#BelBi2026 · Belgrade, Serbia

BOOK OF ABSTRACTS

6th

Belgrade Bioinformatics Conference

International Conference

JUNE, 8-10, 2026

EDITORS
Dr. Valentina Đorđević
Dr. Ivana Morić

belbi.bg.ac.rs

Title 6th Belgrade Bioinformatics Conference
BOOK OF ABSTRACTS

Publisher Institute of Molecular Genetics
and Genetic Engineering,
University of Belgrade
Vojvode Stepe 444a, Belgrade, Serbia
https://www.imgge.bg.ac.rs/

Editors Dr. Valentina Đorđević,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia
Dr. Ivana Morić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Technical editors Dr. Dušan Radojević,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia
Saša Todorović,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

ISBN **978-86-82679-17-2**

BelBi2026 Committees

International Advisory Board

Alessandro Treves,
International School for Advanced Studies,
Trieste, Italy

Elena Fimmel
Mannheim University of Applied Sciences,
Mannheim, Germany

Guanglan Zhang,
Boston University,
Boston, Massachusetts, United Stated of America

Lubomir Lou Chitkushev,
Boston University,
Boston, Massachusetts, United Stated of America

Nedeljko Budisa,
University of Manitoba,
Winnipeg, Manitoba, Canada

Nenad Mitić,
Faculty of Mathematics University of Belgrade,
Belgrade, Serbia

Oxana Galzitskaya,
Institute of Protein Research, Russian Academy of Sciences,
Moscow, Russia

Vladimir Uversky,
University of South Florida,
Tampa, Florida, United States of America

Yuriy Orlov,
I.M. Sechenov First Moscow State Medical University,
Moscow, Russia

Zoran Ognjanović,
Mathematical Institute Serbian Academy of Sciences and Arts,
Belgrade, Serbia

International Program Committee

Aleksandar Krstić,
Systems Biology Ireland, School of Medicine, UCD, Dublin, Ireland

Alexandre de Brevern,
INSERM, Université Paris Cité, Université de la Réunion, Paris, France

Andrea Gelemanović,
School of Medicine, University of Split, Split, Croatia

Biljana Stanković,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

Branko Dragovich,
Mathematical Institute Serbian Academy of Sciences and Arts, Belgrade, Serbia

Branko Tomić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

Fotis Psomopoulos,
Centre for Research and Technology Hellas, Thessaloniki, Greece

Gordana Pavlović-Lazetić,
Faculty of Mathematics, University of Belgrade, Belgrade, Serbia

Ivan Jovanović,
Vinča Institute of Nuclear Sciences University of Belgrade, Belgrade, Serbia

Ivana Morić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

Jelena Begović,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

Jovan Pešović,
Faculty of Biology, University of Belgrade, Belgrade, Serbia

Jovana Kovačević,
Faculty of Mathematics University of Belgrade, Belgrade, Serbia

Milana Grbić,
Faculty of Sciences, University of Banja Luka, Banja Luka, Bosnia and Herzegovina

Mirjana Novković,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

Nataša Pržulj,
Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

Peter Tompa,
VIB Structural Biology Research Center, Brussels, Belgium

Saša Malkov,
Faculty of Mathematics, University of Belgrade, Belgrade, Serbia

Sergei Kozyrev,
Steklov Mathematical Institute RAS, Moscow, Russia

Valentina Đorđević,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

Vladimir Babenko,
Institute of Cytology and Genetics, Novosibirsk, Russia

Submitted abstracts were peer-reviewed by the members of the International Program Committee.

Local Organizing Committee

Aleksandar Bisenić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Aleksandra Uzelac,
Institute for Medical Research, University of Belgrade, Serbia

Aleksandra Vitkovac,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Ana Sarić,
Institute for Biological Research "Siniša Stanković", University of Belgrade, Serbia

Anđela Rodić,
Faculty of Biology,
University of Belgrade, Serbia

Branislava Gemović,
VINCA Institute of Nuclear Sciences, University of Belgrade, Serbia

Dušan Radojević,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Igor Davidović,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Ivana Petrović,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Jelena Pantović,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Jovana Perišić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Luka Bojić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Marija Atanasković,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Marko Antonijević,
Institute for Information Technologies Kragujevac, University of Kragujevac, Serbia

Mateja Ilić,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Miodrag Dragoj,
Institute for Biological Research "Siniša Stanković", University of Belgrade, Serbia

Mirjana Maljković-Ružičić,
Faculty of Mathematics,
University of Belgrade, Serbia

Nikola Kotur,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Pavle Erić,
Institute for Biological Research "Siniša Stanković", University of Belgrade, Serbia

Radmila Novaković,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Saša Todorović,
Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

Stefan Milošević,
BIO4 Campus, Belgrade, Serbia

Željko D. Popović,
Faculty of Sciences,
University of Novi Sad, Serbia

FOREWORD

It is our great pleasure to present the Book of Abstracts for the 6th Belgrade Bioinformatics Conference - BelBi 2026. This year's conference was marked by inspiring research, lively discussions, and a strong collaborative spirit. We extend our heartfelt thanks to all speakers, delegates, and contributors, and proudly present this volume as a comprehensive reflection of the high-quality work presented at BelBi 2026. We are particularly proud that BelBi officially became an ISCB Affiliated Conference this year, representing a major milestone that reflects our growing international standing and commitment to global scientific standards.

Jointly organized by leading research institutions, faculties, and scientific societies from Serbia, BelBi 2026 brought together more than 300 participants from 21 countries. Spanning modern computational biology, structural biology, biomedical, and health informatics, BelBi 2026 highlighted the growing impact of advanced informatics, particularly artificial intelligence and machine learning, on biomedicine and biotechnology. Discussions emphasized that the future of modern biomedicine relies not only on the volume of multi-omics data generated, but on ensuring its ethical, trusted, and secure utilization through validated computational frameworks, privacy-preserving federated learning, and cyberbiosecurity practices.

In addition to scientific sessions, BelBi 2026 introduced an exciting new initiative. We were thrilled to host the first BIO4 AI Hackathon, bringing together talented young researchers and students to solve real-world problems in the fields of biodiversity, biotechnology, and biomedicine at the intersection of bioinformatics and artificial intelligence. We also proudly continued our Future Keynotes Program, providing students of all levels with dedicated access to keynote lectures, educational sessions, and practical workshops to empower the next generation of scientists.

We would like to express our deepest appreciation to all members of the International Advisory Board and the International Program Committee for their rigorous peer-review process. Our sincere thanks also go to all session chairs, workshop instructors, hackathon mentors, volunteers, as well as conference friends and corporate sponsors, whose dedication, commitment, and continued support were instrumental in making BelBi 2026 a resounding success.

Finally, we thank all participants. At its heart, the true success of BelBI lies in the connections made, ideas shared, and friendships built during coffee breaks and social gatherings. We hope this Book of Abstracts serves as a valuable scientific reference and lasting inspiration for your research, and we look forward to welcoming you to future BelBi conferences!

Warm regards,
Belgrade, July 2026

Dr. Valentina Đorđević &
Dr. Ivana Morić

On behalf of BelBi 2026
Organizing Committee

ORGANIZER

Institute of Molecular Genetics and Genetic Engineering, University of Belgrade

MAIN CO-ORGANIZERS

Faculty of Mathematics, University of Belgrade

Mathematical Institute of SASA, Belgrade

Faculty of Biology, University of Belgrade

Vinča Institute of Nuclear Sciences, University of Belgrade

Institute for Biological Research "Siniša Stanković", University of Belgrade

Serbian Society for Bioinformatics and Computational Biology

CO-ORGANIZERS

Institute for Medical Research, University in Belgrade

Faculty of Sciences, University of Novi Sad

Faculty of Engineering, University of Kragujevac

Institute of Information Technology Kragujevac

Centre for the Fourth Industrial Revolution in Serbia

Center for the Promotion of Science

Serbian Society for Molecular Biology

Serbian Society for Microbiology

SUPPORTED BY

Ministry of Science, Technological Development and Innovation Republic of Serbia

BIO4 CAMPUS

CHAMBER OF COMMERCE AND INDUSTRY OF SERBIA

PLATINUM² SPONSOR

COMTRADE SYSTEM INTEGRATION

SILVER SPONSOR

SUPERLAB® Your lab – Our passion

BRONZE SPONSORS

ALFA GENETICS

East Diagnostics

Telekom Srbija

SPONSORS

Labena

BGI

LabCore

tecnalia MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

Alfamed

PROMEDIA Laboratory supply specialists

Altium

novos partnership through science

KEFO

RTC

RNIDS

Contents

Keynote Speakers

Invited Speakers

Oral Presentations

Flash Talks

Poster Presentations

Abstract ID: **2** Type: **Keynote Speaker**

Predicting variant pathogenicity by combining protein language models and biological features

Alexandre G. de Brevern[1]

[1] *INSERM, Université Paris Cité, Université de la Réunion, EFS*

Contact e-mail: **alexandre.debrevern@u-paris.fr**

Predicting the pathogenic impact of missense variants is a cornerstone of genetic disease interpretation and diagnosis. In practice, this is addressed using variant effect predictors (VEPs)—from classical tools such as SIFT to recent deep-learning systems such as AlphaMissense. We recently benchmarked 65 VEPs and identified substantial performance variability and systematic biases, with results strongly shaped by dataset composition, curation, and circularity effects, underscoring the critical importance of rigorous evaluation design (Radjasandirane et al., 2025).

VEP methodology has evolved rapidly, with the most recent approaches relying heavily on deep learning. However, relatively few predictors fully exploit protein language models (PLMs), despite their strong performance across a broad range of protein-centric tasks. To address this gap, we developed PATHOS (Radjasandirane et al., 2026), which integrates embeddings from an optimized pair of PLMs (ESM C 600M and Ankh2 Large) together with complementary biological features, including phylogenetic probabilities, allele frequency, and protein annotations. These inputs are fused through a fully connected architecture to yield pathogenicity predictions. Across clinical benchmarks, PATHOS outperforms competing methods, reaching an MCC of 0.591 on a carefully curated clinical dataset and 0.826 on ClinVar, surpassing leading tools. Case studies on the progesterone receptor and the KCNQ1 ion channel further show that PATHOS can pinpoint functionally critical regions and known pathogenic variants that are missed by other state-of-the-art predictors, including AlphaMissense.

More recently, we developed HECATE (Radjasandirane et al., in preparation), a multi-task deep-learning framework that jointly predicts missense-variant pathogenicity, stability change ($\Delta\Delta G$), and functional impact, aiming to provide more mechanistically informative interpretation. HECATE uses paired wild-type vs mutant embeddings from ESM C 600M and Ankh2 Large, processed through a siamese/shared encoder followed by three task-specific heads. For pathogenicity, the model can additionally incorporate external biological features (e.g., allele frequency and protein/network annotations). We report large-scale application of HECATE, with predictions spanning ~140 million possible amino-acid substitutions across ~17,690 human proteins.

Acknowledgement:

These works were supported by the France 2030 program through the Idex Université Paris Cité (ANR-18-IDEX-0001_GREx).

Keywords:

Diseases, variants, Pathogenicity, ClinVar, AlphaMissense

Abstract ID: **57** Type: **Keynote Speaker**

Harnessing real and artificial intelligence in patient-centric medicine

Igor Jurisica[1,2,3]

[1] *Schroeder Arthritis Institute and Data Science Discovery Centre for Chronic Diseases Krembil Research Institute University Health Network*

[2] *University of Toronto*

[3] *Slovak Academy of Sciences*

Contact e-mail: **juris@ai.utoronto.ca**

Data-driven discoveries are becoming transformative. AI and integrative computational biology help improving treatment of complex diseases. However, clinical applications require deeper understanding of the system and its underlying processes. Systematic integration of multi-omics datasets with rich biological networks does help to build explainable models, which in turn address questions of trust, fairness, empathy, and governance. Without proper management of such complex biological and clinical data, AI algorithms cannot be effectively trained, validated and successfully applied.

Despite challenges, there are many opportunities in precision medicine using AI, big data analytics and integrative computational biology workflows. From systematic data curation and analysis to improved biomarkers, drug mechanism of action, and patient selection, such analyses enable patient-centric medicine.

Mapping diverse clinical, multi-omic and multi-spectral data to diverse biological networks, further annotating with ontologies and atlases requires spectrum of algorithms from data mining, machine learning, graph theory and advanced visualization to aid analysis, model building and interpretation. Intertwining computational prediction and modeling with biological experiments and preclinical studies will lead to more useful findings faster and more economically.

Acknowledgement:

This work was supported in part by funding from Natural Sciences Research Council (NSERC RGPIN-2024-04314), CIHR (#519474), Canada Foundation for Innovation (CFI #225404, #30865), Ontario Research Fund (RDI #34876, RE010-020), Krembil Foundation and Ian Lawson van Toch Fund.

Keywords:

Data science, AI, network biology

Abstract ID: **89** Type: **Keynote Speaker**

Integrating strain-level metagenomics in human and food microbiomes

Edoardo Pasolli[1]

[1] *University of Naples Federico II*

Contact e-mail: **edoardo.pasolli@unina.it**

Strain-level metagenomics is transforming our understanding of microbial diversity across human-associated and food-related ecosystems, with major implications for health, food safety, and microbial ecology. This talk highlights recent advances in computational and analytical approaches for large-scale strain-resolved metagenomics in both human and food microbiomes.
Focus is given to machine learning strategies developed for metagenomic applications, including emerging foundation models for microbiome analysis. The presentation will also introduce key resources supporting cross-study integration, such as harmonized reference catalogs, standardized metadata frameworks, and scalable pipelines for both read-based strain profiling and metagenome-assembled genome reconstruction. Efforts aimed at discovering and systematically characterizing underexplored microbial species and strains will also be discussed, with emphasis on improving reproducibility and comparability across datasets and environments.
Examples of integrative analyses connecting human and food microbiomes will illustrate microbial transmission routes, shared reservoirs, strain tracking, and diet-associated microbial exchange. The session will conclude by addressing current limitations and future priorities toward the development of robust, standardized, and generalizable frameworks for strain-resolved microbiome research.

Keywords:

Metagenomics; large-scale analysis; human-food microbiome

Abstract ID: **100** Type: **Keynote Speaker**

Bioinformatics and proteomic studies of biomacromolecules

Oxana Galzitskaya[1]

[1] *Gamaleya Research Centre of Epidemiology and Microbiology*

Contact e-mail: **ogalzit@vega.protres.ru**

Studying the fundamental principles of protein self-organization has enabled us to address important applied problems, such as developing new antimicrobial peptides, determining the composition of amyloid deposits for patients, and participating in a project to develop cancer and HIV vaccines.
Based on the co-aggregation model, new antibacterial peptides have been created against the pathogens Pseudomonas aeruginosa and Staphylococcus aureus. The innovative peptide should interact with the target protein of the model or pathogenic bacteria, forming aggregates and removing this protein, which is essential for the life of the bacteria, from its active state. Synergistic effects have been demonstrated for several peptides.
We discovered a new mutation in lysozyme amyloidosis with the amino acid substitution p.F21L/T88N, found in a Russian family, and correct typing of the amyloidosis saved the lives of two patients.
We studied the influence of physicochemical properties (hydrophobicity, net charge, aliphatic index, instability index, isoelectric point, molecular weight, etc.) on the non-covalent interaction of human epitopes with polymorphic proteins of the major histocompatibility complex class I (MHC I). The structural characteristics of this interaction are important in studying the immunogenicity of epitopes, as well as in the development of multi-epitope peptide vaccines. We developed an AI-based method for determining the binding of potential neoepitopes to a given MHC I complex and assessing its immunogenicity.

Keywords:

amyloids, antimicrobial peptide, neoantigen, vaccine

Abstract ID: **13** Type: **Invited Speaker**

Dissecting the transcriptomic basis of behavioural differences between male morphs of ruffs (Calidris pugnax)

Vladimir Jovanović[1,2], Alex Zemella[3], Jasmine Loveland[3,4,5], Katja Nowick[1], Clemens Küpper[3]

[1] *Institute for Biology, Freie Universität Berlin, Berlin, Germany*

[2] *Bioinformatics Solution Center, Freie Universität Berlin, Berlin, Germany*

[3] *Max Planck Institute for Biological Intelligence, Seewiesen, Germany*

[4] *Department of Cognitive and Behavioral Biology, University of Vienna, Vienna, Austria*

[5] *Messerli Research Institute, University of Veterinary Medicine, Vienna, Austria*

Contact e-mail: **vladimir.jovanovic@fu-berlin.de**

Alternative reproductive tactics in the ruff (*Calidris pugnax*) are orchestrated by an autosomal inversion that determines three distinct male morphs: Independents, Satellites, and Faeders. These morphs exhibit nearly discrete differences in size, plumage, aggression, and circulating androgen levels. To resolve the molecular mechanisms driving these complex phenotypes, we employed an integrative bioinformatics framework combining multi-tissue transcriptomics with structural and functional modeling. We performed RNA-seq on eleven tissues, including steroidogenic glands and nine brain nuclei comprising the social behaviour network and mesolimbic reward pathways. Differential gene expression analysis revealed a significant enrichment of morph-differentiating genes within the inverted supergene region. To investigate cis-regulatory effects, we quantified allele-specific expression from RNA-seq data, identifying widespread allelic imbalance in most inversion-linked protein-coding genes across all examined brain areas.

We identified HSD17B2, encoded within the inversion, as the primary driver of morph-specific androgen variation. This enzyme oxidizes testosterone into the less potent androstenedione. HSD17B2 is significantly upregulated in the blood, hypothalamus, and several brain nuclei of Satellites and Faeders, while its expression is nearly absent in Independent blood. This allows the inversion morphs to rapidly metabolize testosterone in the periphery before it reaches the brain.

Using a branch model selection test on related Charadriiform species, we also demonstrated that HSD17B2 is under strong positive selection in the inversion haplotypes. In silico homology modeling and computer simulations of ruff homodimers further revealed that derived isozymes possess higher binding affinities for testosterone and NAD compared to the ancestral variant.

We expected HSD17B2 to likely have a strong pleiotropic effect on the neural expression of several genes across different brain areas, through changes in the circulating sex steroid hormone levels. Therefore, we integrated results from various transcriptomic analyses to characterise the morph-specific neural molecular signatures underlying stereotypical mating tactics of male morphs. Overall, this study demonstrates how structural and regulatory innovations in a single gene can induce modest changes in expression across hundreds of genes in different brain areas and result in a morph-specific mating behaviour.

Keywords:

Behavioural genomics, intraspecific selection, supergene

Abstract ID: **14** Type: **Invited Speaker**

Identifying regulatory RNA networks behind therapy-induced senescence (TIS) in breast cancer cells

Éva Schád[1], Beáta Szabó[1], Eszter Bajtai[1], András Füredi[1,2,3]

[1] *HUN-REN Research Centre for Natural Sciences, Institute of Molecular Life Sciences*

[2] *HUN-REN Centre of Energy Research, Institute of Technical Physics and Materials Science*

[3] *Obuda University, University Research and Innovation Center, Physiological Controls Research Center*

Contact e-mail: **tantos.agnes@ttk.hu**

Therapy-induced senescence (TIS) is considered a permanent cell cycle arrest following DNA-damaging treatments; however, its irreversibility has been challenged in our recent publication, where we demonstrated that escape from TIS is universal across breast cancer cells. Moreover, TIS provides a reversible drug resistance mechanism that ensures survival of the population, and could contribute to relapse.
The experimental setup was to induce TIS in four different breast cancer cell lines with high-dose chemotherapy and culture the cells until they escape TIS. Parental (CTR), TIS and repopulating (REPOP) cells were analysed by bulk and single-cell RNA sequencing and surface proteomics. Cells in the TIS state showed a remarkable resistance to a broad panel of anticancer agents and different senolytic drugs. Our results revealed that TIS, as a transient drug resistance mechanism, could contribute to overcome the immune response and to relapse by reverting to a proliferative stage.
In order to better understand the molecular mechanisms driving TIS induction and escape, we performed a deep analysis of the transcriptomic changes that accompany these processes, with specific focus on the non-coding RNAs. Our main goal in the present work was to identify regulatory RNA interactions that orchestrate large-scale transcriptomic deregulation of genes necessary for the survival and later repopulation of cancer cells undergoing TIS.
Regulatory RNA networks consist of lncRNA sponges that bind a large amount of miRNAs, releasing the miRNA-target mRNAs from translational repression. Using the transcriptomic data from the bulk RNA sequencing of the different states, we reconstructed the regulatory RNA networks that exist in CTR, TIS and REPOP cells. These revealed cell type-specific and generally existing networks and showed a strong upregulation in the TIS state. Comparison of the different networks revealed general, but also specific RNA networks that are exclusively present in the TIS cells. The lncRNAs with the highest connectivity in these networks may serve as new target candidates in a possible TIS-specific treatment. Importantly we also identified RNA networks that remain active in the REPOP cells –these may be used as TIS history markers and may facilitate the escape from TIS during a repeated treatment cycle.

Acknowledgement:

The project received funding from National Research, Development and Innovation Office (ADVANCED 152119, STARTING 149496 and 2023-1.2.4-TÉT-2023-00107), the European Commission's Marie Skłodowska-Curie Actions (Grant agreement ID: 101065044), the Research Grant Hungary (RGH- 151536).

Keywords:

ceRNA networks, TIS, breast cancer

Abstract ID: **16** Type: **Invited Speaker**

AI-driven bioinformatics: advancing chronic disease management and aging assessment

Ming Chen[1]

[1] *College of Life Sciences, Zhejiang University, Hangzhou, China*

Contact e-mail: **mchen@zju.edu.cn**

Chronic diseases and aging have become a global public health burden, demanding innovative approaches to improve early assessment, decode hallmark mechanisms, and advance personalized treatment. With the explosion of multi-omics data and the advancement of artificial intelligence (AI) technologies, bioinformatics engineering has emerged as a transformative force bridging biomedical discoveries and clinical applications for chronic diseases.

This talk introduces innovations in AI-driven biomedical and bioinformatics engineering for chronic disease management. We developed BioAgeX and HALDxAI, which integrate multi-omics data and clinical records to enable early risk stratification, accurate diagnosis of biological age, and even mortality risk prediction, overcoming the limitations of traditional single-dimensional assessment. We propose AI-driven knowledge graph approaches, including decoding disease-related molecular regulatory networks, identifying novel biomarkers, providing intervention treatment responses for chronic conditions, and optimizing precision therapeutic regimens. Furthermore, we discuss challenges and future directions in this interdisciplinary field, such as data integration, model interpretability, and the translation of AI-driven discoveries.

Keywords:

aging, AI, integration, knowledge graph

Abstract ID: **19** Type: **Invited Speaker**

Zero-shot prediction of thermodynamic properties of proteins

Gábor Erdős[1], Zsuzsanna Dosztányi[2]

[1] *ELTE Protein Dynamics Research Group*

[2] *MTA-ELTE Momentum Bioinformatis Research Group*

Contact e-mail: **gabor.erdos@ttk.elte.hu**

The thermodynamic properties of proteins are fundamental to understanding their function, dysfunction, and evolution. However, experimental characterization of these properties, especially for intrinsically disordered proteins (IDPs) that exist as dynamic conformational ensembles, remains a significant challenge. Computational methods have emerged as a powerful alternative, yet they often require extensive training on protein-specific data, limiting their ability to generalize.

Here, we present a novel transformer based message parsing graph neural network (trMPNN) architecture for the zero-shot prediction of protein thermodynamic properties. By representing proteins as graphs our model learns the underlying physicochemical principles governing protein thermodynamics while retaining speed that allows for the analysis of complete proteomes. The network was trained by maximizing the probability of the native structure against a set of decoys, guided by the Boltzmann distribution, allowing it to learn a transferable energy function. This approach enables accurate predictions on proteins not seen during training, overcoming a major limitation of previous methods.

We demonstrate the power of our network highlighting two key areas. First, we show its ability to predict ensemble-averaged thermodynamic properties of IDPs, providing insights into their unique conformational landscapes. Second, we showcase its accuracy in predicting absolute protein stability (ΔG values), a critical factor in protein engineering and disease pathogenesis. Our model achieves state-of-the-art performance in both tasks, with predictions in excellent agreement with experimental data.

The zero-shot capability of our GNN opens up exciting avenues for high-throughput screening of protein stability, the design of novel proteins with desired thermodynamic properties, and a deeper understanding of the complex interplay between sequence, structure, and thermodynamics in the proteome.

Keywords:

Statistical mechanics, Protein structures

Abstract ID: **24** Type: **Invited Speaker**

Nextflow and nf-core: a virtual infrastructure for the next generation of bioinformatics pipelines

Cedric Notredame[1]

[1] *Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, United Arab Emirates*

Contact e-mail: **cedric.notredame@mbzuai.ac.ae**

Biology generates unprecedented volumes of data, pushing computational infrastructures beyond present capacities. Turning this raw material into knowledge requires workflows that are reproducible, portable, and community-owned. Nextflow, together with the nf-core initiative, has become the de facto framework for this purpose, providing a virtual infrastructure on which thousands of groups now build, share, and run their analyses.

In this talk, I will first retrace how Nextflow and nf-core evolved into a cooperation platform for bioinformatics, as described in our recent Genome Biology paper (Langer et al., 2025). The strength of the framework lies less in its workflow engine than in the community it has brokered: a distributed network of scientists coordinating their efforts through shared tooling, common standards, and joint governance. A central development has been the move to DSL2 and the emergence of an extensive library of modules and subworkflows. This granular layer has fundamentally changed how cooperation takes place —contributors no longer exchange full workflows but reusable building blocks, which are independently tested, versioned, and maintained across institutions. The result is a progressive convergence toward FAIR workflows, adopted by consortia as diverse as EuroFAANG, Darwin Tree of Life, and Genomics England.

This modular foundation naturally called for the next step: a hub. Pipeline hubs turn workflows into versioned, peer-reviewed scientific objects, but their defining contribution is to make alternative ways of doing the same thing directly confrontable —running side by side, on the same data, under the same conditions. I will illustrate the concept with nf-core/multiplesequencealign (Santus et al., 2025), which systematically confronts alignment strategies against structural and evolutionary references.

Pipeline hubs are also, in my view, the natural substrate on which AI will operate in biology. Because these workflows are standardized, ontology-ready, and continuously benchmarked, they can be coordinated, extended, and eventually co-developed with AI agents —opening the way to pipelines adaptable to shifting scientific questions and self-aware enough to follow targeted objectives semi-autonomously.

Acknowledgement:

The author thanks the nf-core community for its collective contribution to the infrastructure described in this talk, and MBZUAI for supporting his participation in BelBi.

Keywords:

Nextflow, nf-core, MSA, FAIR

Abstract ID: **26** Type: **Invited Speaker**

Rethinking protein complex discovery from network topology to dynamic graph learning

Milana Grbić[1]

[1] *Faculty of Natural Science and Mathematics, University of Banja Luka*

Contact e-mail: **milana.grbic@pmf.unibl.org**

Protein complexes play a central role in cellular processes, yet their reliable identification from protein–protein interaction (PPI) networks remains challenging due to incompleteness and noise in experimentally derived interaction data. Traditional approaches primarily rely on static network representations and topology-based assumptions, which often fail to fully capture the dynamic nature of protein interactions. More recently, increasing attention has been directed toward dynamic PPI networks, as they are expected to provide richer and more informative representations of protein interactions.

The evolution of computational approaches for protein complex identification reflects a shift from classical network analysis to modern representation learning techniques. Structural limitations of PPI networks were investigated by examining the extent to which known protein complexes are supported as connected subgraphs. To address missing interactions, a variable neighborhood search (VNS) metaheuristic was applied to minimally augment networks and improve complex connectivity. The obtained results demonstrate that widely used PPI networks and complex standards exhibit incomplete support, indicating systematic gaps in interaction data. This observation is further supported by the tendency of community detection methods, which typically identify dense structures in networks, to yield inconsistent results when applied to protein complex identification.

In addition to topology-based approaches, machine learning methods were examined for predicting protein membership in complexes using protein-specific biological and sequence-derived features. While these methods improved predictive performance, their effectiveness remained dependent on the underlying network representation. To address this limitation, more recent approaches have focused on graph representation learning, with dynamic graph embedding techniques applied to model temporal aspects of protein interaction networks. By combining embedding-based representations with clustering methods, improved identification of protein complexes has been observed in dynamic settings compared to static ones.

Overall, the results indicate that the transition from static PPI networks toward learned and dynamic graph representations significantly enhances the ability to capture biologically meaningful protein complexes. This progression also suggests promising directions for integrating representation learning with biological prior knowledge in future computational models.

Keywords:

protein–protein-interaction-networks, protein-complexes, graph-representation learning, dynamic-networks

Abstract ID: **28** Type: **Invited Speaker**

Towards causal and interpretable virtual cells

Mikel Hernaez[1]

[1] *CIMA University of Navarra*

Contact e-mail: **mhernaez@unav.es**

Recent advances in single-cell and perturbational genomics have created new opportunities to model cellular behavior under genetic and disease-associated perturbations. However, most representation learning approaches remain difficult to interpret biologically and often capture statistical associations rather than mechanistic (causal) structure. Here, we develop interpretable machine learning frameworks for modeling transcriptional dysregulation through biologically grounded and causal representation learning of cell states.

We first introduce NetActivity, an interpretable autoencoder that incorporates prior biological knowledge into its architecture to infer robust gene-set activity scores from transcriptomic data. By constraining latent representations around biological processes, NetActivity enables pathway-level interpretation of disease-associated transcriptional programs. We then extend this framework to single-cell perturbation data through SENA-discrepancy-VAE, a biologically informed causal representation-learning model that integrates Perturb-seq interventions with pathway-aware latent factors. This approach aims to recover causal pathway archetypes that explain how genetic perturbations propagate through transcriptional programs and reshape cellular phenotypes.

Together, these methods provide a step toward interpretable virtual cells: predictive and perturbable computational models that connect gene expression, biological processes, and causal mechanisms of disease.

Keywords:

virtual cells, Causality, perturb-seq, single-cell

Abstract ID: **35** Type: **Invited Speaker**

Noninvasive biomarker discovery by metabolomics of fingertip sweat to guide individualized anti-PD-1 immunotherapy in melanoma

Verena Paulitschke[1]

[1] *Department of Dermatology, Medical University Vienna*

Contact e-mail: **verena.paulitschke@meduniwien.ac.at**

Reliable biomarkers to predict or monitor response to anti-PD-1 therapy in metastatic melanoma remain unavailable. Noninvasive analysis of fingertip sweat has recently been proposed as a promising method to detect inflammatory signatures and tumor-related metabolic changes.

Finger sweat is collected at baseline, 3 weeks, and 3 months after initiation of therapy, alongside clinical data. In a preliminary study, patients treated with 480 mg nivolumab provided samples immediately before and after therapy. Sweat was analyzed by liquid chromatography–mass spectrometry (Orbitrap Exploris 480), followed by bioinformatic and statistical processing.

In prior work, we identified a serum signature of 10 proteins (CRP, LYVE1, SAA2, C1RL, CFHR3, LBP, LDHB, S100A8, S100A9, SAA1) associated with poor response to anti-PD-1 therapy. Extending this concept, we now apply fingertip sweat metabolomics as a noninvasive biomarker strategy. Preliminary analyses revealed elevated kynurenine (tryptophan metabolism), p-cresol sulfate (microbial dysbiosis), and markers of mitochondrial stress in melanoma patients'sweat samples. These metabolites may serve as pharmacodynamic indicators during immunotherapy. A larger cohort is currently being evaluated at the predefined time points to validate these findings and explore additional metabolites. Plasma proteomics within our longitudinal cohort revealed a distinct regulation of immune-related proteins between responders and non-responders after three months of therapy. Responders were characterized by a significant upregulation of immunoglobulins and the Golgi-associated protein GLIPR2. Conversely, non-responders exhibited elevated levels of SPINK1, a marker closely associated with epithelial-mesenchymal transition (EMT) and therapeutic resistance. Protein set enrichment analysis (PSEA) further confirmed a selective enrichment of immunological functions exclusively in responders, indicating effective systemic immune activation. Integrating these proteomic findings with fingertip sweat and plasma metabolomics will provide a comprehensive, multi-omic perspective on treatment efficacy.

This study seeks to establish fingertip sweat profiling as an innovative, noninvasive platform for biomarker discovery in melanoma immunotherapy. Capturing metabolic processes at the patient level may enable early identification of non-responders, improve risk stratification, and provide mechanistic insights into resistance, all without imposing significant procedural burden.

Keywords:

melanoma, immune therapy, biomarker

Abstract ID: **45** Type: **Invited Speaker**

From fatty liver to scar: Exploring the rewiring of gene networks across the human chronic liver disease continuum.

Thomas Mohr[1]

[1] *Institute of Analytical Chemistry, University of Vienna*

Contact e-mail: **thomas.mohr@univie.ac.at**

In 2024, Gribben et al published a paper, "Acquisition of epithelial plasticity in human chronic liver disease" . This paper includes a single-nucleus RNASeq experiment spanning all stages of chronic liver disease, from healthy tissue to Metabolically Dysfunctional-Associated Steatotic Liver Disease (MASLD), Metabolic Dysfunction-Associated Steatohepatitis (MASH), liver cirrhosis, and end-stage liver cirrhosis. Samples from at least 4 patients represent each stage; MASH encompasses samples from 27 patients. The size of the dataset (>80.000 cells, with ~20.000 genes sequenced) poses special challenges but also enables the application of state-of-the-art network-based reengineering of gene regulatory networks.
In this talk, we will explore workflows available for these data and examine co-expression network-based determination of gene modules and structural equation model-based pathway perturbation analysis. This talk will give a brief overview of available methods, their strengths and limitations, and will present initial results on gene modules associated with disease progression, hub genes and pathways that change during chronic liver disease progression.

Acknowledgement:

We want to thank Christopher Gerner and Samuel Meyer-Menches (both University of Vienna) for valuable input and Vorenica Moreno-Viedma (Medical University of Vienna) for pointing out the dataset,

Keywords:

MASH, MASLD, SEM, WGCNA

Abstract ID: **46** Type: **Invited Speaker**

Exploiting the linearity of joint embedding spaces to mine cancer precision medicine knowledge

Nataša Pržulj[1], Noël Malod-Dognin[1]

[1] *Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates*

Contact e-mail: **noel.malod@mbzuai.ac.ae**

Cancer is a leading cause of death worldwide. It is a multifactorial disease that results from multiple alterations, which are only partially captured by any biological data type studied in isolation (e.g., somatic mutation profiles or dysregulated genes). Hence, novel multi-omics data-fusion and analytics methods are needed to holistically mine the wealth of all available clinical and molecular data for new cancer precision medicine discovery.
Network embedding is a cornerstone of modelling and analysis of such complex, biological datasets. In the field of NLP, it was observed that the embeddings of words capture semantic relationships linearly, allowing for efficient mining using simple linear vector operations rather than by using computationally intensive, black-box downstream analysis methods. Since then, this observation has been made in other fields, yielding the so-called linear representation hypothesis in vision language models. The question is whether this observation holds and can be exploited in the context of multi-omics data-fusion for cancer precision medicine.
To assess this, we defined a Non-Negative Matrix Tri-Factorization based framework to jointly embed the multi-omics datasets of all cancer patients from TCGA (somatic mutation profiles and gene expression profiles of 9,134 cancer patients from 33 TCGA projects), the protein interactions from BioGRID and the drug data from DrugBank into the same embedding spaces. By comparing the performances of linear and non-linear methods for downstream classification and clustering tasks, we demonstrate that our joint embedding spaces are indeed linearly organized. Then, we define linear vector operations that enable prioritizing new pan-cancer or cancer-specific genes and drug repurposings, paving the way to new cancer therapies.

