HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

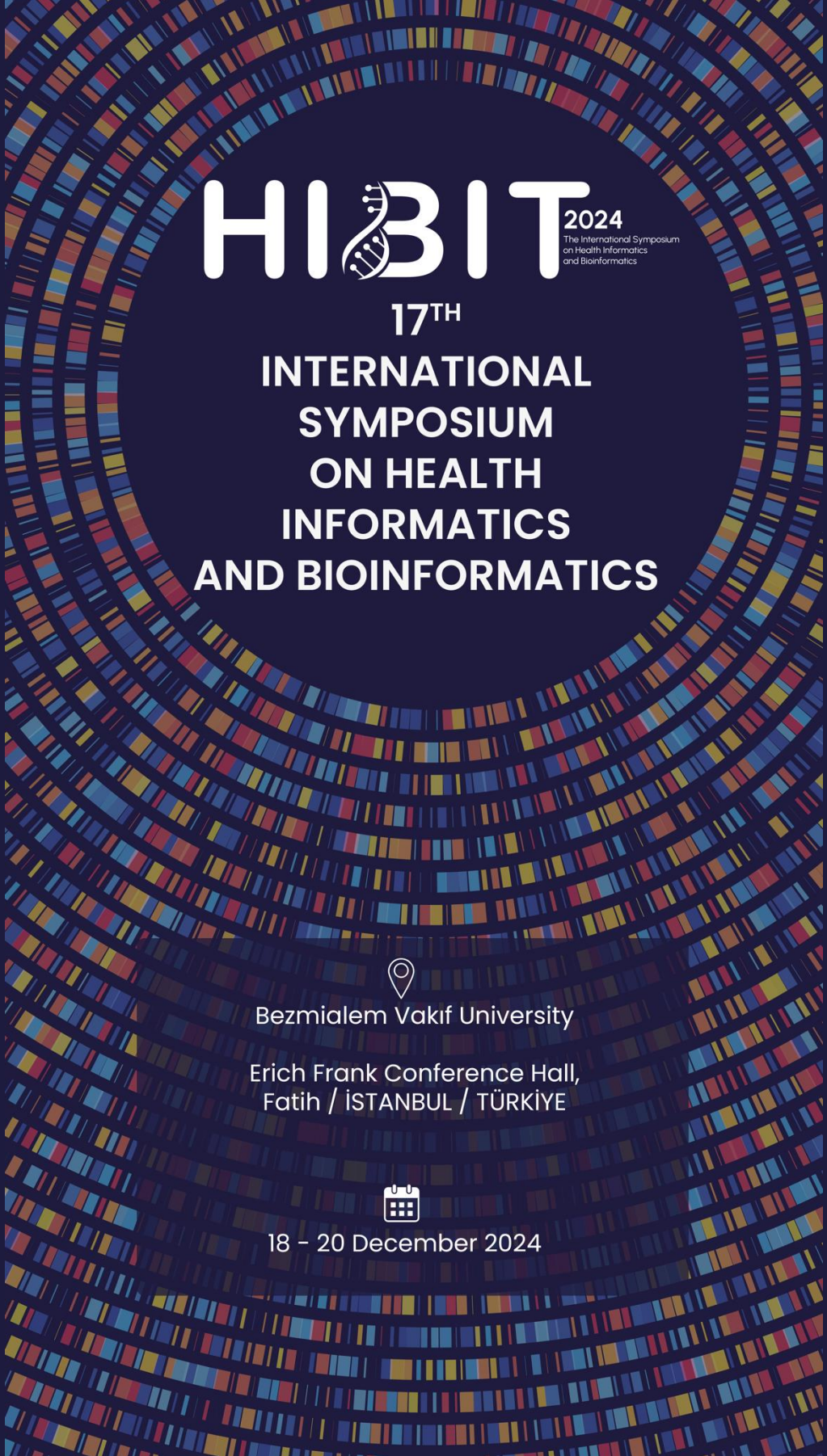
17TH INTERNATIONAL SYMPOSIUM ON HEALTH INFORMATICS AND BIOINFORMATICS

Bezmialem Vakıf University

Erich Frank Conference Hall,
Fatih / İSTANBUL / TÜRKİYE

18 - 20 December 2024

ABSTRACT BOOK


HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics
17TH
INTERNATIONAL SYMPOSIUM ON HEALTH INFORMATICS AND BIOINFORMATICS
Bezmialem Vakıf University
Erich Frank Conference Hall,
Fatih / İSTANBUL / TÜRKİYE
18 - 20 December 2024

