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

Simone Avesani
simone.avesani@univr.it
University of Verona
Verona, Italy

Vincenzo Bonnici
vincenzo.bonnici@unipr.it
University of Parma
Parma, Italy

Simone Pernice
simone.pernice@unito.it
University of Turin
Turin, Italy

Marco Beccuti*
marco.beccuti@unito.it
University of Turin
Turin, 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 to enable biological discoveries. The workshop will be sponsored by the InfoLife CINI National Laboratory (https://www.consorzio-cini.it/index.php/en/).

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; **Computational biology**; **Systems biology**.

KEYWORDS

bioinformatics, data integration, modelling, multi-omics data analysis

*Both authors contributed equally to this research.

1 A BRIEF INTRODUCTION ON CINI AND ITS INFOLIFE NATIONAL LABORATORY

CINI (National Interuniversity Consortium for Informatics) is the main point of reference for the Italian national academic research in the fields of Computer Science, Computer Engineering, and Information Technology. Established on 6.12.1989, CINI is under the supervision of the competent Italian Ministry for University and Research. It includes public universities, only. According to its by-laws, it is an in–house subject w.r.t. its founders, participants, and legitimately customers. The Consortium is submitted to the periodic Quality Evaluation of its research activities by ANVUR, the Italian National Agency for the Evaluation of the University System and Research. The Consortium involves 1,300+ professors of both Computer Science (Italian SSD INF/01) and Computer Engineering (Italian SSD ING-INF/05), belonging to 49 Italian public universities. The CINI InfoLife laboratory includes 39 groups from Italian Universities for a total of more than 250 professors and researchers plus an equal amount of PhD students and post-docs. Each group is a "node" of the lab. The number of research topics covered by the lab nodes is huge and allows them to provide high-level research contributions in practically any area of the Systems Biology and Bioinformatics domain.

2 THEME AND TOPICS

The theme of the workshop is multi-omics data integration for modelling biological systems. The relevant topics include the following (but 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
- modelling 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 OBJECTIVES, GOALS, AND EXPECTED OUTCOME

The main objective of the workshop is to create a link between bioinformatics and the 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 is to bring skills that are peculiar to the CIKM community to 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 is to build interdisciplinary networks of scientists in order to implement innovative methodologies for integrating and analysing multi-omics data at the system level.

4 AUDIENCE

The target audience is composed of doctoral students, scientists, and academics whose research interests regard the extraction of biological information from multi-omics sources and the management of its integration. The estimated number of attendees is 40.

5 RELEVANCE

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. 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. New theoretical approaches and their combination with existing tools are needed for implementing such instruments.

The workshop covers the following topics of the CIKM conference:

- integration and aggregation
- efficient data processing
- special data processing
- analytics and machine learning
- understanding multi-modal content
- data presentation
- applications

6 RELATED WORKSHOPS

Since the first edition of MODIMO was held in 2021, a few workshops on multi-omics research have been proposed in other venues. Among them, the "Workshop on Standards for Microbiome and Multi'Omics Measurements" hosted by the National Institute of Standards and Technology of the USA (https://www.nist.gov/news-events/events/2022/08/2022-nist-hosted-workshop-standards-microbiome-and-multiomics), and the "Multi-omics in Health and Disease: Current Applications, Challenges, and Future Directions" hosted by the National Human Genome Research Institute (https://www.genome.gov/event-calendar/multi-omics-in-health-and-disease). However, as far as we know, none of the past and current workshops regarding multi-omics research has been held at a conference in the Computer Science field, thus previous workshop focus on the biological results rather than on the details of the computational approaches.

7 PROGRAM FORMAT

The workshop will be opened with a 45-minute-long keynote, followed by questions. Then, the contribution of long papers will be presented. 30 minutes (including questions) for each. After, short paper contributions will be given (15 minutes for each). Subsequently, a 1-hour long discussion panel will be given with the aim of discussing the presented contributions from an integrated perspective. Lastly, participants will be invited to a 2-hours long group session. One group for each main research topic that emerged from contributions will be created by the organizers.