Keywords:

AI/ML, Multi-omics Data-Fusion, Precision Medicine

Abstract ID: **52** Type: **Invited Speaker**

AI/ML in the search for novel antivirals

Alexey Kuzikov[1], Anastassia Rudik[1], Dmitry Filimonov[1], Dmitry Karasev[1], Leonid Stolbov[1], Nadezhda Biziukova[1], Olga Tarasova[1], Tatiana Filippova[1], Victoria Shumyantseva[1], Vladimir Poroikov[1]

[1] *Institute of Biomedical Chemistry, Russian Academy of Medical Sciences (RAMS)*

Contact e-mail: **olga.a.tarasova@gmail.com**

The risk of interspecies virus transmission, the emergence of poorly understood pathogens, and the rising prevalence of multidrug-resistant viral variants lead to the necessity to discover novel, more safe and effective antiviral agents. Our objective is to develop an integrated AI/ML computational framework and curated data resource that can be helpful in the development of new antiviral agents.

The developed framework comprises the Antivir database, specifically constructed for this purpose, together with a computational module for predicting antiviral activity, off-target interactions, and adverse drug reactions (ADRs). The database integrates heterogeneous data on antiviral compounds, standardized according to the nomenclature of the International Committee on Taxonomy of Viruses (ICTV). The web resource also includes modules for off-target prediction, ADR evaluation, and ternary classification models for HIV drug resistance; all estimations are based on the Naïve Bayesian approach implemented in the PASS algorithm.

The Antivir database currently contains data on approximately 76,000 compounds with biological activity profiles against more than 20 viruses, integrating relationships among compounds, diseases, resistance-associated mutations, and gene polymorphisms. Predictive models are available for multiple viruses, including influenza, dengue, West Nile, hepatitis B and C, HIV-1, SARS-CoV, and SARS-CoV-2. The ROC AUC values characterizing predictive performance range from 0.86 to 0.99 for antiviral activities and reach 0.94 for adverse drug reactions (ADRs). For HIV drug resistance prediction, ternary classification models demonstrate average performance metrics of 0.913 sensitivity, 0.894 specificity, 0.741 precision, and 0.953 ROC AUC. For SARS-CoV-2 main protease (Mpro) inhibitory activity, the ROC AUC is approximately 0.93. To identify novel SARS-CoV-2 Mpro inhibitors, we conducted virtual screening of 1.6 million compounds from the ChemRar library. A hybrid ligand- and structure-based approach was applied, combining PASS models, pharmacophore mapping, and molecular docking. Based on this screening, 32 compounds were prioritized for experimental validation; four have already been tested, and two exhibited activities in the micromolar to submicromolar range.

The data and AI/ML methods integrated into the Antivir web-resource (https://way2drug.com/viruses/), which provides a robust environment for antiviral research. It combines high-quality, curated data with advanced machine learning models to accelerate the discovery of safe and effective therapeutic agents.

Acknowledgement:

The study is supported by the Program of Fundamental Scientific Research in the Russian Federation for the long-term period (2021–2030) (No. 124050800018-9).

Keywords:

antivirals, AI, database/web-resource, structure-activity relationship

Abstract ID: **56** Type: **Invited Speaker**

Context-dependent remodeling of aging-related and neurodevelopmental pathways in human cerebral organoids across APOE genotypes

Minja Belic[1], Joanna Bons[1], Sydney Alderfer[1], Mikayla Hady[1], Tyne L. M. McHugh[1], Kizito-Tshitoko Tshilenge[1], Maria A. Sanchez[1], Kevin Schneider[1], Long Wu[1], Nikola Markov[1], Birgit Schilling[1,2], David Furman[1], Lisa M. Ellerby[1,2]

[1] *Buck Institute for Research on Aging*

[2] *University of Southern California, Leonard Davis School of Gerontology*

Contact e-mail: **mbelic@buckinstitute.org**

Apolipoprotein E (*APOE*) genotype is a major determinant of late-onset Alzheimer's disease risk and a modulator of neural aging, yet the effects of *APOE* isoforms on developmental trajectories in human neural tissue remain incompletely understood. Here, we combined single-cell RNA sequencing with quantitative proteomics in induced pluripotent stem cell-derived isogenic human cerebral organoids carrying *APOE ε2/ε2, ε3/ε3*, or *ε4/ε4* alleles to characterize genotype- and stress-dependent remodeling across cellular, transcriptional, and secretory axes.

We profiled ~46,000 cells across control and irradiation-induced senescence conditions using single-cell RNA-seq. Raw counts were denoised with CellBender to remove ambient-RNA contamination, and scVI was applied to learn a latent embedding while adjusting for stress-related and technical covariates before clustering and pseudotime trajectory analysis. Cells were annotated using snapseed with marker panels adapted from the Human Neural Organoid Cell Atlas (HNOCA), and cell-type composition differences were tested with scCODA.

Cell-type composition differed significantly across *APOE* genotypes at baseline, with *APOE4* organoids showing a higher relative abundance of glial and excitatory neuronal populations. After adjusting for these compositional differences, residual genotype-dependent differences in differentiation trajectories persisted, indicating that *APOE4* affects both the cell types and their progression through development. Following irradiation, organoids exhibited genotype-dependent transcriptional responses and changes in inferred regulatory activity beyond the baseline compositional differences.

In parallel, data-independent acquisition mass spectrometry quantified >6,500 proteins across organoid lysates and secretomes under basal and senescence conditions. Single-sample gene-set enrichment analysis (ssGSEA) showed *APOE* isoforms differentially regulated aging-associated pathways, with *APOE4* exhibiting increased senescence signatures and an inflammatory senescence-associated secretory phenotype (SASP), while *APOE2* showed reduced senescence and a SASP enriched in antioxidant enzymes and chaperones.

Together, these results suggest that *APOE4*-associated risk biology already emerges in developmental tissue composition and stress-response programs prior to aging-related insults.

Acknowledgement:

This work is supported by the NCRR shared instrumentation grant 1S10 OD028654 (PI: Birgit Schilling) for the Orbitrap Eclipse Tribrid system and by the NIH grant P01AG066591 (PI: Lisa M. Ellerby). T.L.M.M was supported by the NIA grant T32 AG052374. MS and LW were supported by T32-AG000266 (PI: Lisa M. Ellerby).

Keywords:

scRNA, proteomics, cerebral_organoids, APOE, senescence

Abstract ID: **65** Type: **Invited Speaker**

The role of local energetic frustration across conformational ensembles of disordered proteins

Franco Simonetti[1], Edgar Chacon[1], Maria Freiberger[2], Leandro Radusky[3], Alexander Monzon[4], Diego Ferreiro[5], R. Gonzalo Parra[1]

[1] *Barcelona Supercomputing Center*

[2] *Institut de Biologie Paris Seine, Sorbonne Université*

[3] *Mimark Diagnostics*

[4] *Universidad de Padova*

[5] *Universidad de Buenos Aires*

Contact e-mail: **parra.gonzalo@gmail.com**

Globular protein domains satisfy the principle of minimal frustration, and thus their folding landscapes are strongly funneled towards their native state with residual energetic conflicts often associated with function. In contrast, intrinsically disordered proteins populate heterogeneous conformational ensembles as a consequence of competing interactions and therefore flatter energy landscapes. Still, the specific energetic signatures that lead to protein disorder are poorly understood. Here we characterized the local frustration patterns of 3,176 proteins and 59,627 conformers spanning different flavours of intrinsic disorder. We found that in the case of direct residue-residue interactions, except for a minor contribution of specific interaction types, disordered residues are not excessively highly frustrated; instead, they are enriched in neutrally frustrated interactions and a reduced prevalence of minimally frustrated ones. Furthermore, we found that the most distinguishable feature between ordered and disordered residues comes from long range interactions where disordered residues are depleted of stabilizing interactions and enriched in energetic conflicts. When disordered regions are analysed in complex with known protein partners, their energetic profiles resemble those of ordered residues. We conclude that intrinsic disorder is not a consequence of excessive energetic conflicts as believed for the case of statistical random heteropolymers. Instead, disordered proteins exist at the edge of foldability, where a small number of stabilizing interactions with other proteins or substrates can shift their equilibrium towards more structured, functional conformations.

Keywords:

protein folding, function, disorder

Abstract ID: **72** Type: **Invited Speaker**

Predictive modeling of B- and T-cell epitopes in allergen discovery and immunotherapy design

Marija Gavrović-Jankulović[1]

[1] *University of Belgrade –Faculty of Chemistry*

Contact e-mail: **mgavrov@chem.bg.ac.rs**

Predictive modeling of immune epitopes has become an important component of computational allergology, enabling systematic analysis of allergenicity and cross-reactivity of proteins. This work presents an integrated in silico framework for the identification of B- and T-cell epitopes in model of food and inhalant allergenic proteins, combining sequence homology analysis, motif-based approaches, and machine learning–driven predictors. Linear B-cell epitopes are identified using sequence-derived features and propensity scores. In parallel, T-cell epitopes are prioritized based on predicted binding to major histocompatibility complex (MHC) class II molecules, incorporating peptide binding affinity and population coverage analyses.

In this study key focus is on the identification of conserved epitope patterns across homologous proteins, enabling the detection of cross-reactive allergens and panallergen families. Such approaches facilitate the interpretation of IgE cross-reactivity observed between phylogenetically related and unrelated allergen sources. The predictive approach further supports the rational design of hypoallergenic variants by selectively modifying IgE-binding regions while preserving T-cell immunogenicity. In addition, large-scale screening of protein datasets enables the identification of novel allergen candidates based on epitope signatures associated with sensitization risk.

Integration with experimental IgE-binding and immunological validation enhances the robustness and translational relevance of these predictive approaches.

Acknowledgement:

MGJ acknowledge financial support to the Ministry of Science, Technological Development and Innovation (contract numbers: 451-03-136/2025-03/200168, 337-00-216/2023-05/133), the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101072377.

Keywords:

computational allergology, protein allergenicity prediction

Abstract ID: **76** Type: **Invited Speaker**

Genetic and phenotypic heterogeneity of type 2 diabetes across Russian ancestry groups

Ekaterina Markelova[1], Yulia Medvedeva[2]

[1] *Research Center of Biotechnology, Institute of Bioengineering, Russian Academy of Sciences, Moscow, Russia*

[2] *MBZUAI*

Contact e-mail: **ju.medvedeva@gmail.com**

Type 2 diabetes (T2D) is a highly polygenic disease involving multiple pathways. Genetic ancestry influences the predominant mechanisms driving T2D, as evidenced by population-specific risk alleles and metabolic adaptations. However, research remains heavily biased toward Western populations, leaving a profound knowledge gap regarding the immune, genetic, and epigenetic architecture of T2D in most ancestries worldwide, with diverse Russian ethnic groups—Turkic, North Caucasian, and East Siberian—remaining absent from multi-omic profiling despite the central role of inflammation and cluster-specific mechanisms in T2D pathogenesis. Understanding how genetic background shapes T2D risk is crucial for personalized prevention and treatment. To directly address this gap, we personally collected and generated a unified multi-omic dataset comprising genetics, scRNA-seq, and scATAC-seq on PBMCs from individuals with T2D and matched controls across multiple understudied Russian ancestries, including Chechens, Tatars, and Yakuts. We analyzed ancestry-specific T2D mechanisms by integrating genetic data with cluster-specific polygenic scores and single-cell chromatin accessibility and transcriptional profiles to assess T2D-associated genetic clusters across ancestry groups. This unique, first-hand unified dataset enabled high-resolution, multi-omic characterization of ancestry-associated immune states unobtainable from any published resource, mapping population-specific transcriptional signatures, epigenetic landscapes, and cell-type-specific perturbations. Furthermore, cluster-specific polygenic scores varied significantly between populations. Yakuts exhibited higher scores for β-cell dysfunction, hyperinsulin secretion, and lipid metabolism alterations, whereas Chechens and Tatars had higher scores for obesity-related mechanisms. Our findings provide the first description of genetic, transcriptional, and epigenetic immune diversity in ancestrally distinct Russian populations, establishing a critical multi-omic reference for global diabetes research. Predominant T2D mechanisms differ across populations, with some groups showing greater susceptibility to β-cell dysfunction while others show obesity-driven pathways. These ancestry-specific differences should guide public health recommendations and personalized medicine, underscoring the necessity of ancestry-aware, primary multi-omic data collection.

Keywords:

T2D, multi-omics, PRS, personalized medicine

Abstract ID: **79** Type: **Invited Speaker**

A universal toolkit for single-cell and spatial long-read transcriptomic sequencing data

Andrey Przhibelskiy[1], Lieke Michielsen[2], Careen Foord[2], Iman Hajirasouliha[3], Hagen U. Tilgner[2], Alexandru I. Tomescu[4]

[1] *University of Helsinki*

[2] *Feil Family Brain and Mind Research Institute, Weill Cornell Medicine*

[3] *Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine of Cornell University*

[4] *Department of Computer Science, University of Helsinki,*

Contact e-mail: **andrey.przhibelskiy@helsinki.fi**

Recent advances in single-cell and spatial transcriptomics combined with long-read sequencing have opened new possibilities for isoform-level profiling and studying alternative splicing. However, upstream processing of such data poses distinct algorithmic challenges, including detecting barcodes in error-prone reads using extremely large whitelists, resolving molecule concatenation artifacts and PCR duplicates, accurately quantifying and discovering complex isoforms, as well as a lack of methods for detecting spatially variable isoforms.

Here we present Spl-IsoQuant —a universal toolkit for processing long-read RNA-seq data obtained with various single-cell and spatial protocols. Our software extends the IsoQuant framework with high-precision barcode calling that also handles cDNA concatenation typical for some protocols and ONT sequencing. Spl-IsoQuant performs UMI-based PCR deduplication and accurately quantifies genes and isoforms in individual cells, cell types, and user-defined conditions. It supports multiple protocols, such as 10x single-cell, 10x Visium and Visium HD, Stereo-seq and Curio. Moreover, it can process long-read data from any custom protocol via a user-specified molecule description format. We complement the toolkit with Spl-IsoFind, which discovers spatially variable isoforms using Moran's I spatial autocorrelation. Combined with region-wise isoform tests, this framework detects both known patterns and signals not aligned with predefined anatomical regions.

Applying our toolkit to several mouse brain spatial long-read Stereo-seq and 10x Visium HD datasets with 500nm vs 2µm resolution respectively, we demonstrate spatially variable isoform discovery at single-cell level and reproducibility across protocols, establishing a general-purpose pipeline for single-cell and spatial long-read RNA-seq analysis.

Keywords:

transcriptomics, long reads, single-cell, spatial

Abstract ID: **80** Type: **Invited Speaker**

Unraveling the molecular effects of nanoplastics on human blood cells using microfluidic chip-based single-cell RNA sequencing

Vladimir Kovacevic[1]

[1] *School of Electrical Engineering, University of Belgrade*

Contact e-mail: **vladimir.kovacevic@etf.bg.ac.rs**

The widespread use of plastics in various industries has led to the accumulation of its micro and nano particles in the environment, raising concerns about their potential impact on human cellular processes and health. In particular, the interaction of nanoplastics with biological systems remains largely unexplored. In the conducted study, we exposed human peripheral blood samples to carboxylated polystyrene nanoparticles of different particle sizes and performed single-cell RNA sequencing to examine their effects on gene expression profiles. By utilising microfluidic chip-based exposure platform with physiologically relevant cell culture, controlled exposure dynamics, and preservation of native cell–cell interactions we achieved closer approximation to human in vivo conditions. Our results reveal distinct gene expression patterns and cellular responses to nanoplastic exposure: monocytes and CD4+ T cells show particularly high numbers of differentially expressed genes with a strong bias toward downregulation (3.3-5.7 times more than upregulated), and smaller nanoparticles elicit broader transcriptional changes across all immune cell types. While monocytes undergo size-dependent, pathway-coherent state remodeling, B cells and $CD4^+$ T cells display distributed, lineage-preserving transcriptional tuning without discrete state transitions. These findings could contribute to better understanding of potential risks to human health and well-being.

Keywords:

Nanoplastics, single-cell RNA sequencing

Abstract ID: **81** Type: **Invited Speaker**

Integration and standardization of gene–disease associations in ProDiGenIDB

Jovana Kovacevic[1]

[1] *Faculty of Mathematics University of Belgrade, Belgrade, Serbia*

Contact e-mail: **jovana.kovacevic@matf.bg.ac.rs**

Comprehensive analysis of gene–disease associations is fundamental for biomedical and translational research; however, relevant data remain dispersed across multiple databases, often using heterogeneous formats and inconsistent disease terminology. To address these challenges, we developed ProDiGenIDB (*Pro*tein *Di*sease *Gen*e *I*ntegrated *D*ata*b*ase), an integrated resource designed to harmonize gene–disease relationships from diverse, publicly available databases while providing consistent gene and disease annotations.

ProDiGenIDB integrates more than 400,000 gene–disease associations collected from established sources including DisGeNet, COSMIC, HumsaVar, Orphanet, ClinVar, HPO, and DISEASES. For each association, the database provides standardized gene identifiers (Gene Symbol, Entrez ID, UniProt ID, Ensembl ID), curated disease descriptors, Disease Ontology identifiers (DOIDs), and references to the original data sources, enabling traceability and interoperability.

A key component of the database construction is the normalization and harmonization of disease terminology. To achieve robust mapping of disease names to standardized ontology terms, we employed Natural Language Processing (NLP) techniques based on advanced text representation models. This approach improves consistency across heterogeneous datasets and enhances the reliability of downstream analyses.

By unifying gene–disease associations and applying ontology-based standardization supported by NLP methods, ProDiGenIDB provides a scalable and extensible platform for integrative studies. The database facilitates systematic exploration of disease mechanisms and supports data-driven research in genomics, disease classification, and precision medicine.

Keywords:

gene–disease associations, data harmonization

Abstract ID: **85** Type: **Invited Speaker**

Multi-omics integration approaches for addressing cardio-renal-metabolic multimorbidity

Andrea Gelemanović[1]

[1] *Mediterranean Institute for Life Sciences (MedILS), University of Split*

Contact e-mail: **agelemanovic@medils.unist.hr**

Chronic non-communicable diseases (NCDs), such as cardiovascular disease, chronic kidney disease, type 2 diabetes, and metabolic syndrome, are the leading causes of mortality and disability worldwide. These conditions are frequently observed as interconnected multimorbidity, reflecting shared molecular, physiological, and environmental determinants.

Here I am introducing MultiOM4CRM, a systems medicine initiative designed to unravel the shared biological architecture underlying cardio-renal-metabolic (CRM) multimorbidity. This project integrates multi-omics data with environmental, lifestyle, and clinical factors to move beyond disease-centric models toward a network-based understanding of chronic diseases and their progression. Advanced bioinformatic approaches, including multi-omics factor analysis, Mendelian randomization, and multimodal machine learning, are applied to identify shared biological pathways and regulatory mechanisms behind CRM. Recent insights into type 2 diabetes and metabolic syndrome emphasize the importance of pleiotropy, shared genetic architecture, gene-environment interactions, and chronic inflammation as central drivers of multimorbidity. Ongoing work is directed toward the development of predictive machine learning models for the identification of clinical and lifestyle signatures driving cardio-renal-metabolic multimorbidity.

Overall, MultiOM4CRM aims to shift the focus from individual diseases to shared biological mechanisms, enabling improved risk stratification and a more unified understanding of chronic disease progression.

Acknowledgement:

This work is supported by the Croatian Science Foundation under the project MultiOM4CRM (UIP-2025-02-9908).

Keywords:

multimorbidity, multi-omics integration, personalised medicine

Abstract ID: **86** Type: **Invited Speaker**

Unlocking complex diseases through federated data governance, FAIR multi-modal data, and AI

Venkata Satagopam[1]

[1] *University of Luxembourg*

Contact e-mail: **venkata.satagopam@uni.lu**

Neurodegenerative diseases (e.g., Parkinson's and Alzheimer's) and immune-mediated diseases (e.g., Inflammatory Bowel Disease, Multiple Sclerosis, and Rheumatoid Arthritis) are highly complex in their etiology, presenting major challenges for early diagnosis, patient stratification, and the discovery of robust biomarkers. Addressing these challenges requires access to high-quality, longitudinal health data and well-characterised clinical cohorts —yet such data are typically distributed across hospitals and research centres in different countries, under heterogeneous legal, ethical, and technical constraints.

This talk will present federated approaches to data governance, management, and analysis, and how translational medicine informatics enables the integration of multi-modal data —clinical records, multi-omics, imaging, and sensor/mobile data —while keeping sensitive data at its source. A central theme will be the substantial effort required for data curation, harmonisation, and FAIRification (making data Findable, Accessible, Interoperable, and Reusable) to enable cross-cohort and cross-border analysis.

I will showcase how statistical and Machine Learning (ML) methods applied to such multi-layered data can stratify patients by disease severity and progression, identify biomarkers, and underpin clinical decision support models. Complementing these data-driven approaches, disease maps will be discussed as a means to uncover the underlying mechanistic models and co-morbidities of these complex diseases.

Together, these efforts illustrate how federated, FAIR, and AI-ready ecosystems —aligned with European research infrastructures and regulatory frameworks —can accelerate translational research and improve patient outcomes in complex disease areas.

Keywords:

Federated data analysis, FAIR data

Abstract ID: **106** Type: **Invited Speaker**

Gut microbiota profiling as a tool for predicting and enhancing immunotherapy response

Dušan Radojević[1], Marina Bekić[2], Sergej Tomić[2], Jelena Đokić[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Institute for the Application of Nuclear Energy, University of Belgrade, Serbia*

Contact e-mail: **dusan.radojevic@imgge.bg.ac.rs**

Immune-mediated diseases, including cancer and autoimmune disorders, represent a growing global health burden. Although immune checkpoint inhibitors and immunogenic cell-based therapies achieved substantial success in oncology, immunosuppressive cell therapies are increasingly investigated for autoimmune diseases. Accumulating evidence identified the gut microbiota as an important factor influencing immunotherapy efficacy and immune-related adverse effects. We therefore investigated how gut microbiota profiles correlate with cell-based immunotherapies in two models.

In healthy human donors, we profiled fecal microbial communities and identified taxa associated with distinct dendritic cell (DC) phenotypes. Donors with higher microbial diversity and abundant short-chain-fatty-acid-producing bacteria generated monocyte-derived DCs with a regulatory phenotype characterized by high ILT-3 expression, low CD1a expression, and weaker maturation potential. In contrast, donors enriched in Bifidobacterium and Collinsella produced DCs with increased CD1a expression, upregulated co-stimulatory molecules (CD40, CD86), and enhanced pro-inflammatory cytokine responses, including TNF-α, IL-6, IL-8, and IL-12. These findings indicated that specific gut microbiota configurations may promote differentiation of highly immunogenic DCs and potentially improve the efficacy of DC-based anticancer vaccines. We further evaluated microbiota-associated effects of immunosuppressive therapy in a rat model of experimental autoimmune encephalomyelitis (EAE). Untreated EAE animals developed gut-barrier disruption, reduced microbial diversity, and pronounced alterations in microbial community structure. In contrast, treatment with bone-marrow-derived myeloid-derived suppressor cells (MDSCs), particularly following PGE_2 conditioning, preserved microbial diversity and maintained a microbiota profile associated with immunosuppressive metabolic functions. These changes were accompanied by expansion of IL-10-producing regulatory T and B cell populations and suppression of pathogenic Th1/Th17 immune responses, resulting in attenuation of disease severity.

Together, these findings demonstrated that specific gut microbiota configurations were closely associated with immune-cell immunogenicity and therapeutic responsiveness. The results highlighted the translational potential of microbiota-informed strategies and suggested that interventions based on probiotics, dietary fibres, and faecal microbiota transplantation could improve the efficacy of anticancer and anti-inflammatory immunotherapies.

Keywords:

/

Acknowledgement:

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, and by the Science Fund of the Republic of Serbia, PROMIS, #6062673, Nano-MDSC-Thera project.

Abstract ID: **119** Type: **Invited Speaker**

Resolving variants of uncertain significance at scale with calibrated computational and experimental evidence

Predrag Radivojac[1]

[1] *Khoury College of Computer Sciences, Northeastern University, Boston MA, USA*

Contact e-mail: **predrag@northeastern.edu**

Resolving variants of uncertain significance (VUS) is a key challenge in genomic medicine, with about most of the variants in ClinVar classified as VUS. This limits the clinical utility of genetic testing for patients with suspected Mendelian disease and is a hindrance to improving human health and well-being. Two major sources of evidence hold promise for resolving VUS at scale: (1) variant effect predictors, which generate computational scores for any variant, and (2) high-throughput experimental assays, which directly measure the impact of variants on gene function. Realizing the clinical potential of these evidence types requires systematic data generation, rigorous calibration, and data sharing. We will mostly focus on new calibration strategies for computational and experimental evidence, and demonstrate their combined power to dramatically reduce the burden of VUS.

Towards the goal of reducing VUS, we generated experimental measurements for over 60,000 variants across ten genes using multiplexed assays and curated about 200,000 existing variant effect measurements across 30 genes from the literature. Using the new calibration framework, we developed a scalable classification workflow relying solely on experimental and predictive evidence that enables reclassification of about 75% VUS across these genes as pathogenic or benign, with error below 1%. The framework was also applied prospectively. Among more than 90,000 variants not yet observed in clinical databases, 62% carried sufficient evidence to be "preclassified" before ever appearing in a patient record. This approach establishes a blueprint for how coordinated data generation, calibration, and open dissemination can transform genomic medicine.

Keywords:

calibration strategies, uncertain significance variants

Abstract ID: **120** Type: **Invited Speaker**

Monitoring heart health at home: Health informatics challenges

Vladimir Brusic[1]

[1] *Metropolitan College, Boston University, Boston MA, USA*

Contact e-mail: **vbrusic@bu.edu**

This study examines the key informatics challenges involved in developing and deploying home-based heart-health monitoring systems. These systems use wearable sensors for continuous heart-rate measurement. Although heart rate is the most accessible and widely collected physiological signal, using it to assess cardiovascular status requires advanced engineering and informatics methods. The monitoring systems require reliable sensors and sensor networks, as well as solutions for signal processing, heart-rate-variability modeling, and anomaly detection.

The first challenge is the technical complexity and reliability of such systems, which must ensure high-quality data collection, secure data transmission, and home-based data governance that guarantees trustworthiness, safety, security, access control, decentralized protocols, and privacy-preserving analytics. The second challenge is medical usefulness: home-based heart-monitoring systems must produce digital biomarkers that are accurate, reliable, clinically interpretable, and validated across populations—meeting the same standards of safety, performance, and reproducibility required for medical-device certification. The third challenge is that system design must meet ethical and social requirements, ensuring autonomy, transparent data governance, and fair access. Adoption depends on public trust, digital literacy, and clinician confidence in home-generated biomarkers. Affordability, reimbursement, regulatory alignment, and workflow integration ultimately determine whether such systems achieve broad, routine clinical use.

Smart Health Home uses continuous heart-rate monitoring to detect early hypertension, arrhythmias, sleep disorders and apnea, postoperative complications, and emerging cardiac deterioration. Its performance is enhanced through federated learning, which improves models without sharing personal data, and through community-level monitoring that aggregates privacy-preserving trends to support broader public-health insight.

Successful implementation of home-based heart-health monitoring requires the integration of biomedical engineering, informatics architectures, ethical frameworks, and clinical validation, with particular emphasis on reliability, interoperability, and data protection in real-world conditions.

Keywords:

Heart-rate monitoring, Digital biomarkers

Abstract ID: **122** Type: **Invited Speaker**

Cyberbiosecurity threats in the time of AI: Digital vulnerabilities with biological consequences

Lubomir Lou Chitkushev[1]

[1] *Boston University, Boston, USA*

Contact e-mail: **ltc@bu.edu**

Cyberbiosecurity threats are rapidly evolving at the intersection of artificial intelligence, biotechnology, and digital infrastructure. As AI systems become increasingly embedded in biological research, healthcare, agriculture, and bio-manufacturing, they introduce new classes of vulnerabilities that extend beyond data breaches to real-world biological consequences. This talk examines how AI-enabled tools—ranging from generative models for protein design to automated laboratory platforms—expand the attack surface for malicious actors, enabling risks such as the manipulation of genomic data, the corruption of biofoundry workflows, and the design or synthesis of harmful biological agents.

We explore emerging threat scenarios, including adversarial attacks on biological datasets, exploitation of cloud-connected lab equipment, and the misuse of AI models trained on sensitive biological information. Particular attention is given to the convergence of cyber and physical systems, where compromised digital inputs can directly alter experimental outcomes or production pipelines. The talk also assesses current gaps in governance, security practices, and cross-disciplinary awareness that leave critical systems exposed.

Finally, we outline a framework for mitigating cyberbiosecurity risks in the age of AI, emphasizing secure-by-design principles, robust access controls, model auditing, and collaboration between cybersecurity experts and life scientists. By highlighting both technical and policy dimensions, this talk aims to inform a proactive approach to safeguarding biological systems against digitally mediated threats.

Keywords:

Cyberbiosecurity, mitigating cyberbiosecurity risk

Abstract ID: **6** Type: **Oral Presentation**

Metaplasia as a transitional cell-of-origin state revealed by epigenome-informed cancer prediction

Mohamad D. Bairakdar[1,2,3], Wooseung Lee[1,2,3], Bruno Giotti[1,2,3], Akhil Kumar[1,2,3], Paula Stancl[4], Elvin Wagenblast[3], Dolores Hambardzumyan[3], Paz Polak[5], Rosa Karlic[4], Alexander M. Tsankov[1,2,3]

[1] *Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai (ISMMS), New York, NY, USA*

[2] *Lipschultz Precision Immunology Institute, ISMMS, New York, NY, USA*

[3] *Tisch Cancer Institute, ISMMS, New York, NY, USA*

[4] *Bioinformatics Group, Division of Molecular Biology, Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia*

[5] *Haystack Oncology, Quest Diagnostics, Baltimore, MD, USA*

Contact e-mail: **pstancl@bioinfo.hr**

Understanding the cellular origin of cancer is essential for elucidating tumorigenesis and improving early detection and therapeutic strategies. Here, we leverage large-scale whole-genome sequencing and single-cell chromatin accessibility data to develop a machine learning framework for predicting the cell of origin (COO) across multiple cancer types. By integrating mutational landscapes with cell-type-specific epigenomic profiles, our approach accurately assigns tumors to their originating cellular populations at high resolution, demonstrating that somatic mutation patterns retain a record of the pre-malignant chromatin state.

Our model reveals both expected and previously unrecognized cellular origins, including a predominant basal cell origin for small cell lung cancer, challenging traditional assumptions of a neuroendocrine origin. Importantly, beyond static COO prediction, we identify dynamic cellular trajectories during tumorigenesis, highlighting the presence of intermediate metaplastic states in several gastrointestinal cancers. These findings support a model in which metaplasia represents a transitional and potentially permissive state linking normal tissue and malignant transformation, rather than a terminal differentiation endpoint.

This work demonstrates that integrative, epigenome-informed machine learning enables robust prediction of cancer cell of origin while uncovering metaplastic intermediates as critical components of tumor evolution. These insights have important implications for understanding cancer initiation, refining early detection strategies, and targeting pre-malignant cellular states.

Acknowledgement:

This project was funded by the Chan Zuckerberg Initiative (CZI) Data Insights Grant 2022-249299 to P.P., R.K., and A.M.T. This work was also supported in part by the Croatian National Science Foundation Project PREDI-COO (Project number: HRZZ IP-2019-04-9308) to P.S. and R.K and through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences.

Keywords:

cell-of-origin, single-cell, cancer, genomics, metaplasia

Abstract ID: **7** Type: **Oral Presentation**

EffRank: a bioinformatics scoring pipeline for prioritizing novel type III secreted effectors in the Pseudomonas syringae species complex

Iva Rosić[1,2], Marina Sokić[1,2], Tamara Ranković[2,3], Olja Medić[2,3], Tanja Berić[1,2,3], Slaviša Stanković[2,3], Ivan Nikolić[2,3]

[1] *Institute of Physics Belgrade*

[2] *Center for Pathogen Biocontrol and Plant Growth Promotion - Faculty of Biology, University of Belgrade*

[3] *Faculty of Biology, University of Belgrade*

Contact e-mail: **riva@ipb.ac.rs**

The identification of novel type III secreted effectors (T3SEs) remains challenging because these proteins exhibit strong sequence diversity and frequently lack the canonical sequence signatures used by existing prediction algorithms. As a consequence, individual machine-learning or motif-based predictors often generate large sets of candidate effectors that include numerous false positives, making downstream experimental validation inefficient. To address this limitation, we developed an integrated bioinformatics workflow designed to improve confidence in the in silico discovery of novel T3SEs and applied it to the *Pseudomonas syringae* species complex.

The workflow combined multiple complementary computational approaches, including predictions from Effectidor2, Bastion3, and EffectiveT3, similarity searches against curated effector databases, all-vs-all BLAST analyses, genomic context evaluation, plant subcellular localization prediction, and structural modeling. These outputs were integrated within a newly developed scoring pipeline, EffRank, which assigns quantitative scores to each effector candidate using structural, genomic, and functional criteria.

Application of this workflow to 57 complete *P. syringae* genomes initially produced 283 putative novel T3SE candidates. Sequential filtering steps integrating applied predictors consensus, homology exclusion, and localization analysis reduced this dataset to 15 high-confidence candidates lacking similarity to previously described effectors. EffRank analysis further prioritized three out of 15 candidates as the most promising novel T3SEs. Transcriptomic data obtained under T3SS-inducing conditions further supported the expression of orthologues corresponding to the highest-scoring candidates. In addition, the workflow identified several systematic false positives, including proteins related to YenB-like toxins, highlighting biases inherent to current prediction tools.

In conclusion, this study demonstrated that integrating multiple prediction tools with a transparent scoring framework and comparative genomics improved the reliability of novel T3SEs discovery. The proposed workflow provides a reproducible bioinformatics strategy for prioritizing candidate T3SEs and guiding targeted experimental validation.

