# Unraveling the coordinated dynamics of protein- and metabolite-mediated cell-cell communication

Erick Armingol[1,2], Reid O. Larsen[3,4], Martin Cequeira, Hratch Baghdassarian[1,2], Nathan E. Lewis[2,5*]

[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] Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA
[4] Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA
[5] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA

*Correspondence: nlewisres@ucsd.edu

## Abstract

Cell-cell communication involves multiple classes of molecules, diverse interacting cells, and complex spatiotemporal dynamics. While this communication can be inferred from single-cell RNA-seq, no computational methods can account for both protein and metabolite ligands simultaneously, while also accounting for the temporal dynamics. We adapted Tensor-cell2cell here to study several time points simultaneously and jointly incorporate both ligand types. Our approach detects temporal dynamics of cell-cell communication during brain development, allowing for the detection of the concerted use of key protein and metabolite ligands by pertinent interacting cells.

## Main

Cell-cell communication (CCC) is multimodal and dynamic, involving both small molecule and protein signals integrated with fine spatiotemporal coordination to drive multicellular functions. Although single-cell transcriptomics has enabled computational methods to infer cell-cell communication from gene expression[1], these approaches have mainly focused on protein-protein interactions defining ligand-receptor (LR) pairs that cells can use to communicate. Thus, computational approaches are needed to untangle complex dynamics of CCC involving both proteins and metabolites simultaneously.

Two recent efforts have helped to incorporate the use of non-protein ligands or metabolites for inferring CCC[2,3], enabling the study of these molecules from single-cell transcriptomics. However, it has remained unclear how to integrate protein- and metabolite-mediated CCC to perform analyses accounting for both simultaneously. Unlike CCC mediated by protein ligands, which can be inferred directly from the expression levels of the encoding genes encoding, CCC mediated by metabolite ligands is indirectly inferred from the corresponding enzymes that could either produce or consume them. Thus, the resulting communication scores of proteins and metabolites extracted from gene expression represent distinct activities of LR pairs, being only informative in isolation due to differences in the scale of values between protein- and metabolite-based communication scores. In addition, these tools do not account for temporal coordination of LR pairs; rather, only up to two samples can be accounted for at a time, preventing insights of protein and metabolite ligands acting concertly and limiting the analysis to be differential between up to two time points.

To address these challenges, here we adapt Tensor-cell2cell[4] to integrate both proteins and metabolites as the ligands mediating the intercellular interactions. Briefly, Tensor-cell2cell is a tool for studying complex CCC patterns and dynamics. Distinct context-dependent patterns are summarized by signatures or factors that identify LR pairs which act in specific combinations of sender-receiver cell types. To integrate the protein- and metabolite-based scores for joint assessment of changes in CCC, we generate a 4D-communication tensor wherein the mediator molecules include proteins and metabolites as ligands, and their communication scores are transformed to be comparable (Fig. 1). First, scores are separately inferred using *cell2cell*[5] for protein-mediated CCC and MEBOCOST[2] for metabolite-mediated CCC. Since each method is designed for the distinct types of molecule, they output different communication score distributions (Supplementary Fig. 1a). Then, we built a separate 4D-communication tensor for the protein-protein and metabolite-protein LR pairs (Figs. 1a,b). Next, we transform their values using a regularization approach[6] (see Methods). After this transformation, the communication scores between the two classes of LR pairs were comparable in terms of scale of values (Supplementary Fig. 1b). They further became more similar in terms of distribution types, but with differences in their skewness and kurtosis (see Supplementary Notes). Thus, the resulting 4D tensor with both classes of ligands (Fig.1c) allows Tensor-cell2cell to jointly assess the signaling activities coming from these two kinds of molecules, enabling a better understanding of how they are interconnected and the extent to which they follow similar patterns.

