Causal network analysis of single-cell multiomics data to investigate how injection of induced regulatory B cells can cure multiple sclerosis

Louise Dupuis[1,2], Benoît Manfroi[2], Vincent Cabeli[1], Mohamed El Behi[2], Simon Fillatreau[2,3,4], Hervé Isambert[1]

[1]Laboratoire Physico Chimie Curie, Institut Curie, PSL Research University, CNRS UMR168, Paris, 75005, France; [2]Institut Necker Enfants Malades, Institut National de la Santé et de la Recherche Médicale (INSERM) U1151, Centre National de la Recherche; [3]Université Paris Cité, Faculté de Médecine, Paris, France; [4]AP-HP, Hôpital Necker-Enfants Malades, Paris, France

Abstract

Single-cell analyses have led to many breakthrough discoveries on the identification of celltypes and functional states. However, downstream analyses aiming to infer causal gene regulatory networks driving cell differentiation and cell-cell interactions remain challenging. More generally, causal networks are usually difficult to learn and interpret regardless of the data type, as most algorithms (both constraint-based and score-based approaches) are structurally stringent and data-restrictive. Our team has been developing a novel approach called MIIC (Multivariate Information-based Inductive Causation) that efficiently combines constraint-based and information-theoretic frameworks, which greatly improves the precision of inferred causal networks. Here we showcase MIIC on single-cell transcriptomic data coming from an in-depth study of the induction of regulatory B cells called iBregs to cure central nervous system autoimmunity. We identified that B cells can control autoimmunity through interleukin(IL)-10 production, and prevented the development of a multiple sclerosis mice model called EAE as well as cured recipient mice within a few days upon administration of the iBregs at the peak of clinical signs. However, the cellular interactions of the injected B cells causing the mice to heal remain unknown. The inferred causal network of microglia samples using MIIC uncovered gene regulation activities that will be tested in vitro, in an effort to reconstruct a causal model of the EAE recovery in mice following iBregs injection.

Methods

To reconstruct the MIIC causal single-cell networks, we first preprocessed the transcriptomic datasets obtained from brain and spinal cord EAE mice, D+2 and D+14 after treatment, with classical steps of quality control, normalization, scaling, PCA (Seurat [1]), batch correction (Harmony [2]), UMAP reduction and clustering, resulting in 30000 (resp. 22000) QC-controlled cells for the brain (resp. spinal cord). To extract features of interest from the dataset, numerous downstream analyses were carried out: multiple differentially expressed genes (DEG) analyses using Wilcox tests and Benjamini-Hochberg correction, CellPhoneDB cell-cell interactions analysis [3] to identify combined expression of multi-subunit receptor-ligand complexes, Gene Set Enrichment Analysis (GSEA) analysis with the EnrichR pipeline [4] (focusing on GO Biological Process 2015 database) and NicheNET analysis ([5]) on iBreg-microglia interactions to find top bona fide ligands and their targets.

The MIIC algorithm [6] [7] was performed on microglia cells with the resulting list of genes and Treatment as a contextual variable. MIIC main steps, which combines constraint-based and information-theoretic frameworks, are represented in Figure 1.a : first the removal of dispensable edges based on mutlivariate information scores; second, the orientation of 'V-structure' (with reliable orientations and latent common causes shown as bidirected edges); and optionally the propagation of orientation. Recently, three fundamental and methodological developments were introduced by MIIC, which greatly improve its causal discovery performance in respects to other constraint based methods: the reliability of inferred orientations (based on a general information-theoretic principle), the distinction of "genuine" causes from "putative" and "latent" causal effects and the quantification of indirect effects, while ensuring their consistency and interpretability with the global network structure. These unique enhancements as well as the scalability to very large datasets [8] open up new avenues to discover reliable and interpretable causal gene regulatory networks and single-cell multiomics causal networks. Treatment subnetworks are shown in Figure 1.b but whole networks can be seen on the MIIC server website ([9], [10]):

Brain - MG D+2 Brain - MG D+14 Spinal Cord - MG D+2 Spinal Cord - MG D+14

Results

Single-cell downstream analyses revealed microglia as main candidate targets of iBregs and important actor in this intercellular communication. MIIC reconstructed networks helped visualize the impact of the treatment on microglia: the inferred MIIC network indicated an effect of iBregs (Treatment node) on their antigen-presenting capacity (various MHC-I and MHC-II molecules, *Cd74, Tap1*), lipid metabolism (*Apoe, Apoc1, Apoc2, Apoc4*), lysosome M (*Lyz2*), complement (*C3*), chemokines (*Ccl3, Ccl4*) and Fos-controlled genes (Figure 1.a). The GSEA on different microglial clusters indicated that iBregs also regulated pathways related

to cytokine and inflammation as well as aging and lipid metabolism. Analysis of central nervous system myeloid cells by flow cytometry confirmed MIIC results and showed that iBregs rapidly acted on microglia, reducing their expression of MHC-II as well as CD68 proteins. Furthermore, as found in MIIC networks, microglia from iBregs-treated mice showed restored expression of P2RY12, which is expressed in resting microglia but reduced after their activation [11], while P2RY12 remained low on microglia from untreated mice with chronic disease.