Keywords:

T3SE, Pseudomonas, novel effector prediction

Abstract ID: **20** Type: **Oral Presentation**

An integrated single-nucleus landscape of cardiovascular disease

Dušan Ušjak[1], Cong Feng[2], Valentina Đorđević[1], Ming Chen[2]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *College of Life Sciences, Zhejiang University*

Contact e-mail: **dusan.usjak@imgge.bg.ac.rs**

Cardiovascular diseases (CVDs) are the leading cause of morbidity and mortality worldwide, accounting for approximately 32% of all deaths annually. Despite this substantial burden, their underlying molecular mechanisms remain poorly understood. The aim of this study was to construct an integrated single-nucleus RNA-seq (snRNA-seq) atlas using data from heart samples from different pathological conditions and to analyze gene expression patterns across cell types. A systematic search and curation of publicly available snRNA-seq datasets were performed to extract count matrices originating from left ventricular tissue samples. A total of 155 samples from eight independent datasets were included, comprising 75 unaffected controls and 80 samples from diseased hearts (cardiomyopathy, myocardial infarction, aortic stenosis, and related conditions). Data integration was performed using the Harmony algorithm with batch correction applied to preprocessed and concatenated data, resulting in an integrated object containing 1,100,688 nuclei and 25,217 genes. Cell clusters were annotated as cardiomyocytes, fibroblasts, endothelial cells, pericytes, myeloid cells, lymphocytes, smooth muscle cells, endocardial cells, neural cells, lymphatic endothelial cells, adipocytes, mast cells, and epicardial cells. Gene expression patterns were analyzed using high-dimensional weighted gene co-expression network analysis (hdWGCNA) and relative expression analysis to identify differential pathways. The constructed integrated snRNA-seq atlas provides a valuable platform for the identification of cell type-specific gene markers and pathways associated with major CVDs.

Keywords:

CVD, Single-nucleus, RNA-seq, Data-mining, hdWGCNA

Abstract ID: **33** Type: **Oral Presentation**

Omics-driven agent-based modelling of hepatocellular carcinoma tumours

Mehmet Izmirli[1], Pinar Pir[1]

[1] *Gebze Technical University*

Contact e-mail: **mehmettizmirli@gmail.com**

Agent-based modelling (ABM) is a computational approach in which system-level dynamics are simulated by modelling each component as individuals, defined by a relatively simple set of rules. The ABM method is well-suited for modelling heterogeneous tumour microenvironments (TME), in which cells interact with their environment and with other cells. The progression of the disease is significantly affected by the features of the resident tissue, such as architecture and physiological conditions. This is especially true for metabolically complex organs like the liver. Most cancer ABMs, however, are built based on simplifications that limit biological realism.

In this study, an omics data-driven workflow was established to guide ABM development, with a focus on hepatocellular carcinoma (HCC). First, a single-cell RNA sequencing (scRNAseq) dataset of HCC, including tumour and healthy liver samples, was processed and integrated to remove batch effects between samples. Using the integrated data, major cell types, like malignant cells, were annotated. Then, spatial deconvolution of a spatial transcriptomics (ST) dataset was performed using the scRNAseq dataset as a reference, to predict the cell-type composition and spatial distribution over a tissue slice. Lastly, these predictions, along with the physical measurements of the tissue slice, were used to calculate the spatial coordinates of individual cells.

The resulting coordinates were then used to initialise the agent positions in the ABM, while the behaviour of individual agents was modelled based on the literature and transcriptomic features obtained from scRNAseq data. Finally, to validate the credibility of the tumour growth patterns, preliminary simulations were run, and a qualitative comparison was made with observed tumour regions, including the tumour core and the growing edge. This established framework allows characterisation of emergent tumour behaviours during malignant growth and spread in the context of tissue-specific spatial heterogeneity introduced with spatial data integration.

Thus, an agent-based model of hepatocellular carcinoma, informed by integrated scRNAseq and ST datasets, was developed, promising more realistic and data-driven simulations of cancer dynamics and metastatic potential.

Acknowledgement:

This project is generously funded by TÜBİTAK UIDB 1071 (ERA-NET TRANSCAN-3 , 122N905)

Keywords:

Agent-Based Modelling, scRNAseq, Spatial Transcriptomics

Abstract ID: **36** Type: **Oral Presentation**

Detection of mosaic copy number alterations in PTA-amplified single cells

Abhiram Natu[1], Alexej Abyzov[2], Flora M Vaccarino[1], Milovan Suvakov[2]

[1] *Yale University*

[2] *Mayo Clinic*

Contact e-mail: **suvakov.milovan@mayo.edu**

Somatic mosaicism encompasses a wide spectrum of genomic alterations, yet the landscape of copy number variation in individual cells remains poorly characterized due to amplification biases in traditional single-cell genomics. Here, we present our results on detecting mosaic copy number variations (CNVs) utilizing Primary Template-directed Amplification (PTA) to sequence single genomes across different normal human tissues. By achieving bulk-like uniform genome coverage, PTA enables the detection of copy number changes ranging from massive megabase-scale alterations to highly focal events at single-cell resolution. In this presentation, we detail the broader landscape of overall mosaic CNVs, capturing frequent sub-chromosomal copy-number changes and whole-chromosome aneuploidies, such as the loss of chromosome Y. We further highlight the capacity of PTA to reconstruct a striking example of a complex, multi-break rearrangement between several chromosomes in a single cell. Using orthogonal evidence from both sequencing depth and discordant read fragments, we capture a complex mutational event entirely invisible to bulk sequencing. Furthermore, we demonstrate that it is possible to accurately identify locus-restricted deletions within the T-cell receptor (TCR) loci. These stepwise focal copy-number losses faithfully reflect canonical somatic V(D)J immune rearrangements, validating our ability to detect biologically expected alterations. Together, these results demonstrate that single-cell PTA provides an expansive view of somatic mosaicism, uncovering extensive copy number diversity and localized genomic changes in normal tissues.

Keywords:

mosaicism, single-cell, PTA, CNV, CNVpytor

Abstract ID: **37** Type: **Oral Presentation**

Structured generative AI for interpretable analysis of complex biomedical data

Ivan Lorencin[1], Sandi Baressi Šegota[1], Damir Bartolin[2], Zvonimir Babić[2], Domagoj Frank[2]

[1] *Juraj Dobrila University of Pula, Faculty of Informatics*

[2] *University North*

Contact e-mail: **ilorencin@unipu.hr**

Biomedical data are increasingly heterogeneous, multimodal, and inconsistently structured, limiting their effective use in scalable and reproducible artificial intelligence systems. In practice, such data originate from diverse experimental, clinical, and observational settings, often lacking standardized formats and consistent integration. Existing approaches frequently rely on rigid preprocessing assumptions, constraining their ability to generalize across real-world biomedical data sources.

This work introduces an AI-centric framework for adaptive data processing and interpretation in biomedical contexts. The framework operates on arbitrarily structured inputs, enabling flexible ingestion and integration without requiring predefined schemas. Instead, representations are learned directly from the data, allowing the system to adapt to varying formats, modalities, and acquisition conditions commonly encountered in biomedical applications.

At its core, the approach transforms heterogeneous biomedical inputs into structured representations that capture underlying relationships and enable consistent downstream analysis. These representations are processed through analytical models, supporting robust and scalable inference across diverse datasets while preserving domain-relevant context.

To enable controlled and interpretable reasoning, generative components are employed as structured transformation modules, producing schema-constrained outputs (e.g., JSON-based representations) rather than free-form text. This allows deterministic integration of generative reasoning into the analytical pipeline while maintaining interpretability and consistency.

For human-facing interpretation, a retrieval-augmented generation (RAG) layer is applied at the reporting stage. Retrieved intermediate results, domain-relevant context, and metadata are used to guide generative models, ensuring that outputs remain grounded in the underlying data. This improves factual consistency, reduces hallucinations, and enables traceable reasoning through the inclusion of provenance and confidence indicators.

The framework supports iterative refinement of both representations and models, enabling continuous adaptation to evolving biomedical datasets. By combining adaptive data processing, structured generative reasoning, and retrieval-grounded reporting, the approach enables scalable, reproducible, and interpretable AI systems for complex biomedical data analysis.

Acknowledgement:

This work was (partially) supported by the EC Digital Europe Programme EDIH Adria 2.0 (101256325); SPIN projects IP.1.1.03.0120, IP.1.1.03.0028 and IP.1.1.03.0039; and NextGenerationEU University grants: uniri-iz-25-6, uniri-iz-25-220, IIP_010144, IIP_010136, and UNIN-TEH-25-1-8.

Keywords:

Adaptive Learning,Explainability,Multimodal Integration,Retrieval Generation

Abstract ID: **42** Type: **Oral Presentation**

Federated learning in Healthcare across Europe: Challenges and Insights from the BETTER Project

Fu-Sung Kim-Benjamin Tang[1], Mehrshad Jaberansary[1], Ana Grönke[2], Nelly Barret[3], Pietro Pinoli[4], Carlo Alberto Bono[4], Elia Broggio[5], Alessandra Colli[5], Michele Compare[5], Adrián García Andreu[6], Diana Martínez-Minguet[6], Oya Beyan[1], Mattia Sabella[7]

[1] *Institute for Biomedical Informatics, University of Cologne, Medical Faculty and University Hospital Cologne, Germany*

[2] *Medical Data Integration Center (MEDIC), University of Cologne, Faculty of Medicine and University Hospital Cologne, Germany*

[3] *INSA Lyon - Institut National des Sciences Appliquées de Lyon*

[4] *Politecnico di Milano Milan, Italy*

[5] *Datrix S.p.A., Foro Buonaparte 71, 20121, Milan, Italy*

[6] *PROS Group, Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València, Camí de Vera, València, Spain*

[7] *Politecnico di Milano*

Contact e-mail: **fu-sung.tang@uk-koeln.de**

Healthcare advances through machine learning for personalized patient treatments require access to large-scale, diverse, and high-quality patient datasets. However, strict privacy regulations on such sensitive data cause data fragmentation across institutions, limiting its potential for broader clinical insights. Platforms such as the "Platform for Analytics and Distributed Machine Learning for Enterprises"(PADME) address this challenge by enabling federated learning on sensitive patient data. In PADME, patient data remains at its source, while only derived outputs, such as model parameters, are shared. Within the EU Horizon project "BETTER", PADME facilitates three real-world clinical use cases for privacy-preserving federated learning without requiring patient data exchange across participating institutions.

For the federated learning real-world use cases in BETTER, we structured the implementation along three layers: (1) a data layer for integration and harmonization of pseudonymized patient data via an extract-transform-load (ETL) pipeline, developed according to FAIR principles, mapping heterogeneous clinical, genomic, and laboratory data into a shared, standardized representation; (2) an infrastructure layer enabling the privacy-preserving federated learning orchestration through PADME; and (3) an algorithm layer supporting use-case-specific data pre-processing, model training and validation. Established data governance mechanisms within the BETTER Consortium, including privacy risk assessments and multi-stakeholder review of code and documentation, ensure compliance with GDPR requirements.

Insights gained throughout the BETTER project revealed critical requirements for successful federated learning across heterogeneous clinical environments. Harmonizing data of differing local data schemas and data quality proved essential for federated learning. Limited data visibility for model developers necessitated the creation of a central metadata catalogue under consideration of utility-privacy trade-offs. From a data governance perspective, robust procedures for algorithm review and approval were essential for transparency, including structured documentation templates addressing privacy risks and vulnerabilities. Use-case-specific experiments with synthetic clinical data demonstrated that our approaches perform on par with centralized approaches and outperform single-institution approaches.

Our findings demonstrated that federated learning in real-world healthcare across Europe is feasible, but necessitates coordinated cross-institutional technical and organizational solutions for transparency and data privacy compliance. Future work should further improve interoperability, automated data harmonization, and standardized governance frameworks to broaden the usability of federated learning across Europe.

Acknowledgement:

The project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101136262. The communication reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

Keywords:

healthcare, federated-learning, privacy, distributed-analytics, medical

Abstract ID: **44** Type: **Oral Presentation**

A pan-tissue multi-omics framework for aging analysis via non-negative matrix tri-factorization

Aleksandr Matsun[1], Branislava Jankovic[1], Natasa Przulj[1], Noel Malod[1], Stevan Milinkovic[1], Tarmo Nurmi[1]

[1] *Mohamed Bin Zayed University of Artificial Intelligence*

Contact e-mail: **aleksandr.matsun@mbzuai.ac.ae**

Aging is a complex biological process that affects almost all systems of an organism and is known to be a major driver of numerous diseases. Despite a lot of research on aging in general, few studies look at it from a pan-tissue perspective, thus availability of systemic anti-aging interventions remains limited. In this work we present a novel approach that integrates age- and tissue-specific bulk RNA sequencing data with prior knowledge in the form of bulk proteomics and metabolomics data using non-negative matrix tri-factorization. By analyzing the dynamics of genes'learned embeddings across all the age brackets, our model curates an atlas of tissue-specific aging-associated genes, which we use afterwards to perform a systematic selection of pan-tissue aging-related genes. Applying this framework to the multi-omics data that includes gene expressions in 31 tissues from GTEx database, we derive and validate a set of 21 genes that undergo significant changes with age in the vast majority of tissues. Finally, out of those 21 genes we consider 7 to be promising therapeutic targets, due to their involvement in such molecular processes as inflammation, metabolism and stress response, and propose potential intervention strategies that use drug repurposing and de novo drug generation. Our method provides a scalable approach for integrative multi-omics analysis of aging that can be further extended to additional datasets.

Keywords:

Aging, Multi-omics, Pan-tissue, Integration

Abstract ID: **51** Type: **Oral Presentation**

Targeting resistance and survival mechanisms in Pseudomonas aeruginosa using data fusion and machine learning

Branislava Jankovic[1], Aleksandr Matsun[1], Noël Malod-Dognin[1], Natasa Przulj[1]

[1] *Mohamed Bin Zayed University of Artificial Intelligence*

Contact e-mail: **branislava.jankovic@mbzuai.ac.ae**

Pseudomonas aeruginosa is one of the most dangerous antibiotic-resistant bacteria in the world, classified by the WHO as a critical priority pathogen. It causes severe and often fatal infections in hospitalised and immunocompromised patients, and it is notoriously difficult to treat. Its resistance arises from multiple mechanisms working simultaneously —a near-impermeable outer membrane that blocks drugs from entering, powerful efflux pumps that actively expel antibiotics before they can act, and the capacity to rapidly acquire new resistance genes —leaving clinicians with very few therapeutic options.

To address this at a systems level, we applied Non-negative Matrix Tri-Factorization (NMTF), a machine learning method that integrates multiple biological datasets simultaneously, including drug chemical similarities, drug-target interactions, pathogen resistance profiles, point mutations, and protein co-expression networks. By learning hidden patterns across all data types at once, NMTF predicts novel drug-target associations and identifies bacterial vulnerabilities that are not visible in any single dataset alone, giving us a data-driven map of resistance mechanisms and potential new targets. Guided by these predictions, we selected five targets in P. aeruginosa covering both resistance and essential survival processes: MexA and MexB, the components of the MexAB-OprM efflux pump that actively pumps antibiotics out of the cell; MurA, which catalyses the first committed step in peptidoglycan biosynthesis and is essential for cell wall integrity; PonA, a penicillin-binding protein responsible for cell wall cross-linking; and LpxA, which initiates lipid A biosynthesis and is indispensable for outer membrane assembly.

For each target, new drug candidates were generated and evaluated through virtual screening, molecular docking, and ADMET profiling. Experimental validation of these compounds holds considerable therapeutic potential —efflux pump inhibitors targeting MexA and MexB would restore the efficacy of existing antibiotics by disrupting the bacterium's primary resistance mechanism, while compounds directed against MurA, PonA, and LpxA would compromise essential biosynthetic processes indispensable for bacterial viability.Together, this multi-target strategy reduces the likelihood of resistance developing to any single drug, and offers a genuine computational path toward treating infections that are currently untreatable.

Keywords:

Pseudomonas aeruginosa, NMTF, antibiotic resistance

Abstract ID: **53** Type: **Oral Presentation**

Deeply phenotyped and multi-omics data integration with the Human Phenotype Project

Tarmo Nurmi[1], Noël Malod-Dognin[1], Luiz Maniero[1], Eduardo da Veiga Beltrame[1], Nataša Pržulj[1]

[1] *School of Digital Public Health, Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates*

Contact e-mail: **tarmo.nurmi@mbzuai.ac.ae**

Large-scale, deeply phenotyped and multi-omics datasets are becoming more and more accessible through initiatives such as the UK Biobank, the Human Phenotype Project, Our Future Health, the Danish Precision Health Initiative DELPHI, All of Us, and the Emirati Genome Program. These rich datasets hold great potential for biomedical research. However, individual data types are often studied in isolation, or statistical associations are mined between only two data types at a time. In contrast, multifaceted datasets enable new biomedical insights that can be achieved only through a joint consideration of multiple data types. Yet, integrating these large-scale and highly heterogeneous data into a holistic view for effective data mining remains a challenge. An effective analysis framework should flexibly include diverse data types, from molecular and genetic data to phenotypic features and electronic health records and linkages between them. In addition, multiple data sources should be combined in order to take into account existing knowledge about human biology, molecular interactions, and drug properties. Recently, such flexible data integration and embedding frameworks based on treating the different data types as matrices and combining multiple matrices into one model using joint non-negative matrix tri-factorization (NMTF) have achieved success in identifying novel gene-disease associations and molecule candidates for drug repurposing from a wide variety of heterogeneous data.

In this work, we constructed a multi-omics NMTF-based integration framework on the Human Phenotype Project deeply phenotyped multi-omics data set combined with external data sources. The framework integrates gene expression data (RNA-Seq), medication use, protein-protein interaction data, drug-target interaction data, and drug chemical similarity data, with unrestricted possibilities to include additional data types such as microbiome data and many others. The gene-participant-drug embeddings produced by the framework, together with participant electronic health records, have implications for uncovering novel gene-disease associations, patient stratification, drug repurposing, and exploration of semantic relationships between different biomedical entities using linear vector operations.

Keywords:

Multi-omics, deep phenotyping, data integration

Abstract ID: **54** Type: **Oral Presentation**

Cell-type specific expression and structural features of transcriptional isoform LDLRAD4-217 in rectal cancer

Anastasija Bubanja[1], Tamara Babic[1], Aleksandra Nikolic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **anastasija.bubanja@imgge.bg.ac.rs**

Cancer is characterized by a highly dynamic transcriptional landscape, in which alternative promoter usage plays a key role in regulating gene expression and contributes to transcript diversity. Recent large-scale transcriptomic studies have demonstrated differential promoter activity between malignant and non-malignant rectal tissues, including pronounced activation of the promoter driving the *LDLRAD4-217* transcript (ENST00000592657.1).

In silico mutagenesis was performed by generating all possible single nucleotide variants across the transcript, followed by RNA secondary structure prediction and minimum free energy (MFE) calculation using RNAfold. Structural sensitivity was assessed as ΔMFE relative to the wild-type sequence, enabling identification of regions with high structural impact. Potential functional targets were identified by correlating *LDLRAD4-217* expression with other transcripts in the Rectal Adenocarcinoma dataset from The Cancer Genome Atlas. Significantly correlated transcripts were selected, and RNA–RNA interactions were predicted using IntaRNA, with binding strength assessed based on interaction energy (ΔG). Cell-type deconvolution of TCGA READ data was performed using CIBERSORT (LM22 and custom single-cell-derived signatures) followed by correlation with *LDLRAD4-217* expression and stromal stratification.

In silico mutagenesis revealed a non-uniform distribution of structural sensitivity across the transcript sequence, suggesting the presence of structurally important domains. Notably, prominent hotspot was identified within the region spanning nucleotides 238-302 and two small hotspots around position 165 and 460. Interaction hotspots were predominantly localized around 200 and 300 nt, corresponding to the unique exon of *LDLRAD4-217*. Among candidates a transcript *DUOXA2-203*, showed the strongest predicted interaction. Deconvolution results revealed a significant positive correlation with epithelial cells and immune cell populations, with divergent patterns observed within immune subsets, including positive and negative associations for mast cells, NK cells and CD4 T cells.

Our results suggest that *LDLRAD4-217* represents a structurally constrained lncRNA with potential regulatory interactions and cell-type-specific expression, implicating its role in shaping the tumor microenvironment in rectal cancer.

Keywords:

Rectal Cancer, LncRNA, Prediction, Transcriptome

Abstract ID: **59** Type: **Oral Presentation**

Permafrost-preserved Late Pleistocene wolf from Yakutia: palaeogenomic and metagenomic perspectives

Alina Guliaeva[1], Albert Protopopov[2], Maksim Cheprasov[3], Gavril Novgorodov[3], Alexander Rakitko[1], Fedor Sharko[4], Artem Nedoluzhko[3,4]

[1] *ITMO University, Saint Petersburg, Russia*

[2] *Academy of Sciences of the Republic of Sakha, Yakutsk, Russia*

[3] *Mammoth Museum, North-Eastern Federal University, Yakutsk, Russia*

[4] *Laboratory of Palaeogenomics, European University, Saint Petersburg, Russia*

Contact e-mail: **auguliaeva@gmail.com**

Wolves (*Canis lupus*) served as key predators in Pleistocene ecosystems, regulating large herbivore populations across the mammoth steppe. Permafrost-preserved remains offer rare opportunities for direct genomic and metagenomic analysis of extinct fauna and their associated microbiota. We therefore aimed to reconstruct the phylogenetic position, gut microbiome composition, and diet of a Late Pleistocene wolf from Yakutia.

A well-preserved wolf carcass recovered in 2021 from the Tirekhtyakh River, Yakutia, radiocarbon dated to approximately 44 ka, was subjected to ancient DNA extraction from hair, skin, and gastrointestinal tract (GIT) contents using the GENECLEAN® Ancient DNA kit, followed by sequencing on the Illumina NovaSeq 6000 platform. Reads were processed in PALEOMIX, and the mitochondrial genome was constructed de novo using SPAdes, followed by annotation with MitoZ. A maximum-likelihood phylogenetic tree (IQ-TREE) was inferred from 138 mitogenomes of modern and ancient canids. GIT reads were assembled (metaSPAdes), binned (metabat2, comeBIN, maxbin2), refined with DAS Tool, quality-filtered (CheckM2), and taxonomically assigned (GTDB-Tk).

We reconstructed a complete mitochondrial genome (16,727 bp; mean coverage 195×). Phylogenetic analysis placed the Yakutian wolf within a strongly supported clade (bootstrap >95) of northeastern Siberian Pleistocene wolves (41–67 ka) that have no direct descendants among modern populations. From the GIT contents, we obtained 15 metagenome-assembled genomes (MAGs), including a high-quality MAG of a pathogenic bacteria *Erysipelothrix* (completeness 91.4%, contamination 1.2%). This MAG encodes type III secretion system components, virulence factors, antioxidant proteins, antimicrobial resistance determinants, and mobile genetic elements. Preliminary taxonomic profiling of GIT contents also revealed diverse eukaryotic taxa, providing initial insight into the diet of this Late Pleistocene predator.

This permafrost-preserved specimen recorded both the phylogenetic identity of an extinct wolf lineage and a snapshot of its gut microbiome and diet at the time of death. We additionally uncovered ancient origins of Erysipelothrix virulence and antimicrobial resistance, demonstrating that permafrost-preserved gastrointestinal material can recover ancient pathogen genomes with functional traits and extend the known timeline of antimicrobial resistance determinants to >40 ka.

Keywords:

Pleistocene fauna, paleogenomics, canids, mitogenome

Abstract ID: **60** Type: **Oral Presentation**

Pathogenic variations illuminate functional constraints in intrinsically disordered proteins

Norbert Deutsch[1], Gábor Erdős[2], Zsuzsanna Dosztányi[1]

[1] *Eötvös Loránd University*

[2] *ELTE Protein Dynamics Research Group*

Contact e-mail: **norbert.deutsch@ttk.elte.hu**

Intrinsically disordered regions played key roles in cellular signaling and regulation, yet their contribution to human disease remained poorly understood. To address this, the present study analyzed nearly one million ClinVar missense variants, focusing on those located within intrinsically disordered regions. Pathogenic variants showed significant enrichment in short linear motifs and disordered binding regions, aligning with their central functional importance. To extend these insights beyond existing annotations, subsequent analysis utilizing the *AlphaMissense* tool uncovered localized "island-like"patterns of elevated pathogenicity within these functional regions. Leveraging these signals, a newly developed classifier prioritized predicted eukaryotic linear motifs (PEMs), revealing thousands of candidate functional sites linked to major disease classes, including neurological, cardiovascular, and cancer-associated genes. Specific case studies, including the genes *POLK* and *FOXP2*, demonstrated how this approach successfully connected genetic variation to molecular mechanisms. Ultimately, this framework provided a scalable strategy to interpret variants of uncertain significance and defined the functional constraints governing the disordered proteome.

Keywords:

IDRs, Variants, AlphaMissense, SLiMs, Pathogenicity

Abstract ID: **62** Type: **Oral Presentation**

Integrated transcriptomic analysis reveals coordinated network reprogramming and p53 pathway inactivation in temozolomide-resistant gioblastoma A172 cell line

Milica Pajović[1], Ana Podolski-Renić[1], Jelena Dinić[1], Marija Grozdanić[1], Milica Pešić[1], Sofija Bjeletić[1], Sofija Jovanović Stojanov[1], Ema Lupšić[1], Miodrag Dragoj[1]

[1] *Institute for Biological Research "Siniša Stanković" –National Institute of Republic of Serbia, University of Belgrade*

Contact e-mail: **milica.pajovic@ibiss.bg.ac.rs**

Glioblastoma is the deadliest primary brain tumor, with a median survival of 14 months, largely due to resistance to temozolomide (TMZ), the standard first-line therapy. Therefore, robust preclinical models of TMZ resistance are critical for the development of clinically translatable therapeutic strategies. We established an *in vitro* TMZ resistance system from the sensitive human glioblastoma A172 cell line using cyclic TMZ exposure mimicking the Stupp protocol (5-day treatment/3-week recovery). Besides a stable resistant line (A172-TMZ-R), intermediate populations (A172-TMZ-C2 and A172-TMZ-C4) captured progressive stages of resistance acquisition. Cells were characterized by RNA sequencing, bioinformatics analyses, real-time proliferation monitoring, and flow cytometry.

We performed RNA sequencing on each cell line with and without TMZ treatment. The computational pipeline integrated differential expression analysis (DESeq2), gene set enrichment analysis (GSEA), gene set variation analysis (GSVA), and weighted gene co-expression network analysis (WGCNA). DESeq2 analysis revealed a progressive transcriptomic remodeling trajectory, with 1,285 DEGs at A172-TMZ-C2 *vs.* A172 escalating to 2,306 at A172-TMZ-R *vs.* A172, with downregulated genes predominating at all stages. GSVA uncovered a p53 pathway phase transition as a central mechanistic switch: TMZ-induced p53 activation was maintained through intermediate resistance stages but collapsed abruptly at full resistance (GSVA score was +0.58 at A172-TMZ-C4 and went to −0.05 at A172-TMZ-R), coinciding with complete loss of G2/M checkpoint responsiveness and senescence escape in A172-TMZ-R cells observed by flow cytometry. WGCNA identified resistance-correlated co-expression modules capturing distinct programs: constitutive cell cycle activation (black module: *E2F1*, *UHRF1*; pink module: *BIRC5*, *AURKA*, *CCNB1*), EMT/stemness with epigenetic remodeling (yellow module: *LGR5*, *CNTN1*), p53-IFN inactivation (turquoise module: *STAT2, CDKN1A*), and IFN/immune silencing (blue module: *CHI3L1*, *FN1*). Treatment-insensitive, progressive upregulation of DNA methylation, HDAC, and PRC2 pathways –maintained by the UHRF1-DNMT1 axis –provides the epigenetic basis for resistance consolidation.

Our integrated analysis reveals that TMZ resistance in glioblastoma involves coordinated multi-module transcriptomic reprogramming that uncouples DNA damage from checkpoint arrest and senescence. Hub genes *BIRC5*, *AURKA*, and *LGR5* represent candidate therapeutic targets, while the identified epigenetic circuits support early combination strategies with DNMT or HDAC inhibitors to prevent the consolidation of irreversible resistance.

Acknowledgement:

This research was funded by the Ministry of science, technological development and innovation of the Republic of Serbia, grant number 451-03-136/2025-03/200007. The results presented in this abstract are in line with Sustainable Development Goal 3 (SDG3: Good Health and Well-being) of the United Nations 2030 Agenda.

Keywords:

glioblastoma, temozolomide, drug resistance

Abstract ID: **68** Type: **Oral Presentation**

Understanding the interplay between positive and negative selection in the human genome

Yury Barbitoff[1], Polina Malysheva[1], Polina Bogaichuk[1], Nadezhda Pavlova[1], Alexander Predeus[1]

[1] *Institute of Bioinformatics Research and Education, Belgrade, Serbia*

Contact e-mail: **barbitoff@bioinf.institute**

Natural selection is arguably the most important driving force of microevolution which has a significant impact on the patterns of genetic variation across the genome. In loci with important cellular and organismal functions, negative selection acts to eliminate deleterious derived alleles, leading to evolutionary conservation of sequences. At the same time, positive selection enhances the transmission and fixation of adaptive alleles, leaving a characteristic footprint called a "selective sweep". While many methods have been developed to detect signals of both positive and negative selection, little is known about the interplay of these two forces. In this work, we tried to comprehensively identify loci bearing combinations of negative and positive selection signals across the human genome. To this end, we integrated a wide array of metrics, including; (i) alignment-based measures of evolutionary conservation across vertebrate genomes (phastCons, phyloP, GERP++); (ii) selective constraint estimates obtained from standing genetic variation in human populations (pLI, LOEUF, z-scores for observed-to-expected missense variant ratios); and (iii) measures of recent positive selection covering different evolutionary timescales (iHS, DRC150, SDS). In total, we identified a set of 163 genes bearing a combined selection signal (CSS) according to at least one negative and positive selection metric. Intriguingly, we found that these genes have a broader range of molecular functions despite having a relatively restricted expression profile. Overrepresentation analysis showed a strong enrichment of genes involved in nervous system function, including synaptic transmission and neurogenesis. Of note, CSS genes were also overrepresented among highly pleiotropic genes affecting multiple independent traits (across a collection of over 1,000 complex and Mendelian traits in humans and mice). Taken together, our results emphasize the complex network of connections between natural selection, gene function, and phenotype, and indicate that mutations in genes controlling high-level organismal functions might have stronger fitness effects.

Acknowledgement:

We thank JetBrains Ltd. for providing computational resources for the project.

Keywords:

natural selection, pleiotropy, complex trait

Abstract ID: **70** Type: **Oral Presentation**

A Bayesian framework for quantifying allele count evidence for variant pathogenicity

Fedor Konovalov[1]

[1] *Independent Clinical Bioinformatics Laboratory, Moscow, Russia*

Contact e-mail: **fk@clinbio.ru**

Allele counts in affected individuals and population controls are central to assessing the pathogenicity of genetic variants in human medical genetics, yet their meaning is typically mediated through discrete thresholds and heuristic criteria. This work introduces a quantitative Bayesian population-genetic framework that summarizes allele count evidence as a single continuous measure: the Bayes factor comparing explicitly defined pathogenic and neutral models.

The framework is formulated for autosomal dominant rare diseases and incorporates disease prevalence, penetrance, cohort composition, and sequencing error rates as explicit parameters. Under the pathogenic model, allele frequency is treated as a latent variable governed by mutation-selection balance, approximated by a gamma distribution. Under the neutral model, allele frequencies are represented by a simulated site frequency spectrum of a reference population, obtained via forward Wright-Fisher simulations and calibrated to publicly available data from the Genome Aggregation Database (gnomAD). Observed allele counts in affected and control cohorts were evaluated under both models, with uncertainty handled through marginal likelihood calculation.

Application to publicly available gnomAD data, specifically non-Finnish European cohorts, showed that the Bayes factor varies smoothly across allele count configurations, without reliance on predefined thresholds. The strength of evidence depends strongly on affected cohort size and on false positive rates in large control datasets, both of which materially influence inference at current sample scales. Across representative scenarios, the model reproduces qualitative behavior of ACMG/AMP allele-frequency criteria while providing a continuous, model-based alternative that makes evidence directly comparable across studies and clinical contexts.

Numerical evaluation is performed via direct integration over allele frequency and mutation rate, in log-space with explicit cancellation of low-frequency singular behavior. Robustness analyses showed that realistic deviations in the neutral site frequency spectrum produce bounded changes in the Bayes factor, remaining within supporting-level evidence shifts.

The framework provides a mathematically grounded and flexible approach to allele count interpretation, supporting ongoing efforts toward quantitative variant classification.

Keywords:

Bayesian, pathogenicity, population, monogenic, human

Abstract ID: **74** Type: **Oral Presentation**

A statistical perspective on the Last Universal Common Ancestor (LUCA)

Miljenko Huzak[1], Pavle Goldstein[1], Tonka Rimac[1]

[1] *Department of Mathematics, Science Faculty, University of Zagreb*

Contact e-mail: **payo@math.hr**

We introduce an analysis-of-variance statistical framework for analyzing multiple sequence alignments of protein families shared between two groups of organisms. When applied to present-day bacteria and archaea, the method yields a ranking score, estimating likelihood for their presence in LUCA population. By varying bacterial and archaeal subsamples, we assess the robustness of these inferences and gain some information on metabolic pathways potentially present in LUCA. Finally, our results provide a perspective on evolutionary processes associated with the archaeal–bacterial divergence.

Keywords:

LUCA, statistics, Pfam

Abstract ID: **78** Type: **Oral Presentation**

Predicting drug synergy under realistic scenarios

Uxia Veleiro[1], David Mendez[2], Noël Malod-Dognin[3], Natasa Przulj[4], Mikel Hernaez[5]

[1] *Mohamed bin Zayed University of Artificial Intelligence*

[2] *University of Granada*

[3] *School of Digital Public Health, Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates*

[4] *Mohamed Bin Zayed University of Artificial Intelligence*

[5] *CIMA University of Navarra*

Contact e-mail: **uxia.veleiro@mbzuai.ac.ae**

Drug combination therapies are a promising therapeutic strategy for complex diseases, with particular relevance in oncology. By targeting multiple pathways simultaneously, such combinations can overcome drug resistance and improve treatment outcomes. However, given the combinatorial size of the search space, experimentally screening all possible drug combinations remains prohibitively expensive. In this context, machine learning models have emerged as a powerful approach for prioritizing potentially synergistic drug pairs. Yet, their performance is highly sensitive to the evaluation protocol used. While recent work has highlighted that evaluation with random data splits yields overly optimistic performance estimates, most methods either continue to rely on random splits or lack comparisons under stricter evaluation schemes.

In this work, we first introduced a reproducible evaluation framework for drug synergy prediction that goes beyond random splits by defining increasingly difficult generalization scenarios. The framework is easily adaptable to new models and enabled systematic comparison of model performance across prediction settings. We then evaluated both off-the-shelf machine learning baselines and deep learning architectures, with a particular focus on whether different drug and cell-line featurizations led to meaningful differences in generalization performance. More specifically, we investigated whether external biological information, including pathway-level knowledge, can provide useful inductive biases.