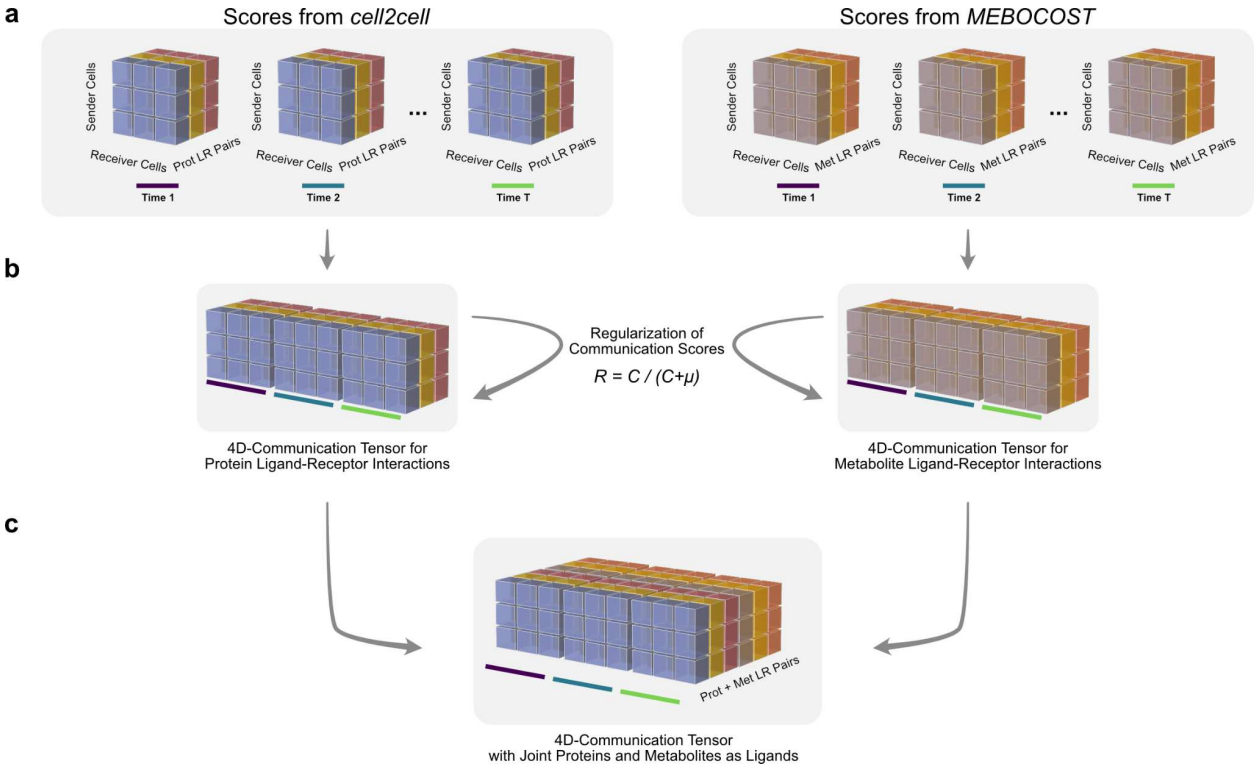



**Fig. 1. Integration of communication scores from protein- and metabolite-based inference.** (**a**) Communication scores are inferred separately for protein- and metabolite-mediated interactions using *cell2cell* and MEBOCOST, respectively. A 4D-communication tensor is built in each case using the resulting scores. (**b**) Each 4D-communication tensor is transformed by a regularization approach as described in Methods, scaling scores between 0 and 1. (**c**) Then, both tensors are concatenated on the ligand-receptor pairs axis, generating a joint 4D-communication tensor.

To illustrate the utility of our approach, we applied it to study brain development. This complex temporal process involves a sequence of stages that generate a wide variety of neural and non-neural cell types. These cell types must be produced in the correct number, and with the proper spatiotemporal dynamics[7]; dynamics that are interconnected with the protein and metabolite ligands transmitted between the different cell types in the brain. Hence, using multiple time points that represent different stages of development of a brain cortical organoid model[8], Tensor-cell2cell detected six main signatures of the dynamics of CCC occurring during brain development (factors in Fig. 2a). Each signature distinguishes CCC between different cell types and involves key protein and metabolite ligand-receptor interactions depending on the development stage that these dynamics represent.

