

21th International Workshop on Data Mining in Bioinformatics (BIOKDD 2022)

Da Yan
yanda@uab.edu
Department of Computer Science
The University of Alabama at
Birmingham
Birmingham, AL, USA

Catia Pesquita
clpesquita@fc.ul.pt
LASIGE
Faculty of Sciences
The University of Lisbon
Lisbon, Portugal

Carsten Görg
carsten.goerg@cuanschutz.edu
Department of Biostatistics and
Informatics
Colorado School of Public Health
University of Colorado Anschutz
Medical Campus
Aurora, CO, USA

Jake Chen
jakechen@uab.edu
Informatics Institute
The University of Alabama at
Birmingham
Birmingham, AL, USA

Mohammed J. Zaki
zaki@cs.rpi.edu
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA

ABSTRACT

The goal of the 21th International Workshop on Data Mining in Bioinformatics (BIOKDD 2022) is to encourage KDD researchers to solve the numerous problems and challenges in Bioinformatics using Data Mining technologies. Based on the organizers' expertise and communities, BIOKDD 2022 features 2 closely related themes "Biomedical Ontologies" and "Biological Data Visualization". These themes encourage the knowledge graph community and the data visualization community to work together to solve the many challenging problems in Bioinformatics with new perspectives. The key goal is to accelerate the convergence between Data Mining and Bioinformatics communities to expedite discoveries in basic biology, medicine and healthcare.

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences; Bioinformatics**; • **Computing methodologies** → **Artificial intelligence**; • **Information systems** → **Data mining**.

KEYWORDS

Bioinformatics, visualization, ontology, health informatics

ACM Reference Format:

Da Yan, Catia Pesquita, Carsten Görg, Jake Chen, and Mohammed J. Zaki. 2022. 21th International Workshop on Data Mining in Bioinformatics (BIOKDD 2022). In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, August 14–18, 2022, Washington, DC, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3534678.3542888>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD '22, August 14–18, 2022, Washington, DC, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9385-0/22/08.

<https://doi.org/10.1145/3534678.3542888>



Figure 1: BIOKDD 2022 Workshop Logo

1 INTRODUCTION

Bioinformatics is the science of managing, mining, and interpreting information from biological data, and data mining plays an essential role in addressing the fundamental problems in Biomedical Informatics. The annual BIOKDD workshop is an established forum for researchers and practitioners of Biomedical Informatics and Data Mining to explore cutting-edge techniques and to exchange ideas and experiences. BIOKDD has been successfully held in conjunction with SIGKDD for 20 years.

The goal of the 21th International Workshop on Data Mining in Bioinformatics (BIOKDD 2022) is to encourage KDD researchers to work on the numerous problems and challenges in Bioinformatics using Data Mining technologies. Based on the organizers' expertise and communities, BIOKDD 2022 features 2 closely related themes "Biomedical Ontologies" and "Biological Data Visualization". These themes encourage the knowledge graph community and the data visualization community to work together to solve the many challenging problems in Bioinformatics with new perspectives. The key goal is to accelerate the convergence between Data Mining and Bioinformatics communities to expedite discoveries in basic biology, medicine and healthcare. Figure 1 shows the logo of the 21th BIOKDD Workshop (BIOKDD 2022).

2 TOPICS OF INTEREST

Besides workshop papers on original research contributions, we also solicit 1-page abstracts that introduce preliminary research outcomes to increase research impact and to nurture collaborations, and late-breaking research that introduces published work with exceptionally high practical significance. We invite submissions on research contributions in (but not limited to) the areas listed below:

- Semantic web, biomedical ontologies and ontology-driven data integration methods
- Biological network visualization, information visualization and visual analytics for biomedical data
- Deep learning methods for biological and clinical data
- Mining and integrating big biological data
- Discovering biological networks and pathways underlying biological processes and diseases
- Analysis, discovery of biomarkers and mutations, and disease risk assessment
- Comparative genomics
- Metagenome analysis using sequencing data
- RNA-seq and microarray-based gene expression analysis
- Genome-wide analysis of non-coding RNAs
- Genome-wide regulatory motif discovery
- Structural bioinformatics
- Automated annotation of genes and proteins
- Discovery of structural variations from next-generation sequencing (NGS) data
- Correlating NGS with proteomics data analysis
- Discovery of genotype-phenotype associations
- Building predictive models for complex phenotypes
- Functional annotation of genes and proteins
- Cheminformatics
- Special biological data management techniques
- Privacy and security issues in mining genomic databases
- Predictive modeling for personalized treatment
- Text mining for biomedical literature and clinical notes
- Information retrieval for healthcare and biomedical applications
- Biomedical signal analysis and processing
- Intelligent medical data management
- Collaboration technologies for biomedicine
- Social networks for biomedicine
- Bioimage analysis, single-cell analysis

