Multi-Omics Data Integration: Advancing Community-Oriented Tools to Meet Tomorrow's Needs

Tiberiu Totu[1-3]
[1]*Molecular Ecology Research Group, Competence Division Method Development and Analytics Unit, Agroscope, Zurich, Switzerland*
[2]*Bioinformatics and Proteogenomics, Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland*
[3]*Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland*
tiberiu.totu@proton.me

With recent advancements in biomedical research, we have reached a turning point where system-level characterization of biological systems has transitioned from being merely desirable to becoming standard practice. In cancer research, tumor profiling is now increasingly conducted alongside microenvironment analysis for a more comprehensive understanding of the system as a whole. Reducing the level of complexity to individual cellular phenotypes, omics profiling, which assesses the abundances of cellular components such as proteins and transcripts, has proven invaluable for characterizing key functions within biological systems. However, due to the complex nature of these systems, analyzing a single omics profile often falls short of capturing the most important molecular traits and interactions[1-3]. The latest breakthroughs in sequencing technologies have made it possible to measure multiple molecular components simultaneously, allowing cells to be viewed as dynamic systems composed of interacting omics layers.

Measuring multiple omics profiles for the same cellular phenotypic state generates an abundance of complementary and unique information about key features and signaling pathways that govern molecular functions, driving the need for various computational tools to integrate multi-omics datasets[4]. During the lecture, we will explore some of these tools that employ either data-driven approaches, relying solely on measured data to infer relationships, or knowledge-based tools, leveraging previously validated interactions to enhance biological interpretability[5]. While data-driven methods can be applied to almost any omics dataset and rely entirely on mathematical frameworks, knowledge-based integrative approaches leverage the continuous expansion of mapped biological interactions, enabling greater interpretability by relying on existing knowledge rather than inferred relationships. However, current knowledge-based multi-omics platforms often employ inference frameworks that are not specifically tailored for biological applications and impose limitations on both the number of datasets that can be integrated and the selection of underlying interaction databases.

Additionally, I will introduce during the talk NOODAI[6], a software platform developed to integrate the outcomes of distinct omics analyses into a joint framework by merging the differentially expressed elements obtained from each omics profile into protein-protein interaction networks that are analysed collectively. The user-friendly webtool allows for the identification of the most important proteins within the joint integrative network by considering a wide range of network centrality metrics. Besides identifying top central proteins, network neighbourhoods are obtained using the MONET[7] tool dependency. The identified clusters are associated with specific signalling pathways that collectively characterize the samples under study, considering simultaneously all input omics profiles. As one of the few tools that allow researchers to jointly analyse and interpret different omics profiles, NOODAI facilitates the identification of robust molecular traits of the systems under study.

1 Manicka, S., Johnson, K., Levin, M. & Murrugarra, D. The nonlinearity of regulation in biological networks. *NPJ Systems Biology and Applications* **9**, 10 (2023).
2 Milzman, J. M. *Dynamics, Networks, and Information: Methods for Nonlinear Interactions in Biological Systems*, University of Maryland, College Park, (2021).
3 Liao, J. C. *et al.* Network component analysis: reconstruction of regulatory signals in biological systems. *Proceedings of the National Academy of Sciences* **100**, 15522-15527 (2003).
4 Joyce, A. R. & Palsson, B. Ø. The model organism as a system: integrating'omics' data sets. *Nature reviews Molecular cell biology* **7**, 198-210 (2006).
5 Wörheide, M. A., Krumsiek, J. & Arnold, M. Multi-omics integration in biomedical research–A metabolomics-centric review. *Analytica chimica acta* **1141**, 144-162 (2021).
6 Totu, T., Tomasoni, M., Bolck, H. A. & Buljan, M. NOODAI: A webserver for network-oriented multi-omics data analysis and integration pipeline. *BioRxiv*, 2024.2011. 2008.622488 (2024).
7 Tomasoni, M. *et al.* MONET: a toolbox integrating top-performing methods for network modularization. *Bioinformatics* **36**, 3920-3921 (2020).