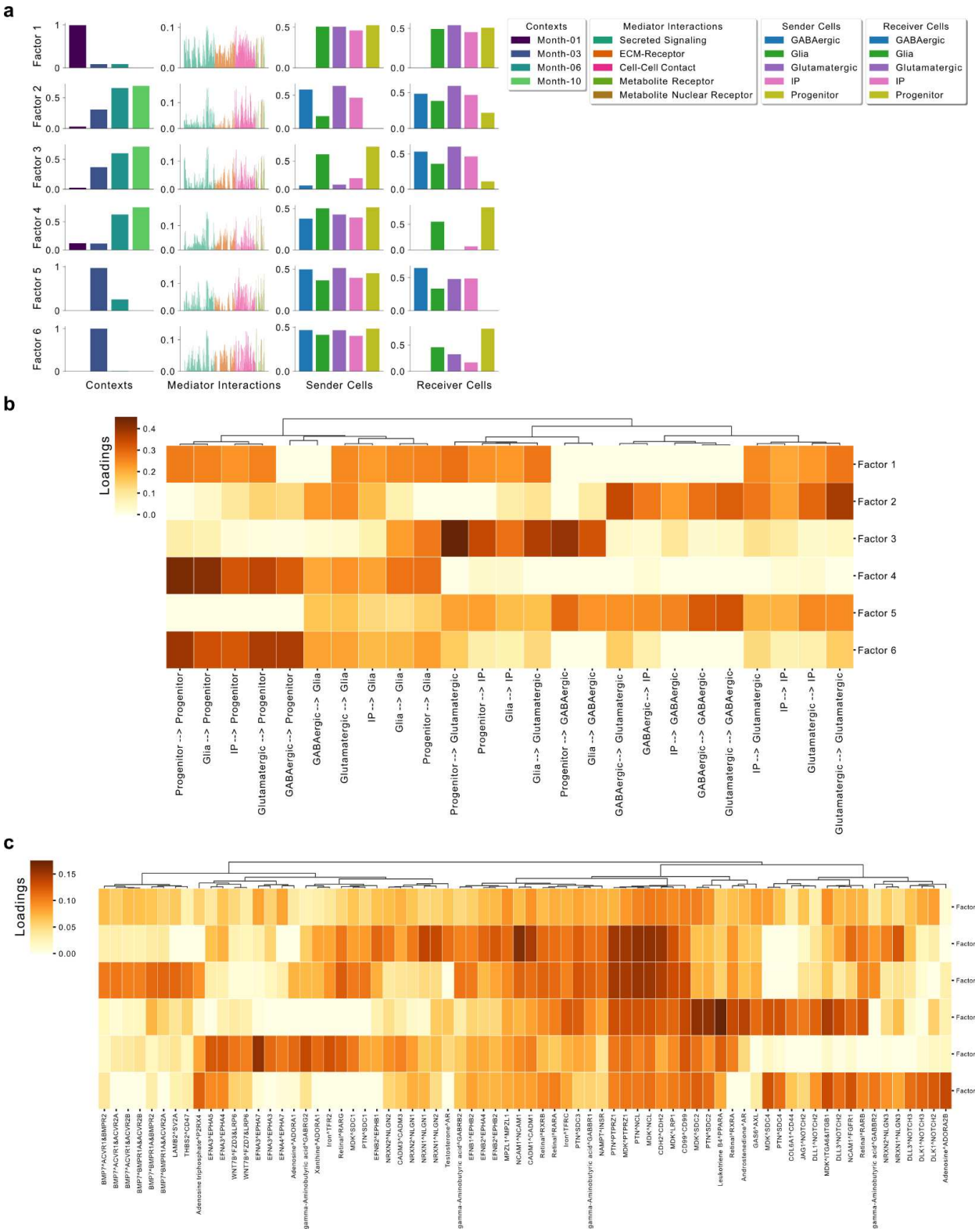



**Fig. 2. Main coordinated dynamics of protein- and metabolite-mediated cell-cell communication in brain development.** (**a**) Tensor decomposition of the joint 4D-communication tensor including both proteins and metabolites as the ligands. Each row indicates a factor obtained with Tensor-cell2cell, wherein the y-axis represents the loadings (conceptually, the importance) of the elements (x-axis) included in the respective dimensions (columns). Bars are colored as indicated in the legend. The number of factors to consider was evaluated through an elbow analysis

(Supplementary Fig. 4). (**b**) Heatmap summarizing the joint loadings for sender-receiver cell pairs in each of the factors. Higher values indicate that a given directed pair is one of the main actors in the factor. A hierarchical clustering was performed on the cell-cell pairs given their loadings across factors. These joint loadings for the directed sender-receiver cell pairs were computed as the outer product of the loadings of sender and receiver cells in (**a**). (**c**) Heatmap showing protein and metabolite LR pairs that are the main mediators per factor. Only LR pairs with a loading value greater than 0.1 in at least one factor are shown. A hierarchical clustering was performed on the LR pairs given their loadings across factors.