8 KEYNOTE SPEAKER

Professor Elham Azizi, from Columbia University, USA (`https://www.bme.columbia.edu/faculty/elham-azizi`), will be the speaker of the invited talk.

Azizi holds a BSc in Electrical Engineering from Sharif University of Technology (2008), an MSc in Electrical Engineering (2010), and a PhD in Bioinformatics (2014) from Boston University. She was a postdoctoral fellow at Columbia University and Memorial Sloan Kettering Cancer Center (2014-2019). She joined the faculty of Columbia Biomedical Engineering and Irving Institute of Cancer Dynamics in 2020. She is also affiliated with the Department of Computer Science, Data Science Institute, and the Herbert Irving Comprehensive Cancer Center.

Characterizing various interacting cell types in the tumour microenvironment, and unraveling their underlying mechanisms can guide the development of improved and personalized cancer treatments. Azizi's approach involves leveraging genomic profiling at single-cell resolution and developing machine learning and statistical method to analyze and integrate high-dimensional genomic data.

9 SCHEDULE/IMPORTANT DATES

August 18, 2023: Paper submission deadline.
September 15, 2023: Paper acceptance notification.
October 22, 2023: Workshop date.

10 PROGRAM COMMITTEE

The program committee is composed of professors, assistant professors, and researchers who have a strong curriculum in the field of Bioinformatics, especially in the topics that are the focus of this workshop. Moreover, members of the CINI InfoLife laboratory will also participate tot he program committee.

11 PARTICIPATION AND SELECTION PROCESS

Participation in the workshop is open to any student, scientist, and academic. Proposed papers will be evaluated by the PC, and authors of selected papers will be automatically participants in the workshop. At least one author of each paper must participate in the workshop.

Authors are invited to submit original, unpublished research papers that are not being considered for publication in any other forum. Manuscripts should be submitted electronically in PDF format following the conference instructions.

Full papers: Submissions must describe substantial, original, completed, and unpublished work. Full papers may consist of up to eight pages. Full papers will be presented at the workshop.

Short papers: Short paper submissions must describe original and unpublished work. Short papers will be presented at the workshop and will be given four pages in the proceedings.

Selected papers will be invited for a special issue of selected journals in the fields of Computer Science and Bioinformatics.

12 ORGANISERS' BACKGROUND

Simone Avesani. Simone Avesani is a PhD student in Computer Science at the University of Verona, Italy. He graduated in Bioinformatics in 2018 and during his Master's Degree in Medical Bioinformatics, successfully achieved in the autumn of 2020, he started working in the InfOmics Lab lead by Professor Rosalba Giugno. In his PhD. program, Simone Avesani is co-supervised by Professor Rosalba Giugno of the University of Verona and Professor Katja Kanninen of the University of Eastern Finland, and his research is mainly focused on the analysis of multi-omics data, with emphasis on single-cell and spatial transcriptomics analysis. He is currently partner of the European Union's Horizon 2020 project TUBE focused on studying the adverse health effects of pollutants on brain health and their possible involvement in the Alzheimer's disease development.

Marco Beccuti. He is an associate professor at the 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, three assistant professors, one postdoctoral researcher, two research fellows, three PhD students. Moreover, he is the director of the national laboratory CINI InfoLife () which involved more than 39 Italian Universities. He is also the technical coordinator of ELIXIR Node of the Università degli Studi di Torino (link: https://www.elixir-europe.org/), scientific coordinator of the laboratory "HPC for biomed and AI" nel "Centro interdisciplinare dell'Università di Torino (IcxT) (link:https://icxt.di.unito.it/)" and scientific coordinator of the research centre "High-Performance Computing for Artificial Intelligence" (HPC4AI) at the University of Turin (https://hpc4ai.unito.it/). His research is currently mainly focused on computational modelling and simulation of complex systems. In particular, he is interested in (1)Stochastic and hybrid modelling languages; (2)Exact and approximated techniques to analyze the behaviour of complex systems; (3)Applications to computational Systems Biology. Moreover, he works on the design of bioinformatics algorithms and workflows for the analysis of deep