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

17th The International Symposium on Health Informatics and Bioinformatics

The International Symposium on Health Informatics and Bioinformatics (HIBIT), first held in 2005, aims to bring together academics, researchers, and practitioners who work in these popular and fulfilling areas and create the much-needed synergy among medical, biological, and information technology sectors. HIBIT is one of the few conferences emphasizing such synergy. HIBIT provides a forum for discussion, exploration, and development of theoretical and practical aspects of health informatics and bioinformatics and a chance to network between students, academia, and other stakeholders.

ISCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY
AFFILIATED CONFERENCE

International Society for Computational Biology (ISCB) Affiliated Conference

TÜBİTAK

Conference was supported by TUBITAK BIDEB 2223-B Grant Program

Chair's Message

Dear Colleagues and Guests,

It is with great pleasure that I welcome you to the 17th International Conference on Health Informatics and Bioinformatics (HIBIT 2024) through the pages of this abstract book.

Reflecting on the success of this year's event, I extend my deepest gratitude to Prof. Dr. Rümeyza Kazancıoğlu, Rector of Bezmialem Vakif University, for hosting us in the beautiful Erich Frank Conference Hall. I also offer my heartfelt thanks to our distinguished speakers for sharing their invaluable expertise and to our enthusiastic participants, whose curiosity and engagement enriched every session. This conference would not have been possible without the tireless efforts of the organizing and scientific committees, to whom I am profoundly grateful.

Celebrating the 17th year of HIBIT, we witnessed the incredible synergy of medicine, biology, and information technology, with discussions spanning from artificial intelligence applications to cutting-edge advancements in genomics and beyond. HIBIT 2024 once again highlighted its essential role as a platform for fostering collaboration, sharing knowledge, and inspiring innovation.

The pre-conference workshops provided a solid foundation for the main event, with engaging sessions on high-performance computing, NGS workflows, and pathway analysis. These workshops set the tone for the meaningful exchanges and connections that followed throughout the conference.

As you explore this abstract book, I hope it serves as both a record of the remarkable contributions presented at HIBIT 2024 and a source of inspiration for future research and collaboration. Thank you for being part of this event and for contributing to its success.

With warm regards,

Onur Emre Onat, Ph.D.
Institute of Life Sciences and Biotechnology
Bezmialem Vakif University
Email: onur.onat@bezmialem.edu.tr

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

Organizing Committee

Onur Emre Onat	Bezmialem Vakıf University
Arda Çetinkaya	Hacettepe University
Arzu Karahan	Middle East Technical University
Ayhan Serkan Şık	Medipol University
Günseli Bayram Akçapınar	Acıbadem University
Mehmet Baysan	Istanbul Technical University
Özge Şensoy	Medipol University
Şeref Gül	Bezmialem Vakıf University
Tunca Doğan	Hacettepe University
Yesim Aydın Son	Middle East Technical University

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

Organizing Team

Volunteer Students

Levise Tenay	Bezmialem Vakıf University
Vildan Saraç	Bezmialem Vakıf University
Sumeyye Cimeli	Bezmialem Vakıf University
Faruk Ustunel	Bezmialem Vakıf University
Ebru Sarsılmaz	Bezmialem Vakıf University
Nisa Çelebi	Bezmialem Vakıf University
Aleyna Çelebi	Bezmialem Vakıf University
Hande Çubukçu	Hacettepe University
Ntefne Chasan Oglou	Istanbul Technical University
Cahit Burduroğlu	Middle East Technical University
Demet İnci	Bezmialem Vakıf University
Arda Baran Nişan	Middle East Technical University
Zeynep Özbiltekin	Middle East Technical University
Barış Yıldırım	Middle East Technical University
Useyma Gülhan	Middle East Technical University