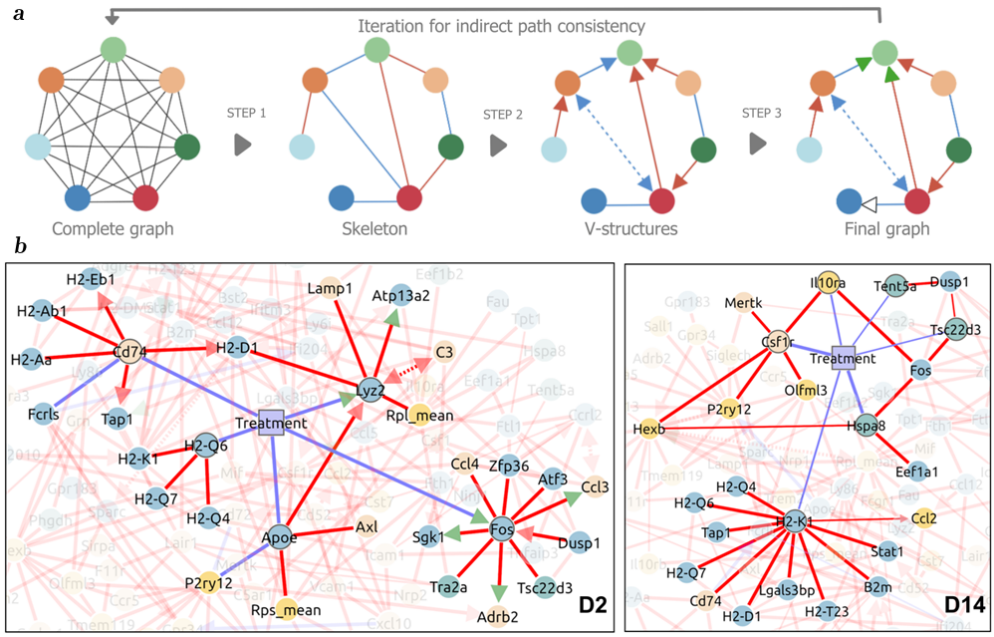

Figure 1: application of MIIC to single-cell multiomics data

Figure 1: application of MIIC to single-cell multiomics data
(a) General scheme of constraint-based methods including MIIC newest advanced features: Step 1, removal of dispensable edges (guaranteeing indirect path consistency); Step 2, 'V-structure' orientation (with reliable orientations and latent common causes shown as bidirected edges); Step 3, optional propagation of orientation shown with white arrowhead (and distinction between 'putative' and 'genuine' causes, green arrowheads). (b) MIIC inferred causal networks on the microglia population at D+2 (left panel) and D+14 (right panel) post iBregs injection. The highlighted "Treatment" variable subnetwork shows inferred causal and correlated effects of the iBregs treatment on microglia genes expression.

References

[1] Yuhan Hao et al. Integrated analysis of multimodal single-cell data. *Cell*, 2021.

[2] Ilya Korsunsky et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods*, 2019.

[3] Mirjana Efremova, Miquel Vento-Tormo, Sarah A Teichmann, and Roser Vento-Tormo. Cellphonedb: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protocols*, 2020.

[4] Zhuorui Xie et al. Gene set knowledge discovery with enrichr. *Current Protocols*, 2021.

[5] Robin Browaeys, Wouter Saelens, and Yvan Saeys. Nichenet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods*, 2020.

[6] Louis Verny, Nadir Sella, Séverine Affeldt, Param Priya Singh, and Hervé Isambert. Learning causal networks with latent variables from multivariate information in genomic data. *PLOS Computational Biology*, 2017.

[7] Vincent Cabeli, Louis Verny, Nadir Sella, Guido Uguzzoni, Marc Verny, and Hervé Isambert. Learning clinical networks from medical records based on information estimates in mixed-type data. *PLOS Computational Biology*, 2020.

[8] Marcel da Câmara Ribeiro-Dantas, Honghao Li, Vincent Cabeli, Louise Dupuis, Franck Simon, Liza Hettal, Anne-Sophie Hamy, and Hervé Isambert. Learning interpretable causal networks from very large datasets, application to 400,000 medical records of breast cancer patients. *preprint*, 2023.

[9] Nadir Sella, Louis Verny, Guido Uguzzoni, Séverine Affeldt, and Hervé Isambert. MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data. *Bioinformatics*, 34(13):2311–2313, 2018.

[10] Nadir Sella, Anne-Sophie Hamy, Vincent Cabeli, Lauren Darrigues, Marick Laé, Fabien Reyal, and Hervé Isambert. Interactive exploration of a global clinical network from a large breast cancer cohort. *NPJ Digit Med*, 2022.

[11] Sharon E Haynes, Gunther Hollopeter, Guang Yang, Dana Kurpius, Michael E Dailey, Wen-Biao Gan, and David Julius. The p2y12 receptor regulates microglial activation by extracellular nucleotides. *Nat. Neurosciences*, 2006.