; Supplementary Figure S3) or optimized (for example, when a binary search is used), multiplying the speed-up we reported here by ~10-fold and ~3-fold, respectively. Remarkably, the memory usage of Tensor-cell2cell never surpassed 16GB in any of the tested scenarios, even when using 60 samples as individual contexts (Figure 3b); meanwhile, CellChat surpassed 16GB when aggregating 12 samples (Figure 3d) or using 24 samples as individual contexts (Figure 3b). Overall,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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 is more efficient in both time and memory, indicating it can more readily be run on multiple contexts simultaneously in a personal computer or laptop. Moreover, Tensor-cell2cell can run on a GPU when available, which can substantially improve the computational time of the analysis (up to 19- and 790-fold faster than the “functional” and “structural” methods of CellChat, respectively, when analyzing 60 PBMC samples).

We next evaluated the accuracy of Tensor-cell2cell and CellChat in classifying individual samples after predicting context-driven CCC (Figures 3e-h). It is important to consider that the outputs coming from each tool are extremely different due to the scope of their analyses, so a direct comparison is not feasible. Hence, we instead used an intermediary approach that uses a classification model to evaluate how well each tool separates contexts given their outputs. In particular, we measured how well each tool separates samples by COVID-19 severity (Figures 3e-f) and disease state (Figures 3g-h). For this, we trained a classifier to predict severity (control, mild/moderate vs severe/critical) and a disease state (healthy vs COVID-19) in two different COVID-19 datasets, one containing PBMC samples[27] and the other bronchoalveolar lavage fluid (BALF) samples[28]. We next measured their accuracy with the area under the receiver operating characteristic curve (AUC). Tensor-cell2cell outperformed CellChat when classifying PBMC samples by severity (Figure 3e), and performed similarly when classifying samples by disease state (Figure 3g). Moreover, Tensor-cell2cell performed better than CellChat in all classification tasks associated with BALF samples (Figures 3f,h). Surprisingly, all methods performed better (highest AUC) when classifying BALF samples than when classifying PBMC samples, possibly due to a more evident severity-driven variation of the immune response in the infection site rather than in the periphery. Thus, these results show that Tensor-cell2cell can successfully find signatures of CCC that differentiate between contexts in a computationally efficient manner.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

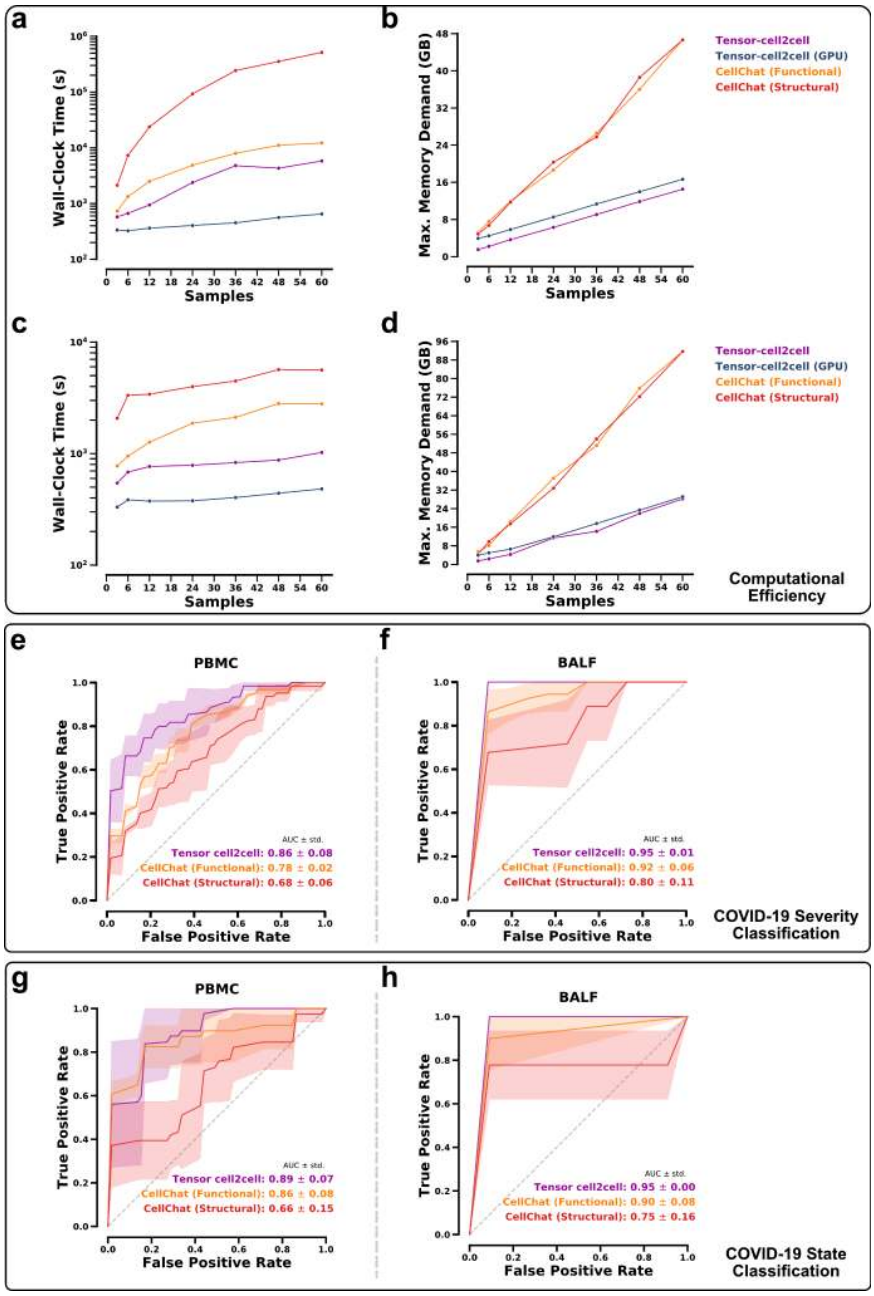


Figure 3. Benchmarking Tensor-cell2cell. (**a**) Running time of Tensor-cell2cell and CellChat for analyzing CCC when the contexts correspond to individual patient samples. (**b**) Memory usage of each method to perform analyses in (a). (**c**) Running time of Tensor-cell2cell and CellChat for analyzing CCC when the contexts correspond to COVID-19 severity (individual samples were aggregated by severity: control, mild/moderate and severe/critical COVID-19). (**d**) Memory usage of each method to perform the analyses in (c). In (a) and (c), Tensor-cell2cell was benchmarked when running with or without a GPU, a feature unavailable in CellChat. In addition, CellChat was benchmarked by using the two approaches it has for pairwise comparisons (functional and structural similarities, see Methods). (**e-h**) Receiver operating characteristic (ROC) curves of random forest models for classifying individual samples from the outputs of Tensor-cell2cell and the two CellChat approaches (functional and structural). These models predict specific severities of patients (control, mild/moderate or severe/critical) for (**e**) PBMC and (**f**) BALF samples, and disease state (healthy or COVID-19) for (**g**) PBMC and (**h**) BALF samples. For each classifier, the mean (solid line) ± standard deviation (transparent area) of the ROCs were computed from the 3-fold cross validations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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 successfully associates differences in intercellular communication with varying severities of COVID-19