The signatures of CCC dynamics in brain development captured by Tensor-cell2cell involve either specific time points (factors 1, 5, and 6) or changes across time points (factors 2-4), which are linked to different ligand-receptor and cell-cell interactions (see a summary of these dynamics in Supplementary Table 1, and further details in the Supplementary Notes). Particularly, factor 1 captures communication in the first month including all cell types except GABAergic neurons (Figs. 2a,b), a cell type that is virtually absent in the first month (Supplementary Fig. 2c). Mediators in factor 1 encompass a wide range of molecules (Fig. 2c), but the main actors correspond to midkine (MDK), pleiotrophin (PTN), ephrins (EFNA and EFNB), and CD99. Meanwhile, factors 5 and 6 capture communication in the third month. Factor 5 is associated with receiver cells that are not progenitors, wherein GABAergic neurons are important (highest loadings in the receiver dimension, Fig. 2a). Tensor-cell2cell also links this factor with the use of ephrin/Eph and Wnt by neurons (Fig. 2c). In contrast, factor 6 is associated with progenitors as the main receiver cells, excluding GABAergic neurons as main receivers (Figs. 2a,b). Here, delta-like/Notch and GABA are important actors on progenitors during the third month of development (Fig. 2c). Additionally, MDK, PTN, CD99, retinal and adenosine act on factors 5 and 6 (Fig. 2c).

Regarding CCC inferred to increase across time, factor 2 is associated with communication of GABAergic and Glutamatergic neurons, and intermediate progenitors (IPs) as both senders and receivers, and glial cells as additional receivers (Figs. 2a,b). Important molecular mediators correspond to MDK, PTN, NCAM1, CADM1, NRXN/NLGN, ephrin/Eph, retinal, iron, GABA, and testosterone. Factors 3 and 4 are primarily associated with progenitors and glial cells as the main sender and receiver cells, respectively. Progenitors and glial cells send signals to all cells except progenitors in factor 3, including mediators such as MDK, PTN, NCAM1, CADM1, ephrins, retinal, iron, GABA, and BMP7. Meanwhile, factor 4 identifies signaling from all cell types to progenitors and glial cells, where leukotriene has a crucial role (Fig. 2c). MDK, PTN, and CD99 are important in this factor as well (Fig. 2c), showing their central role in brain development[9].

Our adapted approach is uniquely capable of detecting concerted activities of specific protein and metabolite ligands. For example, interactions of MDK and retinal with their cognate receptors cluster together across factors (Fig. 2c). MDK is encoded by a retinoic acid (RA)-responsive gene[9], and retinal is the precursor of RA, a molecule that has a crucial role during brain development[10], aligning with our data-driven results that suggest these two ligands act in concert. Moreover, Tensor-cell2cell reveals that delta-like/Notch and GABA are important mediators of CCC involving progenitors as the main receiver cells in the third month of development (factor 6 in Fig. 2). Notch signaling at this point maintains a proliferative state of progenitors[11] that help generate a proper number of cells during early stages while blocking differentiation into neurons[7,12]. Similarly, GABA plays an important role in promoting proliferation of undifferentiated

neural progenitors[13]. Thus, our approach reveals biologically meaningful coordination of protein- and metabolite-mediated CCC by which both ligand classes together deliver synergistic biological information.