3 WORKSHOP ORGANIZERS

• **Program Chairs.** **Da Yan** is an Assistant Professor at the Department of Computer Science of the University of Alabama at Birmingham (UAB). He was the sole winner of Hong Kong 2015 Young Scientist Award in Physical/Mathematical Science, and he is a senior member of ACM and IEEE. Dr. Yan's research expertise lies in developing scalable systems and algorithms for Big Data analytics. He frequently publishes in conferences such as SIGMOD, VLDB, ICDE, SIGKDD and he also regularly serves in the PC of these conferences, and serves as reviewers of journals such as TODS, VLDBJ, TPDS TKDE. Dr. Yan organizes BOKDD since 2018.

Catia Pesquita is an Assistant Professor at the University of Lisbon, where she leads the Research Line of Excellence in Health

and Biomedical Informatics at LASIGE. She has made significant contributions in data analytics and integration with ontologies and knowledge graphs, producing over 50 peer-reviewed publications in high impact venues including PLoS Computational Biology, BMC Bioinformatics, Journal of Biomedical Semantics and the International Semantic Web Conference. She is an Associate Editor at BMC Bioinformatics, founder and co-organizer of the VOILA workshop at ISWC, and has held Program Chair and Track Chair roles at ISWC and ESWC conferences. Her research team and collaborators develop AML, an award-winning software for ontology matching.

Carsten Görg is an Assistant Professor in the Department of Biostatistics and Informatics at the Colorado School of Public Health. He is a co-Director of the MS and PhD Biostatistics Graduate Programs and also a faculty member in the Computational Bioscience Program at the University of Colorado's School of Medicine. His research interests are focused on data visualization and visual analytics in the biological and biomedical domains. He has published over 50 peer-reviewed articles in journals such as BMC Bioinformatics, IEEE TVCG, and Information Visualization, and also reviewed for these journals. He has served as a PC member of IEEE VIS, EuroVis, BioVis. He served as a Papers co-Chair of ACM SoftVis 2010 and was also the lead organizer of the first Dagstuhl Seminar on Biological Data Visualization in 2012.

• **General Chairs.** **Jake Chen** is the Chief Bioinformatics Officer of Informatics Institute at the University of Alabama at Birmingham and a Professor of Genetics. Dr. Chen is President-elect of the Midsouth Computational Biology and Bioinformatics Society. He has over 25 years of R&D experience in biological data mining and systems biology with >180 peer-reviewed publications and >200 invited talks worldwide on bioinformatics methodologies and biomedical applications. He is an elected fellow of the American College of Medical Informatics (ACMI) and the American Institute of Medical and Biological Engineering (AIMBE).

Mohammed J. Zaki is a Professor of Computer Science at RPI. He is also the associate department head and the graduate program director for the CS department at RPI. He received his Ph.D. degree in computer science from the University of Rochester in 1998. His research interests focus on developing novel data mining and machine learning techniques, especially for applications in text mining, social networks, bioinformatics and personal health. He has over 250 publications (and 6 patents), including the widely used textbook "Data Mining and Machine Learning" (2ed, Cambridge University Press, 2020). He is the founding co-chair for the BOKDD series of workshops. He was a recipient of the NSF CAREER Award and the DOE Early Career Principal Investigator Award. He is a Fellow of the IEEE and a Fellow of the ACM.

• **Program Committee.** We are very grateful to each of our Program Committee members for their reviews and service. Each submission received at least 3–4 reviews.

The PC members of BOKDD 2021 were: Asa Ben-Hur, Debswapna Bhattacharya, Christina Boucher, Serdar Bozdog, Rita Casadio, Zechen Chong, Yuanqi Du, Alexandre P. Francisco, Jingshan Huang, Mehmet Koyutürk, Stefan Kramer, Keren Li, Shengyu Li, T. M. Murali, William Stafford Noble, Tony Pan, Steve Qin, Miguel Rocha, Saeed Salem, Huiyan Sun, Yingfeng Wang, Yuying Xie, Kevin Yip, Wei Zhang, Zhongming Zhao, Jie Zheng, Cuncong Zhong.