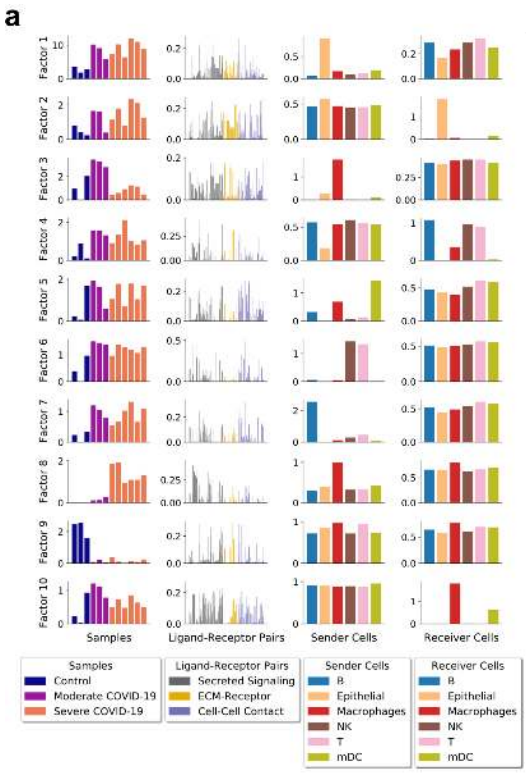
Great strides have been made to unravel molecular and cellular mechanisms associated with SARS-CoV-2 infection and COVID-19 pathogenesis. Thus, we tested our method on a single-cell dataset of BALF samples from COVID-19 patients[28], to see how many cell-cell and LR pair relationships in COVID-19, could 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 that distinguish features of the different disease stages (Supplementary Figure S4), in both cases consistent with the high performance that the classifier achieved for this dataset (Figures 3f,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 (Supplementary Figure S4). Consistently, a correlation between COVID-19 severity and the airway epithelium-immune cell interactions has been reported elsewhere[18], coinciding with our observation (Figure 4c). Specifically, epithelial cells are highlighted as the main sender cells in factor 1 (Figure 4a), involving top ranked signals such as APP, MDK, MIF and CD99 (Figure 4b). These molecules were reported to be produced by epithelial cells[29–35] and participate in immune cell recruiting[31–33], in response to mechanical stress in lungs[34] and regeneration of the alveolar barrier during viral infection[35]. Importantly, MDK can exacerbate leukocyte recruitment to sites of inflammation[36], consistent with the immune response that Tensor-cell2cell deconvolved in severe COVID-19 patients[18,37–39]. 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 determined important for tissue repair and inflammation during

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

lung injury previously[40–42]. Remarkably, a new technology for tracing experimentally CCC revealed that SEMA4D-PLXNB2 interaction promotes inflammation in a diseased central nervous system[43]; a role consistent with the pattern in factor 2, which captured this interaction as the top-ranked LR pair correlating with COVID-19 severity (Figures 4a,b).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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	0.266	**SEMA4D - PLXNB2**	0.261
MDK - NCL	0.244	**SEMA4A - PLXNB2**	0.249
MIF - CD74 & CD44	0.229	**MDK - SDC4**	0.248
MDK - ITGA4 & ITGB1	0.225	**COL9A2 - SDC4**	0.229
CD99 - CD99	0.218	F11R - F11R	0.228

Factor 3		Factor 4	
SIGLEC1 - SPN	0.191	**MIF - CD74 & CD44**	0.413
RETN - CAP1	0.179	**LGALS9 - CD44**	0.371
MDK - NCL	0.174	**COL9A2 - CD44**	0.315
FN1 - ITGA4 & ITGB1	0.173	**LAMB3 - CD44**	0.311
FN1 - ITGA4 & ITGB7	0.170	**LAMB2 - CD44**	0.301

Factor 5		Factor 6	
CD99 - CD99	0.271	**CD99 - CD99**	0.491
ITGB2 - CD226	0.268	**CCL5 - CCR5**	0.450
CD86 - CTLA4	0.267	**CCL5 - CCR1**	0.432
ITGB2 - ICAM2	0.255	**GZMA - F2R**	0.355
ITGB2 - ICAM1	0.244	**PTPRC - MRC1**	0.327

Factor 7		Factor 8	
CD99 - CD99	0.321	**CCL2 - CCR2**	0.442
MIF - CD74 & CXCR4	0.253	**CCL3 - CCR1**	0.410
CD22 - PTPRC	0.243	**CCL8 - CCR1**	0.388
MIF - CD74 & CD44	0.240	**CCL3 - CCR5**	0.354
SELL - MADCAM1	0.200	**CCL3L1 - CCR1**	0.354

Factor 9		Factor 10	
FN1 - CD44	0.296	**CD99 - PILRA**	0.257
MIF - CD74 & CD44	0.291	**LGALS9 - HAVCR2**	0.230
APP - CD74	0.288	**ANXA1 - FPR1**	0.220
PTPRC - MRC1	0.284	**MDK - LRP1**	0.208
RETN - CAP1	0.262	**PTPRC - MRC1**	0.207

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). 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 joint distribution of cell pair loadings (to measure the imbalance of cell-cell communication). Important values are highlighted in red (higher absolute Spearman coefficients represent stronger correlations; while smaller Gini coefficients represent distributions with similar overall communication values).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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 S4). 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 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[18,44]. 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 S5) 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 in previous independent work[45–47], also associating them with mild but not severe myalgic encephalomyelitis[48], and with the immune response to SARS-CoV-2[45,46,49–51]. Remarkably, the source article of the BALF dataset[28] reported that M2-like (anti inflammatory) macrophages were present with higher frequency than M1-like (pro-inflammatory) macrophages in healthy and moderate COVID-19 patients, while M1-like macrophages were more frequent in severe COVID-19 patients[28,52], supporting the results of Tensor-cell2cell.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

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 communication scores between sender and receiver cells are evenly distributed across cell pairs. Smaller Gini coefficients show more even distributions, i.e., more equally weighted communication scores across sender and receiver cell pairs (see ***Methods***). This indicates that the intercellular communication represented by factor 9 is ubiquitous across cell types. 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 S4). Interestingly, the interaction between MIF and the complex CD74-CD44 is a top-ranked LR pair in factor 9 (Figure 4b), which is consistent with ubiquitous expression of MIF across tissues and its protective role in normal conditions[35,53]. 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 can be important for therapeutic applications[54]. For example, our results suggest that COVID-19 severity might be reduced by delivering a drug to control the ratio of M1/M2 macrophages in lungs[55].

