MODIMO: Workshop on Multi-Omics Data Integration for Modelling Biological Systems

Marco Beccuti
beccuti@di.unito.it
University of Turin
Turin, Italy

Vincenzo Bonnici
vincenzo.bonnici@univr.it
University of Verona
Verona, Italy

Rosalba Giugno
rosalba.giugno@univr.it
University of Verona
Verona, Italy

ABSTRACT

Multi-omics analysis aims at extracting previously uncovered biological knowledge by integrating information across multiple single-omic sources. Past approaches have focused on the simultaneous analysis of a small number of omic data sets. Current challenges face the problem of integrating multiple omic sources into a unified complex model, or of combining already available tools for two-by-two omics analyses and merging their outcomes. By doing so and leveraging integrated system-level knowledge, multi-omic approaches ought to enable the development of better qualitative and quantitative models for descriptive and predictive analyses. To move this area forward, new statistical and algorithmic frameworks are needed, for example for generalizing classical graph theory results to heterogeneous networks and applying them to diverse problems such as drug repurposing or understanding the immune response to infections. Thus, in short, this workshop aims at investigating novel methodologies for providing crucial insights into multi-omics data management, integration, and analysis in order to enable biological discoveries.

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences**; • **Computing methodologies** → **Modeling and simulation**; • **Information systems** → **Information integration**.

KEYWORDS

multi-omics, data integration, modelling, biological systems

1 INTRODUCTION

The currently available huge amount of biological data is the primary source of information for systematically investigating biological phenomena. Present computational challenges regard the extraction and management of a secondary level of knowledge that classical single-omics approaches can not provide [2]. Novel and innovative computational approaches are required to increase the number of integrated sources. Such instruments have been used to study and treat cancer, and they will be a key tool in facing the current and the next pandemics [1, 3]. New theoretical approaches and their combination with existing tools are needed for implementing such instruments.

2 OBJECTIVES AND TOPICS OF INTEREST

The main objective of the workshop was to create a link between bioinformatics and CIKM communities. Such a link is necessary for addressing current and future challenges in bioinformatics which require the development of outperforming systems. Therefore, the goal of the workshop was to bring skills that are peculiar to CIKM community to the bioinformatics research, and vice versa, to lay the foundations to make bioinformatics one of the main domains of application for CIKM research. The expected outcome of the workshop was to build interdisciplinary networks of scientists in order to implement innovative methodologies for integrating and analysing multi-omics data at the system level.

The relevant topics included (but were not limited to): mutli-omics data analysis; multi-omics data integration; system biology; computational biology; biological network construction and analysis; biological knowledge graphs; heterogeneous networks; modeling of biochemical pathways and biological networks; big bio- or clinical- data analytics; integration of structured and unstructured resources for biomedical applications; mining multi-relational data; information retrieval from large biomedical data collections.

3 WORKSHOP PROGRAM AND ORGANIZATION

MODIMO takes place as a full-day virtual workshop due to the COVID19 pandemic. The program opens with a keynote talk and continues with the original contributions of research papers. Subsequently, a 1-hour long discussion panel takes place with the aim of discussing the presented contributions from an integrated perspective. Lastly, participants are invited to a 2-hours long group session. One group for each main research topic that emerged from contributions is created by the organizers. The workshop also hosts a meeting of the CINI (Consorzio Interuniversitario Nazionale per l'Informatica) InfoLife Laboratory.

3.1 Keynote Speaker

Patrick Aloy, Institute for Research in Biomedicine, Spain. He obtained his BSc in Biochemistry (1996), MSc in Biotechnology (1998) and PhD in Biochemistry & Molecular Biology (2000) from the

Autonomous University of Barcelona. He then joined the Structural Bioinformatics Group at the European Molecular Biology Laboratory (2001) where he did his postdoctoral training and became Staff Scientist (2003). Since April 2006, Patrick Aloy is an ICREA Research Professor and Principal Investigator of the Structural Bioinformatics Lab in the Institute for Research in Biomedicine (IRB Barcelona). His main scientific interests are in the field of structural bioinformatics, in particular, the use of high-resolution three-dimensional structures to reveal the molecular details of how macromolecular complexes and cell networks operate.

3.2 Accepted Papers

Authors were invited to submit original work in the form of short and full papers, describing research results that are unpublished and not under consideration for publication elsewhere. All paper submissions were reviewed by members of the program committee with respect to scientific quality, novelty, technical quality, potential impact, clarity, and reproducibility. Among submitted works, four long papers (9 pages) and four short contributions (4 pages) were selected for oral presentation.