Technical

Şehlem Ertekin	Bezmialem Vakıf University
Esra Bekaroğlu	Bezmialem Vakıf University
Hilal Başyeğit	Bezmialem Vakıf University
Şehnaz Yüceer	Bezmialem Vakıf University
Mahir Aka	Bezmialem Vakıf University
Melek Mil	Bezmialem Vakıf University
Onur Kalyoncu	Bezmialem Vakıf University

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

Scientific Committee

A.Matteen Rafiqi	Bezmialem Vakıf University
Şükrü Anıl Doğan	Bezmialem Vakıf University
İdil Yet	Hacettepe University
Gülşah Merve Kılınç	Hacettepe University
Ahmet Acar	Middle East Technical University
Arzucan Özgür	Bogazici University
Aslı Suner	Ege University
Atakan Ekiz	Izmir Institute of Technology
Athanasia Pavlopoulou	Izmir Biomedicine and Genome Center
Aybar Acar	Middle East Technical University
Ayhan Yılmaz	Middle East Technical University
Barbaros Yet	Middle East Technical University
Barış Ethem Süzek	Mugla Sitki Kocman University
Burçak Otlu	Middle East Technical University
Burcu Bakir-Gungor	Abdullah Gul University
Çağdaş Son	Middle East Technical University
Carlos Trenada	Heinrich Heine University
Ceren Sucularlı	Hacettepe University
Ceylan Polat	Hacettepe University
Emrah Kırdök	Mersin University
Emre Dayanç	Izmir University of Economics
Emre Keskin	Ankara University
Erdal Cosgun	Microsoft
Eren Yüncü	Middle East Technical University
Ergin Murat Altuner	Kastamonu University
Gökhan Karakülah	Izmir Biomedicine and Genome Center

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

Scientific Committee (cont.)

Gökmen Zararsız	Erciyes University
Handan Melike Dönertaş	Leibniz Institute on Aging
Hatice Zengin	Hacettepe University
Hilal Kazan	Antalya Bilim University
İnci M. Baytas	Bogazici University
Kaya Bilguvar	Acıbadem University
Muzaffer Arıkan	İstanbul University
Nilüfer Rahmioğlu	University of Oxford
Nurcan Tunçbağ	Koc University
Oktay İ. Kaplan	Abdullah Gul University
Özgül Doğan	Sivas Cumhuriyet University
Özkan Özdemir	Acıbadem University
Özlem Keskin	Koç University
Özlen Konu	Bilkent University
Öznur Taştan	Sabanci University
Remzi Çelebi	Maastricht University
Sait Can Yücebaş	Çanakkale Onsekiz Mart University
Şebnem Eşsiz	Kadir Has University
Serap Erkek Özhan	Izmir Biomedicine and Genome Center
Süreyya Özcan	Middle East Technical University
Tuğba Önal Süzek	Mugla Sitki Kocman University
Tunahan Çakır	Gebze Technical University
Vahap Eldem	İstanbul University
Vilda Purutçuoğlu	Middle East Technical University
Volkan Atalay	Middle East Technical University
Yavuz Oktay	Izmir Biomedicine and Genome Center

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

17TH INTERNATIONAL SYMPOSIUM ON HEALTH INFORMATICS AND BIOINFORMATICS

18 - 20 December 2024

KEYNOTE SPEAKERS

Yuval Itan,
Icahn School of Medicine
at Mount Sinai

Altuna Akalin,
Berlin Institute for Medical Systems
Biology (BIMSB), Max Delbrück Center
for Molecular Medicine