sequencing data (i.e. genomic, transcriptomic and single-cell data) with particular emphasis on reproducibility aspects. He is a co-author of 96 reviewed papers published in the proceedings of the most important international conferences of the field and in highly recognized scientific journals. He has been a member of the program committee of many international conferences and he is on the editorial board of Frontiers in Bioinformatics. E.d.Frontiers. He was successfully involved in several large-scale collaborative research projects. These include the unit leader of the EUMaster4HPC (EC H2020 RIA, EuroHPC-2020-03) project, the WP leader of SUS-MIRRI.IT (PNRR IR - European Commission's NextGenerationEU programme), and PI for two projects supported by the CRT foundation.

Vincenzo Bonnici. Vincenzo Bonnici is a temporary Assistant Professor in Computer Science with tenure track at the University of Parma, Italy. He received his PhD in 2015 with a thesis in Bioinformatics. He has been Visiting Researcher Scholar at the 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 supporting his studies are data science, information theory, and graph theory. He is a teacher in charge of the Bachelor's and Master's degrees in Computer Science at the University of Parma. He is an editorial board member of the journal Healthcare Analytics (Elsevier) and Plos ONE, a reviewer editor for the Frontiers journals in Bioinformatics, Genetics and Applied Mathematics and Statistics, lead guest editor for 4 special issues in MDPI journals, and a reviewer for 26 journals of Computer Science. He is a junior member of the IEEE Technical Committee on Computational Life Sciences (TCCLS), a member of the Bioinformatics Italian Society (BITS) and a co-founder of the InfOmics laboratory (University of Verona, Italy) and the Ahead laboratory (University of Parma, Italy). He was a member of the Program Committee of the WORKSHOP on Challenges and Opportunities in Large Scale Network Analysis in Systems Biology (COLNASB'18). IEEE BIBM 2018, BIOTECHNO 2021, LOD 2021, BIOTECHNO 2022, LOD2023, IEEE ICDH 2023, IEE SERVICES 2023, and ECAI 2023. He has been the organizer and chair of the Annual Meeting of the Italian Bioinformatics Society 2020, IEEE PhD school in Imaging Genetics 2019, and International Proteomics & metabolomics conference and advanced school 2018.

Rosalba Giugno. Rosalba Giugno is an Associate professor In Medical Bioinformatics, Computer Science, and Tenure Track, of the University of Verona, Italy. She is the referent of the International Master Degree in Medical Bioinformatics, Computer Science Department of the University of Verona; and she was the director of the Italian National Laboratory of Bioinformatics, coordinating 35 Italian academic groups of Bioinformatics researchers. She is an editor for Information Systems Elsevier Journal and editor for Frontiers in Bioinformatics, and Frontiers in Genetics. Since she moved to Verona in 2016, she leads a research group, named InfOmics lab (https://infomics.github.io/InfOmics/), of 2 post-docs, 4 PhD students, and 16 undergraduate master thesis students. 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 for mining biological networks, integrating heterogeneous data, analyses omics, and classifying patients. She uses theory coming from machine learning, data science, mathematics, and graph theory. She organizes every year J. T. Schwartz International School for Scientific Research in Bioinformatics (https://www.liparischool.it/), Lipari, Italy.

Simone Pernice. Simone Pernice is currently a research fellow (rtd-A) at the Computer Science Department of the University of Turin. He is a member of "Quantitative Biology" (q-Bio) group, of "Bioinformatics ITalian Society" (BITS), and coordinator of the Young-InfoLife group of the InfoLife National Laboratory of CINI. He is an academic editor for Frontiers in Bioinformatics. In 2016 he received his Master degree in Mathematics from the University of Turin, Italy, and in 2021 his Ph.D. degree in the Ph.D. program of "Computer Science" with a thesis entitled "Modeling and analysis of biological complex systems". His current research interests are in the area of computational modeling and simulation of complex systems. Moreover, he works on developing workflows to analyze omics data, emphasizing reproducibility and FAIR aspects.