Discussion

Here we present Tensor-cell2cell, a novel computational approach that finds modules or patterns of cell-cell communication and their changes across contexts (e.g., across subjects with different disease severity, multiple time points, different tissues, etc.). We show the power of Tensor-cell2cell with both simulated and real datasets to extract complex patterns of intercellular communication in an unsupervised fashion. Crucially, our approach can rank LR pairs based on their importance in each module of communication, and it connects these signals with variations in phenotype (e.g., increasing use of LR pairs with COVID-19 severity, as shown in factors 1,2 and 8 in Figure 4a). These features distinguish Tensor-cell2cell as a novel tool in contrast to state-of-the-art tools that are either unaware of

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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 context driving CCC[5,20,23,56] or require analysis of each context separately and perform pairwise comparisons in posterior steps[10,11]. Thus, Tensor-cell2cell is the first tool that facilitates the study of CCC in a context-aware manner, simultaneously integrating any number of samples, and helping generate testable hypotheses about the biological role of molecules across multiple conditions.

Tensor-cell2cell is fast for analyzing multiple contexts simultaneously, providing up to 790-fold improvement in running time and reduced memory requirements with respect to analyses involving multiple pairwise comparisons (Figures 3a-d). It is also more accurate, resulting in a marked 10-20% higher classification accuracy of subjects with COVID-19 when compared to CellChat (Figures 3e-h), the only available tool that summarizes multiple pairwise comparisons of contexts. However, it is important to consider that benchmarking tools for predicting CCC is challenging due to the lack of a ground truth[5], and it is hard to compare and evaluate tools because of the diversity of scoring approaches[57]. Indeed, the outputs and details offered by CellChat and Tensor-cell2cell differ. CellChat reports context-associated UMAP embeddings of signaling pathways, while Tensor-cell2cell outputs TCA embeddings for contexts, ligand-receptor pairs, and interconnected sender and receiver cells. By training classifiers that accept these differing outputs, in most cases, Tensor-cell2cell greatly outperformed CellChat (Figures 3e,f,h). While in a few scenarios we observed qualitatively comparable performance (Figure 3g, Tensor-cell2cell and the functional method of CellChat), Tensor-cell2cell always performed better quantitatively. Although context classification is a useful approach for comparison, this strategy cannot evaluate how well these methods infer CCC due to their distinct scopes and the vast differences in the tools' outputs. In this regard, the outputs of Tensor-cell2cell seem valuable for identifying specific molecular targets and involved cells of a context-dependent module of communication, encompassing information beyond the scope of CellChat, which largely focuses on pair-wise, context-specific differences between signaling pathways.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

Meaningful biological roles are easily identifiable from the patterns captured by Tensor-cell2cell. For example, we detected a previously reported correlation between the interactions of the lung epithelium with the immune cells and COVID-19 severity[18] and differences between moderate and severe COVID-19, especially associated with modules of macrophage communication[28]. Tensor-cell2cell recapitulated molecular findings such as the role of SEMA4D-PLXNB2 interaction promoting inflammation[43], interaction that Tensor-cell2cell revealed to be stronger in cases with more lung inflammation (severe cases) (Figure 4), and the role of CCL2, CCL3, CCR1 and CCR5 as proinflammatory molecules, which makes them potential therapeutic targets for diminishing COVID-19 severity[18], proteins that our strategy associated with CCC in severe cases (Figure 4b). Additionally, we identified novel CCC patterns and mechanisms regarding COVID-19 pathogenesis. For example, Grant *et al.* reported that CD206^{hi} alveolar macrophages participate in the immune response to SARS-CoV-2 infection[51], but the underlying mechanisms mediating this response remain unclear. Factor 10 seems to extend the results presented by Grant *et al.* by showing that macrophage-expressed MRC1 (CD206) interacts with PTPRC (CD45) expressed by other cells (Figure 4a-b). Interestingly, the MRC1-PTPRC interaction mediating macrophage communication can promote immune tolerance[58], which is consistent with factor 10 being associated with moderate cases, wherein anti-inflammatory macrophages (M2-like phenotype) seem to be characteristic. Another example is a recent GWAS study that reported 13 significant loci associated with SARS-CoV-2 infection[59], wherein ICAM1 popped up as an involved gene. Remarkably, Tensor-cell2cell assigned a high loading to the ITGB2-ICAM1 interaction in a communication pattern that seems to be associated with antigen presentation (factor 5, Figures 4a-b), providing further insights of its potential mechanism.

One limitation of Tensor-cell2cell is that it is not intended for analyzing the behavior of specific pairs of cells or ligands and receptors of interest across contexts. Instead, Tensor-cell2cell finds data-driven modules of communication, which may result in factors with small loadings for cells and LR pairs of

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

interest. In such a scenario, those cells and LR pairs may not participate in a latent context-dependent module but may still follow certain behavior across contexts; thus, applying Tensor-cell2cell would not be appropriate. Instead, using other tools may be recommended (Table 1). Another 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 allows the inclusion of cell types missing in some conditions, the implemented tensor decomposition algorithm is not specifically designed for handling or imputing missing values. For that reason, we built a 4D-communication tensor that contains only the cell types that are shared across all samples in our COVID-19 study case. Thus, further improvements are needed for analyzing situations with missing values to include all possible members of communication (i.e., LR pairs and cell types that may be missing in certain contexts).