Asif M. Khan,
University of Doha for
Science and Technology

BEZMIALEM VAKIF UNIVERSITY
1845

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

17TH INTERNATIONAL SYMPOSIUM ON HEALTH INFORMATICS AND BIOINFORMATICS

18 - 20 December 2024

INVITED SPEAKERS

Seref Gül,
Bezmialem University

Günseli Bayram Akçapınar,
Acıbadem University

Kaya Bilguvar,
Acıbadem University

Gülşah Merve Kılınç,
Hacettepe University

Mehmet Baysan,
İstanbul Technical University

Melike Dönertaş,
Leibniz Institute

Ruçhan Karaman,
Genomize

Ali Çakmak,
İstanbul Technical University

BEZMIALEM VAKIF UNIVERSITY
1845

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

17TH INTERNATIONAL SYMPOSIUM ON HEALTH INFORMATICS AND BIOINFORMATICS PROGRAM

BEZMIALEM VAKIF UNIVERSITY
1845

18 - 20 December 2024
Bezmialem Vakıf University,
Erich Frank Conference Hall

hibit2024.bezmialem.edu.tr

Day 1 - December 18, 2024, Wednesday

Time	Event	Speakers	Title
13:00	Welcome & Coffee		
13:30	Workshop 1	Faruk Üstünel, Bezmialem University	High-Performance Computing (HPC) Systems
13:50	Workshop 2 (online)	Emrah Akkoyun, TUBITAK	Efficient and Scalable Data Analysis on HPC: Harnessing Snakemake and Conda for Reproducibility
14:10	Workshop 3	Cahit Burduroğlu, METU	Accessing TRUBA and creating a workflow
14:30	Workshop 4	Kübra Narcı, The German Human Genome-Phenome Archive	Standardizing and harmonizing NGS analysis workflows
17:00	Workshop 5	Hatice Kübra Kibar, Massive Bioinformatics	Massive Analyzer and Mini Analyzer for Genetic Data Analysis
17:20	Workshop 6	Tolga Aslan, Genomize Inc	Use of Secondary and Tertiary Analysis Products in Clinical Genomics, A Use Case With Genomize-SEQ Platform
17:40	Workshop 7	Can Koşukcu, Qiagen	IPA: Ingenuity Pathway Analysis, Interpreting RNA-Seq Results: The secrets to pathway analysis
18:00	Closure		

Day 2 - December 19, 2024, Thursday

09:00	Opening Speech	Rümeyza Kazancıoğlu (Rector), Bezmialem Vakıf University	
09:05	Who we are?	Mehmet Ziya Doymaz (BILSAB Director), Bezmialem Vakıf University	
09:10	HIBIT2024 Welcome Speech	Onur Emre Onat (HIBIT2024 Chair), Bezmialem Vakıf University	

Session 1 - Human Genetics
Moderators: Onur Emre Onat & Asif M. Khan

09:15	Keynote Talk 1	Yuval Itan, Icahn School of Medicine at Mount Sinai	Novel Machine Learning Approaches for Predicting Functional Consequences of Human Genetic Variants
10:05	Invited Talk 1	Gülşah Merve Kılınç, Hacettepe Üniversitesi	Complex trait evolution in the light of palaeogenomics: Computational approaches for reconstructing the health in past
10:35	Coffee Break		

Session 2 - Artificial Inteligence for Health Informatics
Moderators: Günseli Bayram Akçapınar & Tunca Doğan Onat & Asif M. Khan

11:00	Keynote Talk 2	Altuna Akalın, Max Delbrück Center	How AI Will Reshape Life Sciences Research: Examples and Demos from the Frontlines
11:50	Selected Talk 1	Atabey Ünlü, Hacettepe University	Autoregressive Design of Target-Centric Drug Candidate Molecules with Chemical and Protein Language Models
12:05	Selected Talk 2	Melih Yiğit, METU & Hacettepe University	Flowdrug: Target-Specific Drug Candidate Molecule Generation with Latent Flow Matching Models
12:20	Selected Talk 3	Eyyüb Ünlü, Wellcome Sanger Institute	Network-Based Heterogeneity Clustering in Primary Open-Angle Glaucoma Patients Reveals Ancestry-Specific Gene Clusters with Known and Novel Genes
12:35	Lunch Break		

Session 3 - Multiomics Approaches
Moderators: Gülşah Merve Kılınç & Arda Çetinkaya