Our results showed that conclusions drawn from random evaluation settings do not necessarily hold under stricter generalization scenarios, and, further, some widely used feature construction methods rely on transductive dependencies that complicate their correct assessment. Further, predicting synergy on unseen drugs is a major bottleneck, making model rankings less stable under stricter generalization settings. In addition, we showed how incorporating biological information can reduce the need for highly parameterized models, pointing toward more efficient and biologically grounded approaches.

Overall, our framework provides a reproducible way to benchmark drug synergy models in controlled generalization scenarios, allowing systematic comparison across models. Within this setting, we revealed differences that remain hidden under simpler splits, highlighting that robust prediction depends as much on evaluation design and biological knowledge as on model complexity.

Keywords:

drug synergy, benchmarking, generalization evaluation

Abstract ID: **87** Type: **Oral Presentation**

Exploring the role of the ZBTB14 transcription factor in U937 macrophages

Anja Petrović[1], Vera Pavić[1], Marija Stanković[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **anjap5301@gmail.com**

A human *ZBTB14* protein belongs to a large conserved family of transcriptional regulators which are crucial in the regulation of physiological and pathological processes such as development, differentiation, metabolism, apoptosis and carcinoma. Transcription factor *ZBTB14* was found to regulate the expression of several genes (*e.g. c-MYC, FMR1, DT, IL6* and *LIF*), and also to be involved in the maintenance of genome stability. So, *ZBTB14* transcription factor affects multiple pathways and cellular functions, but its role in the regulation of macrophage development is still largely unknown.

The main objective of this study was to generate *ZBTB14* knockout (KO) using U937 monocytes that can be further employed in research of *ZBTB14* function.

To generate *ZBTB14* KO clone, CRISPR-Cas9 approach was applied in U937 monocytes. Potential U937 ZBTB14KO clones were tested by PCR and Sanger sequencing. Selected U937 ZBTB14KO clone and U937 cells were subjected to Whole Genome Sequencing (WGS) and bioinformatics analysis. DNA library preparation and sequencing were performed by using the MGI technology. Clean reads were aligned to the GRCh38 human reference genome using the Burrows–Wheeler aligner and processed with SAMtools, MANTA and IGV tool. Off-targets were analysed using BLASTn algorithm. The effect of *ZBTB14* KO was examined on potential gen-target using RT-PCR and ELISA method.

A homozygous deletion of 12.902 bp was found at the chromosome 18 of the U937 ZBTB14KO clone. The deletion encompassed 5′ UTR with the full promoter, the first exon, intron and a part of the second exon of the *ZBTB14* gene, completely preventing the transcription of this gene. Potential off-target sequences showed no changes in the genome of U937 ZBTB14KO clone. Interestingly, the expression and secretion of *MMP9* was significantly higher ($p=0.0043$ and $p=0.0031$, respectively) in PMA-induced U937 ZBTB14KO macrophages than in U937 macrophages.

Herein, we showed that the application of WGS approach was necessary to successfully confirm the homozygous deletion of 12.902 bp in U937 ZBTB14KO clone. The absence of changes in off-target sequences indicated the precision in design and application of CRISPR-Cas9 approach. Also, we showed a novel role of *ZBTB14* transcriptional factor in the regulation of *MMP9* in macrophages.

Keywords:

CRISPR-Cas9, WGS, ZBTB14, transcription regulation

Abstract ID: **88** Type: **Oral Presentation**

Mechanistic insights into ageing biology from single-cell omics

Cyril Lagger[1]

[1] *Kyiv School of Economics*

Contact e-mail: **clagger@kse.org.ua**

Ageing is the main risk factor for cancer, neurodegenerative diseases and cardiovascular diseases. Deciphering the mechanisms underlying this complex phenomenon may have tremendous biomedical and global health implications. Despite impressive advances over the last few decades, a major challenge remains to connect the diverse hallmarks of ageing across biological scales.

Technological and computational advances in single-cell omics, notably scRNA-seq, have enabled the characterisation of ageing tissues at unprecedented resolution. Yet, most analyses remain largely data-driven, focusing on identifying statistical patterns in high-dimensional data, often with limited integration of mechanistic principles. We argue that complementary approaches grounded in biophysical modelling and prior biological knowledge may help bridge descriptive observations and causal, interpretable mechanisms of ageing.

Along these lines, using published single-cell datasets, we investigated age-associated alterations in cell-cell communication networks and identified widespread rewiring of ligand-receptor interactions across tissues and cell types, including increased immune and inflammatory signalling and reduced extracellular matrix organisation. We then leveraged spike-in normalisation in the Tabula Muris Senis dataset to quantify age-related changes in absolute transcript abundance independently of transcriptome composition. This analysis revealed a widespread decline in total mRNA abundance across non-immune cell types, in contrast to increased transcript abundance in many immune populations, consistent with a broad decline in transcriptional and metabolic activity during ageing.

To further characterise changes in transcriptional dynamics, we are currently developing physics-informed inference methods of burst kinetic parameters from omics data. On one hand, we leverage established stochastic telegraph models to infer gene expression dynamics from ageing datasets. In parallel, we are developing novel regulatory frameworks inspired by stochastic thermodynamics to move beyond independent-gene models and integrate epigenetic regulation.

Together, these approaches illustrate how single-cell omics, combined with mechanistic and physics-informed modelling, may help bridge ageing mechanisms across molecular, cellular and tissue scales.

Keywords:

ageing, scRNA-seq, communication, transcription, physics-informed

Abstract ID: **90** Type: **Oral Presentation**

Generalized discovery of chromatin structural patterns across developmental and evolutionary scales

Aleksandra Galitsyna[1]

[1] *Independent Researcher, previous position: Massachusetts Institute of Technology*

Contact e-mail: **galitsyn@mit.edu**

Chromatin in the eukaryotic nucleus is organized into regulatory structures that are visible in chromosome-conformation maps, including domains, loops, and other patterns that remain less systematically characterized. Although 3D genome organization is closely linked to gene regulation, development, and disease-associated misregulation, current approaches often rely on predefined structural classes, limiting the discovery of new chromatin patterns and their connection to underlying regulatory determinants.

Here, we present a general deep-learning framework for unbiased discovery and comparison of chromatin structural patterns in Hi-C and Micro-C data. Rather than searching only for known features, our approach identifies recurrent spatial patterns directly from contact maps and enables their systematic analysis across cell types, developmental stages, and species.

Using this framework, we identify "fountains," a previously undercharacterized chromatin contact pattern visible as a signal extending perpendicular to the main diagonal of Hi-C maps. Fountains are detected independently in zebrafish early embryos at the onset of zygotic genome activation and in mouse dendritic cells, and are reproducible across biological replicates and datasets generated by different laboratories.

By integrating multi-omics data, we show that fountains are associated with active enhancer chromatin rather than promoters. They coincide with enhancer marks, pioneer transcription factor binding, cohesin enrichment, increased chromatin accessibility, and nearby early-transcribed genes. Fountains weaken or disappear after cohesin perturbation, and polymer simulations of loop extrusion reproduce the observed contact patterns, supporting a model in which fountains reflect facilitated cohesin loading at active enhancer regions.

In mouse dendritic cells, fountain-associated cohesin activity is linked to immune-cell activation and transcriptional responses to external stimuli. Loss of cohesin impairs enhancer-gene communication and reduces dendritic cell capacity to respond to pathogens and tumor-related challenges, suggesting that enhancer-associated 3D genome structures contribute to context-specific regulatory programs.

Finally, we apply the pattern-discovery framework to 17 available Hi-C and Micro-C maps of diverse species and identify conserved chromatin structural patterns across evolutionary scales. Together, our results establish a generalizable computational approach for exploratory analysis of 3D genome organization and reveal conserved enhancer-associated chromatin structures that connect genome folding, regulatory activity, and biological function.

Acknowledgement:

We thank the laboratory of Prof. Leonid Mirny at MIT for valuable guidance and discussions on the polymer simulations. We are grateful to the laboratory of Prof. Boris Reizis at the University of Chicago, and especially Nick Adams, for providing and supporting the dendritic-cell datasets. We thank the laboratory of Dr. Daria Onichtchouk at the University of Freiburg for providing the zebrafish datasets. We also acknowledge the laboratory of Prof. Mikhail Gelfand, and especially Aleksei Shkolikov, for implementing the neural-network approach central to this study.

Keywords:

3D-genome, Deep learning, Enhancer regulation

Abstract ID: **93** Type: **Oral Presentation**

CardioSafe: Multi-task prediction of cardiac ion channel activity with reverse-leak audited benchmarking

Aakaash Meduri[1], Emre Ulgac[1], Lukas Weidener[1], Marko Brkic[1], Mihailo Jovanovic[1]

[1] *Applied Scientific Intelligence*

Contact e-mail: **mihailo@appliedscientific.ai**

Drug-induced inhibition of the hERG potassium channel is the leading cause of cardiac safety-related drug attrition, but the Comprehensive in Vitro Proarrhythmia Assay (CiPA) framework requires activity data on multiple cardiac ion channels to assess proarrhythmic risk. We present CardioSafe, a three-branch multi-task neural network with cross-attention fusion that integrates chemical fingerprints, ChemBERTa embeddings, and predicted L1000 transcriptomic features to predict blocker status and potency for hERG, Nav1.5, and Cav1.2, with an exploratory IKs head. CardioSafe was trained on the largest publicly reported multi-channel cardiac ion channel dataset, combining ChEMBL 36 with the hERGCentral database (331127 hERG, 3160 Nav1.5, 1138 Cav1.2, and 115 IKs compounds), curated under a pharmacology-aware policy that retains censored measurements and inhibition-percentage votes. Under Tanimoto-similarity-controlled splits, CardioSafe outperforms the leading published comparators (CToxPred2 and CardioGenAI) on the data-rich hERG head; on the smaller Nav1.5 and Cav1.2 heads the standard evaluation is statistically inconclusive. A reverse-leak audit revealed that 22% of Nav1.5 and 21% of Cav1.2 test compounds were present in published comparators' training data (92% as exact compound matches); after removing these contaminated compounds, CardioSafe's lead on Nav1.5 and Cav1.2 also reaches statistical significance, demonstrating that prior cross-publication benchmarks for these channels were inflated by training-data overlap.

Keywords:

cardiotoxicity, hERG, CiPA, ion channels

Abstract ID: **103** Type: **Oral Presentation**

Time- and cost-efficient parameter optimization for 3D bioprinting of composite bioinks via active machine learning with Gaussian Processes

Evgeniya Borkovskaya[1], Katherine Vilinski-Mazur[2], Dmitry Kolomenskiy[3], Bogdan Kirillov[4]

[1] *Moscow Institute of Physics and Technology, Moscow, Russia*

[2] *Sechenov First Moscow State Medical University, Moscow, Russia*

[3] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[4] *Sechenov First Moscow State Medical University*

Contact e-mail: **bogdan.kirillov@skoltech.ru**

3D bioprinters fabricate functional tissue constructs using cell-laden bioinks made out of composite hydrogels with tunable physical properties. The fidelity of resulting prints is governed by a multidimensional parameter space (e.g. concentration of each component, temperature, extrusion multiplier, speed and layer height). Grid search for optimal parameters is impractical and wasteful of both material and instrument time, however, most current approaches to address the parameter search task amount to empirical tuning based on the practitioner's experience. A well-established field of active machine learning shows how to build algorithms to efficiently traverse the search space for new experimental parameters by controlling the model training with intelligent iterative selection and labeling of new data. We propose the application of Active Learning to accelerate the optimization of printing parameters for composite bioinks. A robust body of existing research in Active Learning shows reductions in dataset size requirements ranging from 50% to over 90% in comparable applications for computational metamaterial design, drug discovery and protein engineering. In our method, the Gaussian Processes surrogate model approximates the connection between printing parameters and geometric fidelity while a two-stage experimental design algorithm selects the next most informative experiment, maximizing data utility and minimizing trial costs. Ongoing empirical study performed using a custom low-cost extrusion bioprinter, diverse kappa-carrageenan-based compounds and deliberately suboptimal starting parameters demonstrates improvements in fidelity of the extruded lines during an iterative active learning loop. Our method is straightforward to adapt for new formulae of composite hydrogels and offers a practical, extensible workflow for tissue engineering applications.

Keywords:

active machine learning, 3d bioprinting

Acknowledgement:

The study was supported by grant No. 25-45-02119 from the Russian Science Foundation, https://rscf.ru/project/25-45-02119/.

Abstract ID: **104** Type: **Oral Presentation**

CTCF transcription factor as a potential therapeutic target in osteosarcoma

Luka Bojic[1], Jovana Kostic[1], Aleksandra Medić[1], Milena Milivojevic[1], Jelena Pejic[1], Mina Peric[1], Isidora Petrovic[1], Danijela Stanisavljevic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **luka.bojic@imgge.bg.ac.rs**

Osteosarcoma is an aggressivetumor, and therapeutic approaches for its treatment have n't significantly changed over the past several decades, resulting in persistently poor clinical outcomes for affected patients. Although transcriptomic analyses have contributed to a better understanding of osteosarcoma biology, many studies have been conducted on cohorts that included metastatic samples and samples obtained after chemotherapy treatment. Analyses of such heterogeneous samples may obscure transcriptional patterns associated with the intrinsic aggressiveness of the tumor. To achieve a clearer understanding of these transcriptional programs, we analyzed transcriptomic profiles of osteosarcoma samples that hadn't been previously exposed to therapy and with no metastases. By focusing on these samples, we aimed to obtain a more accurate insight into the biological characteristics associated with aggressive osteosarcoma behavior. Our analysis identified the transcription factor CTCF as one of the key regulators in osteosarcoma. CTCF was strongly associated with transcription programs linked to unfavorable clinical outcomes. The CTCF regulon identified in our analysis was enriched for transcriptional programs associated with proliferation and metabolic reprogramming, indicating a potentially important role for CTCF in regulating these processes. In addition, CTCF regulon activity showed a strong association with patient survival. Given the role of CTCF in chromatin organization and transcriptional regulation, alterations in its regulatory activity may contribute to gene expression changes associated with tumor aggressiveness. To investigate the possibility of pharmacological modulation of this regulatory network, we analyzed the interaction between the natural compound hesperidin and the DNA-binding domain of CTCF using molecular docking and molecular dynamics simulations. Our results of indicated a potentially stable interaction between hesperidin and CTCF. Furthermore, transcriptomic analysis of osteosarcoma cells treated with hesperidin demonstrated changes in the composition of the CTCF regulon, reduced CTCF activity, and enrichment of pathways related to apoptosis and negative regulation of various metabolic processes. Additional analyses confirmed that hesperidin treatment led to a statistically significant increase in the proportion of SAOS-2 cells undergoing early and late apoptosis. Our findings highlight the CTCF regulatory network as a potential driver of aggressive osteosarcoma biology. Moreover, this approach may contribute to the identification of novel prognostic biomarkers and therapeutic opportunities.

Keywords:

Osteosarcoma, CTCF regulon, Hesperidin, Transcriptomics

Acknowledgement:

Ministry of Science, Technological Development and Innovation of the Republic of Serbia, institutional funding - 200042 (University of Belgrade, Institute of Molecular Genetics and Genetic Engineering) (RS-MESTD-inst-2020-200042) and the Science Fund of the Republic of Serbia, (grant no. 7503).

Abstract ID: **105** Type: **Oral Presentation**

Meta-analysis of oral shotgun metagenomes for prediction of caries and periodontitis

Serafim Dobrovolskii[1], Aleksandra Denisova[2,3], Polina Kuznetsova[1], Layal Shaheen[2,4], Dmitrii Kharitonov[1,2], Anna Ilinskaya[5], Michael Agami[6], Valery Ilinsky[5], Alexander Rakitko[1,2,3]

[1] *Genotek Center: AI in Personalized Medicine, ITMO University, Saint-Petersburg, Russia*

[2] *Genotek Ltd., Moscow, Russia*

[3] *National Research University Higher School of Economics, Russian Federation*

[4] *Moscow Center for Advanced Studies, Moscow, Russia*

[5] *Eligens SIA, Riga, Latvia*

[6] *Michael Agami's Family Dentistry*

Contact e-mail: **alexandraa.denisova@gmail.com**

The transition from microarrays to whole-genome sequencing (WGS) enables simultaneous analysis of the host genome and oral microbiome. Oral dysbiosis has been linked to systemic diseases through immune and metabolic mechanisms, highlighting microbiome features for disease risk assessment.

In this work, publicly available shotgun metagenomic saliva datasets were collected for caries (K02) and periodontitis (K05), as other dental diseases lacked sufficient data for integrative analysis. After retrieval from SRA/ENA, data from 14 independent WGS salivary studies, comprising a total of 545 samples, were selected and subjected to quality filtering, host read removal, and taxonomic and functional profiling using MetaPhlAn and HUMAnN.

At the taxonomic level, samples were more strongly separated by bioproject than by clinical group. In Bray–Curtis space, PERMANOVA showed that bioproject explained 28.6% of the variation ($R^2 = 0.286$, $p < 0.001$), whereas the clinical group accounted for 4.1% ($R^2 = 0.041$, $p < 0.001$). Despite the strong batch effect, disease-associated differences remained significant across all cohorts.

Characteristic microbial patterns were reproduced for periodontitis: enrichment of *Porphyromonas gingivalis*, *Treponema denticola*, *Tannerella forsythia*, and representatives of Saccharibacteria, as well as enhancement of biosynthetic pathways of polyamines and L-ornithine. Caries was characterized by *Streptococcus mutans*, *Veillonella parvula*, pathways of carbohydrate metabolism, and biofilm formation.

Predictive models were developed using taxonomic, functional, and combined feature sets with CLR, ILR, and rank-based transformations. The best performance for periodontitis was achieved by the ridge logistic regression model using CLR-transformed taxonomic data (AUC = 0.903 ± 0.011), comparable to previously reported plaque-based shotgun metagenomic models (AUC ~0.93–0.97). In contrast, the highest predictive performance for caries reached AUC = 0.802 ± 0.045. Leave-one-study-out (LOSO) validation for periodontitis demonstrated good generalizability, with AUC values reaching 0.876 and 0.934, while performance decreased moderately for clinically heterogeneous cohorts (AUC = 0.778 and 0.735).

Overall, the resulting models demonstrated robust and stable performance in LOSO validation, supporting their potential application for predicting oral diseases in a large unlabeled cohort and representing, to our knowledge, one of the largest integrative shotgun metagenomic analyses of saliva microbiomes for oral disease prediction.

Keywords:

metagenomics, saliva microbiome, caries, periodontitis

Abstract ID: **107** Type: **Oral Presentation**

Genetic architecture and AI-assisted phenotyping of alopecia in a large Russian genomics cohort

Aleksandra Denisova[1,2], Layal Shaheen[1,3], Dmitrii Kharitonov[1,4], Anna Ilinskaya[5], Saleem Mansour[1,3], Iskandar Hweijeh[1,3], Grigorii Travin[1], Valery Ilinsky[5], Alexander Rakitko[1,2,4]

[1] *Genotek Ltd., Moscow, Russia*

[2] *National Research University Higher School of Economics, Russian Federation*

[3] *Moscow Center for Advanced Studies, Moscow, Russia*

[4] *Genotek Center: AI in Personalized Medicine, ITMO University, Saint-Petersburg, Russia*

[5] *Eligens SIA, Riga, Latvia*

Contact e-mail: **rakitko@genotek.ru**

Androgenetic alopecia and related hair loss disorders are highly heritable conditions, yet their genetic architecture remains insufficiently characterized in Eastern European populations. Here, we performed a large-scale genomic study of alopecia in one of the largest Russian consumer genomics cohorts using an integrated phenotyping framework combining ICD-10 diagnoses, antiandrogen therapy records, and AI-assisted assessment of baldness severity from participant photographs.

To minimize population stratification, controls were selected using ancestry-aware propensity score matching. We analyzed both individual phenotypes and a composite phenotype integrating clinical, therapeutic, and image-derived features.

Genome-wide association analysis of androgenetic alopecia (2,436 cases; 24,304 controls) identified significant loci on chromosomes 1, 2, 5, 7, 18, 20, and X, including established risk regions previously implicated in male pattern baldness. The strongest association was observed at *20p11.22* (p=1.48×10-86). In alopecia areata (517 cases; 5,177 controls), six significant loci were detected, including the immune-associated *6q25.1* region near *RAET1M* and *ULBP3* (*rs12205199*; p=5.15×10-15). Pathway enrichment analysis highlighted genes involved in T-cell activation and immune regulation, including *CTLA4*, *ICOS*, and *IL2RA*, supporting the autoimmune basis of the disease. AI-derived severe baldness phenotypes identified one association at the *2q14.3* locus (p=2.62×10-8). Analysis of androgen receptor CAG repeat variant (*rs746853821*) demonstrated that longer alleles (>23 repeats) were associated with reduced alopecia risk (OR=0.69, 95% CI 0.56–0.86). Polygenic risk models showed strong predictive performance for androgenetic alopecia (AUC up to 0.705), while the best result for external PRS (PGS003558) had an AUC of 0.725 on our cohort. An even better score was obtained when applying the latter score to the phenotype of severe baldness (stage VII), where an AUC of 0.728 was achieved.

Our study integrates diagnostic records and AI-based image phenotyping, enabling scalable discovery of alopecia genetics in the understudied Eastern European population.

Keywords:

alopecia, GWAS, PRS, AI-based phenotyping

Abstract ID: **108** Type: **Oral Presentation**

Improving spatial transcriptomics data processing and annotation under suboptimal image-based cell segmentation

Igor Davidović[1], Mateja Ilić[1], Jelena Kusic-Tisma[1], Nevena Vezmar[1], Aleksandra Vitkovac[1], Marija Lazić[1], Mila Ljujić[1], Aleksandra Divac Rankov[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **igor.davidovic@imgge.bg.ac.rs**

Spatial transcriptomics, in addition to providing information about the cell from which an RNA molecule originates, also provides information about its spatial position. In this study, we analyzed spatial transcriptomics data from 3-day-old zebrafish larvae, which were generated using the Stereo-seq platform (BGI STOmics; 1 × 1 cm chip) and imaged via a Leica Thunder fluorescent imaging system. One of the main challenges in spatial analysis is cell segmentation, which is image-based and automatically performed using the Stereo-seq Analysis Workflow (SAW) and StereoMap. Therefore, low-quality images and differences in tissue variability in staining affinity can lead to inaccurate cell boundary definitions.

This limitation can affect clustering, annotation and downstream analyses. To address these issues, spatial smoothing was performed using a square grid-based approach. The optimal observation (square) size was determined by balancing the number of counts per bin with the empirically determined average cell size. The Stereo-seq platform provides a dense grid of spatial bins with a 500 nm spacing between the centres of adjacent bins. For one observation, a bin 20 was selected as the observation unit, corresponding to a 10 × 10 μm observation area. This does not strictly correspond to the average cell size because it slightly exceeds it. The resulting h5ad file was used to benchmark multiple bioinformatic tools, including StereoPy, Scanpy, Tangram, Squidpy, STAGATE, Seurat and GraphST.

Following benchmarking, Scanpy and StereoPy are the most suitable tools for preprocessing and postprocessing under low-quality segmentation. However, square-grid-based smoothing produces observations that approximate real-cell boundaries. As a result, some observations within each cluster do not accurately reflect the cell, and tools that use cell-by-cell annotation are unreliable. Therefore, cluster-level annotation was a better approach. Marker genes are determined for each cluster by comparing each cluster against all remaining clusters. Obtained marker genes are used for manual or semi-automatic annotation based on anatomical enrichment analysis using BgeeDB and topGO.

Collectively, these strategies improved spatial-data processing and annotation under suboptimal imaging conditions, yielding a high-quality dataset suitable for downstream analysis.

Keywords:

spatial transcriptomics, Stereo-seq, zebrafish

Acknowledgement:

This study is part of the Enhancing Non-Communicable Disease Research Excellence Through Zebrafish Capacity Building (ZeNCure) project, supported by the European Union under the Horizon Europe programme Widening Participation and Spreading Excellence, HORIZON-WIDERA-2023-ACCESS-02, Grant Agreement number 101160259.

Abstract ID: **111** Type: **Oral Presentation**

Study of circulating blood linear RNA nominates CD55 and DLD as novel causal genes and early-stage biomarkers for Parkinson's disease

Aleksandra Beric[1], Sarp Sahin[1], Santiago Sanchez[1], Zining Yang[1], Ravindra Kumar[1], Isabel Alfradique-Dunham[1], Jessie Sanford[1], Daniel Western[1], Bridget Phillips[1], John P. Budde[1], Richard J. Perrin[1], Paul T. Kotzbauer[1], Joel S. Perlmutter[1], Scott A. Norris[1], Carlos Cruchaga[1], Laura Ibanez[1]

[1] *Washington University in St. Louis*

Contact e-mail: **aberic@wustl.edu**

Identifying non-invasive Parkinson's disease (PD) biomarkers is crucial for improving disease diagnosis and monitoring. We analyzed transcriptomic profiles from four major PD studies to identify circulating PD-associated transcripts that could serve as biomarkers.

Differential expression (DE) analyses were conducted in four independent PD whole-blood RNAseq datasets: PDBP (N=1,604), PPMI (N=1,943), BioFIND (N=213), and WashU-MDC (N=583). Results were consolidated through meta-analysis and biologically contextualized via pathway analyses and integration with brain transcriptomic, and plasma and CSF proteomic datasets. The available PD GWAS resources were utilized to verify whether the DE transcripts originated from known PD-risk loci. We examined the clinical relevance by correlating DE transcripts with UPDRS-III and MoCA scores. Finally, we leveraged our findings to develop predictive models to distinguish between PD and healthy controls.

We identified 296 DE transcripts, 28 of which were transcribed from known PD-associated loci. Our findings showed significant overlap with dysregulated brain transcripts (p=3.60×10-19). Nearly half of the identified transcripts were dysregulated before symptom onset. Over a third (105) of the DE transcripts correlated significantly with both UPDRS-III and MoCA scores. Three DE-transcript based predictive models distinguished PD from healthy controls with a ROC AUCs of 0.727-0.733. and were capable of detecting PD transcriptomic signatures before symptom onset. Two transcripts, DLD and CD55, emerged as promising early stage PD biomarkers that were differentially expressed in whole blood, brain and CSF and were included in all predictive models.

Keywords:

transcriptomics, Parkinson's-disease, predictive-modeling, biomarkers

Abstract ID: **117** Type: **Oral Presentation**

IsoChecker: multisample analysis and interactive browsing of long-read transcript isoforms

Alla Mikheenko[1], Andrey Przhibelskiy[2]

[1] *Department of Computational Biology, Mohamed bin Zayed University of Artificial Intelligence*

[2] *University of Helsinki*

Contact e-mail: **alla.mikheenko@mbzuai.ac.ae**

Long-read RNA sequencing routinely generates full-length transcript catalogues for each sample of an experiment. Combining these catalogues across samples and conditions presents two major challenges: constructing a non-redundant consensus annotation and identifying which novel transcripts are reliable enough for downstream analysis. Expression alone is not an optimal filter, as experimentally validated isoforms can show both low abundance and low cross-sample reproducibility. Therefore, robust isoform prioritization requires integrating multiple sources of evidence.

We developed IsoChecker, a pipeline combined with an interactive browser for multisample long-read transcriptome analysis. IsoChecker takes per-sample GTF files and abundance estimates, merges them into a consensus annotation, and matches novel transcripts across samples using a configurable splice-junction tolerance. Each consensus transcript is assigned a structural category and a composite reliability score that integrates transcript category, detection frequency, abundance, junction support, and canonical splice-site usage. When available, IsoChecker can additionally incorporate orthogonal evidence such as short-read splice-junction support and CAGE/QuantSeq end peaks, enabling more confident prioritization of biologically relevant isoforms.

Beyond transcript scoring, IsoChecker provides a broad overview of reproducibility across samples and condition-specific transcriptomic changes. The framework evaluates detection consistency across replicates; TSS and TTS support relative to reference annotation, transcript- and gene-level structural complexity, and sample similarity through heatmaps. Differential isoform usage and differential splicing are assessed per gene using Mann-Whitney statistics.

Results are presented through an interactive isoform-focused gene browser designed to simplify interpretation of genes with complex isoform structure. The browser visualizes all isoforms of a gene together with per-condition read coverage, supports within-condition normalization for comparisons across libraries of different depth, and highlights shifts in isoform composition between conditions. IsoChecker also evaluates transcript stability across replicates, measures concordance with alternative long-read isoform discovery tools, and generates ranked candidate lists of novel isoforms for downstream validation. Overall, IsoChecker provides both a practical reliability-scoring framework and an interactive environment for exploring and prioritizing isoforms in multisample long-read transcriptome studies.

Keywords:

transcriptomics, long-reads, genome browser

Abstract ID: **1** Type: **Flash Talk**

Rare variant pharmacogenomic analysis identifies pathways associated with vedolizumab response in Crohn's disease patients

Biljana Stankovic[1], Aleksandar Toplicanin[2], Bojan Ristivojevic[1], Vladimir Gasic[1], Ivana Grubisa[1], Branka Zukic[1], Aleksandra Sokic Milutinovic[2,3], Sonja Pavlovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Clinic for Gastroenterohepatology, University Clinical Center of Serbia*

[3] *School of Medicine, University of Belgrade*

Contact e-mail: **biljana.stankovic@imgge.bg.ac.rs**

Vedolizumab (VDZ), a monoclonal antibody targeting α4β7 integrin, is used in Crohn's disease (CD) management, yet patients' responses vary, underscoring the need for pharmacogenomic (PGx) markers. This study aimed to identify PGx pathways associated with suboptimal VDZ response using a rare-variant analytical framework.

DNA from 63 CD patients treated with VDZ as first-line advanced therapy underwent whole-exome sequencing. Clinical response at week 14 classified patients as optimal responders (ORs) or suboptimal responders (SRs). Sequencing data were processed using GATK Best Practices, annotated with variant effect predictors, and filtered for rare damaging variants (damaging missense and high-confidence loss-of-function; minor allele frequency < 0.05). Variants were mapped to genes specific for SRs and ORs, and analyzed for pathway enrichment using the Reactome database. Rare-variant burden and composition differences were assessed with Fisher's exact test and SKAT-O gene-set association analysis.

Suboptimal VDZ response was associated with pathways related to membrane transport (ABC-family proteins, ion channels), L1–ankyrin interactions, and bile acid recycling, while optimal response was associated with pathways involving MET signaling. SKAT-O identified lipid metabolism-related pathways as significantly different —SRs harbored variants in pro-inflammatory lipid signaling and immune cell trafficking genes (e.g., PIK3CG, CYP4F2, PLA2R1), whereas ORs carried variants in fatty acid oxidation and detoxification genes (e.g., ACADM, CYP1A1, ALDH3A2, DECR1, MMUT).

This study underscores the potential of exome-based rare-variant analysis to stratify CD patients and guide precision medicine approaches. The identified genes and pathways are potential PGx markers for CD patients treated with VDZ.

Keywords:

gene-set burden analysis, pharmacogenomics

Abstract ID: **4** Type: **Flash Talk**

Confidence-guided lightweight CNN ensemble for OrganMNIST classification

Sandi Baressi Šegota[1], Ivan Lorencin[1], Domagoj Frank[2], Nikola Anđelić[3]

[1] *Faculty of Informatics Pula - Juraj Dobrila University of Pula*

[2] *Department of Computer Science and Informatics, University North*

[3] *Faculty of Engineering - University of Rijeka*

Contact e-mail: **sandi.baressi.segota@unipu.hr**

Accurate classification of medical imagery presented a critical challenge in modern diagnostic workflows, particularly when operating under constrained computational resources. This study addressed the need for high-performance, lightweight diagnostic tools by developing and evaluating an ensemble of small-scale Convolutional Neural Networks (CNNs). The primary objective aimed to demonstrate that a confidence-based combination of computationally undemanding architectures achieved robust classification accuracy comparable to larger, resource-intensive models.

Researchers utilized the OrganAMNIST dataset from the widely standardized MedMNIST collection, a comprehensive repository of abdominal CT images classified into eleven distinct organ categories. The methodology involved designing, training, and validating five unique, lightweight CNN architectures. These models included a basic three-layer SimpleCNN, a reduced VGG8, a compact ResNet10, a custom InceptionMini, and a TinyCNN. Investigators trained each base model independently from scratch for ten epochs using the Adam optimizer and cross-entropy loss. Following individual model optimization, the study implemented a confidence-based ensemble technique. This approach aggregated the predictive probabilities from all five architectures, and it weighted the ultimate classification decision based on the varying confidence levels each model exhibited for specific classes.

Our empirical evaluation revealed significant individual performance among the baseline models. During the validation phase, individual architectures consistently achieved high predictive accuracy, with initial basic models like SimpleCNN and VGG8 surpassing 96% and 97% accuracy, respectively. Subsequently, the confidence-based ensemble method effectively synthesized these individual strengths. By dynamically leveraging the highest confidence predictions across the varied feature-extraction paradigms of the five subset models, the ensemble framework reduced generalization errors and improved overall predictive stability across all eleven target classes.

In conclusion, this research validated the efficacy of employing a confidence-weighted ensemble of minimized CNN architectures for complex medical image classification tasks. The implemented methodology successfully mitigated the typical trade-off between computational efficiency and diagnostic precision. The findings demonstrated that intelligently combined lightweight neural networks offered a highly scalable, accurate, and resource-efficient solution for automated abdominal CT scan analysis, which facilitated broader deployment of artificial intelligence in diverse clinical environments.

Acknowledgement:

This work was (partially) supported by the EC Digital Europe Programme EDIH Adria 2.0 (101256325); SPIN projects IP.1.1.03.0120, IP.1.1.03.0158 and IP.1.1.03.0039; and NextGenerationEU University grants: uniri-iz-25-6, uniri-iz-25-220, IIP_010144, IIP_010136, and UNIN-TEH-25-1-8.

Keywords:

CNN, Ensemble, Performance Testing

Abstract ID: **8** Type: **Flash Talk**

Predicting interaction-specific protein–protein interaction perturbations by missense variants with MutPred-PPI

David Cooper[1], Florent Laval[2], Georges Coppin[2], Kerstin Spirohn-Fitzgerald[2], Marc Vidal[2], Matthew Mort[1], Maxime Tixhon[2], Michael Calderwood[2], Predrag Radivojac[3], Ross Stewart[4], Tong Hao[2]

[1] *Institute of Medical Genetics, School of Medicine, Cardiff University*

[2] *Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute*

[3] *Khoury College of Computer Sciences, Northeastern University, Boston MA, USA*

[4] *Khoury College of Computer Sciences, Northeastern University*

Contact e-mail: **stewart.ro@northeastern.edu**

Disruption of protein–protein interactions (PPIs) is a major mechanism of a variant's deleterious effect. Computational tools are needed to assess such variants at scale, yet existing predictors rarely consider loss of specific interactions, particularly when variants perturb binding interfaces without significantly affecting protein stability. To address this problem, we present MutPred-PPI, a graph attention network that predicts interaction-specific (edgetic) effects of missense variants by operating on AlphaFold 3-based protein complex contact graphs with protein language model embeddings imposed upon nodes. We systematically evaluated our model with stringent group cross-validation as well as benchmark data recently collected within the IGVF Consortium. MutPred-PPI outperformed all baseline methods across all evaluation criteria, achieving an AUC of 0.85 on seen proteins and 0.72 on previously unseen proteins in cross-validation, demonstrating strong generalizability despite scarce training data. To demonstrate biomedical relevance, we applied MutPred-PPI to variants from ClinVar, HGMD, COSMIC, gnomAD, and two *de novo* neurodevelopmental disorder-linked datasets. Disease-associated variants from ClinVar and HGMD showed strong enrichment for both quasi-null and edgetic effects, whereas population variants from gnomAD increasingly preserved interactions with higher allele frequencies. Notably, we observed a strong edgetic disruption signature in highly recurrent cancer variants from both the full COSMIC dataset and a subset of variants from oncogenes. Recurrent tumor suppressor gene variants and autism spectrum disorder-associated variants exhibited moderate quasi-null enrichment, whilst neurodevelopmental disorder-linked variants showed a weak edgetic disruption signature. These results indicate distinct PPI perturbation mechanisms across disease types and show that MutPred-PPI captures functionally relevant molecular effects of pathogenic variants.