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. In addition to single cell data shown here, Tensor-cell2cell also accepts bulk transcriptomics data (an example of a time series bulk dataset of *C. elegans* is included in a CodeOcean 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. Moreover, we have also included Tensor-cell2cell as a part of our previously developed tool cell2cell[60], enabling previous functionalities such as employing any list of LR pairs (even 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[23,61], which could lead to different biological interpretations[57]. Lastly, we demonstrated that Tensor-cell2cell stands

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

as a fast, low-memory and accurate method (Figure 3), which can be substantially accelerated when a GPU is available (Figure 3a). Thus, these attributes make Tensor-cell2cell valuable for identifying key cell-cell and LR pairs mediating complex patterns of cellular communication in future studies, especially when considering the response time needed to deal with, for example, the COVID-19 pandemic.

Methods

RNA-seq data processing

RNA-seq datasets were obtained from publicly available resources. One dataset corresponds to a large-scale single-cell atlas of COVID-19 in humans[27], while the other to a COVID-19 dataset of single-cell transcriptomes for BALF samples[28]. These datasets were collected as raw count matrices from the NCBI's Gene Expression Omnibus[62] (GEO accession numbers GSE158055 and GSE145926, respectively). In total, the first dataset contains 1,462,702 single cells, while the second 65,813. 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 both datasets, we used the cell type labels included in their respective metadata. We aggregated the gene expression from single cells 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[63], as usually happens with genes encoding surface proteins[64].

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

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.

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 both COVID-19 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 both cases we further considered cell types that were present across all samples. Thus, the 4D-communication tensor for the PBMC and BALF datasets resulted in a size of 60 x 1639 x 6 x 6 and 12 x 189 x 6 x 6, respectively (that is, samples x ligand-receptor pairs x sender cell types x receiver cell types).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

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[65,66], 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[17,67]:

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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:

$$||\,\chi\,||_F^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} \chi_{ijkl}{}^2 \tag{4}$$

All the described calculations were implemented in Tensor-cell2cell through functions available in Tensorly[68], 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:

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

Computing Gini coefficients of joint distributions of cell pair loadings

To represent an overall communication state of cells in a factor ***r***, the outer product can be computed between the vectors s^r and t^r, which contains the loadings of sender and receiver cells for that factor, respectively. This outer product represents a joint distribution of cell pair loadings, and each value represents the overall communication between a pair of sender-receiver cells. A Gini coefficient[69] ranging between 0 and 1 can be computed on this distribution to detect the extent of equality in the overall communication values within a factor ***r***. In other words, how evenly distributed is the potential of communication across cell pairs. A value of 1 represents maximal inequality of overall communication

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

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).

Statistical analyses

In the COVID-19 dataset of BALF samples, we compared the loadings of samples as well as the fraction of macrophages with non-zero expression for a set of ligands and receptors across the categories representing healthy patients and varying severities of COVID-19. To perform these comparisons, we used an independent t-test followed by a Bonferroni's correction to indicate significance of the change (Supplementary Figures S3 and S4).

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, signalling 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 are active in two cellular communication networks, while the structural approach computes a network dissimilarity[70] to compare the topology of two signaling networks (see REF[10] for further details). Finally,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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 performs a manifold learning approach on sample similarities and returns UMAP embeddings for each signaling pathway in each different context (e.g. if CellChat evaluates 10 signaling pathways in 3 different contexts, it will return embeddings for 30 points) which can be used to rank the similarity of shared signalling pathways between contexts in a pairwise manner.

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

Training and evaluation of a classification model

A Random Forests[71] (RF) model was trained to predict disease status based on both COVID-19 status (healthy-control vs. patient with COVID-19) and severity (healthy-control, moderate symptoms, and severe symptoms). The RF model was trained using a Stratified K-Folds cross-validation (CV) with 3-Fold CV splits. On each CV split a RF model with 500 estimators was trained and RF probability-predictions were compared to the test set using the Receiver Operating Characteristic (ROC). The mean and standard deviation from the mean were calculated for the area under the Area Under the Curve (AUC) across the CV splits. This classification was performed on the context loadings of Tensor-cell2cell, and the two UMAP dimensions of the structural and functional joint manifold learning of CellChat, for both the BALF and PBMC COVID-19 datasets. All classification was performed through Scikit-learn (v. 0.23.2)[72].

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

Code and data availability

Tensor-cell2cell is implemented in our cell2cell suite[60], 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.v1). 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 September 23, 2021. 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.* 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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. Noël, F. *et al.* Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089 (2021).
22. Wang, Y. *et al.* iTALK: an R Package to Characterize and Illustrate Intercellular Communication. *Cancer Biology* (2019) doi:10.1101/507871.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

23. 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.

24. 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.

25. 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.

26. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).

27. Ren, X. *et al.* COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 1895–1913.e19 (2021).

28. Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).

29. Schmitt, T. L., Steiner, E., Klingler, P., Lassmann, H. & Grubeck-Loebenstein, B. Thyroid epithelial 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).

30. 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).

31. 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).

32. 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).

33. Pasello, M., Manara, M. C. & Scotlandi, K. CD99 at the crossroads of physiology and pathology. *J. Cell Commun. Signal.* **12**, 55–68 (2018).

34. Sanino, G., Bosco, M. & Terrazzano, G. Physiology of Midkine and Its Potential Pathophysiological

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

Role in COVID-19. *Front. Physiol.* **11**, 616552 (2020).

35. 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).
36. Weckbach, L. T., Muramatsu, T. & Walzog, B. Midkine in inflammation. *ScientificWorldJournal* **11**, 2491–2505 (2011).
37. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* **20**, 363–374 (2020).
38. Merad, M. & Martin, J. C. Author Correction: Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat. Rev. Immunol.* **20**, 448 (2020).
39. Chen, Z. & John Wherry, E. T cell responses in patients with COVID-19. *Nat. Rev. Immunol.* **20**, 529–536 (2020).
40. Xia, J. *et al.* Semaphorin-Plexin Signaling Controls Mitotic Spindle Orientation during Epithelial Morphogenesis and Repair. *Dev. Cell* **33**, 299–313 (2015).
41. Nikaido, T. *et al.* Serum Syndecan-4 as a Possible Biomarker in Patients With Acute Pneumonia. *J. Infect. Dis.* **212**, 1500–1508 (2015).
42. 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).
43. Clark, I. C. *et al.* Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science* **372**, (2021).
44. Zhang, F. *et al.* IFN- γ and TNF- α drive a CXCL10 + CCL2 + macrophage phenotype expanded in severe COVID-19 and other diseases with tissue inflammation. *bioRxiv* (2020) doi:10.1101/2020.08.05.238360.
45. Martinez, F. O., Combes, T. W., Orsenigo, F. & Gordon, S. Monocyte activation in systemic Covid-19 infection: Assay and rationale. *EBioMedicine* **59**, 102964 (2020).
46. Ocaña-Guzman, R., Torre-Bouscoulet, L. & Sada-Ovalle, I. TIM-3 Regulates Distinct Functions in

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

Macrophages. *Front. Immunol.* **7**, 229 (2016).