3.3 Program Committee

We thank the program committee for their on-time and thorough reviews and valuable support: **Patrick Aloy**, Institute for Research in Biomedicine, Spain; **Andrè Altman**, University College London, UK; **Martin Hemberg**, Wellcome Sanger Institute, UK; **Pietro Lio'**, University of Cambridge, UK; **Stefano Lonardi**, University of California, Riverside, USA; **Sara Madeira**, University of Lisbon, Portugal; **Alberto Policriti**, University of Udine, Italy;

3.4 Organizing Committee and Chairs

Pierre Baldi, University of California, Irvine, USA. He earned M.S. degrees in mathematics and psychology from the University of Paris, France, in 1980, and Ph.D. degree in mathematics from Caltech, CA, USA, in 1986. He is currently a Distinguished Professor with the Department of Computer Science, Director with the Institute for Genomics and Bioinformatics, and Associate Director with the Center for Machine Learning and Intelligent Systems at the University of California, Irvine, CA, USA. His research interests include understanding intelligence in brains and machines. He has made several contributions to the theory of deep learning, and developed and applied deep learning methods for problems in the natural sciences. He has written 4 books and over 300 peer-reviewed articles. Dr. Baldi was the recipient of the 1993 Lew Allen Award at JPL, the 2010 E. R. Caianiello Prize for research in machine learning, and a 2014 Google Faculty Research Award.

(OC)**Marco Beccuti**, University of Turin, Italy. He is an Associate Professor at Computer Science Dep. of the Università degli Studi di Torino, where he co-leads "Quantitative Biology" (q-Bio) research group composed of two associated professors, one assistant professor, one postdoctoral researcher, two research fellows, two Ph.D. students. He is also technical coordinator of ELIXIR Node of the Università degli Studi di Torino, scientific coordinator of the laboratory "HPC for biomed and AI" nel "Centro interdisciplinare dell'Università di Torino (IcxT) (https://icxt.di.unito.it/)"

and scientific coordinator of research center "High-Performance Computing for Artificial Intelligence" (HPC4AI) at the University of Turin (https://hpc4ai.unito.it/). His research is currently mainly focused on computational modeling and simulation of complex systems. In particular, he is interested in: Stochastic and hybrid modeling languages; Exact and approximated techniques to analyze the behavior of complex systems; Applications to computational Systems Biology. Moreover, he works on the design of bioinformatics algorithms and workflows for the analysis of deep sequencing data with particular emphasis on reproducibility aspects.

(OC)**Vincenzo Bonnici**, University of Verona, Italy. He is a temporary Assistant Professor in Computer Science at the University of Verona. He received his Ph.D. in 2015 with a thesis in Bioinformatics. He has been Visiting Researcher Scholar at Institute for Genomics and Bioinformatics, University of California, Irvine (USA), and he has done an internship at Fondazione per la Ricerca e la Cura dei Linfomi nel Ticino, Istituto Oncologico della Svizzera Italiana, Bellinzona (CH). His research activities regard the development and application of AI methods in computational genomics, and biological network analysis. The main instruments in supporting his studies are data science, information theory and graph theory. He is a teacher in charge of the Bachelor's degree in Bioinformatics, and a lecturer at the Master's degree in Medical Bioinformatics. He is member of the InfOmics laboratory.

(OC)**Rosalba Giugno**, University of Verona, Italy. She is an Associate Professor In Medical Bioinformatics, Computer Science, Tenure Track, of University of Verona, Italy. She is the referent of the International Master Degree in Medical Bioinformatics, Computer Science Department of University of Verona; director of the Italian National Laboratory of Bioinformatics, coordinating 35 Italian academic groups of Bioinformatics researchers. She is editor for Information Systems Elsevier Journal and editor for Frontiers in Bioinformatics and Computational Biology, and Frontiers in Genetics. Since she moved to Verona in 2016, she leads a research group, named InfOmics lab (https://infomics.github.io/InfOmics/). The strategic rationale of the lab is to bring together all of the investigative strategies developed in Bioinformatics and related computer science research community, going by the algorithms, models, up to the formal systems, to form the bulk required to address critical issues in medical applications with new and efficient computational solutions. In particular, she develops new methods to mining biological networks, integrate heterogeneous data, analyse omics, and classify patients. She uses theory coming from machine learning, data science, mathematics and graph theory.

REFERENCES

[1] Debmalya Barh, Sandeep Tiwari, Marianna E Weener, Vasco Azevedo, Aristóteles Góes-Neto, M Michael Gromiha, and Preetam Ghosh. 2020. Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19. *Computers in Biology and Medicine* 126 (2020), 104051.
[2] Sijia Huang, Kumardeep Chaudhary, and Lana X Garmire. 2017. More is better: recent progress in multi-omics data integration methods. *Frontiers in genetics* 8 (2017), 84.
[3] Indhupriya Subramanian, Srikant Verma, Shiva Kumar, Abhay Jere, and Krishanpal Anamika. 2020. Multi-omics data integration, interpretation, and its application. *Bioinformatics and biology insights* 14 (2020), 1177932219899051.