Acknowledgement:

We thank the VarChAMP group from the IGVF Consortium for providing access to the benchmark dataset. We acknowledge the use of AlphaFold\,3 (Google DeepMind) and the Catalogue of Somatic Mutations in Cancer (COSMIC, Wellcome Sanger Institute) under academic licenses. This work was supported by the NIH awards U01HG012022 (P.R.), R01GM145937 (P.R.), and UM1HG011989 (M.V.).

Keywords:

protein–protein interaction, mutation

Abstract ID: **15** Type: **Flash Talk**

CellAnalyst: delivering single-cell transcriptomic insights from bioinformatics analysts to experimental biologists

Cong Feng[1], Ming Chen[1]

[1] *College of Life Sciences, Zhejiang University, Hangzhou, China*

Contact e-mail: **ventson@zju.edu.cn**

The complexity of single-cell transcriptomic data traditionally required advanced programming skills, which limited independent data exploration by experimental biologists. This barrier motivated the authors to develop CellAnalyst, a software tool that explicitly separated routine bioinformatics analysis from subsequent data interpretation. The primary objective was to empower non-computational scientists to explore and visualize single-cell transcriptomes independently. To achieve this, the authors structured the workflow to require only a simple initial setup executed via a command in the R language. Following this brief setup, the platform provided multiple visualization and analytical functions within an intuitive graphical user interface. This approach allowed users to interactively explore dimensionality reduction, cellular clustering, differential gene expression, and functional enrichment without writing additional code. Furthermore, the developers incorporated an optional split view feature that segregated data by metadata categories, which facilitated side-by-side comparative insights across different experimental conditions or time points. The software successfully enabled experimental biologists to navigate pre-processed datasets, accurately identify cellular subpopulations, and generate customizable, high-quality visualizations. This strategic separation of computational processing and data exploration significantly reduced the time and effort required for data interpretation. To maximize accessibility, the developers designed the tool to support cross-platform use, functioning seamlessly across Windows, Mac, and Linux web servers. Ultimately, CellAnalyst provided an accessible and effective solution that democratized complex bioinformatics workflows and accelerated cellular research by bridging the gap between sophisticated analytical algorithms and the broader biological community. The software was made freely accessible at https://bis.zju.edu.cn/cellanalyst.

Keywords:

single cell, visualization, bioinformatics, software

Abstract ID: **22** Type: **Flash Talk**

Integrative multi-layer bioinformatics of transcriptomic data for elucidating molecular mechanisms of action of polyphenol-rich extract on cardiometabolic health

Tatjana Ruskovska[1], Dragan Milenkovic[2]

[1] *Goce Delcev Univesity, Stip*

[2] *North Carolina State University*

Contact e-mail: **tatjana.ruskovska@ugd.edu.mk**

Polyphenols are bioactive compounds of plant origin, widely present in plants. Importantly, several hundreds of them are relevant for human nutrition. Numerous studies have shown that adequate daily polyphenol intake is associated with a beneficial effect on cardiometabolic health in humans. Despite the abundance of scientific evidence, the molecular mechanisms underlying these effects are not completely understood, which results from the diversity of polyphenols in the human diet, their complex metabolism in the human body, and finally, the complexity and interindividual variability in their cellular effects.

Advancements in our understanding of polyphenols and their metabolism in the human body, should be complemented with detailed, in-depth studies aimed at investigating their cellular and molecular targets. In this context, modern omics technologies offer considerable exploratory advantages over classical targeted analytical approaches. Here, the latest advancements in global transcriptomics represent a particular challenge since they provide a wealth of data that includes not only the protein-coding transcriptome, but also non-coding RNAs such as miRNAs, lncRNAs, and circRNAs.

Recent human intervention study with a polyphenol-rich extract has demonstrated modulation of more than 200 protein-coding genes, six miRNAs, and more than 200 lncRNAs. By analyzing the targets of non-coding RNAs and integrating these data with the data on the modulation of global gene expression of protein-coding RNAs, we identified the key affected cellular pathways, including the cAMP signalling pathway, MAPK signalling pathway, Ras signalling pathway, adrenergic signalling in cardiomyocytes, and retrograde endocannabinoid signalling. We further built and visualized an integrative network of differentially expressed protein-coding genes, TFs that potentially regulate their expression, miRNAs and lncRNAs whose expression was modulated by the polyphenol-rich extract, and their targets. With this analysis, we identified key genes that are modulated at all three levels: mRNA, miRNA, and lncRNA.

In conclusion, integrative multi-layer bioinformatics of transcriptomic data has the capacity to provide evidence for key molecular targets underlying the beneficial health effects of polyphenols. In future studies, these data can be utilized for the identification of relevant gene polymorphisms that are potentially involved in determining interindividual variability in the effects of polyphenols, thereby paving the way towards personalized dietary recommendations.

Keywords:

polyphenols, transcriptomics, integrative network analysis

Abstract ID: **32** Type: **Flash Talk**

Identification and functional characterization of tumor-suppresive microRNAs with potential key roles in breast cancer progression and metastasis

Aleksandra Nikezić[1], Jovana Jovankić[1], Jovana Stevanović[2], Miloš Brkušanin[3], Olgica Nedić[2], Sanja Goč[4], Zorana Dobrijevic[2]

[1] *Faculty of Science, University of Kragujevac*

[2] *Institute for the Application of Nuclear Energy*

[3] *University of Belgrade - Faculty of Biology*

[4] *University of Belgrade, Institute for the Application of Nuclear Energy, INEP*

Contact e-mail: **zorana.dobrijevic@inep.co.rs**

Downregulation of tumor-suppressive microRNAs in breast cancer (BC) triggers molecular disturbances underlying cancer initiation, disease progression and metastasis. However, key microRNAs players acting as vital breaks on BC aggressiveness, with crucial regulatory roles in cell migration, epithelial-mesenchymal transition (EMT), invasion, and metas-tasis, remain underinvestigated. Our goal was to uncover candidate tumor-suppresive microRNAs associated with metastatic potential of BC via an in silico approach, which combined data from microRNA profiling studies on cell lines relevant to BC invasiveness, as well as on primary vs. metastatic BC. A minimal set of candidate microRNAs was determined from the intersection of downregulated sets from multiple eligible studies, confirmed by qPCR, assessed for potential clinical relevance in BC and functionally characterized. Predicted and experimentally valdated targets of these microRNAs were identified (MiRTargetLink) and subjected to gene ontology (GO), KEGG pathway, and cancer hallmark enrichment analysis. A microRNA-mRNA interaction network was constructed (MiRNet), and mRNA profiling data was integrated, in order to enable the identification of hub genes and upregulated microRNA targets mediating BC EMT, invasion, and metastasis. Three microRNAs emerged as plausible BC metastasis-related candidates: miR-141-3p, miR-200c-3p, and miR-205-5p. The downregulation of these microRNAs in cell lines with higher invasiveness was confirmed by qPCR, while an association with negative clinical features was also noted. Both members of miR-200 family demonstrated a correlation with the expression of key regulators of EMT in BC tissues. Hub genes from DE-microRNA-mRNA network showed enrichment in pathway and GO terms related to malignant phenotype, while "tissue invasion and metastasis"was one of the top enriched cancer hallmarks. These results support the hypothesized functional relevance of analyzed microRNAs in BC.

Acknowledgement:

This research was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Agreement nos. 451-03-33/2026-03/200019 and 451-03-33/2026-03/200122) and aligns with United Nations Sustainable Development Goal 3 (Good Health and Wellbeing) of the 2030 Agenda.

Keywords:

microRNAs, invasion, EMT, metastasis, BRCA

Abstract ID: **43** Type: **Flash Talk**

Diet reconstruction of a woolly mammoth calf using multiple metagenomic approaches

Kseniia Poliakova[1], Maxim Cheprasov[2], Gavril Novgorodov[2], Alexander Rakitko[1,3], Fedor Sharko[4], Artem Nedoluzhko[2,4,5]

[1] *Genotek Center: AI in Personalized Medicine, ITMO University, Saint Petersburg, Russia*

[2] *Mammoth Museum, North-Eastern Federal University, Yakutsk, Russia*

[3] *Genotek Ltd., Moscow, Russia*

[4] *European University at Saint Petersburg, Saint Petersburg, Russia*

[5] *National Research University Higher School of Economics, Moscow, 117418, Russia*

Contact e-mail: **poliakovakseniaa1@gmail.com**

Reconstructing diet from ancient metagenomic data is complicated by short DNA fragments, contamination of reference genome sequences, and ambiguous taxonomic classification. In this study, we examined plant DNA preserved in the rectal contents of a woolly mammoth (*Mammuthus primigenius*) calf to characterize its diet using multiple computational approaches.

Paired-end sequencing reads were trimmed to remove adapters and aligned to the human genome to eliminate contamination. The remaining reads were analyzed with three main strategies: k-mer-based taxonomic classification, alignment to a reference database of plastid genomes, and lowest common ancestor (LCA) assignment based on multiple alignments. To further improve specificity, reads mapping to bacterial genomes were filtered out.

Consistent signals across all methods were observed for taxa in the family Cyperaceae, particularly the genus *Carex*, and for *Comarum palustre*. These taxa emerged as the most reliable components of the dataset. In contrast, other taxa showed variability depending on the analytical method used. For instance, sequences initially attributed to *Arisaema ringens* were substantially reduced after bacterial filtering, indicating that these signals were likely present due to reference database contamination rather than actual dietary components. Multiple signals related to Poaceae were also detected but were distributed across multiple taxa, likely due to the limited representation of relevant species in reference databases, which hindered confident identification at lower taxonomic levels. A similar pattern was observed for *Carex*, where reads mapped to multiple related species, reflecting the absence of an exact reference genome.

The results indicate that sedges were a major part of the mammoth's diet. This study highlights the importance of employing multiple analytical approaches and carefully considering contamination and limitations of reference databases to minimize taxonomic misclassification in ancient metagenomic studies.

Keywords:

ancient DNA, metagenomics, Mammuthus primigenius

Abstract ID: **49** Type: **Flash Talk**

A multi-omic digital twin network reveals C3-glycosylation as an actionable driver of latent metaflammation

Maxym Shmatkov[1], Janko Diminic[1], Jurica Zucko[1], Olga Gornik[2], Valentina Borko[2], Toma Keser[2], Andrija Karačić[3], Robert Keser[4], Antonio Starcevic[1]

[1] *University of Zagreb Faculty of Food Technology and Biotechnology*

[2] *University of Zagreb Faculty of Pharmacy and Biochemistry*

[3] *Centar Mikrobiom d.o.o.*

[4] *In silico d.o.o.*

Contact e-mail: **astar@pbf.hr**

Metabolic syndrome represents a progressive continuum of systemic dysfunction, yet current diagnostic paradigms rely on late-stage clinical biomarkers (e.g., fasting glucose, HOMA-IR) that fail to capture latent metaflammation. Furthermore, standard population-level multi-omics analyses are frequently confounded by immutable physical and demographic baseline variances. Here, we present a graph-based, synthetic digital twin architecture applied to a well-characterized multi-omic clinical cohort (N=300). By employing a K-nearest neighbor (K-NN) sex-stratified matching algorithm that strictly anchors patients to their exact physical baselines, we mathematically isolated pure metabolic pathophysiology from background biological noise.

Applying this computational framework, we identified a highly prevalent latent-risk group—traditionally misclassified as healthy by standard clinical criteria—that exhibits near-complete phenotypic overlap with the manifest disease cohort. We demonstrate that this latent metabolic decline is characterized and driven by profound shifts in the glycosylation profile of Complement Component 3 (C3), a central protein in the innate immune system. Principal component analysis isolated a discrete C3-glycan inflammatory axis dominated by five specific glycoforms (CIIIGIRMN1N2H10, CIIIGIRMN1N2H8, CIIIGIRMN1N2H9, CIIIKTVLT1N2H5, and CIIIKTVLT1N2H6). Crucially, this C3-glycosylation shift precedes traditional dysglycemia and correlates directly with downstream systemic immune activation and targeted gut microbiome perturbations, specifically the depletion of butyrate-producing Lachnospirales.

To translate these high-dimensional omics findings into clinical utility, we integrated an AI-assisted semantic ontology to generate a deterministic Clinical Decision Support System (CDSS). This engine autonomously separates structural psychological traits from actionable physiological behaviors, generating personalized, biologically realistic intervention targets derived directly from a patient's healthy synthetic twins. Ultimately, our findings establish specific C3-glycan modifications as critical, early-stage biomarkers of metabolic decline, and validate the synthetic digital twin network as a robust, scalable framework for precision preventive medicine.

Acknowledgement:

This research was funded by the European Union (NextGenerationEU) project "The glycome and microbiome as markers of dietary impact on the health of women of reproductive age"within the program 'Targeted Scientific Research'(NPOO.C3.2.R3-I1.04.0073). The work presented has been filed as a EU patent application EP26173207.7 "METHOD AND SYSTEM FOR EARLY DETECTION OF LATENT METABOLIC RISK BASED ON COMPLEMENT C3 GLYCOSYLATION AND HEALTHY-TWIN RECOMMENDATION ANALYSIS."

Keywords:

Digital twin, glycomics, metaflammation

Abstract ID: **69** Type: **Flash Talk**

A compact genotype sketch for cross-platform sample matching

Giulgaz Muradova[1], Fedor Konovalov[1]

[1] *Independent Clinical Bioinformatics Laboratory, Moscow, Russia*

Contact e-mail: **gm@clinbio.ru**

Reliable comparison of sequencing datasets across laboratories has become increasingly important as patients often undergo repeated exome or genome sequencing on different platforms. Existing sample-matching approaches usually require dense genome-wide genotypes, tolerate sparse or uneven exome data poorly, generate impractically large signatures, or preserve enough genotype information to support unintended identity or relatedness analyses. We therefore developed a compact, lossy genotype sketch designed to combine exome compatibility, tolerance to no-calls and limited genotyping errors, small hash-string footprint, and deliberate reduction of recoverable variant-level information.

We selected 462 common, high-coverage, population-broad single-nucleotide polymorphisms from exome-accessible regions. Candidate variants were filtered for high allele count and intermediate allele frequency across multiple populations, restricted to protein-coding exons, and pruned by genomic distance to reduce linkage between markers. Genotypes at the selected loci were transformed using a custom many-to-one encoding scheme. The method shuffled the genotype string, encoded adjacent SNP pairs into three ambiguity-preserving classes, and compressed the resulting class string into a fixed-length representation suitable for rapid comparison and two-dimensional barcode workflows. In parallel, we evaluated established locality-sensitive hashing (LSH) approaches on the same genotype representations and found them either impractical due to signature size or less effective in discriminating matching from non-matching samples. This design preserved similarity between samples while rendering direct recovery of individual SNP genotypes practically infeasible.

We evaluated the approach using publicly available paired whole-genome and whole-exome data from the 1000 Genomes Project and multiple publicly available sequencing datasets of the reference sample NA12878 generated on different platforms. The resulting sketches consistently separated matching from non-matching samples and remained comparable under substantial missingness, supporting their use with heterogeneous exome data. The compact output enabled practical storage in laboratory reports and allowed straightforward report-to-sample verification without sharing raw genotypes.

This approach provided a practical external quality-control layer for sample continuity verification, detection of sample swaps, and cross-platform comparison of sequencing datasets. Because the representation was intentionally lossy, it should be interpreted as a QC checksum rather than a forensic identifier. Further evaluation of relatedness leakage and privacy properties is ongoing.

Keywords:

genotyping, exome, LSH, privacy

Abstract ID: **73** Type: **Flash Talk**

Analysis of total transcriptome from brain tissue of progressive multiple sclerosis patients identifies ferroptosis-related genes relevant for distinction of inactive from chronic active white mater lesions

Milan Stefanović[1], Ivan Jovanović[1], Nađa Trklja[1], Jovana Kuveljić[1], Aleksandra Stanković[1], Maja Živković[1]

[1] *Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, University of Belgrade, Department of Radiobiology and Molecular Genetics*

Contact e-mail: **nadja.trklja@vin.bg.ac.rs**

Chronic active lesions (CAL) are the main hallmarks of progressive multiple sclerosis (progMS) and are characterized by a demyelinated, inactive core surrounded by an iron-enriched rim. Change of lesion phenotype into stable inactive lesions (IL), gliotic white mater scars, relates to iron loss. Recently, ferroptosis, a form of iron-dependent programmed cell death characterized by iron overload, has been linked to MS. Therefore, we aimed to identify genes associated with ferroptosis relevant to separate CAL from IL, by analyzing total transcriptome data from post-mortem brain lesions of progMS patients. Whole RNA data (containing mRNA and miRNA counts) from GEO database GSE138614 (16 CAL and 14 IL lesions from 10 progMS patients) were analyzed. Weighted Gene Coexpression Network Analysis (WGCNA) was performed, identifying expression modules most correlated with IL (Tan: r=0.6, p=0.0033, ngenes=301; Blue: r=-0.66, p=0.0009, ngenes=2861). Using DESeq2 (Log2FC>|0.6|, p.adj<0.05) we've identified 2728 differentially expressed genes (DEGs) coding both proteins and miRNAs in the IL vs. CAL comparison. We've intersected DEGs, ferroptotsis-related genes from human biological sources (FerrDbV3: n=1196) and the selected WGCNA module genes to extract ferroptotsis-related genes associated with IL (n=43; 17 ferroptosis drivers, 19 suppressors and 5 with dual roles). STRING (interaction score≥0.4; all interactions) was used to identify protein interactions between the 43 identified ferroptotsis-related genes. Interaction networks were analyzed via Cytoscape plug-in cytoHubba. *PKM*, *MDH2*, *GOT1*, *PPARGC1A*, *PGAM1*, *KCNA1* and *CKB* were identified as the most interconnected, hub, protein-coding genes. By employing miRTarBase-v9.0, we've identified miRNAs coded by DEGs *MIR612*, *MIR6852* and *MIR17HG* as potential regulators of key ferroptosis-related genes. *MIR17HG*-coded miRNAs had the greatest regulatory potential (targeting 14 key genes) while being a key ferroptosis-related gene and known suppressor of ferroptosis. Every hub gene and potential miRNA regulator had a Receiver Operating Characteristic Area Under Curve value ≥0.8, meaning that they were valid discriminators between CAL and IL. We conclude that the protein-coding genes *PKM*, *MDH2*, *GOT1*, *PPARGC1A*, *PGAM1*, *KCNA1*, and *CKB*, along with the miRNA host gene *MIR17HG*, potentially contribute to lesion phenotype change to IL in progMS through ferroptosis-related mechanisms. Further experiments are needed to validate this finding.

Keywords:

Ferroptosis, multiple sclerosis, lesions, progression

Abstract ID: **92** Type: **Flash Talk**

RefusalBench: Why refusal rate misranks frontier LLMs on biological research prompts

Lukas Weidener[1], Marko Brkic[1], Mihailo Jovanovic[1]

[1] *Applied Scientific Intelligence*

Contact e-mail: **mbrkic16@gmail.com**

Frontier large language models are increasingly deployed as orchestration backbones for biological research workflows, yet no shared evidence base exists for comparing their refusal behaviour on legitimate research prompts. We introduce RefusalBench, a matched-triple benchmark of 141 prompts in 47 bundles that holds task framing constant while varying only biological risk tier (benign, borderline, dual-use), enabling tier-conditioned comparisons robust to subdomain confounding. A 15-prompt should-refuse positive-control module establishes per-model calibration floors; three models fail to refuse even these prompts. Evaluating 19 frontier models in the inaugural May 2026 snapshot, strict refusal rates span 0.1% to 94.6% on identical prompts. Jurisdiction does not predict refusal in this snapshot (Mann–Whitney U, p = 0.393; EU n = 1, US bimodal); provider identity does, with Anthropic's API stack predicting refusal at OR = 21.03 (95% CI: 14.58–30.34 prompt-clustered; 5.70–77.55 under model-clustered GEE). This effect is best read as access-path-level rather than model-weight-level: 99.8% of Anthropic's strict refusals carry the same safety_policy adjudicated reason code, consistent with a small set of canonical refusal templates rather than case-by-case model reasoning. Strict refusal rate misranks safety calibration: Grok 4.20 achieves the highest tier discrimination (Youden's J = 0.787) while ranking only seventh by overall refusal rate, and Claude Opus 4.7's J dropped 65% from prior versions with no improvement in dual-use detection. Nine of 18 frontier models exhibit a hedge-but-help partial-compliance pattern at dual-use tier that binary refusal metrics cannot detect.

Keywords:

LLMs, biosecurity, safety benchmarking

Abstract ID: **96** Type: **Flash Talk**

Cross-workflow signal reporting in pan-viral metagenomics

Milana Djonovic[1], Andrew Norgan[1], Alexej Abyzov[1]

[1] *Mayo Clinic*

Contact e-mail: **milana.djonovic@gmail.com**

Available viral detection tools provide broad taxonomic coverage but often overreport low-abundance viral targets and fail to address contamination-derived signals. We developed and evaluated a contamination-aware pan-viral metagenomic workflow focused on contamination removal and interpretable taxonomic reporting. A clinical control cohort of 62 samples with enriched viral DNA/RNA across multiple tissue types was analyzed using both a clinical benchmark workflow and our in-house workflow, which removes contamination-associated signals through mapping against artificial vector sequences. Initial benchmarking demonstrated strong concordance between workflows. However, read counts frequently differed because results were reported at different taxonomic levels (strain, species, genus, etc.), highlighting the need for taxonomic collapsing. In general, clinical interpretation is often better served by genus- or group-level identification than strain-level resolution, thus taxonomic collapsing can improve interpretability and also amplify detection signal. We therefore identified the taxonomic levels at which the workflows converged and implemented targeted realignment using RefSeq references for the corresponding categories. This enabled aggregation of read counts within groups and quantitative comparison of read support across workflows. Together, these findings support a two-step framework using a compact reference database for rapid screening followed by targeted realignment for signal refinement and improved taxonomic resolution.

Keywords:

clinical metagenomics, contamination, databases, viruses

Abstract ID: **112** Type: **Flash Talk**

Temporal dynamics of codon usage in SARS-CoV-2 proteins

Biljana Stojanović[1], Saša Malkov[2], Miloš Beljanski[3], Nenad Mitić[2]

[1] *Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia*

[2] *Faculty of Mathematics, University of Belgrade, Belgrade, Serbia*

[3] *Institute for General and Physical Chemistry, Belgrade, Serbia*

Contact e-mail: **bstojanovic@mi.sanu.ac.rs**

Codon usage bias (CUB) reflects the non-random usage of synonymous codons and represents an important feature associated with viral evolution, host adaptation, and translational efficiency. We analysed CUB in approximately 5 million SARS-CoV-2 protein-coding sequences collected between 15 December 2019 and 22 May 2023. The analysis was performed on the complete dataset of protein amino acid sequences and corresponding coding sequences, as well as on 20 temporal subsets obtained by partitioning isolate collection dates into discrete time intervals. Four proteins —*surface glycoprotein*, *nucleocapsid phosphoprotein*, *ORF1a polyprotein*, and ORF1ab polyprotein —were selected for detailed analysis, each exhibiting more than 3% unique coding sequences relative to the total number of sequences.

For each protein type, we analyzed (i) codon usage patterns across all corresponding coding sequences using global codon frequencies and Relative Synonymous Codon Usage (RSCU) values as codon-specific CUB measures, (ii) data derived data derivedfrom alignments of individual amino acid sequences against the corresponding amino acid sequence of the reference isolate, and (iii) coding sequences corresponding to aligned amino acid sequences.

Low-frequency codons were identified across the analyzed datasets. Specific positions within each reference protein at which significant changes in codon frequencies were observed over time were identified. The analysis also examined codon usage patterns in relation to the WHO-defined lineage associated with each isolate. In addition, CUB was analyzed for each protein type at the level of complete coding sequences using the gene-specific measures Effective Number of Codons (ENC) and Relative Codon Bias Score (RCBS). These measures provided an objective characterization of CUB for each analyzed protein type and complemented the codon-specific analysis based on global codon frequencies and RSCU values.

The results indicate that the most pronounced changes in codon usage over time were observed in *surface glycoprotein* and *nucleocapsid phosphoprotein*. Some synonymous codons show relatively stable frequencies over time, whereas others exhibit decreases in codon abundance at different stages (early, intermediate, or late) and with varying magnitudes.

Keywords:

SARS-CoV-2, codons, dynamic of change

Abstract ID: **114** Type: **Flash Talk**

Cobalt chloride affects neural precursor cell differentiation: A transcriptional study

Stefan Lazic[1], Marija Lazic[1], Luka Bojic[1], Jelena Pejic[1], Vanda Balint[1], Simona Lapcevic[1], Mirjana Novkovic[1], Mila Ljujic[1], Milena Stevanovic[2], Danijela Drakulic[1], Danijela Stanisavljevic Ninkovic[1], Marija Mojsin[1], Milena Milivojevic[1], Isidora Petrovic[1], Marija Schwirtlich[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Serbian Academy of Sciences and Arts*

Contact e-mail: **stefan.lazic@imgge.bg.ac.rs**

Various pathological conditions, resulting from oxygen deprivation during development and adulthood, impact neurogenesis, which in turn impairs different cognitive functions and contributes to the onset of age-related neurodegenerative diseases. Therapeutic options remain limited, and many neuroprotective agents that appear promising in preclinical research have not succeeded in human clinical trials. The present study aimed to evaluate how the hypoxia-mimicking agent CoCl2 influences the neurogenic potential of human neural precursor cells (NPCs), reflecting early embryonic development stages and adult neurogenesis.

We employed next-generation sequencing and bioinformatic analysis to functionally characterize the transcriptome and assess treatment effects in NPCs. Six cDNA libraries were generated, with three from each experimental group of cells treated with either 300 μM CoCl2 or deionized water for 48 hours. Transcriptomic analysis identified 2,925 differentially expressed genes, of which 1,065 were upregulated and 1,860 downregulated in treated NPCs. Functional analysis showed enrichment of upregulated genes in processes such as hypoxia-related pathways, metabolic reprogramming, and transmembrane transport. Furthermore, downregulated genes were associated with cell cycle progression and axon guidance. Notably, genes related to immunity and apoptotic cell death were unaffected. Additionally, downregulated genes were identified in gene sets from neural stem and progenitor cells across different regions of the human brain, suggesting that hypoxia may affect specific characteristics and functions of these cells. Finally, by combining results from quantitative real-time PCR and double immunocytochemical analysis of cells differentiated for 10 days post-treatment, we showed that cobalt chloride affects the neuronal differentiation of NPCs in multiple ways: it decreases the number of differentiated neurons but also enhances their maturity.

Taken together, our findings provide a basis for understanding the molecular mechanisms in the human brain in response to hypoxic stress. The key hub genes, core regulatory programs, and pathways identified in our study are designated as potential targets for neuroprotective strategies following hypoxia-related brain injuries.

Keywords:

Hypoxia, neurogenesis, neurodegeneration, transcriptome, NT2/D1.

Acknowledgement:

This research was supported by Ministry of Science, Technological Development and Innovation of the Republic of Serbia (No: 451-03-33/2026-03/200042). The authors would like to thank Sasa Todorovic and Aleksandra Vitkovac for all the help and support with RNA sequencing and initial data processing.

Abstract ID: **121** Type: **Flash Talk**

Predicting nephrotoxicity from cell morphological perturbations using high content image-based profiling

Natacha Cerisier[1], Tahreem Abbasi[1], Marlène Wedler[2], Shu Liu[3], Olivier Taboureau[1]

[1] *Université Paris Cité, INSERM U1133, CNRS UMR 8251, Paris, France*

[2] *German Centre for the Protection of Laboratory Animals (Bf3R) and Experimental Toxicology, German Federal Institute for Risk Assessment (BfR)*

[3] *German Centre for the Protection of Laboratory Animals (Bf3R) and Experimental Toxicology, German Federal Institute for Risk Assessment (BfR), Berlin, Germany.*

Contact e-mail: **natacha.cerisier@inserm.fr**

Drug-induced kidney injury is a leading cause of clinical trial attrition, highlighting the urgent need for early, reliable nephrotoxicity prediction tools. Current *in vitro* and *in silico* approaches often fell short in capturing the complex cellular responses underlying renal toxicity; a gap that morphological profiling may help bridge.
Here, we leveraged *Cell Painting PLUS*, a high-content image-based profiling assay, to investigate whether compound-induced morphological perturbations can serve as predictive signatures of nephrotoxic potential. By integrating nephrotoxicity annotations from multiple curated sources with rich Cell Painting PLUS profiles, we explored a novel application of phenotypic profiling beyond its traditional use in mechanism-of-action studies.
Starting from a raw dataset, rigorous feature selection and compound curation pipelines were applied to maximize data quality, leading to 90 annotated compounds. Unsupervised exploration via hierarchically-clustered heatmaps (*clusterMap*) revealed structured patterns distinguishing nephrotoxic from non-nephrotoxic compounds, and identified morphological features driving this separation.
Building on these insights, we trained and benchmarked several machine learning classifiers. A Random Forest model achieved the best predictive performance, with an AUC of 82%, a sensitivity of 87%, and a specificity of 63%. These results provided a proof of concept that high-content image-based profiles encode biologically meaningful signals relevant to nephrotoxicity —and laid the groundwork for scalable, image-driven toxicity screening pipelines.

Keywords:

Nephrotoxicity, Cell-Painting, Machine-learning, high-content-imaging

Acknowledgement:

This project receives funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement No. 964537 (RISK-HUNT3R), and it is part of the ASPIS cluster.

Abstract ID: **123** Type: **Flash Talk**

Cytokine profiles as promising predictive biomarkers for glucocorticoid therapy response in inflammatory bowel disease

Nikola Kotur[1], Sanja Dragašević Vučićević[2], Uršula Prosenc Zmrzljak[3], Saša Šterpin[3], Katarina Krstajić[1], Marina Jelovac[1], Bojan Ristivojevic[1], Branka Zukic[1], Biljana Stankovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Medical Faculty, University of Belgrade*

[3] *Labena company*

Contact e-mail: **nikola.kotur@imgge.bg.ac.rs**

Glucocorticoids (GCs) are widely used for the induction of remission in inflammatory bowel disease (IBD), but many patients exhibit resistance or dependency. This highlights the need for biomarkers that can predict therapeutic response. Cytokines are central mediators of intestinal inflammation and may contribute to variability in GC responsiveness. Machine learning modeling such as penalized linear models that use multi-variable biological data could be particularly useful for selecting predictive features and identifying novel biomarkers of GC response.

Adult patients with active Crohn's disease or ulcerative colitis initiating systemic prednisone therapy were prospectively recruited. Plasma samples collected before therapy were analyzed using a Bio-Plex Pro Human Cytokine 48-Plex panel. After quality control, 27 cytokines were retained for analysis. Patients were classified as GC good responders or GC poor responders, which include GC-resistant and GC-dependent patients.Penalized logistic regression models were applied to evaluate the predictive potential of cytokine and clinical variables.

Thirty-seven patients were included in the final analysis. Elevated baseline levels of macrophage migration inhibitory factor (MIF), IL-9, and GRO-α were significantly associated with poor GC response, while TNF-β, MIP-1β, SCF, FGF, and IL-16 showed borderline associations. Increased inflammatory cytokine burden, assessed as the sum of standardized cytokine values, was also associated with poor GC response. Principal component analysis showed cytokine profile differences between response groups. Penalized logistic regression models achieved the best predictive performance using a limited number of variables, particularly combinations including MIF and baseline clinical activity, with mean test accuracies up to approximately 73%.

Elevated pro-inflammatory cytokine profiles, particularly increased MIF, IL-9, and GRO-α, are associated with the poor GC response in IBD. Combining cytokine markers with the baseline clinical activity may support the development of predictive models for personalized GC therapy in IBD.

Keywords:

IBD, cytokines, glucoocrticoids, machine learning

Acknowledgement:

The authors acknowledge Labena d.o.o. (Ljubljana, Slovenia) for providing a grant and expertise to perform cytokine measurements.

Abstract ID: **3** Type: **Poster Presentation**

Transcriptomic reanalysis reveals membrane-associated pathways involved in fullerenol C60(OH)24 cytotoxicity

Mariana Seke[1], Ivan Jovanović[2], Nadja Trklja[2], Nataša Mačak Stefanović[2], Jovana Kuveljić[2], Maja Živković[2], Aleksandra Stanković[2]

[1] *Institute of Nuclear Sciences "Vinca", University of Belgrade*

[2] *"Vinca" Institute of Nuclear Sciences, University of Belgrade*

Contact e-mail: **nadja.trklja@vin.bg.ac.rs**

Fullerenols C60(OH)n are widely regarded as biocompatible nanomaterials at low concentrations, but higher doses may induce cytotoxic effects. Previous studies indicate mitochondrial damage as a mechanism of toxicity, yet the broader molecular pathways involved remain poorly characterized.
We performed a transcriptomic reanalysis of publicly available dataset GSE3364 to investigate gene expression changes in human umbilical vein endothelial cells treated with fullerenol C60(OH)24 (100 μg/ml, 24 h). Differentially expressed genes (DEGs) were identified using a linear model with empirical Bayes statistics in the limma R package. Functional pathway and biological process enrichment analyses were conducted with iPathwayGuide (Advaita Bio), using thresholds of $|\log2 \text{ fold change}| \geq 0.6$ and $p \leq 0.05$.
Analysis identified 2161 DEGs, revealing enrichment of pathways related to membrane-associated processes. Notably, cell adhesion molecules, neuroactive ligand–receptor interaction, and other membrane-linked signaling pathways were affected. These findings suggest that fullerenol-induced cytotoxicity may involve perturbations of membrane integrity, receptor-mediated signaling, and cellular adhesion, complementing previous observations of mitochondrial dysfunction. Additional pathways, including those related to extracellular matrix organization, were also implicated, indicating a multi-faceted cellular response to high-dose fullerenol exposure.
Our results highlight membrane-associated pathways as potential mediators of fullerenol C60(OH)24 cytotoxicity in endothelial cells. Understanding these mechanisms is crucial for optimizing biomedical applications, including dosing strategies, treatment duration, and administration routes. This study provides new insight into the molecular effects of fullerenols and supports further research into their safe use in nanomedicine.