47. Matsuyama, T., Kubli, S. P., Yoshinaga, S. K., Pfeffer, K. & Mak, T. W. An aberrant STAT pathway is central to COVID-19. *Cell Death Differ.* **27**, 3209–3225 (2020).
48. Montoya, J. G. *et al.* Cytokine signature associated with disease severity in chronic fatigue syndrome patients. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7150–E7158 (2017).
49. Kohyama, M. *et al.* Monocyte infiltration into obese and fibrilized tissues is regulated by PILRα. *Eur. J. Immunol.* **46**, 1214–1223 (2016).
50. Saheb Sharif-Askari, N. *et al.* Enhanced expression of immune checkpoint receptors during SARS-CoV-2 viral infection. *Mol Ther Methods Clin Dev* **20**, 109–121 (2021).
51. Grant, R. A. *et al.* Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* **590**, 635–641 (2021).
52. Jafarzadeh, A., Chauhan, P., Saha, B., Jafarzadeh, S. & Nemati, M. Contribution of monocytes and macrophages to the local tissue inflammation and cytokine storm in COVID-19: Lessons from SARS and MERS, and potential therapeutic interventions. *Life Sci.* **257**, 118102 (2020).
53. Florez-Sampedro, L., Soto-Gamez, A., Poelarends, G. J. & Melgert, B. N. The role of MIF in chronic lung diseases: looking beyond inflammation. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **318**, L1183–L1197 (2020).
54. Gracia-Hernandez, M., Sotomayor, E. M. & Villagra, A. Targeting Macrophages as a Therapeutic Option in Coronavirus Disease 2019. *Frontiers in Pharmacology* vol. 11 (2020).
55. Garash, R., Bajpai, A., Marcinkiewicz, B. M. & Spiller, K. L. Drug delivery strategies to control macrophages for tissue repair and regeneration. *Exp. Biol. Med.* **241**, 1054–1063 (2016).
56. 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.
57. Dimitrov, D. *et al.* Comparison of Resources and Methods to infer Cell-Cell Communication from

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

Single-cell RNA Data. *bioRxiv* 2021.05.21.445160 (2021) doi:10.1101/2021.05.21.445160.

58. Schuette, V. *et al.* Mannose receptor induces T-cell tolerance via inhibition of CD45 and up-regulation of CTLA-4. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10649–10654 (2016).
59. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* (2021) doi:10.1038/s41586-021-03767-x.
60. Armingol, E. *et al.* Inferring the spatial code of cell-cell interactions and communication across a whole animal body. *bioRxiv* 2020.11.22.392217 (2020) doi:10.1101/2020.11.22.392217.
61. 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).
62. 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).
63. Sina Booeshaghi, A. & Pachter, L. Normalization of single-cell RNA-seq counts by log(x+1)* or log(1+x)*. *bioRxiv* 2020.05.19.100214 (2020) doi:10.1101/2020.05.19.100214.
64. 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).
65. 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).
66. Harshman, R. A. & Others. Foundations of the PARAFAC procedure: Models and conditions for an' explanatory' multimodal factor analysis. (1970).
67. Anandkumar, A., Ge, R. & Janzamin, M. Guaranteed Non-Orthogonal Tensor Decomposition via Alternating Rank-1 Updates. *arXiv [cs.LG]* (2014).
68. Kossaifi, J., Panagakis, Y., Anandkumar, A. & Pantic, M. TensorLy: Tensor Learning in Python. *arXiv [cs.LG]* (2016).
69. Farris, F. A. The Gini Index and Measures of Inequality. *Am. Math. Mon.* **117**, 851–864 (2010).
70. Schieber, T. A. *et al.* Quantification of network structural dissimilarities. *Nat. Commun.* **8**, 13928

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

(2017).

71. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

72. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

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. 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.