14:00	Invited Talk 2	Melike Dönertaş, Leibniz Institute	Towards a Senescent Cell Atlas Using Machine Learning Classification
14:30	Selected Talk 4	Gamze Maden Müftüoğlu, İstanbul Technical University	Detailed Evaluation of Structural Variant Detection Algorithms Through a Diverse Set of Benchmark Datasets
14:45	Selected Talk 5	Yunus Emre Cebeci, İstanbul Technical University	Efficient Combination of Replicates for Better Somatic Variant Detection in WES
15:00	Selected Talk 6	Abdullah Emul, İstanbul Technical University	VCF Observer: A User-friendly Software Tool for Preliminary VCF File Analysis and Comparison
15:15	Poster Session & Coffee		

Session 4 - Drug Design and Development
Moderators: A.Matteen Rafiqi & Şükrü Anıl Doğan

16:30	Invited Talk 3	Seref Gül, Bezmialem University	Targeting Circadian Clock Components for Therapeutic Advances
17:00	Selected Talk 7	Elif Çevrim, Hacettepe University	Leveraging Hypergraph Neural Networks for Drug-Target Interaction Prediction
17:15	Selected Talk 8	Emre Taha Çevik, Gebze Technical University	Development of a Web-Based Tool for Identification of Druggable Pockets in Protein Conformational Ensembles and Ensemble Docking
17:30	Selected Talk 9	Ardan Yılmaz, Middle East Technical University	Enhancing Drug-Target Interaction Prediction via Multimodal Learning and Domain Adaptation
17:45	Selected Talk 10	Ahmet Arıhan Erözden, İstanbul University	MetaPepticon: An Automated Pipeline for Anticancer Peptide Prediction from Shotgun Metagenomics Datasets
18:00	Closure		
19:30	Gala Dinner		

Day 3 - December 20, 2024, Friday

Session 5 - Genome Research
Moderators: Yeşim Aydın Son & İdil Yet

09:10	Keynote Talk 3	Asif M. Khan, University of Doha for Science and Technology	Bioinformatics Teaching and Training Communities: Addressing Grand Challenges
10:00	Invited Talk 4	Mehmet Baysan, İstanbul Technical University	Clinical NGS Analyses at the Intersection of Information and Genome Technologies
10:30	Selected Talk 11	Mehmet Çay, Acibadem University	LARA: A Novel Multimodal Algorithm Achieves Non-Invasive Multicancer Classification
10:45	Selected Talk 12	Ahmed Abunada, İstanbul Technical University	Turkish Genome Project Data Sharing Portal
11:00	Coffee Break		

Session 6 - Protein Structure & Function
Moderators: Şeref Gül & Özge Şensoy

11:30	Invited Talk 5	Günseli Bayram Akçapınar, Acıbadem University	Integrative Computational Approaches Using the Protein Structure-Function Paradigm: Insights into Cerebral Palsy and Obesity
12:00	Selected Talk 13	Erva Ulusoy, Hacettepe University	Automated Protein Function Prediction Using Biological Knowledge Graphs and Heterogeneous Graph Transformers
12:15	Selected Talk 14	Nurdan Kuru, Sabancı University	Phylogeny-Driven Approaches for Variant Effect Prediction and Co-evolution
12:30	Lunch Break		

Session 7 - Machine Learning Applications
Moderators: Mehmet Baysan & Arzu Karahan

14:00	Invited Talk 6	Ali Çakmak, İstanbul Technical University	Completing the Loop in Metabolism-Oriented Omics Data Analysis: Data Imputation and Integration Challenges
14:30	Selected Talk 15	Elif Kubat Öktem, İstanbul Medeniyet University	Breast Cancer Research Targetting Programmed Cell Death in the Concept of 3PM Medicine
14:45	Selected Talk 16	Mehmet Burak Koca, Gebze Technical University	Multi-omic Single-cell Data Analysis for Biomarker Identification in Viral Pathogenesis
15:00	Poster Session & Coffee		
16:30	Invited Talk 8	Ruçhan Karaman, Genomize	How to Combine Genomics, Deep Phenotyping, and AI to Diagnose Patients with Rare Diseases
17:00	Selected Talk 17	Gül Aydemir, Istanbul Technical University	Impact of Computational Parameters on Somatic Variant Calling in Whole Exome Sequencing
17:15	Selected Talk 18	Abdullah Emul, Istanbul Technical University	SABER: Sequencing Analysis Benchmarker A Comprehensive Tool for Effective Benchmarking
18:00	Award Ceremony Closing Remarks	Onur Emre Onat (HIBIT2024 Chair), Bezmialem Vakıf University	