Using enrichment analysis, we also demonstrate that the factors detected by Tensor-cell2cell represent higher-order functions occurring at different stages of brain development. For instance, we found that neurogenesis, neuron differentiation, axon development, morphogenesis, among other related processes are repeated in factors 2 and 5 (Supplementary Fig. 3), but associated with different time points (Fig. 2a). Factor 5 captures CCC that seems associated with neuron differentiation and earlier synapse formation between neurons, while factor 2 involves signals that promote cell-cell contact and consolidate differentiation and synapse formation (Fig. 2c and Supplementary Fig. 3). Furthermore, factors 4 and 6 encompass development and signaling regulation in general (Supplementary Fig. 3), wherein progenitors and glial cells are the main receivers (Fig. 2a). Interestingly, factor 6 represents month 3 and involves Notch signaling that seems to inhibit differentiation of these cells[7,12], while factor 4 represents months 6 and 10, and mainly involves leukotriene, which has an important role in inducing differentiation into neurons[14] (Fig. 2c). Thus, these two factors seem to capture CCC reflecting a transition of progenitors from a proliferative to a differentiating state.

By adapting Tensor-cell2cell to integrate *cell2cell*[5] and MEBOCOST[2] scoring methods, we demonstrated that the temporal dynamics of both protein- and metabolite-mediated CCC can be jointly inferred and analyzed from single-cell omics. Applying our approach to decipher different types of CCC during brain cortical organoid development[8], Tensor-cell2cell found concerted dynamics of protein- and metabolite-mediated CCC. Key proteins and metabolites were linked to cells and time points in a biologically meaningful manner. While we previously demonstrated that Tensor-cell2cell is robust to different protein-based communication scoring methods[4], we cannot test its sensitivity to metabolite-based communication scores due to the scarcity of such methods. Thus, it is important to consider the differences that future tools for inferring metabolite-mediated CCC could introduce.

Our method recapitulated, in one analysis, multiple results discovered across many studies about brain development. Our analysis revealed proteins such as MDK, PTN, ephrins, CD99, Wnt, NOTCH, and neurexin-neuroligin, and metabolites such as retinal, GABA, and leukotriene to be important in brain development (CCC dynamics summarized in Supplementary Table 1). This is consistent with previous experimental evidence (see Supplementary Notes) and demonstrates the utility of jointly assessing protein- and metabolite- ligands as CCC mediators. We gain further novelty by analyzing CCC temporal dynamics across multiple time points, rather than exploring changes in a static manner. Finally, in addition to brain development, our approach can be applied to diverse datasets, enabling mechanistic insights into the dynamic CCC underlying organ development or other temporal biological processes across normal and dysregulated conditions (e.g. mutations, different growth conditions, etc.).

## Methods

### *Preprocessing of the RNA-seq data*

Raw data (GEO accession number GSE130238) from 1, 3, 6, and 10-month brain organoids[8] were processed using the Seurat v3.0 package[15]. To annotate single cells, the following pipeline was followed: Genes from individual time points that were not detected in at least 3 cells were discarded. Data from each time point were merged and cells with fewer than 200 genes, more than 6000 genes, or more than 7.5% mitochondrial genes were discarded. Then, data were scaled to 10,000 transcripts per cell and log transformed using the NormalizeData function. Most variable genes were found using FindVariableFeatures with default parameters, and used as input of the ScaleData function followed by Principal Component Analysis (PCA) across time points. Using the PCA embeddings, all time points were batch-corrected and integrated with Harmony[16] using the RunHarmony function with default parameters. After integration, clustering analysis was performed with default parameters on the Harmony embeddings. Resolution was set to 0.5 and only the first 20 harmonized components were considered for finding neighbor cells. In this case, 17 clusters were identified and further merged into the 7 main clusters to represent Glutamatergic, GABAergic, Glia, Progenitor, Intermediate Progenitor (IP), Mitotic and Other cell types (Supplementary Figs. 2a,b), based on the markers used in the ref.[8] (Supplementary Fig. 2c). Mitotic and Other cells were excluded from the analyses. To perform the CCC analyses, the expression level of each gene was aggregated at the cell-type level by computing the log1p(CPM) average expression within each cluster. Cell types with less than 20 single cells were excluded from the respective time point.