Acknowledgement:

This research was supported by the Serbian Ministry of Science, Technological Development, and Innovation, Grant No. 451-03-33/2026-03/ 200017

Keywords:

fullerenol, transcriptomics, DEGs, pathways, HUVECs,

Abstract ID: **5** Type: **Poster Presentation**

Long-range chromatin interaction dynamics and molecular aging signatures in schizophrenia

Kirill Ulianov[1], Maria Molodova[1], Diana Zagirova[2], Kirill Morozov[1], Polina Morozova[3], Olga Efimova[4], Sergey Ulianov[5], Sergey Razin[5], Philipp Khaitovich[4], Ekaterina Khrameeva[1]

[1] *Center for Bio- and Medical Technologies, Skoltech, Moscow, Russia*

[2] *Institute for Information Transmission Problems (the Kharkevich Institute), Russian Academy of Sciences, Moscow, Russia*

[3] *Moscow Institute of Psychoanalysis, Moscow, Russia*

[4] *Vladimir Zelman Center for Neurobiology and Brain Rehabilitation, Skoltech, Moscow, Russia*

[5] *Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia*

Contact e-mail: **kirill.ulianov.correspondence@gmail.com**

Schizophrenia is a severe neuropsychiatric disorder affecting cognitive, emotional, and behavioral functions. Despite a long history of research, the genetic basis of schizophrenia remains incompletely understood. The majority of genetic variants associated with schizophrenia risk have small effect sizes and are located in non-coding regions of the genome. Many of these variants are thought to reside within regulatory elements, suggesting that the disorder may involve disturbances of gene regulation. In this context, the regulatory landscape of the genome plays a crucial role in controlling gene activity. Regulatory interactions are strongly constrained by the three-dimensional organization of chromatin. Therefore, studying genome architecture provides an important framework for understanding mental disorders. Here, we investigated chromatin architecture and transcriptomic profiles in schizophrenia.

To obtain information on spatial genome contacts, we applied the Hi-C protocol to postmortem cerebral cortex samples from healthy donors and individuals with schizophrenia. Nuclei were isolated from frozen tissue and separated into neuronal and non-neuronal populations using fluorescence-activated cell sorting. Hi-C libraries were prepared for both populations and subsequently sequenced. In addition to the Hi-C data generated in this study, publicly available RNA-seq datasets from the PsychENCODE consortium were analyzed.

Analysis of Hi-C maps showed that local chromatin structures, including topologically associating domains and chromatin loops, are preserved in schizophrenia. However, analysis of genome-wide contact probability revealed altered age-dependent dynamics of chromatin interactions in neuronal samples of healthy donors. Meanwhile, contact probability did not differ between young and old donors with schizophrenia at any genomic distance. Fluorescence microscopy of cortical sections demonstrated an enlargement of neuronal nuclei in schizophrenia. Transcriptomic analysis indicated aging-related molecular changes in schizophrenia samples. A gene expression-based predictor of chronological age revealed a significant increase in predicted transcriptomic age in schizophrenia compared with controls.

These results indicate that schizophrenia is associated with an enlargement of neuronal nuclei, which may contribute to changes in chromatin interactions. The dynamics of these alterations appear to be age-dependent. Consistent with this observation, transcriptomic analysis also supports the presence of aging-associated processes in schizophrenia.

Acknowledgement:

This study was supported by the Russian Science Foundation (grant number 25-71-20017).

Keywords:

schizophrenia, chromatin, aging

Abstract ID: **9** Type: **Poster Presentation**

AI-driven concordance engines as a decision-support layer for veterinary import compliance

Jovana Stankic[1]

[1] *Dechra, Genera Pharma d.o.o*

Contact e-mail: **jovanastankic93.jj@gmail.com**

Regulatory compliance in the importation of veterinary medicinal products remains one of the most labor-intensive and error-prone activities within Regulatory Affairs, particularly in non-EU markets where harmonized standards, centralized databases, and consistent specification formats are often absent. A significant bottleneck in these workflows is the manual comparison of Certificates of Analysis (CoA) with approved release specifications, complicated further by jurisdiction-specific rules regarding parameter acceptance, shelf-life calculations, and expiry-date tolerances. These inconsistencies introduce operational inefficiencies and increase the risk of non-compliance during product release and importation.

This paper proposes an AI-Driven Concordance Engine as an innovative decision-support layer designed to automate detailed compliance checks within global veterinary import processes. The concept integrates natural language processing to extract analytical parameters from CoAs, rule-based logic to apply regulatory constraints, and machine-learning-supported similarity scoring to identify missing parameters, discrepancies, or values that fall outside jurisdiction-specific requirements. The system flags deviations such as stricter-than-approved limits or country-specific acceptance criteria, directly supporting import eligibility decisions.

Drawing on anonymized real-world regulatory workflows, the study demonstrates that Concordance Engines significantly reduce manual effort, enhance consistency across regulatory assessments, and create a fully traceable and auditable compliance record. Importantly, the framework incorporates a risk-based human-in-the-loop model, ensuring that AI augments rather than replaces regulatory expertise. By functioning as an intelligent control mechanism, the Concordance Engine strengthens regulatory decision-making, shortens operational timelines, and supports scalable compliance across diverse markets.

While the framework is illustrated through its application to non-EU import compliance, its relevance extends to EU regulatory processes as well. Specifically, the model offers value in post-approval variation (PAV) management, where changes to quality attributes may influence approved specifications and subsequently CoA evaluation. By providing systematic, structured assessments of such changes, the Concordance Engine enhances cross-functional alignment between Regulatory Affairs and Quality functions and promotes more efficient regulatory oversight.

Overall, AI-Driven Concordance Engines represent a first-in-class RegTech solution for veterinary Regulatory Affairs, enabling a shift from manual verification to intelligent, technology-supported regulatory control. This approach offers a scalable pathway for modernizing data-intensive regulatory processes while ensuring that human expertise remains central to final decision-making.

Keywords:

Veterinary Regulatory Affairs, Regulatory Technology

Abstract ID: **10** Type: **Poster Presentation**

Comparative evaluation of the regenerative efficacy of Lucilia sericata larval secretions across varied storage conditions.

Kübra Tuğtekin[1]

[1] *İstanbul Cerrahpaşa University Faculty of Medicine, Department of Medical Biology*

Contact e-mail: **tugtekin.kubra@hotmail.com**

Comparative analysis of the in vitro wound healing potential of larval secretions of Lucilia sericata under different torage conditions content.
The study evaluated the effects of various storage temperatures and durations on the bioavailability and in vitro wound healing potential of secretions derived from first-stage Lucilia sericata larvae. To address the logistical limitations of using fresh secretions, the research aimed to determine optimal storage conditions suitable for clinical applications. Larval secretions were maintained at +4 ℃, −20 ℃, and room temperature for durations of 24 and 48 hours. In vitro wound healing activity was assessed using a scratch assay performed on the 3T3 murine fibroblast cell line. Cell migration and wound closure rates were quantified through pixel-based numerical analysis using ImageJ software. Statistical differences between the negative control, fresh secretion (positive control), and experimental groups were evaluated via the Kruskal-Wallis test across 24, 48, and 72-hour intervals. The results demonstrated that secretions stored at −20 ℃ for 48 hours exhibited significantly higher efficacy in wound closure compared to those kept at +4 ℃ or room temperature. Notably, storage at −20 ℃ for 48 hours showed higher biological activity than storage at the same temperature for 24 hours. The findings indicated that L. sericata secretions retained their biological effectiveness under controlled conditions, with the −20 ℃ for 48 hours protocol yielding the highest functional impact. The study suggested that cold chain optimization could reduce dependency on fresh secretions, enabling more sustainable, standardized, and clinically accessible secretion-based therapeutic strategies in maggot debridement therapy.

Acknowledgement:

The authors would like to express their sincere gratitude to Prof. Dr. İlhan Onaran for his guidance and supervision throughout this study. We are deeply grateful to Prof. Dr. Erdal Polat for providing the Lucilia sericata secretions and laboratory facilities, and to Prof. Dr. Fahri Akbaş for granting access to the cell culture laboratory infrastructure. Special thanks are extended to Asst. Prof. Dr. Alime Sarıkaya for her invaluable academic and technical support during the experimental procedures. Furthermore, we would like to acknowledge the Istanbul University-Cerrahpaşa GETAT team and the Institute of Graduate Studies for their institutional contributions to this research.

Keywords:

Cell Scratch Assay, fibroblast migration

Abstract ID: **11** Type: **Poster Presentation**

Assessing the predictive power of intrinsic disorder and sequence features for plant stress response

Dragana Dudić[1], Ana Đokić[2], Darko Grbović[3]

[1] *Alfa BK University, Faculty of Mathematics and Computer Science*

[2] *Information Technology School*

[3] *Alfa BK University, Faculty of Information Technology*

Contact e-mail: **dragana.dudic@alfa.edu.rs**

Intrinsically disordered proteins (IDPs) play central roles in stress-responsive signaling and regulation in plants, yet the extent to which protein disorder is systematically associated with stress conditions across species remains unclear. Here, we investigated the relationship between intrinsic disorder and stress status in four plant species: Arabidopsis thaliana, Zea mays, Oryza sativa, and Solanum lycopersicum, with a specific focus on IDP-rich proteins.

Protein-level disorder scores were computed using IUPred, and proteins were classified as IDP-rich based on their overall disorder fraction. We compared disorder-related features between stress and non-stress conditions within each species using non-parametric tests and effect size estimates, and further assessed their predictive value using generalized linear models.

Across all four species, stress and non-stress proteins exhibited largely overlapping disorder distributions within the IDP-rich subset. While Wilcoxon tests identified statistically significant differences in some species, effect sizes were consistently small, as reflected by negligible Cliff's delta values. Logistic regression analyses indicated that disorder fraction and proline content showed weak but statistically detectable associations with stress status, whereas protein length emerged as a stronger and more consistent predictor. Multivariate analyses did not reveal a clear separation between stress and non-stress proteins based solely on disorder-related features.

Taken together, these results suggest that, among IDP-rich proteins, intrinsic disorder alone does not robustly distinguish stress-associated proteins from non-stress proteins across plant species. Instead, protein length and compositional features appear to contribute more substantially to stress-related differences, underscoring the need to interpret disorder in the context of broader sequence properties rather than as an isolated determinant.

Keywords:

disorder, proteins, stress, plants, prediction

Abstract ID: **12** Type: **Poster Presentation**

Hierarchical deep learning for tandem repeat proteins classification

Nevena Ćirić[1], Jovana Kovacevic[1]

[1] *Faculty of Mathematics, University of Belgrade*

Contact e-mail: **nevena.ciric@matf.bg.ac.rs**

Repeat proteins are a widespread class of mostly non-globular proteins containing repetitive subsequences, a so-called repeat units that often occur in tandem arrangements when observed in 3D structure of the protein. These tandem repeats are considerably diverse, ranging from the repetition of a single amino acid to domains of 100 or more residues.

Improvement of the methods for identification of protein tandem repeats and subsequently the increasing number of the known proteins containing repetitive elements necessitates their classification to facilitate further understanding of their sequence-structure-function relationships. According to Kajava's classification scheme based on the repeat unit's length, general structural arrangement and mode of interaction between the repeat units, tandem repeats are classified into five main classes and further divided into subclasses that reflect repeat unit topology, differing in secondary structure arrangement and/or overall structure within the repeat.

The classical approach to obtain the (sub)class assignment for a newly identified tandem repeat is by simply transferring this information from the "master"repeat unit, that is the repeat unit from database of predetermined tandem repeats with associated (sub)classes found to be most similar to the newly identified tandem repeat. This procedure usually implies some kind of structural search algorithm in order to assign master repeat unit, which further implies known tertiary structure of the protein with newly discovered tandem repeat. With intention to tackle the problem of analyzing proteins with unknown 3D structure and facilitate classification of tandem repeats by using only the sequence information, as well as to explore sequence-structure relationship between repeat units sequence and structural characteristics of their corresponding (sub)class, here we propose neural network based model for classification of tandem repeats based on the multiple sequence alignment of its units sequences. Additionally, this model can be further utilized to create an end-to-end pipeline for identification and classification of tandem repeats only from sequence information.

Keywords:

tandem repeats, classification, neural network

Abstract ID: **17** Type: **Poster Presentation**

Multi-resolution framework for orthology and regulatory region delineation in divergent grass genomes

Bojana Banovic Deri[1], Thomas Hartwig[2], Robert VanBuren[3], Vincenzo Rossi[1]

[1] *Council for Agricultural Research and Economics, Research Centre for Cereal and Industrial Crops*

[2] *Heinrich-Heine University*

[3] *Department of Horticulture, Michigan State University*

Contact e-mail: **bojana.banovicderi@crea.gov.it**

Within the BOOSTER project, maize (*Zea mays*), teff (*Eragrostis tef*), and the desiccation-tolerant wild relative *Eragrostis nindensis* were selected as a comparative system for studying drought-associated variation across species with contrasting levels of tolerance. This system supports the identification of conserved and lineage-specific genes and regulatory features that may inform future transfer of beneficial characteristics from more drought-tolerant to more drought-sensitive species. A prerequisite for such analyses is the robust identification of orthologous genes and their associated regulatory regions across the divergent genomes of these three species.

This remains challenging due to extensive structural variation, transposable element activity, lineage-specific rearrangements, gene family expansions, and past whole-genome duplications, which complicate both orthology inference and comparative analysis of regulatory sequences. To address these challenges, we developed a multi-resolution comparative genomics framework that integrates sequence similarity, phylogenetic relationships, gene collinearity, and whole-genome alignment across maize, teff, and *E. nindensis*.

The workflow combines complementary approaches, including orthogroup inference (OrthoFinder), gene synteny detection (MCScanX), whole-genome alignment (Cactus), and sequence similarity–based methods (BLASTN and BLASTP). Publicly available reference genomes and gene annotations were used to extract coding sequences, proteins, and genes together with their flanking regulatory regions for pairwise and cross-species analyses.

By integrating homology-based, phylogenetic, collinearity-aware, and base-resolution alignment evidence, the framework improves the classification of orthologous relationships and supports the delineation and cross-species mapping of candidate regulatory regions, including in genomic contexts with disrupted collinearity. This approach provides a transferable computational foundation for comparative regulatory genomics in divergent grasses and establishes a basis for downstream integration with trait-focused datasets, including abiotic and biotic stress-related data. In the context of BOOSTER, this framework provides a foundation for further comparative investigation of experimentally identified drought-related genes and their associated candidate regulatory regions across the three species, with relevance to drought tolerance improvement in maize and teff.

Acknowledgement:

This work was supported by the project "Boosting drought tolerance in key cereals in the era of climate change (BOOSTER)". The BOOSTER project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101081770.

Keywords:

comparative genomics, gene orthology, synteny

Abstract ID: **21** Type: **Poster Presentation**

Sex-specific distribution of EGFR rs712829 and rs712830 polymorphisms in East Asian, European and Serbian populations

Jasmina Obradovic[1], Vladimir Jurisic[2]

[1] *Institute for Information Technologies Kragujevac, University of Kragujevac*

[2] *Faculty of Medical Sciences, University of Kragujevac, Kragujevac, Republic of Serbia*

Contact e-mail: **jasmina.obradovic@uni.kg.ac.rs**

This study investigated the genotype and allele distributions in males and females of two epidermal growth factor receptor (EGFR) gene polymorphisms related to non-small cell cancer (NSCLC) risk, rs712829 and rs712830, across East Asian (EAS), European (EUR), and Serbian (SRB) populations. We analysed genotype and allele frequencies for two polymorphisms using data from the 1000 Genomes Project (EAS and EUR populations) and data from a cohort of healthy Serbian volunteers (53) from our previous study, stratified by gender. Chi-squared test or Fisher's exact test were applied. The analyses were performed with R software, version 4.5.1 (R Core Team, 2025) (tidyverse, janitor, readr, openxlsx, ggplot2). All examined populations had the highest frequencies of the wild-type homozygotes (G/G and C/C) for both polymorphisms in males and females. Gender-related genotype and allele distributions of EGFR rs712829 and rs712830 significantly differ between East Asian and European populations (all p-values < 0.01). Frequencies of the T allele rs712829, in EAS population were lower in males and females (5,3%; 6,3%), than in EUR population 32,5% males, and 30,6% females. There were no statistical differences for individual populations in genotype and allele distributions. In the Serbian cohort, distribution of alleles and genotypes corresponded well to their distribution in the European population. Frequencies of the A allele rs712830, in EUR population in males and females were 13,9% and 12,9%, interestingly there was no this allele in EAS population. These findings highlight the importance of considering sex and ethnicity in genetic studies related to *EGFR* and lung cancer.

Acknowledgement:

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia, Agreements No. 451-03-33/2026-03/200378 and 451-03-34/2026-03/200111.

Keywords:

EGFR, polymorphism, rs712829, rs712830

Abstract ID: **23** Type: **Poster Presentation**

Regulatory landscape of TGFB pathway in gastrointestinal tumors

Jelena Karanović[1], Tamara Babić[1], Sandra Dragičević[1], Aleksandra Nikolić[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **jelena.karanovic@imgge.bg.ac.rs**

Transforming growth factor beta (TGFB) signaling, a central pathway regulating cell growth, differentiation, and immune responses, is enrolled in gastrointestinal (GIT) tumors. Non-canonical transcripts significantly contribute to the complexity of TGFB regulation, adding layers of transcriptional and post-transcriptional control. This study aimed to identify common deregulated transcripts of the TGFB pathway across seven GIT tumors and to explore similarities and differences in their regulatory landscapes.

Expression data for coding and non-coding transcripts of six canonical TGFB pathway genes (*TGFBR1*, *TGFBR2*, *SMAD2*, *SMAD3*, *SMAD4*, and *SMAD7*) were retrieved from GTEx (normal tissue) and TCGA (primary tumor and solid normal tissue) databases via the XENA browser. For each tumor type (head and neck, esophagus, stomach, liver, pancreas, colon, and rectum) differential expression analysis was performed using *R* (v4.5.2). Transcripts deregulated in all seven tumors were subjected to principal component analysis (PCA) to identify transcriptional patterns, and to *in silico* annotation of coding potential (*CPC2*) and subcellular localization (*LNClocator*, *iLoc-LncRNA*). Enrichment analysis was performed using the *gprofiler2* package.

Out of 59 significant transcripts, four non-canonical (*SMAD2*: ENST00000586040.5 and ENST00000356825.8; *SMAD3*: ENST00000439724.7; *SMAD4*: ENST00000591126.5) were deregulated across all tumors. They were predicted to be protein-coding, with predominant cytoplasmic localization, while the *SMAD3* transcript could also localize to nuclear compartments. PCA revealed that *SMAD2* transcripts differentiated samples along PC1 (47.7% variance) synergistically, whereas *SMAD3* and *SMAD4* transcripts separated samples along PC2 (31.4% variance) antagonistically, distinguishing normal from patients-derived samples (tumor and solid normal). All GIT tumors displayed broadly similar expression profiles relative to normal tissues, with subtle inter-tumor differences, most pronounced in colon and rectum compared to other tumors. Enrichment analysis highlighted aberrant TGFB signaling, coupled with dysregulated cell cycle control and genomic instability, as a convergent oncogenic mechanism across GIT tumors, integrating developmental pathways, immune modulation, and malignant transformation.

In conclusion, four deregulated non-canonical TGFB transcripts represent a common feature of GIT tumors, with subtle inter-tumor differences most evident in colon and rectum relative to other tumors. These findings highlight their potential as biomarkers for tumor classification and therapeutic targeting, warranting validation in future investigations across diverse patient cohorts.

Acknowledgement:

This work was funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-33/2026-03/200042).

Keywords:

gastrointestinal tumors, TGFB, transcripts

Abstract ID: **25** Type: **Poster Presentation**

Spatial organization of gene expression in 3-day old zebrafish larvae

Jelena Kusic-Tisma[1], Nevena Vezmar[1], Igor Davidović[1], Mateja Ilic[1], Mila Ljujić[1], Aleksandra Divac Rankov[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **jelena.kusic@imgge.bg.ac.rs**

In order to better characterise expression profile of 3-day old zebrafish larvae, we performed spatial transcriptomic analysis using STomics stereoseq technology. Analyses were performed on 4 sagittal cryosections from 2 different larvae (2 sections per sample, ~60 μm apart). Stereoseq is a grid-based technology that aggregates a number of DNA nanoball spots (DNB) and clusters their spatial profile into substructures. Here, we used grid size of 10 μm (bin 20, 20 × 20 DNB), which is equivalent to approximately one cell diameter. The number of transcripts and genes detected among different sections were comparable, with an average of 1064 unique molecular identifiers (UMIs) and 535 genes per unit. Unsupervised clustering was performed by spatially aware clustering algorithm BANKSY using lambda parameter 0.2. Number of clusters varied between sections from 10 to 13, which was in good correlation with expected number of organs/tissues per section of 3-day old zebrafish larvae. Analyses of spatially variable genes (SVG) detected 95 genes with spatial pattern having Global Moran' s I coefficient greater than 0.4. In order to identify co-varying genes showing similar spatial distribution, we applied Hotspot algorithm. Ten spatial modules per section were recognised. Gene ontology (GO) enrichment analysis on the spatially correlated genes of each module indicated the identity of these spatial modules (e.g. heart, epidermis, cartilage, eye, blood, forebrain, notochord, liver). Spatial visualization of modules confirmed that the distribution of detected modules matched their region-specific biological functions. Taken together, our analyses contribute to the gene expression landscape of zebrafish at the 3-day old developmental stage.

Acknowledgement:

This study is a part of the Enhancing Non-Communicable Disease Research Excellence Through Zebrafish Capacity Building (ZeNCure) project, supported by the European Union, under Horizon Europe programme Widening Participation and Spreading Excellence, HORIZON-WIDERA-2023-ACCESS-02, Grant Agreement number 101160259.

Keywords:

spatial transcriptomics, stereoseq, zebrafish

Abstract ID: **29** Type: **Poster Presentation**

Towards protein complex identification in dynamic PPI networks using embeddings

Nenad Vilendečić[1], Milana Grbić[1], Miloš Radovanović[2], Dragan Matić[1]

[1] *Faculty of Natural Sciences and Mathematics, University of Banja Luka*

[2] *Faculty of Sciences, University of Novi Sad*

Contact e-mail: **milana.grbic@pmf.unibl.org**

Proteins are recognized as essential functional elements in a wide range of cellular processes, where they typically act as components of protein complexes. In most network-based studies, protein complex identification has been mainly performed using static protein–protein interaction (PPI) networks. More recently, dynamic PPI networks have been employed for protein complex detection in order to account for temporal variability in interaction patterns between proteins and to enable time-aware analysis of protein complex formation.
In this study, the problem of detecting protein complexes in dynamic PPI networks is investigated. A dynamic PPI network is constructed by integrating a static PPI network with time-aware gene expression data, thereby enabling changes in protein–protein interactions to be represented across successive time points. Based on this representation, dynamic node embeddings are generated using a temporal network embedding approach. This approach encodes both structural and temporal properties of the network into a lower-dimensional vector space. The resulting embeddings are then used for the detection of dynamic communities within the network. These communities are interpreted as candidate protein complexes, which are subsequently subjected to a specially designed filtering procedure aimed at reducing noise and improving biological relevance. Finally, the filtered candidate complexes are compared against a set of reference protein complexes for evaluation purposes.
The proposed approach enables the analysis of temporally varying interaction patterns and provides a basis for protein complex identification in dynamic PPI networks. Preliminary observations indicate the potential advantages of incorporating temporal information compared to static approaches.

Keywords:

protein-complex, protein-protein-interaction-(PPI)-network, dynamic-PPI-network, embedding, dynamic-community

Abstract ID: **34** Type: **Poster Presentation**

Bioinformatics identification of extracellular vesicle - derived microRNAs biomarkers in pancreatic cancer

Sanja Goč[1], Zorana Dobrijević[1], Tamara Janković[1], Filip Janjić[1], Ninoslav Mitić[1], Miroslava Janković[1]

[1] *University of Belgrade, Institute for the Application of Nuclear Energy, INEP*

Contact e-mail: **sanjagoc@inep.co.rs**

Pancreatic cancer is one of the most lethal malignancies worldwide due to the lack of effective early diagnostic tools and its asymptomatic early stages. It is the 6th leading cause of cancer-related mortality, with a 5-year survival rate ranging from 44% in localized disease to 3% in metastatic cases. CA19-9, the most commonly used biomarker, has limited sensitivity and low positive predictive value, restricting its use in screening. Therefore, novel biomarkers are needed, and extracellular vesicle-associated miRNAs (EV-miRNAs) are increasingly recognized for their stability and potential in non-invasive cancer diagnosis. In this study, we performed an *in silico* analysis of plasma EV-miRNA expression profiles using the GEO dataset GSE221185 (BioProject: PRJNA913165), including 16 pancreatic cancer patients and 13 healthy controls. Seventeen differentially expressed miRNAs (5 downregulated and 12 upregulated) were identified, of which nine candidates were prioritized based on network topology metrics, including the number of single-line regulators (NSR) and transcription factor ratio (TFR) ($p < 0.05$). Further refinement using dbDEMC 2.0 meta-profiling identified five miRNAs (miR-218-5p, miR-221-3p, miR-423-3p, miR-505-3p, and miR-766-3p) exhibiting $|\log2FC| \geq 2$. Receiver operating characteristic (ROC) analysis demonstrated that only hsa-miR-766-3p achieved statistically significant diagnostic performance (AUC = 0.750, $p = 0.028$). Network analysis using miRNet 2.0 revealed that the nine miRNAs collectively regulate 3,499 target genes. Based on degree and betweenness centrality, three regulatory clusters were identified: high-centrality hub miRNAs (miR-218-5p, miR-221-3p, miR-423-3p), intermediate regulators (miR-590-3p, miR-766-3p, miR-512-3p), and peripheral miRNAs. Functional enrichment analysis indicated involvement of key oncogenic pathways, including *PI3K–AKT, EGFR/MAPK, JAK–STAT*, cell cycle regulation, apoptosis, angiogenesis, and *TGF-β/EMT* signaling. The four final candidate miRNAs (miR-218-5p, miR-221-3p, miR-423-3p, and miR-766-3p) collectively targeted 1,922 genes, including 14 pancreatic cancer-associated genes. TargetLink2.0 confirmed validated interactions for miR-218-5p (*CDK6, E2F2, EGFR, IKBKB*) and miR-221-3p (*PIK3R1*). External validation using independent serum-based GEO datasets (GSE85589 and GSE59856) demonstrated significant diagnostic performance for miR-221-3p (AUC = 0.734, $p < 0.0001$) and miR-218-5p (AUC = 0.676, $p = 0.016$). Overall, miR-218-5p and miR-221-3p represent promising non-invasive diagnostic biomarkers for pancreatic cancer, warranting further experimental validation.

Acknowledgement:

This work was supported by the Ministry of Science, Technological Development, and Innovation, Republic of Serbia [Contract No: 451-03-33/2026-03/200019]. This research aligns with the Agenda 2030- United Nations Sustainable Development Goal 3, promoting good health and well-being.

Keywords:

Extracellular vesicles, microRNA, Pancreatic cancer

Abstract ID: **38** Type: **Poster Presentation**

Trehalose synthesis is induced in maize roots under low-temperature stress

Manja Božić[1], Dragana Ignjatović Micić[1], Nikola Grčić[1], Marko Mladenović[1], Ana Nikolić[1]

[1] *Maize Research Institute Zemun Polje*

Contact e-mail: **mbozic@mrizp.rs**

Trehalose is an important osmoprotectant associated with tolerance to various types of abiotic stress factors, such as drought and extreme temperatures, particularly low temperatures. Many research works have shown that trehalose synthesis and accumulation were included in the low-temperature (LT) response in maize, but most were not focused on its synthesis in roots. As one of the most promising strategies of avoiding yield loss due to effects of climate change in temperate areas is early sowing, understanding all the mechanisms and pathways responsible for establishing LT tolerance in every part of the maize plant during the early developmental stages becomes a priority.

Herein, plants belonging to two maize genotypes (tolerant - Ltol and susceptible - Lsus to low temperatures) were exposed to LT conditions for 24h (8/10° C) while in the V3 developmental stage (fully developed third leaf). Samples for RNA isolation and cDNA library preparation were taken 6h and 24h after the treatment began, and paired-end transcriptome sequencing was performed (Illumina® Novaseq 6000). The same was done for control plants grown in optimal conditions (25°/20°C). Sequencing was followed by raw-data quality check and filtering, mapping to the reference genome, and read quantification, necessary for the differential expression (DE) analysis.

DE profiling showed that several trehalose-6-phosphate synthase (*TPS*) and trehalose 6-phosphate phosphatase (*TPP*) genes, necessary for trehalose synthesis, were differentially expressed between the control and treatment conditions in both genotypes. *TPP* genes were up-regulated in either of the time-points in both genotypes – *ZmTPP3.2*, *ZmTPP9*, and *ZmTPP12*. On the other hand, several *TPS* genes were down-regulated in either of the time points in Lsus, but were up-regulated after 24h in Ltol –*ZmTPS6.3*, *ZmTPS9.3*, and *ZmTPS11.3*. The results show that while both genotypes seemed to successfully synthesise the trehalose precursor, trehalose-6-phosphate, only in Ltol is the precursor translated into trehalose.

Keywords:

maize, low-temperature stress, RNAseq, trehalose

Abstract ID: **39** Type: **Poster Presentation**

Molecular docking approach in prediction of binding between selected bioactive phenolic compounds and human serum albumin

Vladimir Vlatković[1], Saša Lazović[1], Aleksandra Đukić-Vuković[2], Milica Radan[3], Bojan Marković[4], Jelena Savić[4], Zorica Vujić[4], Darija Obradović[1]

[1] *Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia*

[2] *Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, Serbia*

[3] *Institute for the Study of Medicinal Plants "Dr. Josif Pančić", Belgrade, Serbia*

[4] *Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia*

Contact e-mail: **vlatkovic@ipb.ac.rs**

Molecular docking is used as a virtual screening approach that allows the prediction of ligand-receptor binding modes and ranking ligands. Human serum albumin (HSA) is a key transport protein responsible for the distribution of drugs and small bioactive molecules in the human body. HSA binding is important for the compound's therapeutic potential. Phenolic compounds are recognised for their bioactive potential, including antioxidative, neuroprotective and antidiabetic activities.

The goal of the study was to determine the binding affinities of selected phenolic acids to Sudlow binding site I (warfarin binding site) within HSA and to describe intermolecular interactions governing the binding process. The receptor structure of HSA subdomain IIA containing Sudlow binding site I (PDB ID: 2BXD) and ligands (structurally related chlorogenic acid and isochlorogenic acid A, as well as reference ligand warfarin), were preprocessed before the docking study. The receptor grid box was generated with pinpoint accuracy to include warfarin binding site shape, size and properties. Molecular docking of selected phenolic compounds was performed using Glide molecular modeling software. This software performs docking by examining conformational, positional, and orientational degrees of freedom of ligands within a defined rigid receptor grid, while utilizing the Glide score function to rank binding affinities and visual generation of compound-HSA binding complex environment and interactions.

The results showed that chlorogenic acid (-6.335) and isochlorogenic acid A (-7.438) had similar binding affinity to warfarin (-7.866). Binding interactions include hydrophobic, hydrogen bond, and, in the case of chlorogenic acid, salt bridge interactions. Isochlorogenic acid features fewer hydrogen bond interactions, but is enclosed in a more hydrophobic cavity, and is less exposed to solvent than chlorogenic acid, indicating its stronger and more stable binding to HSA binding site I.

Molecular docking using Glide enabled fast, effective, high-accuracy analysis, showing that tested phenolic acids have comparable binding affinity and similar binding interactions to HSA binding site I as warfarin. The high accuracy of this method also showcased the influence that subtle structural differences play in the binding mechanism and positioning of ligands within the binding site. This approach proved its potential as a virtual screening method in bioactive compound studies.

Acknowledgement:

This research was supported by the Science Fund of the Republic of Serbia, GRANT No 13534, Proof of Concept - DIAMICA. The authors also acknowledge funding provided by the Institute of Physics Belgrade, National Institute of the Republic of Serbia, through the grant from the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia

Keywords:

molecular docking. HSA. binding interactions

Abstract ID: **40** Type: **Poster Presentation**

Double-stranded RNA sequencing and bioinformatic analysis reveal viruses in Verticillium dahliae

Vanja Miljanić[1], Jernej Jakše[1], Sebastjan Radišek[2], Nataša Štajner[1]

[1] *Biotechnical Faculty, University of Ljubljana, Slovenia*

[2] *Slovenian Institute of Hop Research and Brewing*

Contact e-mail: **vanja.miljanic@bf.uni-lj.si**

Verticillium dahliae is a soil-borne phytopathogenic fungus and the causal agent of Verticillium wilt, a destructive vascular disease affecting hundreds of economically important plant species worldwide. Its broad host range, long-term survival and persistence in soil make disease control particularly challenging. Mycoviruses, or fungal viruses, have garnered increasing attention due to their potential role in altering fungal pathogenicity and their suitability for biological control.
In our study, 33 isolates of *V. dahliae* were obtained from the culture collection of the Slovenian Institute of Hop Research and Brewing to test for the presence of mycoviruses. The first step in the search for viral-infected fungi is to look for the presence of dsRNAs, as all RNA viruses have a dsRNA stage during their infection cycle. To isolate dsRNA, fungi were cultured on cellophane membranes overlaying CDA plates. The dsRNA was isolated by the cellulose chromatography method. After digestion with DNase I (New England BioLabs) and S1 nuclease (Thermo Fisher Scientific), the dsRNA segments were analyzed by electrophoresis on a 1% agarose gel. Two out of the 33 tested fungal isolates were found to be virus-infected. Libraries of dsRNAs were constructed using the Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific) and were barcoded using the Xpress™ RNA-Seq Barcode 1–16 Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The yield and size distribution of the amplified cDNA libraries were determined using the Agilent 2100 Bioanalyzer (Agilent Technologies). Sequencing was performed on the Ion GeneStudio S5 Prime System (Thermo Fisher Scientific). The obtained reads were processed with bioinformatics tools CLC Genomic Workbench and Genomics Server (Qiagen). The applied method revealed the presence of bisegmented Verticillium albo-atrum partitivirus 1 (VaaPV1; family *Partitiviridae*) and Verticillium dahliae RNA virus 1 (VdRV1; unassigned non-segmented +RNA virus). Given that some mycoviruses influence the pathogenicity of their fungal hosts, ongoing research focuses on assessing the impact of VaaPV1 and VdRNA1 on host virulence.

Acknowledgement:

This research was funded by the Slovenian Research and Innovation Agency (ARIS): research project J4-70163 awarded to Vanja Miljanić and research program P4-0077 Genetics and Modern Technologies of Crops.