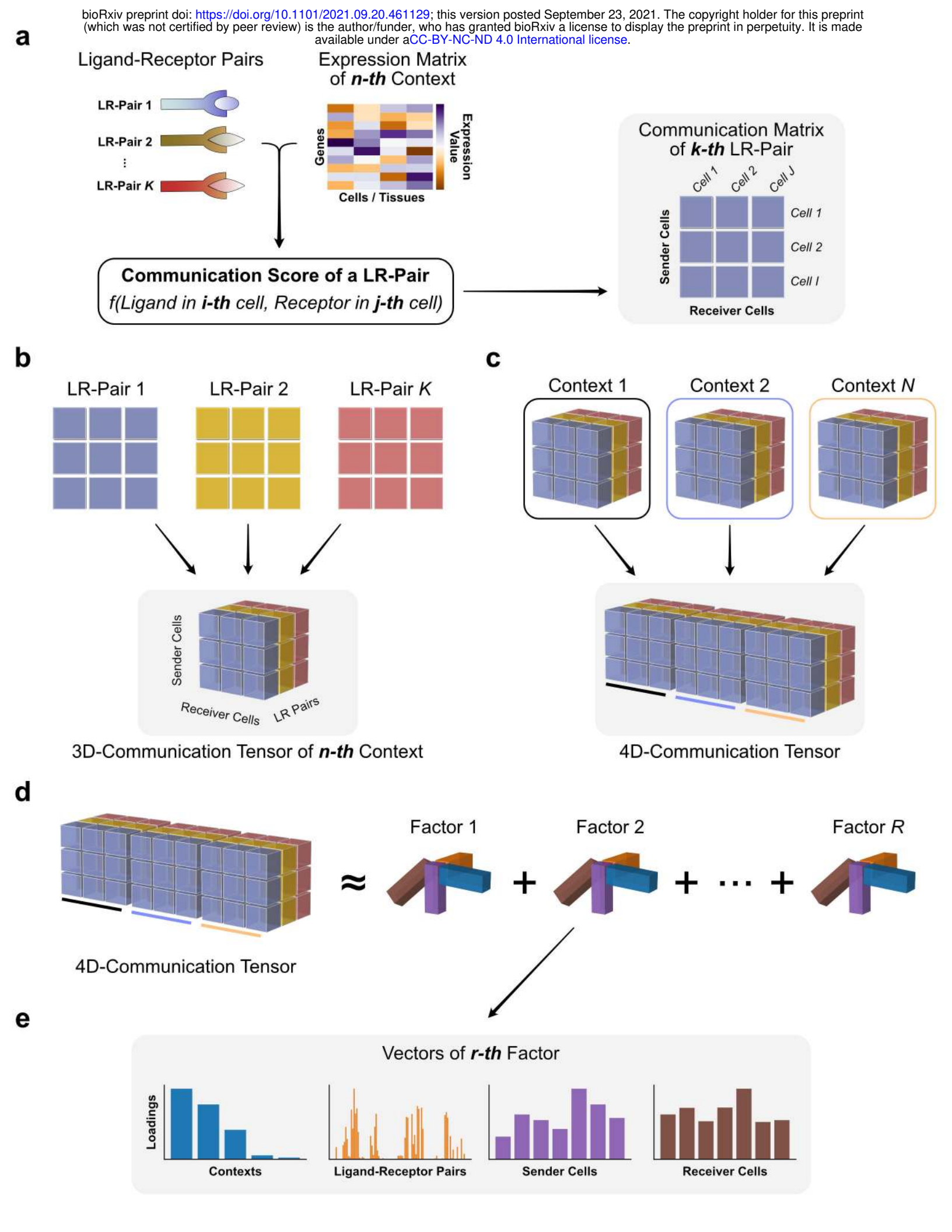
Author contributions

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 *C. elegans*. 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-cell2cel to CellChat. APL helped to interpret the results 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.

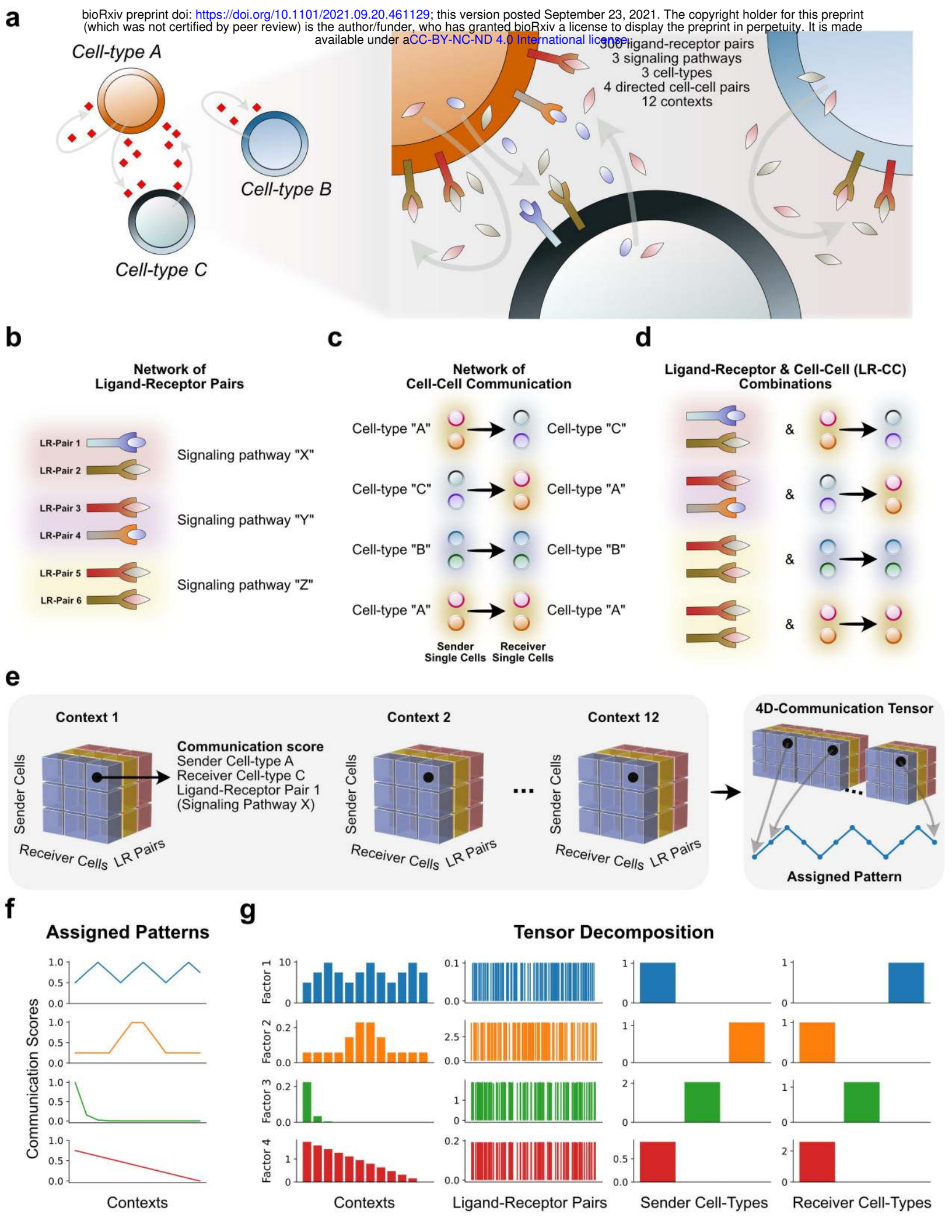
Competing interests

The authors declare no competing interests.