### *Ligand-receptor pairs for inferring cell-cell communication*

The CellChat database of ligand-receptor pairs was used[17], which includes 2,005 human protein-LR interactions and around 48% of these LR pairs include heteromeric-protein complexes. The MEBOCOST database of metabolite-sensor pairs[2] was used, which includes 444 metabolite-LR interactions. Here, we considered only metabolite-receptor interactions (206 total interactions). After integrating these lists with the genes that were present across all time points, a total of 422 protein- and 54 metabolite-based LR pairs were considered.

### *Integrating protein- and metabolite-based interactions*

To integrate protein- and metabolite-mediated CCC, two 4D-communication tensors were built independently and integrated afterwards (Fig. 1). One tensor includes the protein-based LR interactions while the other includes the metabolite-based LR interactions. Each of these tensors is built by three main steps previously detailed[4]: 1) A communication matrix is generated for each ligand-receptor pair of the considered interactions. In this matrix, a communication score is assigned to the respective LR pair, based on the sample-wise expression of the ligand and the receptor by a respective sender and receiver cell. 2) All the resulting communication matrices are joined into a 3D-communication tensor for each time point (Fig. 1a). Steps 1 and 2 are repeated for every sample (time points in this case). 3) Once the 3D-communication tensors for all samples are built, they are then combined in a way that each of these 3D tensors represent a coordinate in the 4th-dimension of the 4D-communication tensor (4th dimension represented by the colored lines in Fig. 1a). To assign the communication scores, *cell2cell* was used for the protein-LR pairs, while MEBOCOST was used for the metabolite-LR pairs.

The communication score for the protein-based LR pairs was computed as the geometric mean of the expression of the ligand in a sender cell type and the receptor in a receiver cell type. The gene expression of multimeric proteins was considered as the minimum gene expression among the subunits. While the communication score for the metabolite-based LR pairs was computed with the default parameters of MEBOSCOST, which considers the average expression of metabolite-consuming and producing enzymes. Negative values were replaced by zeros, and all resulting communication scores were root squared to represent the geometric mean between the metabolite and receptor expression values, similar to the protein-based LR pairs. Permutations were not performed, and all resulting communication scores were considered since we are interested in both background and cell-type-specific communication, instead of just the latter. Then, each communication score (**C**) of both 4D tensors was regularized by using the mean expression level across all genes and cell types (**μ**)[6], resulting in a regularized score (**R**) with lower and upper bounds of 0 and 1, respectively:

$$R = \frac{C}{C + \mu}$$

These regularized tensors were concatenated, resulting in a 4D-communication tensor with joint protein- and metabolite-based LR pairs. This integrated 4D-communication tensor contains Glutamatergic neurons, GABAergic neurons, Glias, Progenitors, and IPs as the sender and receiver cell types. Thus, the different dimensions of this tensor contains 4 time points; 476 LR pairs; 5 sender cells; and 5 receiver cells, resulting in a tensor of shape 4 x 476 x 5 x 5.

***Deconvolution of communication dynamics with Tensor-cell2cell***

To deconvolve the integrated 4D-communication tensor, Tensor-cell2cell runs a non-negative tensor component analysis (TCA)[4], which extracts a specific number of factors that represent different context-driven patterns of cell-cell communication. An elbow analysis was performed to automatically select a proper number of factors to extract (Supplementary Fig. 4), based on a tensor decomposition error[18]. These analyses were done through a robust run of Tensor-cell2cell, by using tol=1e-8, and n_iter_max=500 as parameters of the tensor decomposition. Tensor-cell2cell was run using a NVIDIA RTX A6000 GPU with 48GB of memory and ~10k CUDA cores.

***Gene set enrichment analysis for ligand-receptor pairs***

LR-pair loadings of each factor can be used to run a Gene Set Enrichment Analysis (GSEA)[19]. Thus, biological functions can be assigned to each of the factors. Before running the analysis, we created a list of biological functions associated with each of the LR pairs in CellChat. We considered the gene sets of the Biological Process GO Terms, available at http://www.gsea-msigdb.org/, and annotated each LR pair with all the gene sets that contain all of its interacting genes for the protein-based LR pairs. Metabolite-based LR pairs were annotated with the gene sets associated with the receptor genes. Then, by filtering LR sets to those containing at least 15 LR pairs, we end up with 664 LR sets. As inputs of GSEA, we passed the LR pairs in each factor ranked by their loadings and the LR sets. In this case, we used the PreRanked GSEA function in the package gseapy with the default parameters and 999 permutations.