Keywords:

Verticillium wilt, VaaPV1, VdRNA1

Abstract ID: **41** Type: **Poster Presentation**

Integrated analysis reveals bias in single-tool miRNA profiling during grapevine viral coinfection

Katja Jamnik[1], Hana Šinkovec[1], Jernej Jakše[1], Vanja Miljanić[1], Nataša Štajner[1]

[1] *Biotechnical faculty, University of Ljubljana, Slovenia*

Contact e-mail: **katja.jamnik@bf.uni-lj.si**

Grapevine (*Vitis vinifera* L.) is one of the most important fruit crops worldwide. Its production is threatened by numerous pathogens, including viruses, which can cause significant economic losses. Plants respond to viral infection through multiple regulatory mechanisms, including changes in microRNA (miRNA) expression. miRNAs are a class of small non-coding RNAs that regulate gene expression at the post-transcriptional level. In this study, we compared miRNA predictions generated by three widely used miRNA prediction tools in grapevine under viral coinfection.

Virus-free and virus-infected plants of two grapevine cultivars, Refošk ("Terrano") and Zeleni Sauvignon ("Sauvignon Vert"), infected with Grapevine Pinot gris virus (GPGV), Grapevine rupestris stem pitting-associated virus (GRSPaV), and Grapevine rupestris vein feathering virus (GRVFV), were analyzed. Small RNA sequencing was performed, and three prediction tools (miRador, miRDeep2, and ShortStack) were used to identify and quantify miRNAs. Differential expression analysis was conducted separately for each tool and using an integrated approach combining all three datasets. The expression of selected miRNAs was further validated by stem-loop RT-qPCR.

By small RNA sequencing, between 6,344,090 and 25,041,733 reads per library were obtained. The presence of GPGV, GRSPaV, and GRVFV in infected samples and absence of viruses in the virus-free plants were confirmed through sRNA sequencing data. Each prediction tool identified a distinct number of miRNAs, with miRDeep2 detecting the highest number and miRador the lowest. Consequently, the sets of differentially expressed miRNAs varied substantially between tools. The integrated approach yielded an additional set of differentially expressed miRNAs, most of which overlapped with those identified by at least one individual tool. Stem-loop RT-qPCR supported the differential expression of several selected miRNAs.

Overall, our results demonstrate that miRNA profiling outcomes in grapevine under viral coinfection are strongly dependent on the choice of prediction tool. This highlights the need for integrative analytical approaches combined with experimental validation to achieve reliable and biologically meaningful interpretation of miRNA expression data.

Acknowledgement:

This research was funded by Slovenian Resarch and Innovation Agency (ARIS), grant number P4-0077 and ARIS grant to young researcher Katja Jamnik.

Keywords:

miRNA, prediction tools, differential expression

Abstract ID: **47** Type: **Poster Presentation**

Prothrombin reshapes the temporal transcriptomic trajectory of Caco-2 colorectal cancer cells

Marija Lazić[1], Marija Cumbo Stikić[1], Aleksandra Vitkovac[1], Igor Davidović[1], Mateja Ilić[1], Marija Schwirtlich[1], Martina Mia Mitić[1], Branko Tomić[1], Valentina Đorđević[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **marijal3009@gmail.com**

Emerging evidence highlights the non-hemostatic roles of the coagulation proteins in shaping the immune and stromal landscape of various cancers. However, despite the prevalence of coagulation factors in the tumor microenvironment, the precise gene expression profiles induced by prothrombin remain poorly understood.

To elucidate these mechanisms, we systematically profiled the temporal transcriptional response of Caco-2 colorectal cancer cells to prothrombin treatment. Using a robust experimental design, we generated an in-house RNA-seq dataset (DNBseq technology) comprising cells treated with prothrombin, glycerol-vehicle controls, and untreated counterparts at 24 and 48 hours, thereby allowing us to distinguish genuine treatment-induced signals from time-dependent culture drift. Raw counts were analyzed with DESeq2 using Benjamini–Hochberg false discovery rate control, followed by pre-ranked Gene Set Enrichment Analysis (GSEA) of Gene Ontology Biological Process terms.

GSEA showed that, under control conditions, the 48 h versus 24 h contrasts were characterized by a marked shift toward reduced biosynthetic and proliferative programs: ribosome biogenesis, rRNA metabolism, and DNA replication were negatively enriched (NES ≈ −2 to −3). In untreated cells, autophagy regulation and Golgi/vesicle transport were positively enriched at 48 h, consistent with a time-dependent adaptive shift toward autophagy and vesicular trafficking. In contrast, prothrombin-treated cells displayed a distinct 48 h versus 24 h signature with positive enrichment for endoplasmic reticulum to Golgi and vesicular transport, protein quality control and proteostasis, translational initiation and regulation, microtubule-dependent organelle transport, and cell-cycle regulation (NES ≈ +1.6 to +2.3). The broadly similar decline observed across control conditions supports the interpretation that prothrombin-associated enrichments are not explained by vehicle exposure or time in culture alone.

These results demonstrate that prothrombin can act as a modulator of cellular state, redirecting the transcriptional landscape toward pathways critical for cellular fitness and proliferation. This distinct signature provides a foundation for future studies investigating how prothrombin-mediated signaling influences colorectal cancer.

Keywords:

prothrombin, colorectal cancer, RNA-seq, GSEA

Abstract ID: **48** Type: **Poster Presentation**

DigestedProteinDB: A compact key–value database for in silico peptide digestion and mass-based search

Antonio Starcevic[1], Janko Diminic[1], Jurica Zucko[1], Krešimir Križanović[2], Toni Čvrljak[3]

[1] *University of Zagreb Faculty of Food Technology and Biotechnology, Zagreb*

[2] *University of Zagreb Faculty of Electrical Engineering and Computing, Zagreb*

[3] *University of Zagreb Faculty of Mining, Geology and Petroleum Engineering, Zagreb, Croatia*

Contact e-mail: **jdiminic@pbf.hr**

Efficient comparison of experimental peptide masses with theoretical values represents a central step in mass spectrometry (MS)–based proteomics, microbial biotyping, and MS imaging. As protein sequence databases such as UniProtKB continue to expand, the need for fast, scalable access to large collections of in silico–digested peptides has grown substantially. Existing approaches either relied on computationally demanding on-the-fly digestion or required high-performance server infrastructure for centralized peptide databases, limiting their accessibility for routine laboratory use.

We developed DigestedProteinDB, a compact and high-performance key–value database of peptides generated by enzymatic in silico digestion of UniProtKB/Swiss-Prot and TrEMBL sequences. We implemented the system using RocksDB and incorporated multiple optimization strategies to minimize disk footprint and accelerate mass-range queries. We discretized peptide masses to four decimal places and stored them as 32-bit integer keys, enabling efficient range-based retrieval aligned with the LSM-tree architecture of RocksDB. We encoded peptide sequences using a compact 5-bit scheme per amino acid residue, yielding approximately 37.5% reduction in raw sequence storage compared to standard byte-based encoding. Furthermore, we serialized UniProt accession numbers using Base36 encoding and represented them as 64-bit integers, while maintaining internal integer identifiers via an in-memory array for constant-time resolution. We applied additional compression using Snappy at the RocksDB block level.

We benchmarked the system using 252 million UniProtKB protein sequences digested with trypsin, allowing up to two missed cleavages and peptide lengths of 6–50 amino acids, yielding approximately 5.9 billion peptide entries. The resulting database occupied approximately 250 GB of disk space and required less than 16 GB of RAM during both construction and querying. We completed database construction in approximately two days, with an average mass-range query time of 200–300 ms. We performed all operations on a standard workstation without high-performance server infrastructure.

DigestedProteinDB is publicly accessible via a web interface and REST API. The modular design allowed rapid generation of experiment-specific databases, making it suitable for integration into diverse MS-based analytical workflows.

Acknowledgement:

This research was funded by the European Union (NextGenerationEU) project "The glycome and microbiome as markers of dietary impact on the health of women of reproductive age"within the program 'Targeted Scientific Research'(NPOO.C3.2.R3-I1.04.0073).

Keywords:

peptide, database, RocksDB, UniProtKB, mass

Abstract ID: **55** Type: **Poster Presentation**

PromoterLab: browser-based application for human promoter design, visualization and sequence analysis

Anastasia Matlaeva[1], Evgenii Lunev[2]

[1] *Institute of Transport and Service, Private Educational Institution of Professional Education, Sochi, Russia*

[2] *Laboratory of Modeling and Gene Therapy of Hereditary Diseases, Institute of Gene Biology Russian Academy of Sciences, Moscow, Russia*

Contact e-mail: **anmatlaeva@gmail.com**

In gene therapy and disease modeling, achieving efficient transgene expression is necessary but not sufficient. Vectors must also ensure tissue specificity, which requires carefully selected regulatory elements. In this context, natural promoters of tissue-specific genes are a logical choice for this purpose. Furthermore, when the aim is to recapitulate the expression pattern of a particular gene, using its own natural promoter may offer the most direct path to physiologically relevant expression levels. However, only a limited number of natural promoters have been experimentally characterized, and no standardized framework exists for defining the boundaries of their proximal regions suitable for cloning into expression vectors.

We developed PromoterLab, a browser-based software-as-a-service platform that integrates promoter sequence retrieval, annotation, visualization, prioritization, and comparative sequence analysis within a single workspace. The platform integrates data from multiple curated and public resources, including NCBI, EPD/EPDnew, JASPAR, ENCODE, FANTOM, UCSC Genome Browser, and Ensembl, enabling systematic exploration of proximal regions of human promoters. The platform supports analysis of variable proximal promoter regions around the transcription start site (TSS) and automated annotation of key regulatory features, including CpG islands, GC-rich regions, and transcription factor binding sites. PromoterLab incorporates algorithmic and ML/LLM-based modules that, based on a machine learning model trained on CAGE-derived transcriptional activity using sequence composition, CpG structure, and transcription factor binding features, help prioritize candidate promoter regions according to predicted transcriptional activity, tissue specificity, regulatory feature composition, and preservation of key regulatory constraints. An interactive visualization engine displays promoter maps, transcription factor binding site density, CpG island tracks, GC-content profiles, regulatory clusters, and nucleotide-level sequence views, allowing users to interactively adjust candidate promoter boundaries.

Future development of PromoterLab will include expansion of the list of supported organisms beyond humans, with particular focus on commonly used experimental models. We also plan to implement prediction of transcriptional strength and assessment of how well the selected proximal promoter region preserves tissue-specific expression relative to the endogenous gene. In addition, the platform will be extended with an automated pipeline for vector design and primer generation, enabling direct cloning into standardized or user-defined expression vectors.

Acknowledgement:

The authors would like to acknowledge Oleg Mozhey for his contributions to the development of the PromoterLab platform, including the design and implementation of the core software architecture, cloud resource provisioning, data integration, and the development of multiple analytical modules.

Keywords:

promoter design, sequence analysis

Abstract ID: **58** Type: **Poster Presentation**

Development and validation of a reproducible RNA-seq pipeline for transcriptomic analysis

Nikola Jocic[1], Anita Skakic[1], Marina Parezanovic[1], Marina Andjelkovic[1], Nina Stevanovic[1], Sara Stankovic[1], Kristel Klaassen[1], Vesna Spasovski[1], Milena Ugrin[1], Jovana Komazec[1], Kristina Grujic[1], Maja Stojiljkovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **nikola.jocic@imgge.bg.ac.rs**

RNA sequencing (RNA-seq) has become a standard approach for transcriptomic analysis, yet building reproducible, well-documented pipelines remains a practical challenge for many research groups. Publicly available automated pipelines such as nf-core/rnaseq offer robust solutions, but provide limited opportunity to develop a deep understanding of each tool used in the workflow. Here we describe the development and validation of a custom, modular RNA-seq pipeline built from widely used bioinformatics tools, designed to be reproducible and adaptable to different experimental contexts.

The pipeline was developed and tested using a publicly available RNA-seq dataset (SRP139131) from human T cells stimulated with three anti-CD3 antibodies (OKT3, FvFcM, FvFcR), and an untreated control (Illumina technology, 150 bp paired-end reads). The pipeline consisted of two parts. The first part, implemented in Bash, covered quality control (FastQC, MultiQC), adapter trimming (fastp), splice-aware alignment to the human reference genome GRCh38 (HISAT2), post-alignment quality control (RSeQC), and gene-level quantification (featureCounts). The second part, implemented in R, performed differential expression analysis (DESeq2), followed by functional enrichment analysis using Gene Ontology and KEGG pathway databases (clusterProfiler). We validated pipeline performance by comparing the generated count matrix against that produced by the nf-core/rnaseq pipeline (RSEM quantification) using per-sample Spearman correlation and principal component analysis (PCA).

Per-sample Spearman correlation between count matrices ranged from 0.884 to 0.896 across all detected genes, and improved to 0.950–0.952 when restricted to expressed genes with a count above 5. PCA revealed that the first principal component (32.9% variance explained) separated the two pipelines, reflecting a systematic quantification difference attributable to algorithmic differences between featureCounts and RSEM, while the second principal component (16.3%) captured biological variation consistently across both pipelines, with clear separation of conditions (untreated, FvFcM, FvFcR, OKT3).

The custom pipeline produced biologically consistent results comparable to a well-established reference pipeline, with observed count differences consistent with known tool-specific quantification behaviour. The modular design facilitates adaptation to different experimental designs and provides a framework suitable for researchers seeking to understand each analytical step.

Acknowledgement:

This research was supported by the Horizon Europe Project BRIDGING-RD, HORIZON-WIDERA-2023-ACCESS-02, Grant Agreement N°101160079.

Keywords:

RNAseq, pipeline, bioinformatics, transcriptomics

Abstract ID: **61** Type: **Poster Presentation**

Dynamics in soil microbial community structure under the influence of heavy metal contamination and phytoremediation: a pot trial experiment

Kristina Kokotović[1], Dragan Radnović[1], Marijana Kragulj Isakovski[2], Jelena Beljin[2], Nina Đukanović[2], Stanko Milić[3], Snežana Maletić[2], Dragana Tamindžija[1]

[1] *University of Novi Sad, Faculty of Sciences, Department of Biology and Ecology*

[2] *University of Novi Sad, Faculty of Sciences, Department of chemistry, biochemistry and environmental protection*

[3] *Institute of field and vegetable crops Novi Sad*

Contact e-mail: **kristina.kokotovic@dbe.uns.ac.rs**

This study investigated the dynamics of soil microbial communities under phytoremediation in heavy metal-contaminated soil. Soil used for the pot experiment was collected from Srpski Itebej, a site affected by canal dredging and characterized by elevated concentrations of heavy metals. A controlled pot experiment was established with two plant species, *Sorghum bicolor* (forage sorghum) and *Cannabis sativa* (hemp), combined with commercial plant growth promoting rhizobacteria and fungi (PGPR) treatments. Samples were collected at the beginning and after three months. Soil DNA was extracted, and 16S rRNA gene amplicon sequencing was performed. Sequence data were processed and analyzed in R using established bioinformatics workflows for ASV inference, taxonomic classification, and community analysis. Total counts of heterotrophic, ammonium-oxidizing, denitrifying, and nitrogen-fixing bacteria were higher in the final samples, in contrast to nitrifiers, which exhibited greater abundance at the start of the experiment. Catalase and acid phosphatase activities were more pronounced at the beginning of the study, whereas alkaline phosphatase and dehydrogenase activities were slightly more prominent at the conclusion of the experiment. Alpha diversity analyses revealed higher diversity at the beginning of the experiment, particularly in control samples and those treated with PGPR. Beta diversity analysis indicated that dominant taxa were broadly similar across samples and timepoints; however, Thermoproteota consistently dominated, while Cyanobacteriota were more abundant in samples collected at the end of the experiment. Principal coordinate analysis (PCoA) based on Bray-Curtis distances revealed clear separation of samples according to both treatment and experimental stage. PERMANOVA confirmed that both treatment and time significantly influenced microbial community structure. Principal component analysis (PCA) integrating chemical and microbial parameters showed that heavy metals were the dominant drivers at the initial stage of the experiment. In contrast, reduced metal concentrations at the final stage were associated with recovery and stabilization of functional microbial groups, including denitrifiers and ammonifiers. Overall, the results demonstrate that phytoremediation can promote the recovery and stabilization of soil microbial communities under heavy metal stress, highlighting the importance of integrative bioinformatics approaches for understanding microbe-environment interactions.

Acknowledgement:

This research was supported by the Science Fund of the Republic of Serbia, #GRANT No 6769, Natural based efficient solution for remediation and revitalization of contaminated locations using energy crops –ReNBES.

Keywords:

diversity, taxonomic composition, 16S rRNA

Abstract ID: **63** Type: **Poster Presentation**

Precision and recall in rare disease genomics: benchmarking the IMGGE whole genome sequencing pipeline

Sara Stankovic[1], Nikola Jocic[1], Djordje Pavlovic[1], Marina Parezanovic[1], Nina Stevanovic[1], Kristina Grujic[1], Jovana Komazec[1], Kristel Klaassen[1], Anita Skakic[1], Marina Andjelkovic[1], Milena Ugrin[1], Vesna Spasovski[1], Steven Laurie[2], Maja Stojiljkovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Centro Nacional de Análisis Genómico, Barcelona, Spain*

Contact e-mail: **sara.stankovic@imgge.bg.ac.rs**

Benchmarking of bioinformatics pipelines is the foundation of quality assurance in diagnostics, and is of particular importance in the rare disease field, where sensitivity and specificity directly influence diagnostic success rate. Institute of Molecular Genetics and Genetic Engineering (IMGGE) has an in-house short-read whole genome sequencing (SR-WGS) pipeline for variant detection, without prior cross-institutional benchmarking in the context of rare disease diagnostics performed.

As part of the BRIDGING-RD project, IMGGE SR-WGS pipeline was benchmarked against the NIST HG002/NA24385 Genome in a Bottle gold-standard truth set (v4.2.1), in collaboration with Centro Nacional de Análisis Genómico (CNAG). Both centers used the same 30x coverage Illumina short-read FastQ data generated from HG002 NIST reference material. Benchmarking was performed using Illumina's open-source *hap.py* framework, with evaluation restricted to the ~88.5% of the autosomal GRCh38 genome covered by the NIST high-confidence callable regions BED file. Precision, Recall, and F1 score were calculated for single nucleotide variants (SNVs) and short insertions and deletions (InDels).

Both pipelines performed well against the gold-standard truth set. The IMGGE workflow achieved F1 scores of 0.971–0.981, with SNV Precision of 99.41–99.76% and Recall of 95.35–96.81%, and InDel Precision of 98.99–99.19% and Recall of 95.07–96.42%. The CNAG pipeline achieved slightly higher F1 scores of 0.986–0.991 across all categories, with higher Recall for both variant types. Notably, IMGGE workflow showed higher Precision but lower Recall compared to CNAG, suggesting that post-calling filtration step in the IMGGE pipeline reduces false positives at the cost of increased false negatives.

The IMGGE SR-WGS pipeline demonstrated strong performance, confirming suitability for rare disease genomic testing. Comparison with CNAG pipeline identified optimization opportunities, particularly around filtering strategies and recall sensitivity. This work was conducted through knowledge transfer between CNAG and IMGGE as part of the BRIDGING-RD project, in which IMGGE staff received practical training in benchmarking methodology, genomics file formats, and open-source bioinformatics tools. The capacity to independently benchmark WGS pipelines was thereby established at IMGGE, laying a foundation for quality-assured workflows in rare disease genomics.

Acknowledgement:

This research was supported by the Horizon Europe Project BRIDGING-RD, HORIZON-WIDERA-2023-ACCESS-02, Grant Agreement N°101160079.

Keywords:

WGS, benchmarking, bioinformatics, pipeline, genomics

Abstract ID: **64** Type: **Poster Presentation**

Implementation and clinical utility of the GATK gCNV pipeline for rare disease diagnostics using whole exome sequencing

Djordje Pavlovic[1], Anita Skakic[1], Kristel Klaassen[1], Irena Marjanovic[1], Marina Parezanovic[1], Nina Stevanovic[1], Goran Cuturilo[2], Marija Brankovic[2], Brankica Bosankic[2], Jelena Ruml Stojanovic[2], Bojan Ristivojevic[1], Branka Zukic[1], Maja Stojiljkovic[1], Marina Andjelkovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *University Children's Hospital, Belgrade, Serbia*

Contact e-mail: **djordje5996@gmail.com**

Copy number variants (CNVs) are critical contributors to rare disease etiology. While standardized workflows exist for detecting short variants from exome sequencing, historically, detecting CNVs required separate sequencing approaches or clinical assays. Here, we present the implementation of the GATK gCNV pipeline to detect CNVs using routine whole-exome sequencing (WES) in rare disease diagnostics.

We performed Illumina short-read WES on 646 rare disease patients from the Belgrade University Children' s Hospital over a one year period (Apr 2025 - Apr 2026). Data was processed utilizing standard GATK Best Practices for SNVs alongside the GATK gCNV pipeline for detecting structural variants. The gCNV workflow leverages read-depth coverage across exonic targets, normalizes read counts against a robust cohort model to correct for technical biases like GC content and bait capture efficiency. A Bayesian hidden Markov model (HMM) then infers copy-number states to call specific duplications and deletions. These were then filtered according to quality score, and analyzed for pathogenicity and disease causality in the Varsome Clinical software.

Genetic analysis of both single-nucleotide variants and copy number variants resulted in a positive diagnosis in 209 patients (32%), including 187 with SNVs and 22 with CNVs. An inconclusive result was obtained in 66 patients (10%), while 371 patients (58%) had a negative result. Detected CNVs were subsequently validated by chromosomal microarray analysis, with 21 being confirmed (>95%). Among the 209 positive diagnoses, 21 (approximately 10%) were clinically significant CNVs detected by the gCNV pipeline. These CNVs comprised 11 deletions and 10 duplications, with chromosome 16 being the most frequently affected (7 CNVs).

Integrating the GATK gCNV pipeline into exome sequencing data workflows provides a robust and reliable approach for CNV calling. This allows detecting pathogenic CNVs directly from exome data, which enhances diagnostic yield and streamlines clinical testing of rare diseases.

Acknowledgement:

This work has been realized within the Center for Genetic Diagnostics of Rare Diseases, and funded by BRIDGING-RD, HORIZON-WIDERA-2023-ACCESS-02, N°101160079

Keywords:

CNV, rare disease, GATK

Abstract ID: **66** Type: **Poster Presentation**

FluxInside: A high-performance Flux Balance Analysis engine for spatial agent-based models.

Burak Özyürek[1]

[1] *Gebze Technical University*

Contact e-mail: **b.ozyurek2021@gtu.edu.tr**

Agent-based models (ABMs) are highly effective at simulating multi-scale spatial interactions within biological environments. However, standard ABMs typically lack the capacity to mechanistically simulate intracellular metabolic activity at the single-cell level.
To address this issue, we have developed FluxInside: a novel, high-performance Flux Balance Analysis (FBA) extension designed specifically for agent-based models. Utilising the GNU Linear Programming Kit (GLPK), the module was developed and integrated into the PhysiCell/BioFVM ABM platform through optimised, object-oriented C++ architecture. Our algorithm uses a 'warm start' to update localised environmental changes dynamically, rather than rebuilding the FBA object. This reduces execution time to under 5 milliseconds per FBA of a genome-scale metabolic model. This architecture seamlessly integrates cell-type-specific genome-scale metabolic models (GEMs) into local niches of agent-based environments.
To validate this computational framework, we simulated a highly heterogeneous glioblastoma microenvironment comprising over 50 thousand agents and dozens of diffusible substrates. Additionally, we tested it across multiple GEMs belonging to various organisms (E. coli, S. cerevisiae, and H. sapiens). Driven purely by intracellular FBA calculations and without any hard-coding, the agents naturally recapitulated emergent, systems-level behaviours including the Warburg effect, TCA switching, necrotic core formation and hypoxia-induced cellular motility. Subsequent in silico therapeutic trials successfully captured the dynamics of metabolic competition underlying immunotherapy resistance.
FlowsInside provides a highly scalable, bidirectional link between intracellular metabolic flux and tissue-scale continuum dynamics. By enabling metabolically aware simulations for tens of thousands of individual agents, this platform offers a powerful new framework for systems biology research, in silico experimental design, and the optimization of targeted therapies.

Acknowledgement:

This project is generously funded by TÜBİTAK UIDB 1071 (ERA-NET TRANSCAN-3 , 122N905)

Keywords:

Mathematical modelling, agent-based modeling

Abstract ID: **67** Type: **Poster Presentation**

Transcriptomic analysis reveals suppression of the filamentation-associated virulence program in Candida albicans

Natasa Radakovic[1], Ivana Moric[1], Dejan Opsenica[2,3], Lidija Senerovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Institute of Chemistry, Technology, and Metallurgy, University of Belgrade*

[3] *Centre of Excellence in Environmental Chemistry and Engineering, ICTM*

Contact e-mail: **natasa.radakovic@imgge.bg.ac.rs**

The ability of *Candida albicans* to undergo yeast-to-hypha transition is a key determinant of its virulence and represents an attractive target for anti-virulence strategies. In our previous work, we synthesized and characterized a candidate compound structurally related to 2-heptyl-4-quinolone (HHQ), which inhibits filamentation. Here, we aimed to elucidate its mechanism of action using transcriptomic profiling and functional validation.
RNA-seq analysis was performed in RPMI medium, a condition that supports filamentation, enabling assessment of compound effects on the hypha-associated transcriptional program. Differential gene expression analysis was conducted by comparing treated samples to untreated controls. In addition, the compound's activity was compared to that of HHQ, a reference molecule from the same structural class.
Transcriptomic data revealed significant downregulation of hypha-associated genes, including *ECE1* and *ALS3*, indicating suppression of the filamentation-associated transcriptional program. Functional validation using an *ECE1* reporter strain confirmed a marked reduction in *ECE1* expression, accompanied by inhibition of filamentation at the phenotypic level. Notably, both transcriptional and phenotypic effects of the tested compound were more pronounced than those observed with HHQ.
Together, these findings indicate that the compound exerts a targeted anti-virulence effect by suppressing key regulatory pathways associated with filamentation in *C. albicans*. This study highlights the value of integrating transcriptomic analysis with functional assays to characterize novel anti-virulence agents.

Keywords:

Candida albicans, transcriptomics, ECE1, anti-virulence

Abstract ID: **71** Type: **Poster Presentation**

Contribution of ILCs to the development of different forms of multiple sclerosis

Suzana Stanisavljević[1]

[1] *Institute for Biological Research "Siniša Stanković"*

Contact e-mail: **ssuzana@ibiss.bg.ac.rs**

Innate lymphoid cells (ILCs) are emerging regulators of immune responses, yet their role in autoimmune diseases remains insufficiently characterised. In multiple sclerosis (MS), there are controversial findings regarding the role of ILCs in the pathogenesis of the disease. Some studies in an animal model of MS, experimental autoimmune encephalomyelitis (EAE), are showing that the infiltration of ILCs in the central nervous system can exacerbate the disease. Other studies are exploiting modulation of some ILCs, such as gut ILC3, in order to alleviate the EAE. In order to better understand the potential role of ILCs in MS, a computational re-analysis of publicly available transcriptomic datasets is performed. Using signature genes, ILC subsets (ILC1, ILC2 and ILC3) were identified. Relative enrichment of ILC subsets was quantified in peripheral blood samples of healthy controls and patients with different forms of MS, relapsing-remitting and progressive MS. Differences in transcriptional profile across MS disease forms were analysed and used as a potential indicator of the role of specific ILC subsets in the progression of MS.

Our analysis revealed a consistent shift in ILC-associated signatures between disease stages. Progressive MS samples exhibited increased enrichment of ILC1- and ILC3-related transcriptional profiles, accompanied by a relative reduction in ILC2-associated signatures. This pattern suggests a polarisation toward pro-inflammatory innate lymphoid profiles with disease progression. Notably, these trends were reproducible across independent datasets, supporting their strength despite differences in cohort composition and experimental platforms.

These findings highlight the importance of ILC polarisation as a potential contributor to disease progression in MS. Moreover, our study demonstrates that meaningful insights can be obtained from transcriptomic data through targeted computational approaches.

Keywords:

transcriptomics, multiple sclerosis, ILC

Abstract ID: **75** Type: **Poster Presentation**

Assessment of coumarin derivatives potential as cyclooxygenase-2 inhibitors

Vladimir Vlatković[1], Katarzyna Wicha-Komsta[2], Tomasz Kocki[2], Łukasz Komsta[3], Saša Lazović[4], Darija Obradović[4]

[1] *Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade*

[2] *Institute of Medical Sciences, Faculty of Medicine, The John Paul II Catholic University of Lublin, Lublin, Poland*

[3] *Faculty of Pharmacy, Medical University of Lublin, Lublin, Poland*

[4] *Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia*

Contact e-mail: **vlatkovic@ipb.ac.rs**

Inhibition of the cyclooxygenase-2 (Cox-2) enzyme is important for the mediation of inflammation. The binding of inhibitors to the Cox-2 active site blocks the conversion of arachidonic acid into pro-inflammatory prostaglandins, thus reducing inflammation. This is a well-established and important direction in the development of anti-inflammatory substances. Coumarin derivatives are known to possess Cox-2 inhibitory potential. The goal of this study is to examine the binding affinity of selected coumarin derivatives to the Cox-2 active site and to gain insight into important molecular interactions governing the binding, using molecular docking. A parallel study was performed for celecoxib, a well-known and effective Cox-2 selective inhibitor, for comparison. The structure of the Cox-2 active site (PDB ID: 3LN1) was obtained from the RCSB PDB database. Target protein structure and ligand molecules were preprocessed, optimized and refined before the docking study. Amino acid residues key to binding of celecoxib to the active site were chosen for the generation of the receptor grid box: His 75, Val 102, Arg 106, Gln 178, Val 335, Leu 338, Ser 339, Tyr 341, Leu 345, Phe 367, Leu 370, Tyr 371, Trp 373, Arg 499, Phe 504, Val 509, Ser 516, and Leu 517. The docking study was performed using the Glide molecular modeling tool. Ligand binding affinity was based on the Glide scoring function that accounts for Coulomb-van der Waals energies, hydrogen bond interactions and lipophilic interactions for high-accuracy assessment of binding strength.
The results indicated that the tested coumarin derivatives had a binding affinity to the active site defined in the docking study, similar to celecoxib. Binding affinities are as follows: umbelliferone> 4-metoxyumbelliferone> celecoxib> 7-metoxycoumarin> coumarin> 6,7-dimetoxycoumarin> 6,7-dixydroxycoumarin> 3,3-methylen-bis-hydroxycoumarin. Visualization of the binding complex showed that the studied ligands were located in the binding site consisting of large, aromatic, hydrophobic, as well as positively charged and polar amino acid residues. Tested coumarin derivatives interacted primarily by hydrogen bonds, hydrophobic contacts and Pi-cation (Pi-Pi for celecoxib) interactions. The applied docking method represents a fast and precise approach that can be used for the preliminary screening of novel coumarin derivatives as potential Cox-2 inhibitors.

Keywords:

Cox-2 inhibition. coumarin derivatives. docking

Abstract ID: **77** Type: **Poster Presentation**

Benchmarking of pre-processing steps in Sarek pipeline on MGI sequenced standard human genome

Marina Jelovac[1], Djordje Pavlovic[1], Nevena Vucinic[2], Novak Martinovic[2], Nikola Kotur[1], Aleksandra Vitkovac[1], Saša Todorović[1], Iva Sabolic[3], Aleksandar Mihajlovic[2], Branka Zukic[1], Biljana Stankovic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Persida doo*

[3] *Labena doo*

Contact e-mail: **marina.jelovac@imgge.bg.ac.rs**

The expanding availability of next-generation sequencing (NGS) technologies has enabled the generation of massive genomic datasets. In addition to challenges related to data storage and processing, ensuring reproducibility of genome analyses remains critical. Workflow management systems such as Nextflow have been developed to address these issues. Sarek, a freely available nf-core pipeline for detecting genetic variants in whole-genome sequencing data, integrates widely accepted bioinformatics tools considered the gold standard in life sciences. However, NGS protocols differ in library preparation methods, which may influence the necessity of certain analytical steps. Notably, MGI library preparation does not involve PCR amplification, potentially reducing the need for duplicate marking and base quality score recalibration (BQSR). Eliminating these computationally intensive steps could significantly reduce analysis time and resource usage.

The aim of this study was to assess the necessity of alignment preprocessing steps and evaluate the impact of variant caller choice when applying the Sarek pipeline to MGI sequencing data.

The Sarek pipeline was adapted to include multiple analysis strategies applied to the NA12878/HG001 human genome sample, sequenced using MGI technology and provided by Genome in a Bottle consortium. Reads were aligned to the GRCh38 reference genome using bwa-mem. Preprocessing steps included adapter trimming (fastp), duplicate marking (GATK MarkDuplicatesSpark), and base quality score recalibration (GATK BaseRecalibrator and ApplyBQSR), applied individually, in combination, or omitted. Variant calling was performed using GATK HaplotypeCaller and DeepVariant. Variant call files were benchmarked against a known truth set using hap.py, and performance was evaluated using F1 scores for SNPs and INDELs.

No significant differences in variant calling performance were observed between analyses that included duplicate marking and BQSR and those that omitted them. DeepVariant consistently outperformed HaplotypeCaller, although differences were modest. For INDEL detection, DeepVariant achieved an F1 score of 99.5% compared to 99.1%, while for SNP detection it reached 99.7% versus 99.2%.

In conclusion, duplicate marking and BQSR have limited impact on variant calling accuracy in MGI data and can be omitted to reduce computational cost. DeepVariant showed a consistent advantage and may be the preferred variant caller for this data type.

Acknowledgement:

This work has been funded by EC project HORIZON-WIDERA-2023-ACCESS-07-01: InnoThyroGen (GA No. 101187880)

Keywords:

bioinformatics, sequencing, sarek, pipeline, benchmarking

Abstract ID: **82** Type: **Poster Presentation**

Classification of coronavirus types based on repeats derived from amino acid and nucleotide sequences

Anđela Damnjanović[1]

[1] *Faculty of Mathematics, Belgrade, Serbia*

Contact e-mail: **andjela.damnjanovic13@gmail.com**

The spike protein is one of the four major structural proteins that constitute coronaviruses, in addition to the membrane, nucleocapsid, and envelope proteins. These proteins play important roles in different stages of the infectious cycle. Despite its relatively small size (compared to the entire genome), the spike protein is the most important, as it enables the virus to bind to the host cell receptor and enter the cell, thereby providing a suitable environment for the creation of new viral copies. Since different types of coronaviruses can infect different hosts, there must also be differences in the spike proteins that bind to those hosts.

Therefore, the aim of this study was to develop and evaluate several classification models for predicting coronavirus types based on features extracted from amino acid and nucleotide sequences. Repeated subsequences (repeats) were used as features. In nucleotide sequences, four types of repeats can be identified: direct non-complementary, direct complementary, inverse non-complementary, and inverse complementary. On the other hand, due to the absence of a complementarity concept among amino acids, only direct non-complementary and inverse non-complementary repeats can be found in amino acid sequences.