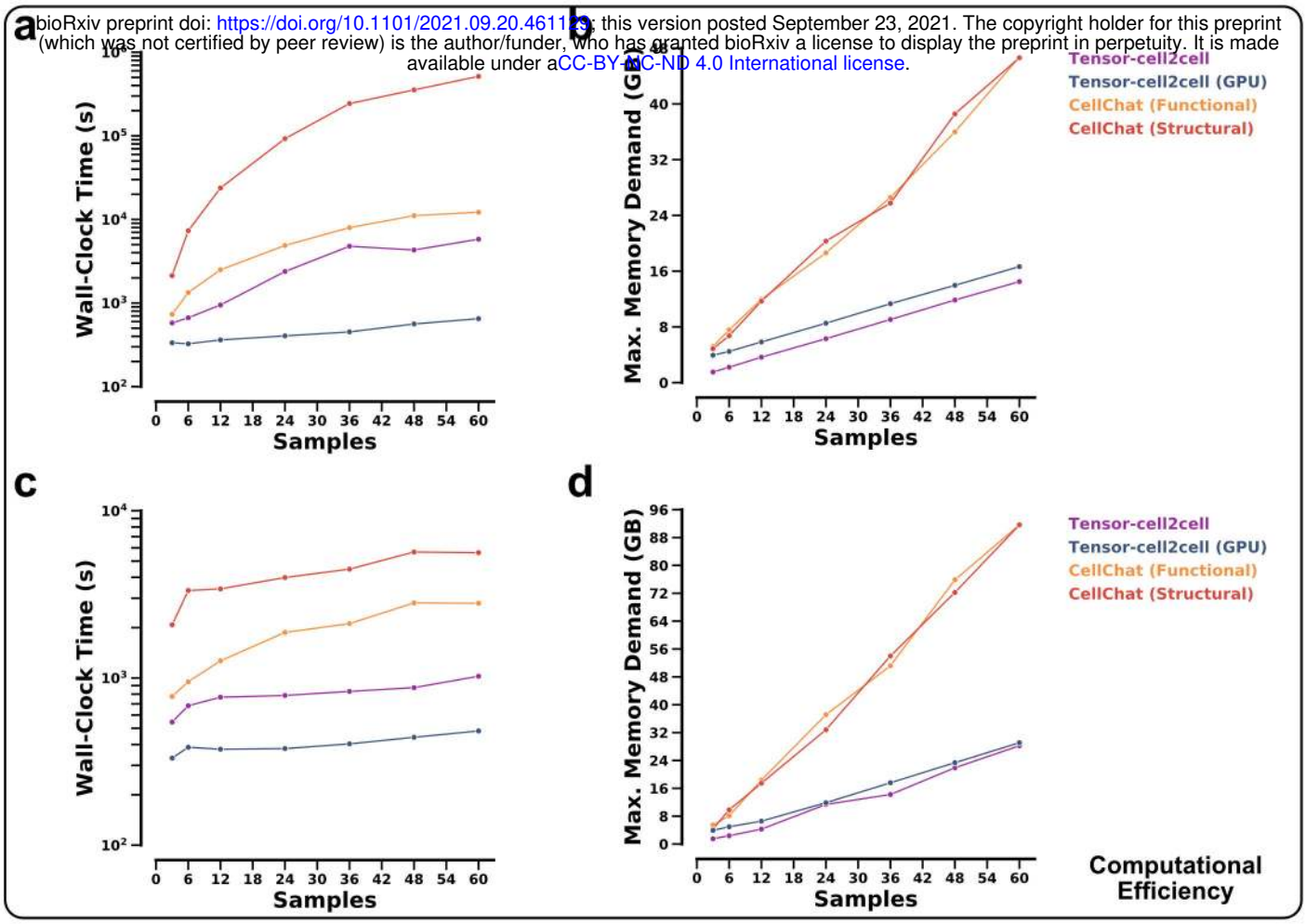
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

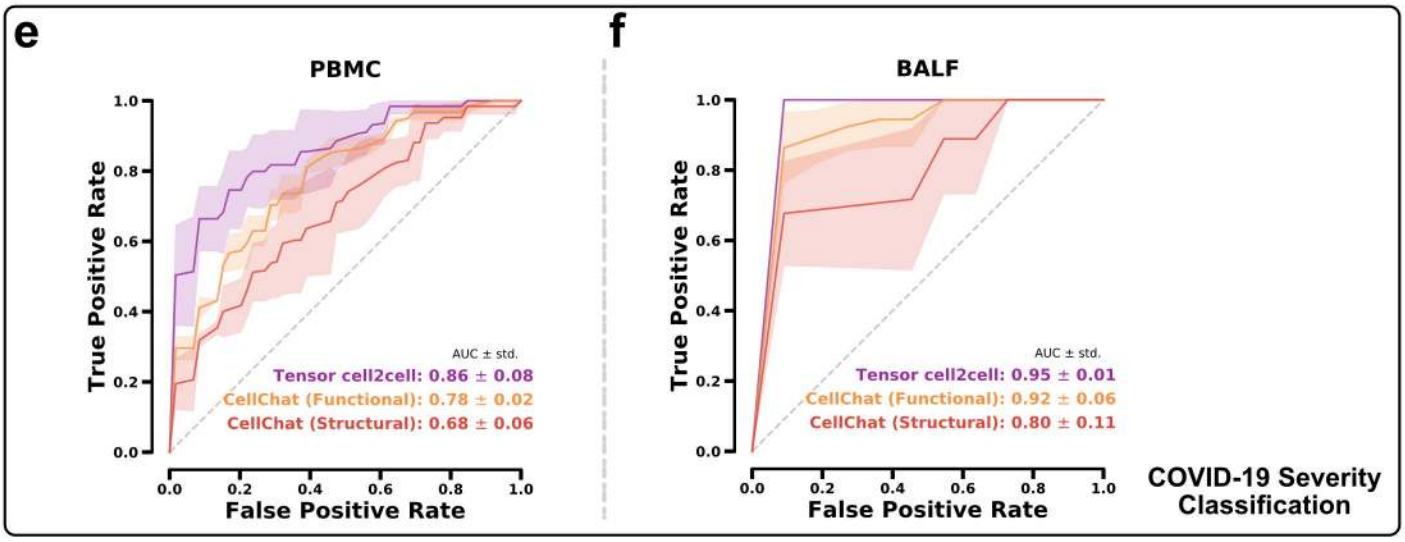


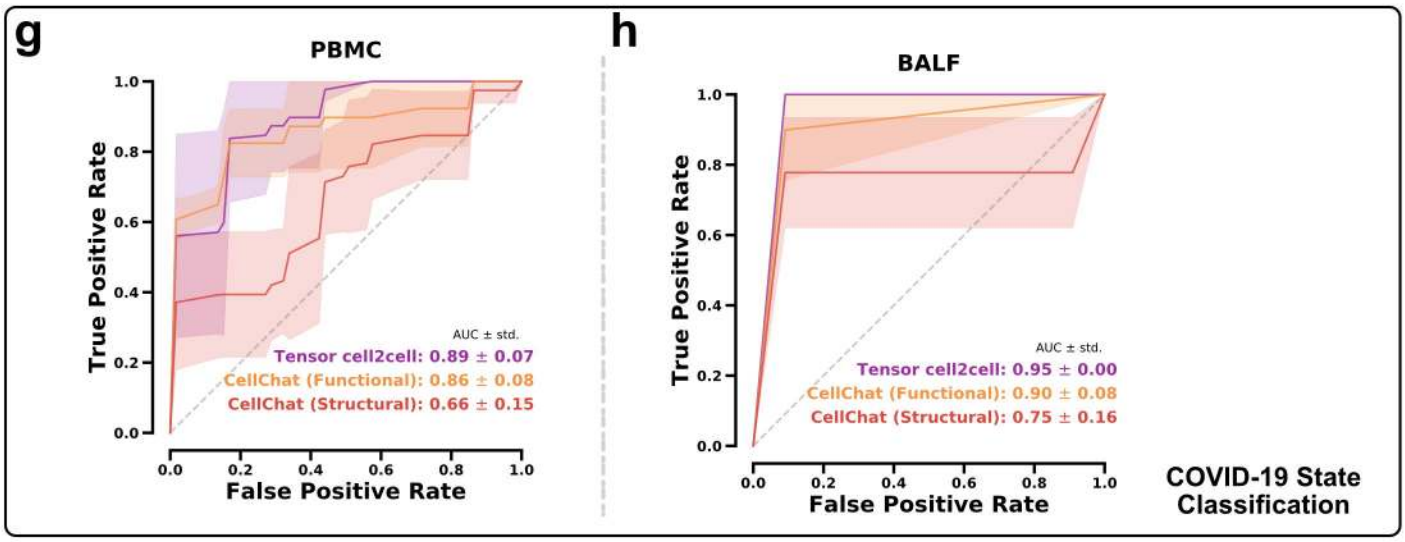
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. 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.