## Data availability

All input data used for the analyses in this work and the result-generated data are available online in a "Code Ocean capsule [https://doi.org/10.24433/CO.7772943.v1]". In particular, single-cell RNA-seq data for the brain cortical organoid[8] was previously deposited in the NCBI's Gene Expression Omnibus database under accession code "GSE130238 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130238]".

## Code availability

The tools *cell2cell* and Tensor-cell2cell are available in a "GitHub repository [https://github.com/earmingol/cell2cell]", and MEBOCOST is available in another "GitHub repository [https://github.com/zhengrongbin/MEBOCOST]". The code and input data used for the analyses are available online in a "Code Ocean capsule [https://doi.org/10.24433/CO.7772943.v1]", which can be used to run all the analyses performed in this work online.

## References

1. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
2. Zheng, R. *et al.* MEBOCOST: Metabolic Cell-Cell Communication Modeling by Single Cell Transcriptome. *bioRxiv* 2022.05.30.494067 (2022) doi:10.1101/2022.05.30.494067.
3. Garcia-Alonso, L. *et al.* Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).
4. Armingol, E. *et al.* Context-aware deconvolution of cell-cell communication with Tensor-cell2cell. *Nat. Commun.* **13**, 3665 (2022).
5. Armingol, E. *et al.* Inferring a spatial code of cell-cell interactions across a whole animal bodys. *bioRxiv* 2020.11.22.392217 (2021) doi:10.1101/2020.11.22.392217.
6. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
7. Jiang, X. & Nardelli, J. Cellular and molecular introduction to brain development. *Neurobiol. Dis.* **92**, 3–17 (2016).
8. Trujillo, C. A. *et al.* Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell* **25**, 558–569.e7 (2019).
9. Kadomatsu, K. & Muramatsu, T. Midkine and pleiotrophin in neural development and cancer. *Cancer Lett.* **204**, 127–143 (2004).
10. Maden, M. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat. Rev. Neurosci.* **8**, 755–765 (2007).
11. Ohtaka-Maruyama, C. & Okado, H. Molecular Pathways Underlying Projection Neuron Production and Migration during Cerebral Cortical Development. *Front. Neurosci.* **9**, 447 (2015).
12. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
13. Fukui, M. *et al.* Modulation of cellular proliferation and differentiation through GABA(B) receptors expressed by undifferentiated neural progenitor cells isolated from fetal mouse brain. *J. Cell. Physiol.* **216**, 507–519 (2008).
14. Wada, K. *et al.* Leukotriene B4 and lipoxin A4 are regulatory signals for neural stem cell proliferation and differentiation. *FASEB J.* **20**, 1785–1792 (2006).
15. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
16. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
17. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

18. 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).

19. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

## Acknowledgements

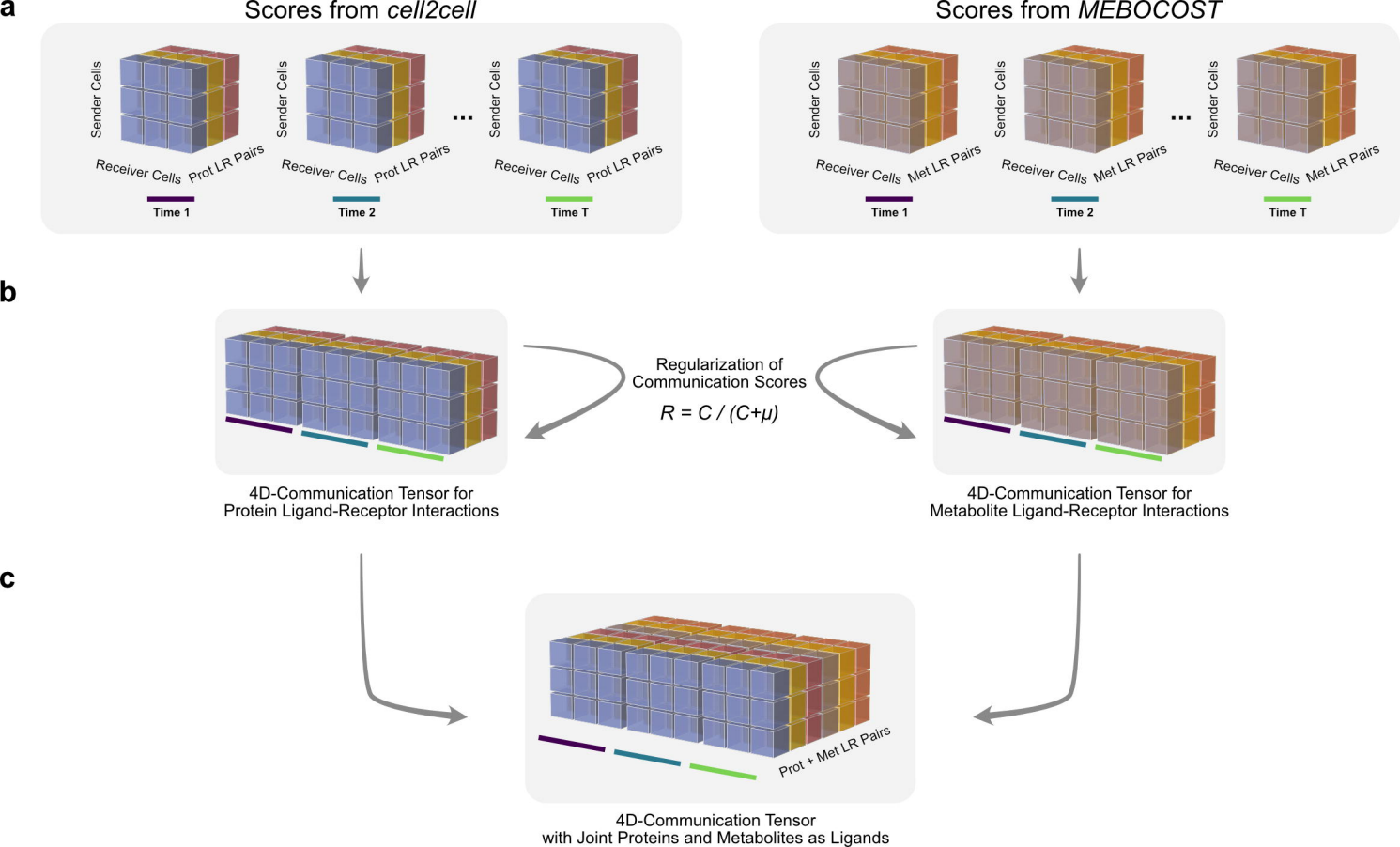
This work was further supported by NIGMS grant R35 GM119850 to NEL, and by the NVIDIA Academic Hardware Grant Program to EA. EA was supported by the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through its scholarship program DOCTORADO BECAS CHILE/2018 - 72190270, the Fulbright Chile Commission, and the Siebel Scholar Foundation. ROL was supported in part by the UCSD Graduate Training Program in Cellular and Molecular Pharmacology through an institutional training grant from the National Institute of General Medical Sciences, T32 GM007752. HMB was supported in part by an appointment to the Research Participation Program at the U.S. Food and Drug Administration administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

## Author contributions

EA and NEL conceived the work. EA, MC, and HMB integrated protein- and metabolite-based communication scores. ROL preprocessed the single-cell RNA-seq data. ROL and EA annotated the single cells based on known markers. EA ran the cell-cell communication analyses, summarized the results, and deposited both the code and data in the Code Ocean capsule. EA wrote the paper and all authors carefully reviewed, discussed and edited the manuscript.

## Competing interests

The authors declare no competing interests.


a
Scores from cell2cell
Sender Cells
Receiver Cells
Prot LR Pairs
Time 1
Time 2
Time T
Scores from MEBOCOST
Met LR Pairs
b
Regularization of
Communication Scores
R = C / (C+μ)
4D-Communication Tensor for
Protein Ligand-Receptor Interactions
4D-Communication Tensor for
Metabolite Ligand-Receptor Interactions
c
Prot + Met LR Pairs
4D-Communication Tensor
with Joint Proteins and Metabolites as Ligands