X

Causal discovery and network inference for biological and biomedical data

Hervé Isambert[0000−0001−9638−8545]

Computational Biology Lab, CNRS-UMR168, Institut Curie, Paris, France
herve.isambert@curie.fr

Discovering causal effects is at the core of scientific investigation but remains challenging when mostly observational data is available. In essence, causal discovery infers cause-effect relations from specific correlation patterns involving at least three variables, which goes beyond the popular notion that pairwise correlation does not imply causation. In principle, causal insights may also help improve the explainability and generalizability of AI methods. Yet, in practice, causal representations have been difficult to learn and interpret, in particular, for high dimensional data such as state-of-the-art biological and biomedical data.

Single-cell transcriptomics and bioimaging techniques produce massive amounts of gene expression and cell imaging data at single cell resolution. However, this wealth of high-dimensional biological data remains largely under-explored due to the lack of unsupervised methods and tools to interpret them without preconceived hypothesis. This highlights the need to develop novel Machine Learning strategies to better exploit the richness and complexity of the information contained in high-dimensional biological and biomedical data.

In this talk, I will outline some network reconstruction methods and applications to a broad range of biological and clinical datasets. In particular, our group has developed a reliable and scalable causal discovery method, MIIC (`https://miic.curie.fr`), to learn cause-effect relationships in a variety of biological or biomedical data, from single-cell transcriptomic data [1,2,3,4] and live-cell imaging data [5] to clinical data from medical records of patients [6,7,8,9]. MIIC combines multivariate information analysis with interpretable graphical models [10,11,12] and outperform other methods on a broad range of benchmark datasets, in particular on complex non-linear, categorical or mixed-type datasets [9,10]. MIIC has also been recently integrated in the CausalXtract tool [5] to analyze live-cell time-lapse images of tumor-on-chip data, as well as in the CausalCCC tool [1] to reconstructs gene interaction pathways across multiple interacting cell types from single-cell or spatial transcriptomic data.

References

1. Dupuis L, Debeaupuis O, Simon F, Isambert H. CausalCCC: a web server to explore intracellular causal pathways enabling cell-cell communication, **Nucleic Acids Res**, gkaf404 (2025).
2. Fusilier Z, Simon F, Calvente I, Crestey L, Clément A, Mathieu M, Jean-Marie R, Piastra-Facon F, Clément JG, Lumineau E, Tonani M, Manriquez V, Lacerda L, de Villemagne P, Piaggio E, Semetey V, Coscoy S, Martini E, Scita G, Gelly J-C, Ivaska J, Isambert H, Goudot C, Pierobon P, Lennon-Duménil A-M, Moreau

HD. Macrophages restrict tumor permissiveness to immune infiltration by controlling local collagen topography through a Tcf4-Collagen3 fibrotic axis, **bioRxiv** 2025.01.17.633527 (2025)

3. Sella N, Verny L, Uguzzoni G, Affeldt S, Isambert H, MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data. **Bioinformatics**, **34**(13):2311-2313 (2018).
4. Verny L, Sella N, Affeldt S, Singh PP, Isambert H, Learning causal networks with latent variables from multivariate information in genomic data. **PLoS Comput Biol**, **13**(10):e1005662 (2017).
5. Simon F, Comes MC, Tocci T, Dupuis L, Cabeli V, Lagrange N, Mencattini A, Parrini MC, Martinelli E, Isambert H. CausalXtract: a flexible pipeline to extract causal effects from live-cell time-lapse imaging data. **eLife**, **13**:RP95485 (2025).
6. Sella N, Guinot F, Lagrange N, Albou L-P, Desponds J, Isambert H. Preserving Information while Respecting Privacy: An Information Theoretic Framework for Synthetic Health Data Generation, **npj Digital Med**, **8**, 49 (2025).
7. Ribeiro-Dantas M, Li H, Cabeli V, Dupuis L, Simon F, Hettal L, Hamy AS, Isambert H, Learning interpretable causal networks from very large datasets, application to 400,000 medical records of breast cancer patients, **iScience**, **27**(5):109736 (2024).
8. Sella N, Hamy AS, Cabeli V, Darrigues L, Laé M, Reyal F, Isambert H, Interactive exploration of a global clinical network from a large breast cancer cohort. **npj Digital Med**, **5**, 113 (2022).
9. Cabeli V, Verny L, Sella N, Uguzzoni G, Verny M, Isambert H, Learning clinical network from medical records based on information estimates in mixed-type data. **PLoS Comput Biol**, **16**(5):e1007866 (2020).
10. Lagrange N, Isambert H. An efficient search-and-score algorithm for ancestral graphs using multivariate information scores for complex non-linear and categorical data, International Conference on Machine Learning (**ICML**) (2025).
11. Cabeli V, Li H, Ribeiro-Dantas M, Simon F, Isambert H, Reliable causal discovery based on mutual information supremum principle for finite dataset in Why21 at Neural Information Processing Systems (**NeurIPS**) (2021).
12. Li H, Cabeli V, Sella N, Isambert H, Constraint-based causal structure learning with consistent separating sets. Advances in Neural Information Processing Systems (**NeurIPS**) **32**, 14257 (2019).