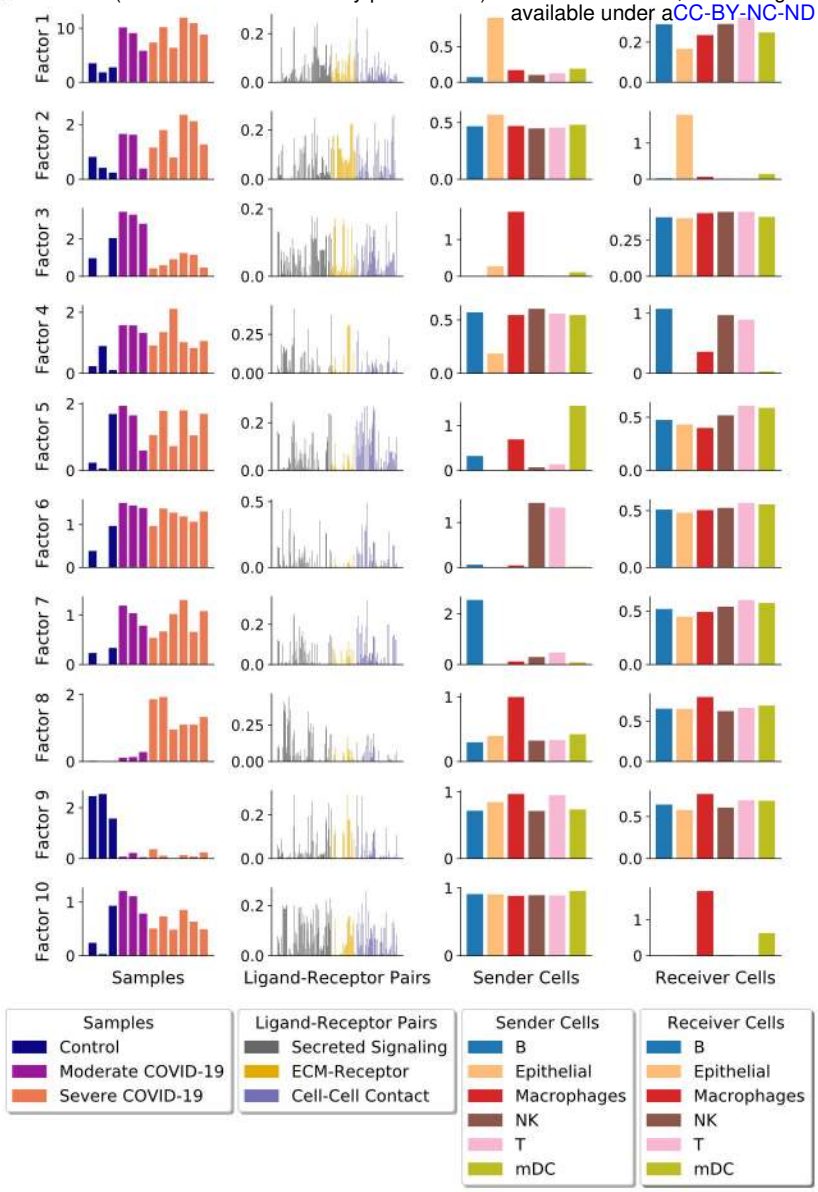






bioRxiv preprint doi: https://doi.org/10.1101/2021.09.20.461129; this version posted September 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a



b

Top-5 Ligand-Receptor Pairs

Factor 1		Factor 2	
APP - CD74	0.266	SEMA4D - PLXNB2	0.261
MDK - NCL	0.244	SEMA4A - PLXNB2	0.249
MIF - CD74 & CD44	0.229	MDK - SDC4	0.248
MDK - ITGA4 & ITGB1	0.225	COL9A2 - SDC4	0.229
CD99 - CD99	0.218	F11R - F11R	0.228

Factor 3		Factor 4	
SIGLEC1 - SPN	0.191	MIF - CD74 & CD44	0.413
RETN - CAP1	0.179	LGALS9 - CD44	0.371
MDK - NCL	0.174	COL9A2 - CD44	0.315
FN1 - ITGA4 & ITGB1	0.173	LAMB3 - CD44	0.311
FN1 - ITGA4 & ITGB7	0.170	LAMB2 - CD44	0.301

Factor 5		Factor 6	
CD99 - CD99	0.271	CD99 - CD99	0.491
ITGB2 - CD226	0.268	CCL5 - CCR5	0.450
CD86 - CTLA4	0.267	CCL5 - CCR1	0.432
ITGB2 - ICAM2	0.255	GZMA - F2R	0.355
ITGB2 - ICAM1	0.244	PTPRC - MRC1	0.327

Factor 7		Factor 8	
CD99 - CD99	0.321	CCL2 - CCR2	0.442
MIF - CD74 & CXCR4	0.253	CCL3 - CCR1	0.410
CD22 - PTPRC	0.243	CCL8 - CCR1	0.388
MIF - CD74 & CD44	0.240	CCL3 - CCR5	0.354
SELL - MADCAM1	0.200	CCL3L1 - CCR1	0.354

Factor 9		Factor 10	
FN1 - CD44	0.296	CD99 - PILRA	0.257
MIF - CD74 & CD44	0.291	LGALS9 - HAVCR2	0.230
APP - CD74	0.288	ANXA1 - FPR1	0.220
PTPRC - MRC1	0.284	MDK - LRP1	0.208
RETN - CAP1	0.262	PTPRC - MRC1	0.207

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