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

SPONSORS

BEZMIALEM VAKIF UNIVERSITY 1845

HIBIT 2024 The International Symposium on Health Informatics and Bioinformatics

ISCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

MAIN SPONSORS

Nucleus GENETICS

MASSIVE BIOINFORMATICS

QIAGEN

ATC TRUSTED PARTNER IN GENOMICS

GIGA SPONSORS

Humanizing Genomics macrogen EUROPE

MSM

MEGA SPONSOR

genomize

ORGANIZATION SPONSOR

etix events

TÜBİTAK

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

AWARDS

Best Oral Presentations

1st Place
Nurdan Kuru, Sabancı University

2nd Place
Atabey Ünlü, Hacettepe University

3rd Place
Erva Ulusoy, Hacettepe University

Best Poster Presentations

Session: Human Genetics
Nura Fitnat Topbas Selcuk, University of Oxford

Session: Multiomics Approaches
Arda Örçen, Acıbadem Üniversitesi

Session: Drug Design and Development
Zeynep Cinviz, Istanbul Medipol University

Session: Genome Research
Faruk Üstünel, Bezmialem Vakif University

Session: Protein Structure & Function
Levise Tenay, Bezmialem Vakif University

Session: Machine Learning Applications
Seyit Semih Yiğitarslan, Hacettepe University

Session: Artificial Intelligence for Health Informatics
Barış Can, İstanbul Technical University

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

KEYNOTE TALKS

How AI will Reshape Life Sciences Research: Examples and Demos from the Frontlines

Altuna Akalin[1,*]

[1]*Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine, Berlin, Germany*

Presenting Author: altuna.akalin@mdc-berlin.de
**Corresponding Author: altuna.akalin@mdc-berlin.de*

Artificial intelligence (AI) is transforming the life sciences, particularly in drug development, disease diagnosis, and data analysis. This talk explores the comprehensive role of AI in these fields, highlighting its potential to transform personalized medicine and healthcare.

We have built a novel AI toolkit designed to enhance cancer care from drug development to treatment selection. This toolkit aids in target discovery, drug discovery, biomarker identification, and personalized treatment strategies. In the target discovery phase, machine learning models are validated using CRISPR screening. This step ensures that the targets are both theoretically plausible and practically actionable, forming a solid foundation for drug development.

In the drug discovery stage, the toolkit uses advanced generative models to match molecules with targets. This approach streamlines the drug discovery process, making it faster and less resource-intensive. For biomarker discovery, the toolkit integrates multi-omics data with deep learning-based data augmentation techniques. This method extracts valuable insights from small-scale clinical trials, reliably identifying biomarkers that indicate treatment response or resistance.

Beyond oncology, AI systems have the potential to play a crucial role in disease diagnosis and data analysis. Our AI assistants for disease diagnosis utilize vast datasets and advanced algorithms to provide accurate and timely diagnoses, improving patient outcomes. Similarly, our AI-driven data analysis tools help researchers and clinicians interpret complex biological data, accelerating scientific discoveries.

Our AI-driven approaches exemplify the integration of AI into the life sciences, offering new ways to enhance precision and effectiveness in cancer treatment and beyond. The talk will explore use cases, results, and demos of the toolkit we have built.