The dataset used in this study consisted of spike proteins from 20 different types of coronaviruses, resulting in over 26,000 instances. The implemented classification models included Decision Trees, k-Nearest Neighbors (kNN), Neural Networks, and the Naive Bayes classifier, as well as ensemble methods such as Random Forest and XGBoost. In addition, a Voting classifier was tested, combining individual classifiers with complementary strengths. Model performance was evaluated using standard metrics: precision, recall, and F1-score.

The models achieved very high performance, reaching up to 99% accuracy, with precision and recall exceeding 90% on most datasets, and macro-averaged F1-scores above 0.9 in the majority of cases, thus showing that repeat-based features provided meaningful information for distinguishing between coronavirus types. Ensemble methods, particularly Random Forest and XGBoost, achieved the best overall performance, while combined classifiers further improved prediction accuracy by reducing individual model weaknesses.

In conclusion, text-inspired feature extraction from biological sequences effectively supports virus classification, while combining models and engineered repeat features improves robustness and accuracy.

Acknowledgement:

The author is partially supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia, through the project 451-03-33/2026-03/200104.

Keywords:

coronavirus, classification, classification methods, repeats

Abstract ID: **91** Type: **Poster Presentation**

Integrated chemical and metagenomic profiling of urban wastewater: antibiotic residues, microbial community and resistome at a Belgrade collector point

Radmila Novakovic[1], Milos Jovicevic[2], Dusan Kekic[3], Svetlana Grujić[4], Ljiljana Tolic Stojadinovic[5], Nemanja Mirkovic[6], Olja Sovljanski[7], Aleksandra Vitkovac[1], Petra Stojanovic[1], Milos Veselinovic[1], Ivan Vicic[8], Milica Mirkovic[9], Nedeljko Karabasil[8], Natasa Opavski[2], Ina Gajic[2]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Medical Faculty, University of Belgrade, Belgrade, Serbia*

[3] *Institute of MicrobioMedical Faculty, University of Belgrade, Belgrade, Serbia*

[4] *Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, Serbia*

[5] *Innovation Centre of the Faculty of Technology and Metallurgy, Belgrade, Serbia*

[6] *Faculty of Agriculture University of Belgrade, Belgrade, Serbia*

[7] *Faculty of Technology University of Novi Sad, Serbia*

[8] *Faculty of Veterinary Medicine, University of Belgrade, Belgrade, Serbia*

[9] *Faculty of Agriculture, University of Belgrade, Belgrade, Serbia*

Contact e-mail: **radmila.novakovic@imgge.bg.ac.rs**

Urban wastewater is a convergence point for antibiotic residues and antibiotic resistance genes (ARGs) drained from human, clinical and environmental sources, and is increasingly used as an integrative matrix for antimicrobial resistance surveillance. Chemical quantification with shotgun metagenomics on the same sampling point strengthens the interpretation of the local resistome and offers a more biologically meaningful readout than either approach in isolation. In this work we combined HPLC–MS/MS quantification of nine antibiotics with shotgun metagenomic characterisation of a single grab sample collected from a major urban collector Dorcol in Belgrade, Serbia. Reads were taxonomically profiled with Kraken2 and the resistome was screened in parallel with two complementary tools, CARD–RGI and ResFinder. HPLC–MS/MS detected six out of nine target antibiotics in the sample (ng/L), dominated by ciprofloxacin (1940.5), followed by doxycycline (77.4), trimethoprim (70.0), sulfamethoxazole (47.7), amoxicillin (20.1) and azithromycin (13.9). Gentamicin, colistin and ceftriaxone were below the limit of detection. Metagenomic sequencing yielded ~6.07 million reads (90.07% bacterial), with Bacillota (27.96%), Pseudomonadota (26.43%) and Bacteroidota (20.98%) dominating, and the top genera (Acinetobacter 8.94%, Prevotella 4.65%, Bacteroides 4.32%, Aeromonas 3.51%) matching taxonomic signatures consistently reported for raw municipal influents. CARD–RGI identified 1,402 ARO terms (159 at ≥80% gene coverage) and ResFinder confirmed 86 high-confidence acquired ARGs across 14 drug classes. The resistome was internally consistent with the chemical profile: the classes detected at the highest concentrations corresponded to the most abundant and most diverse resistance determinants, while colistin, undetected chemically, yielded no resistance genes. Beyond this expected baseline, the sample carried clinically significant last-resort determinants, including extended-spectrum and carbapenem-hydrolysing beta-lactamases (blaVEB-1, blaGES-5, blaOXA-427), the tigecycline-inactivating tet(X), and cfr(C), conferring cross-resistance to linezolid, phenicols and pleuromutilins. The integrated chemical–metagenomic profile of this single collector point falls within the range expected for untreated urban wastewater, while the recovery of last-resort resistance markers confirms that clinically critical determinants are already circulating in the local sewerage system. Although based on a single grab sample, the study establishes a methodological and gene-prioritizing method for subsequent longitudinal, multi-site surveillance of urban wastewater in Serbia.

Keywords:

OneHealth, Wastewater, Metagenomics, Resistome, ResFinder

Abstract ID: **94** Type: **Poster Presentation**

Identification of shared molecular programs and putative role of ankrd1a in zebrafish heart regeneration and skeletal muscle repair using WGCNA

Mirjana Novković[1], Aleksandra Vitkovac[1], Andjela Milicevic[1], Emilija Milosevic[1], Jovana Jasnic[1], Vladimir Jovanovic[2], Snezana Kojic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Human Biology and Primate Evolution Group, Department for Biology, Chemistry and Pharmacy, Freie Universität Berlin, Berlin, Germany*

Contact e-mail: **mirjana.novkovic@imgge.bg.ac.rs**

Zebrafish possess a remarkable capacity to regenerate both cardiac and skeletal muscle tissues, whereas most higher vertebrates, including humans, have largely lost the ability to regenerate the heart, while retaining the capacity for skeletal muscle repair. Thus, the key for unlocking the endogenous regenerative potential of human heart may be hidden in common genes involved in both processes. Zebrafish regenerate heart through coordinated tissue remodeling and regenerative signaling processes which induce cardiomyocyte dedifferentiation and proliferation, enabling restoration of lost myocardium instead of forming a permanent fibrotic scar as in humans. On the other hand, zebrafish repair skeletal muscle through an evolutionarily conserved multistep process that relies on activation and proliferation of tissue-resident stem (satellite-like) cells, which differentiate into muscle cell progenitors and finally to myofibers, leading to functional muscle recovery. To gain deeper insight into common genes and pathways involved in heart regeneration and skeletal muscle repair, and the role of ankrd1a gene, we performed Weighted Gene Co-expression Network Analysis on RNA sequencing data from injured tissues of wild type and ankrd1a-/- zebrafish (GSE285387, GSE277480). Regarding injury, the heart showed 3 positive and 3 negative, while skeletal muscle showed 2 positive and 1 negative correlated modules. Independent network analysis revealed that injury in both cardiac and skeletal muscle tissues induces highly coordinated transcriptional programs enriched for extracellular matrix organization, inflammatory pathways, and cell cycle, while suppressing muscle contraction, oxidative phosphorylation and other metabolic processes. Despite tissue-specific module structures, functional convergence was observed at the pathway level, supported by sharing more than 40% of central genes across the tissues. We further investigated the role of ankrd1a, a well-established stress-responsive mechanosensor and transcription cofactor that is rapidly and consistently upregulated during cardiac and skeletal muscle healing in zebrafish. Genotype-associated transcriptomic changes were reflected in independent co-expression networks by 3 positively and 1 negatively correlated module in both heart and skeletal muscle. Enrichment analyses showed upregulation of pathways associated with spliceosome, transcription, and post-translational protein modification, suggesting activation of compensatory mechanisms that may help zebrafish adapt to the loss of functional ankrd1a protein.

Keywords:

heart, regeneration, muscle, WGCNA, ankrd1a

Abstract ID: **95** Type: **Poster Presentation**

In silico approach to proteomic identification of biologically active proteins in an insect non-model species, Ostrinia nubilalis

Vanja Tatić[1], Éva Schád[2], Miloš Avramov[1], Teodora Knežić Avramov[3], Ágnes Révész[4], László Drahos[4], Ágnes Tantos[5], Željko D. Popović[1]

[1] *University of Novi Sad, Faculty of Sciences, Department of Biology and Ecology, Novi Sad, Serbia*

[2] *HUN-REN Research Centre for Natural Sciences, Institute of Molecular Life Sciences*

[3] *University of Novi Sad, BioSense Institute, Serbia*

[4] *Institute of Organic Chemistry, Research Centre for Natural Sciences, 1117 Budapest, Hungary*

[5] *HUN-REN Research Centre for Natural Sciences*

Contact e-mail: **vanja.tatic@dbe.uns.ac.rs**

Proteomic analyses based on mass spectrometry (MS) are widely used to study biological systems, but most available bioinformatic tools are optimized for well-annotated model organisms such as mice, humans, *Drosophila* or Arabidopsis. Consequently, research on non-model species is hindered by incomplete protein annotation, limited pathway databases, and reduced accuracy in functional enrichment analyses. These limitations restrict the interpretation of MS data and reduce reproducibility across studies. Therefore, there is a growing need for novel and flexible analytical approaches that can accommodate poorly annotated genomes and species-specific proteins. Developing improved computational pipelines and expanding annotation resources would enhance proteomic analysis in non-model organisms, enabling more accurate biological interpretation and supporting broader applications in ecology, agriculture, and evolutionary biology.

Mass spectrometry data from *Ostrinia nubilalis* larvae infected with two bacterial strains, *P. aeruginosa* and *S. aureus*, was analysed with the goal of obtaining a functional annotation of differentially represented proteins in four tissues: epidermis, intestine, fat body and haemolymph. Proteins that showed two-fold increase compared to the control were considered overrepresented, while those that were reduced to at least half were considered underrepresented in each tissue, under each infection treatment. EggNOG-mapper and its database of ortholog groups was used to obtain predicted protein names, KEGG pathways, modules, as well as orthologs and Gene Ontology (GO) labels. The standard for GO enrichment in the R environment, clusterProfilier, was shown to be inadequate for non-model species, so downstream GO analysis was performed using ontologyIndex together with tidyverse. GO annotations were filtered to remove obsolete and unannotated terms, and mapped to broader GO Slim categories using the current Gene Ontology hierarchy. This allowed proteins to be grouped into three major groups: biological processes, molecular functions, and cellular components, and their relative representation across tissues and treatments was quantified and visualised. Overall, this approach provided a simplified functional overview of protein changes in a non-model insect system, making the biological interpretation of large proteomic datasets more accessible.

Keywords:

Proteomics; Mass spectrometry; Non-model organisms

Acknowledgement:

Acknowledgement: The authors gratefully acknowledge the financial support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Grants No. 451-03-33/2026-03/ 200125, 451-03-34/2026-03/ 200125 and 451-03-33/2026-03/ 200358), the Provincial Secretariat for Higher Education and Scientific Research (Grants No. 003801787 2025 09418 003 000 000 001/1 and No. 003956066 2025 09418 003 000 000 001 04 003) and the Serbian-Hungarian joint research project ''Fighting against antibiotic resistance –identification of prospective antimicrobial peptides from alternative sources (ALARM)".

Abstract ID: **97** Type: **Poster Presentation**

IntelliLung: A human-in-the-loop AI-DSS for ICU ventilation settings

Dejan Paunović[1,2], Jakob Wittenstein[1,3], Jason Li[1,4], Jaume Montanya[1], Jens Lehmann[1], Martin Faltys[1], Miguel Ángel Armengol de la Hoz[1], Miloš Nenadović[1,2], Muhammad Hamza Yousuf[1,4], Robert Huhle[1,5], Sahar Vahdati[1], Thea Koch[1], Valentina Janev[1,2], Álvaro Ritoré Hidalgo[1]

[1] *IntelliLung Consortium*

[2] *Institute Mihajlo Pupin*

[3] *University Hospital Carl Gustav Carus*

[4] *Institut für Angewandte Informatik*

[5] *Technische Universität Dresden*

Contact e-mail: **milos.nenadovic@pupin.rs**

IntelliLung is a human-in-the-loop AI decision-support system for optimizing mechanical ventilation in intensive care units (ICUs) using bedside and physiologic monitor streams. Patient data are ingested through a secure on-premises platform that normalizes, stores, and governs data transfer to ensure traceability, auditability, and regulatory compliance. Using a data aggregator that converts inputs into hourly patient-state representations, an AI engine generates hybrid recommendations that jointly propose ventilator mode selection and coordinated adjustments of key parameters. Recommendations are delivered via a clinician-facing interface embedded within existing ICU workflows, where each suggestion can be explicitly accepted or rejected. IntelliLung functions strictly as an advisory system. The demonstration replays realistic ICU patient trajectories to showcase system behavior, recommendation progression under changing physiology, and mechanisms for safe human-AI collaboration in clinical settings.

Keywords:

DSS, ventilation, ICU, human-in-the-loop, AI

Acknowledgement:

Our work is funded by the European Union IntelliLung project with Grant Agreement No. 101057434.

Abstract ID: **98** Type: **Poster Presentation**

Pilosibacter species emerge as leading contributors to glutarate pathway-mediated butyrate synthesis in the human gut

Aleksandar Bisenić[1], Dušan Radojević[1], Emilija Brdarić[1], Hristina Mitrović[1], Jelena Đokić[1], Miroslav Dinić[1], Nataša Golić[1], Sergej Tomić[2], Stefan Jakovljević[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Institute for the Application of Nuclear Energy, University in Belgrade*

Contact e-mail: **aleksandar.bisenic@imgge.bg.ac.rs**

Although most intestinal butyrate is produced through the pyruvate pathway, alternative butyrate synthesis pathways, including the glutarate pathway, are consistently detected across human gut metagenomes. However, the taxonomic basis of this highly prevalent functional capacity remains poorly resolved, as currently recognized glutarate pathway-utilizing gut butyrogens do not adequately explain its near-universal occurrence. Here, we present *Pilosibacter* species as highly prevalent gut commensals that help explain the taxonomic basis of glutarate pathway-mediated butyrate synthesis. Fecal isolate NGB245 was identified as a novel *Pilosibacter* species and characterized as a glutamate-fueled butyrate-producing gut commensal with *in vitro* anti-inflammatory and intestinal barrier-preserving properties. To define its metabolic specialization, we combined hybrid genome sequencing and genome-resolved comparative functional analysis, integrating KEGG module completeness, COG category enrichment, GO term enrichment, and targeted annotation searches across *Pilosibacter* genomes and dominant saccharolytic butyrate producers. This revealed a coherent genomic signature centered on glutamate acquisition and fermentation via the glutarate pathway, including expanded capacities for the acquisition and salvage of diverse glutamate equivalents. These predictions were supported experimentally, as glutamate supplementation enhanced growth and butyrate production in *Pilosibacter* sp. NGB245. Finally, genome-resolved metagenome tracking across 142 healthy adult fecal metagenomes from five geographically distinct cohorts identified *Pilosibacter* sp. NGB245 and Pilosibacter fragilis as highly prevalent human gut taxa, while relative abundance positioned *Pilosibacter* species as the leading identified contributors to glutarate pathway-mediated butyrate synthesis within the analyzed reference panel. This study provides the first reported functional genomics, metagenomics, and in vitro probiotic characterization of *Pilosibacter*. Our findings define *Pilosibacter* as a low-abundance but highly prevalent glutamate-specialized butyrogenic lineage, provide a strong taxonomic basis for the widespread glutarate pathway signal in human gut metagenomes, and expand the current view of prevalent gut butyrate producers beyond saccharolytic species.

Keywords:

Pilosibacter, butyrate, glutamate, comparative genomics

Acknowledgement:

This work was supported by the Science Fund of the Republic of Serbia under Grant IDEAS (7744507, NextGenBiotics); the Ministry of Science, Technological Development and Innovations of the Republic of Serbia under Grant (451-03-66/2024-03/200042).

Abstract ID: **99** Type: **Poster Presentation**

Comprehensive genomic profiling of advanced Non-small Cell Lung Cancer in Serbia: Implications for targeted therapy selection

Katarina Krstajić[1], Djordje Stojanovic[1], Djordje Pavlovic[1], Natasa Tosic[1], Filip Markovic[2], Bojan Ristivojevic[1], Ivana Grubisa[1], Irena Marjanovic[1], Teodora Karan Djurasevic[1], Milica Kontic Jovanovic[2], Branka Zukic[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

[2] *Clinic for Pulmonology, University Clinical Centre,11000 Belgrade, Serbia*

Contact e-mail: **katarina.krstajic@imgge.bg.ac.rs**

Lung cancer is the leading cause of cancer-related mortality worldwide and in Serbia. Despite advances in diagnostics and therapy, the disease is diagnosed in most patients at an advanced or metastatic stage, resulting in extremely low five-year survival rates. Non-small Cell Lung Cancer is characterized by pronounced molecular heterogeneity and the presence of genomic alterations that may serve as therapeutic targets. This study aimed to characterize the genomic profile of advanced NSCLC and identify therapeutically relevant genomic alterations using comprehensive genomic profiling. Genomic profiling of formalin-fixed paraffin-embedded tumor tissue samples from 72 patients with advanced NSCLC was performed using the AVENIO Tumor Tissue comprehensive genomic profiling kit based on next-generation sequencing technology. Data analysis and interpretation were conducted using the NAVIFY® Mutation Profiler software based on the FoundationOne® platform. Only variants with clear or potential clinical significance and a minimum variant allele frequency of 3% were included in the analysis. The most frequent genomic alterations were detected in the genes *TP53* (67%), *KRAS* (40%), *STK11* (32%), *KMT2D* (26%), and *KEAP1* (24%). Additional genomic alterations included single nucleotide variants, insertions/deletions, and copy number variations. Concurrent alterations affecting key tumorigenic signaling pathways, including the RAS/MAPK and PI3K/AKT pathways, indicated substantial molecular heterogeneity. Clinically relevant genomic alterations were identified in a significant proportion of patients, most commonly involving *KRAS* G12C (13.9%), *EGFR* alterations, including exon 19 deletions, L858R, G719C, S768I, and L861Q mutations, and *ALK* rearrangements (6.9%). Less frequent alterations were detected in *MET* and *ERBB2*, including amplifications and exon 20 insertions, representing potential targets for existing or investigational targeted therapies. These findings contribute to a better understanding of the molecular landscape of advanced NSCLC in Serbia and support the implementation of personalized therapeutic approaches in routine clinical practice. Furthermore, the study highlights the importance of integrating comprehensive genomic profiling into clinical workflows for the identification of predictive biomarkers and novel therapeutic opportunities.

Keywords:

NSCLC, CGP, NGS, biomarkers, oncology

Abstract ID: **101** Type: **Poster Presentation**

Sequence diversity and genomic relatedness of blaVIM-positive clinical Proteus mirabilis isolates

Filip Božić[1,2], Mirjana Hadnađev[2], Ljuban Blanuša[2], Jovana Kabić[3], Ina Gajić[3], Miloš Jovićević[3], Dušan Kekić[3], Anika Trudić[2,4]

[1] *Faculty of Sciences, Department of Biology and Ecology*

[2] *Institute for Pulmonary Diseases of Vojvodina, Sremska Kamenica, Serbia*

[3] *Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade, Belgrade, Serbia*

[4] *University of Novi Sad, Faculty of Medicine, Novi Sad, Serbia*

Contact e-mail: **flpbzc@gmail.com**

Carbapenemase-producing Enterobacterales represent a major therapeutic and infection-control challenge, but are most frequently associated with Klebsiella pneumoniae, whereas VIM-producing Proteus mirabilis remains an uncommon finding. In Serbia, VIM-producing P. mirabilis has only sporadically been reported, and genomic data on blaVIM allelic diversity, associated resistance determinants, and hospital-associated clonal relatedness remain limited.

This study aimed to genomically characterize five clinical P. mirabilis isolates recovered from intensive care and pulmonary units of a tertiary care hospital in Serbia, focusing on sequence type, antimicrobial resistance gene repertoire, blaVIM allelic diversity, plasmid replicon content and core-genome relatedness.

Whole-genome sequencing was performed on an Illumina NovaSeq 6000 platform using 150 bp paired-end reads. Quality-filtered reads were assembled with SPAdes and assessed with QUAST. Downstream analyses included Proteus MLST, antimicrobial resistance gene detection using ResFinder v4.1 and ABRicate-based database screening, plasmid replicon detection with PlasmidFinder, blaVIM mapping with BWA-MEM/SAMtools and Snippy-based core-genome SNP analysis for assessment of isolate relatedness.

All isolates belonged to ST135 and carried blaVIM carbapenemase genes together with multiple resistance determinants affecting beta-lactams, aminoglycosides, macrolides, sulfonamides, tetracyclines, trimethoprim and phenicols. Comparative analysis of the 801 bp blaVIM coding sequence identified distinct allelic variants: blaVIM-1 in isolate 5, blaVIM-4 in isolate 1, blaVIM-75 in isolate 4 and blaVIM-78 in isolates 2 and 3. Relative to the blaVIM-1 reference sequence, these variants differed by substitutions at positions 179 A→G and/or 613 A→C. IncQ1 replicons were detected in four isolates; however, they were located on contigs distinct from blaVIM-containing contigs, indicating that short-read assemblies did not resolve their physical linkage. Core-genome SNP analysis separated isolate 3 from a closely related cluster comprising isolates 1, 2, 4 and 5, despite isolates 2 and 3 carrying identical blaVIM-78 coding sequences.

These findings indicate that blaVIM-positive P. mirabilis isolates from the same hospital belonged to a single sequence type, yet blaVIM allelic distribution did not fully mirror core-genome relatedness. This suggests that blaVIM diversity may not be explained solely by clonal expansion of ST135 and may reflect acquisition or rearrangement of blaVIM-harbouring elements, although their precise genetic context could not be resolved using short-read sequencing alone.

Keywords:

Proteus, blaVIM, carbapenemase, multidrug-resistant, WGS

Acknowledgement:

This research was funded by the Science Fund of the Republic of Serbia, Grant No. 7042, Tracking antimicrobial resistance in diverse ecological niches - one health perspective - TRACE

Abstract ID: **102** Type: **Poster Presentation**

Spatial transcriptomic profiling of thrombin-PAR-fibrinogen axis in colorectal cancer

Mateja Ilić[1], Marija Cumbo Stikić[1], Aleksandra Vitkovac[1], Marija Lazić[1], Igor Davidović[1], Branko Tomić[1], Valentina Đorđević[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **mateja.ilic@imgge.bg.ac.rs**

Colorectal cancer (CRC) is a heterogeneous disease, among the most common cancer types and third leading cause of cancer-related mortality in developed countries. Thrombin, along with its targets - protease-activated receptors (PARs) and fibrinogen, can facilitate the progression of CRC. However, the mechanisms behind this are not fully elucidated.
The aim of this study was to investigate the specific contribution of the thrombin-PAR-fibrinogen axis to CRC pathogenesis by analyzing corresponding gene expression patterns within the tumor microenvironment.
In this study, a publicly available spatial transcriptomics dataset was utilized, derived from a stage III-B colorectal cancer sample originating from the sigmoid colon and generated via the Visium HD platform. The raw reads were subsequently analyzed using Space Ranger v3.0 and the resulting gene-barcode matrix was processed using Scanpy v1.9.1. Following standard gene and barcode filtering and basic preprocessing, the dataset was annotated using Celltypist v1.7.1. The annotated dataset was then used to show the expression patterns of our genes of interest (F2, F2R, F2RL1, F2RL2, F2RL3, FGA, FGB, FGG).
Cell annotation recognized 17 types of cells in the analyzed sample. While prothrombin and fibrinogen genes (F2, FGA, FGB, FGG) expression pattern was uniform, PARs genes (F2R, F2RL1, F2RL2, F2RL3) expression was mapped predominately to regions recognized as CMS2 and CMS3 molecular subtype of CRC cells.
These findings indicate that the functional impact of thrombin, which is inherently dependent on the spatial distribution of PARs and fibrinogen, is distinctly manifested within specific CRC subtypes. However, the presented data are obtained from a single sample, and represent a "snapshot"of transcriptome in the moment of sampling. Consequently, future analyses involving a larger cohort encompassing diverse molecular subtypes of CRC are warranted to fully elucidate the broader clinical and biological significance of this axis in tumorigenesis.

Keywords:

Colorectal cancer, thrombin, PAR

Acknowledgement:

The public dataset used in this abstract is provided here: https://www.10xgenomics.com/platforms/visium/product-family/dataset-human-crc
This project was partially funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-33/2026-03/200042).

Abstract ID: **110** Type: **Poster Presentation**

Genome-wide analysis of human respiratory adenoviruses in Russia reveals extensive diversity and recombination

Alexander Perederiy[1], Andrey Komissarov[1], Anna Ivanova[1], Irina Amosova[1], Nikita Yolshin[1]

[1] *Smorodintsev Research Institute of Influenza*

Contact e-mail: **nikita.yolshin@gmail.com**

Human adenoviruses (HAdVs) exhibit high genetic variability. However, genomic data, including from Russia, remain limited, which limits the ability to assess circulating subtypes and to track viral evolution.
To characterize the genomic diversity of HAdVs in Russia with a focus on recombination events and to apply bioinformatic approaches for phylogenetic and genome structure analysis.
Respiratory samples (n > 1200) were screened, and 128 complete genomes were obtained via next-generation sequencing. Bioinformatic analysis included genome assembly, maximum-likelihood phylogenetics, and recombination detection using tools such as MAFFT, RAxML-NG, IDPlot and SimPlot.
Virus isolation followed by next-generation sequencing yielded 128 complete HAdV genomes representing species B, C, and D. The dataset included 27 B3, 9 B7, 44 B55, 12 C1, 16 C2, 4 C5, 7 C89, 5 C108, and one D109 genome. Extensive recombination was identified, particularly within capsid genes (penton, hexon, fiber). Several unassigned recombinant genomes (e.g., p89h5f5, p5h6f6, p5h57f6) were detected, demonstrating novel genomic combinations. Phylogenetic analyses revealed genotype-specific clustering and mosaic genome structures, confirming recombination as a major driver of HAdV evolution.
Bioinformatic analysis revealed widespread recombination shaping adenovirus diversity and provided new data on circulating adenovirus types. The findings highlight limitations of current classification systems and underscore the importance of whole-genome sequencing for accurate surveillance and evolutionary studies.

Keywords:

human adenoviruses, recombinant adenoviruses

Abstract ID: **113** Type: **Poster Presentation**

Coordination geometry-driven selectivity toward tumor-associated carbonic anhydrases in novel Zn(II) dithiocarbamato complexes

Marko Antonijević[1], Ana Antonijević[2], Jovana Bogojeski[3], Milica Kosović Perutović[2], Nedeljko Latinović[4], Zorica Leka[2], Goran Kaluđerović[5]

[1] *Institute for Information Technologies, University of Kragujevac, Jovana Cvijića bb, 34000 Kragujevac, Serbia*

[2] *Faculty of Metallurgy and Technology, University of Montenegro, Podgorica, Montenegro*

[3] *Faculty of Science, University of Kragujevac, Kragujevac, Serbia*

[4] *Biotechnical Faculty, University of Montenegro, Podgorica, Montenegro*

[5] *Department of Engineering and Natural Sciences, University of Applied Sciences Merseburg, Eberhard-Leibnitz-Straße 2, 06217 Merseburg, Germany*

Contact e-mail: **mantonijevic@uni.kg.ac.rs**

Two novel mixed-ligand zinc(II) complexes, NH4[Zn(idadtc)(en)] (1) and NH4[Zn(idadtc)(en)2] (2), incorporating iminodiacetato-dithiocarbamato (idadtc3−) and ethylenediamine (en) ligands, were synthesized under varied stoichiometric and solvent conditions. Structural characterization was performed using FT-IR and NMR spectroscopy, supported by DFT calculations at the B3LYP-D3BJ/6-311++G(d,p) level with the CPCM solvation model. Toxicological profiles were predicted via ProTox III and pharmacokinetic behavior was evaluated using SwissADME. Binding affinities toward human serum albumin (HSA) and calf thymus DNA (CT-DNA) were quantified by spectrofluorometric titration and UV-Vis absorption spectroscopy. Molecular docking was performed using AutoDock 4.2 against HSA (PDB: 2BXD) and CA II (PDB: 1CA2), CA IX (PDB: 3DD0) and CA XII (PDB: 1JD0).

DFT optimization confirmed tetrahedral ZnN2S2 geometry for 1 and octahedral ZnN4S2 geometry for 2; theoretical and experimental NMR chemical shifts showed excellent agreement. Both complexes displayed low predicted toxicity (LD50 ≈ 4860 mg kg−1; GHS class IV) and favorable pharmacokinetic profiles. Spectrofluorometric analyses revealed moderate HSA binding (KSV = 2.2–2.4 × 104 M−1) and CT-DNA affinity (Kb = 1.3–1.5 × 104 M−1). Molecular docking with HSA yielded binding energies of −8.01 and −9.62 kcal mol−1 for 1 and 2, respectively, with key interactions involving residues GLU188, LYS195, LYS436, and GLU292. Docking against three carbonic anhydrase isoforms revealed notable affinities: for CA II, binding energies were −7.85 and −8.68 kcal mol−1; for tumor-associated CA IX, −8.37 and −6.27 kcal mol−1; and for CA XII, −9.69 and −7.04 kcal mol−1 for 1 and 2, respectively. Complex 1 showed preferential binding toward tumor-associated CA IX and CA XII over CA II, consistent with its compact tetrahedral ZnN2S2 geometry enabling access to their sterically restricted active sites, in contrast to the bulkier octahedral 2 (ZnN4S2) which favors the more accessible CA II active site.

The ability to direct coordination geometry through simple variation of ligand stoichiometry and solvent demonstrated the high coordination flexibility of the Zn(II)–dithiocarbamato framework. Combined with high aqueous solubility, low toxicity, moderate HSA and DNA recognition, and selective affinity of complex 1 toward tumor-associated carbonic anhydrase isoforms, these complexes represent promising scaffolds for further bioinorganic and anticancer investigations.

Keywords:

Dithiocarbamates, DFT, molecular docking, anticancer

Acknowledgement:

This research was supported by the German Academic Exchange Service (DAAD HAW International HOME and EURABridge Projekt-IDs: 57561680 and 57656312).

Abstract ID: **116** Type: **Poster Presentation**

Development of the amplification strategy and NGS data processing pipeline for HIV-1 genome assembly and resistance genotyping

Alexey Masharskiy[1], Andrey Komissarov[1], Artem Fadeev[1], Nikita Yolshin[1]

[1] *Smorodintsev Research Institute of Influenza*

Contact e-mail: **nikita.yolshin@gmail.com**

Next-generation sequencing (NGS) is increasingly used for HIV-1 genotyping due to its scalability, sensitivity, and lower cost per sample compared to Sanger sequencing. The aim of this study was to develop and validate a workflow for large-scale analysis of HIV-1 NGS data, including consensus genome assembly, subtype assignment, drug resistance mutation detection, and cell tropism prediction.
A total of 1888 HIV-1 positive plasma samples collected in Russia during 2024–2025 were analyzed. Sequencing libraries were generated from nested PCR amplicons covering HIV-1 pol and env genes and sequenced using Illumina NextSeq 2000 and MGI DNBSEQ-G400 platforms.
A custom analysis pipeline implemented in Python was developed. Raw FASTQ reads were processed using Fastp for quality trimming followed by de novo assembly with MEGAHIT. In parallel, reads were mapped to a reference subtype dataset from the Los Alamos HIV database using BWA mem. The best-fitting reference was selected according to mapped read counts obtained with Samtools idxstats. Consensus generation included iterative remapping and refinement using Minimap2, Samtools, BCFtools, and iVar.
Subtype determination was performed using COMET, Sierra, and Nextclade tools with a consensus-based decision strategy. Drug resistance mutations were identified using a local Docker deployment of Sierra. CXCR4 tropism prediction was performed using a custom Python implementation of empirical V3-loop rules and compared with Geno2pheno [coreceptor].
The pipeline successfully generated consensus pol sequences for 1888 analyzed samples and env sequences for 60% of them. The workflow enabled simultaneous characterization of HIV-1 subtype diversity, resistance mutations, and tropism. The predominant subtype was A6 (72.4%), followed by CRF63_02A6 (22.2%). The most frequent resistance-associated mutations included M184V/I, K65R, K103N, and G190S. Resistance to NNRTIs and NRTIs was most prevalent among ART-experienced patients. CXCR4-associated variants were detected infrequently.
The developed amplification strategy and bioinformatics workflow provide a scalable solution for HIV-1 NGS analysis and molecular surveillance. Iterative consensus refinement and combined reference-guided/de novo assembly improved robustness for genetically diverse HIV-1 variants. The pipeline enables high-throughput resistance and tropism monitoring and can be adapted for routine epidemiological surveillance and clinical genomics applications.

Keywords:

HIV, antiretroviral therapy, drug-resistance mutations

Abstract ID: **118** Type: **Poster Presentation**

Microbial composition of peloids from Banja Koviljača

Jelena Pantović[1], Ivana Morić[1], Lidija Šenerović[1]

[1] *Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia*

Contact e-mail: **jelena.pantovic@imgge.bg.ac.rs**

Healing mud, also known as peloid, is a natural material composed of minerals, organic matter, and water, traditionally used in balneotherapy and medical treatments. Peloids represent complex ecosystems enriched with diverse microorganisms capable of producing primary and secondary metabolites with important biological activities, including anti-inflammatory, antiviral, immunomodulatory, and anticancer effects. Among these microorganisms, bacterial communities play a crucial role in the development and maintenance of the therapeutic properties of thermal muds.

Recent advances in Next Generation Sequencing technologies and bioinformatics have enabled comprehensive characterization of microbial communities in environmental samples. However, despite the growing interest in therapeutic mud microbiology, the dynamics of bacterial communities during the peloid maturation process remain insufficiently explored. Therefore, the aim of this study was to investigate the composition and diversity of bacterial communities in thermal peloids at different maturation stages under real environmental conditions.

Samples of thermal peloids were collected throughout the maturation process, and bacterial community profiling was performed using 16S rRNA amplicon sequencing. Relative abundances of microbial taxa were determined using rarefied datasets and summarized across taxonomic levels. Alpha diversity analyses included observed operational taxonomic units (OTUs), Shannon diversity index, and evenness at the genus level. Beta diversity and differences in community composition among samples were evaluated using principal component analysis (PCA), while statistical analyses were performed to assess significant differences in OTU abundances between maturation stages.

The results demonstrated noticeable shifts in bacterial community structure throughout the maturation process, indicating that maturation strongly influences microbial diversity and composition. These findings contribute to a better understanding of the microbial ecology of therapeutic peloids and may help elucidate the biological mechanisms underlying their beneficial properties.

Keywords:

metagenomics, bacterial diversity, peloid maturation

Author Index

ISBN: 978-86-82679-17-2