Bioinformatics Teaching and Training Communities: Addressing Grand Challenges

Asif M. Khan[1,*]

[1]*University of Doha for Science and Technology (UDST), Doha, Qatar*

Presenting Author: khan.mdasif@gmail.com
**Corresponding Author*

Bioinformatics education faces significant challenges, including the rapid evolution of technologies, the need for interdisciplinary curricula, and a shortage of trained educators. The Asia-Pacific Bioinformatics Network (APBioNET), in collaboration with partners such as ISCB and GOBLET, has actively worked to address these challenges. By fostering collaboration, standardizing training materials, and promoting capacity-building initiatives, these efforts have significantly advanced bioinformatics education across the region. These initiatives have resulted in the development of comprehensive educational programs and the enhancement of bioinformatics competencies among researchers and students. This keynote will delve into the grand challenges in bioinformatics education and training, highlighting the role of APBioNET and its partners while exploring strategies for building resilient teaching and training communities to propel the field forward.

Novel Machine Learning Approaches for Predicting Functional Consequences of Human Genetic Variants

Yuval Itan[1,*]

[1]*Icahn School of Medicine at Mount Sinai, New York, USA*

Presenting Author: yuval.itan@mssm.edu
**Corresponding Author*

Despite the exponential increase in patient genomic data worldwide, key limitations continue to impede the effectiveness of computational tools for variant assessment. We recently developed LoGoFunc, the first machine learning method capable of effectively identifying pathogenic gain-of-function (GoF) and loss-of-function (LoF) variants on a genome-wide scale. Furthermore, our variant-to-phenotype (V2P) machine learning method is the first to estimate specific disease phenotypes for all human genetic variants genome-wide, surpassing the general pathogenicity predictions of current methods which can only predict general pathogenicity. We have developed additional methods to accurately estimate pathogenic genes and variants in patient whole-exome and whole-genome data, leveraging knowledge of the patients' disease, specifically: (1) The human gene connectome (HGC): to identify novel disease-causing genes; (2) The gene damage index (GDI), to filter out non-pathogenic false-positive genes from the analysis; and (3) The mutation significance cutoff (MSC), to apply gene-level cutoffs for methods like CADD, reducing false negatives. Lastly, we utilized major biobanks, incorporating diverse human populations, to conduct extensive phenome-wide association studies (PheWAS). These analyses enabled the estimation of both population-specific and overlapping phenotypes associated with all pathogenic GoF and LoF variants.

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

INVITED TALKS

Complex Trait Evolution in the Light of Palaeogenomics: Computational Approaches for Reconstructing the Health in Past

Gülşah Merve Kılınç[1,*]

[1]*Department of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, Ankara, Türkiye*

Presenting Author: gulsahkilinc@hacettepe.edu.tr
**Corresponding Author*

Ancient DNA offers a unique potential for reconstructing the past including movements, interactions and health of our ancestors through providing spatiotemporal genetic data. Recent years saw an increasing amount of work that has provided an ancient genomic snapshot on evolutionary processes in human populations especially in Europe. Here, I present a computational workflow that we harmonize distinct methods for analyzing ancient genomic data to reconstruct past population demography and health simultaneously, in the context of a set of diseases and traits including those that are age-related and those that are classified as neurological. This workflow includes a maximum-likelihood based method for estimating the allele frequencies, a comparative approach for detecting the selection based on in silico and real genetic data, and combination of a set of summary statistics to explore signatures of selection. I present application of this workflow on new and published ancient genomes and potential signals of selection that is discovered in ancient human genomes. Findings of this current work contains both a picture of dynamic genetic architecture in ancient Anatolia for neurological traits and a main computational approach for paleogenomic reconstruction of human health and movement simultaneously.

Towards a Senescent Cell Atlas Using Machine Learning Classification

Melike Dönertaş[1,*]

[1]*Leibniz Institute on Aging, Jena, Germany*

Presenting Author: Melike.Donertas@leibniz-fli.de
**Corresponding Author*

Cellular senescence, defined as irreversible cell cycle arrest, is challenging to characterize in vivo due to its rarity and heterogeneity across cell types, induction methods, and time courses. Cellular senescence plays a crucial role in development, regeneration, aging, and various age-related diseases, including cancer. This project aims to develop a machine-learning classifier to identify senescent cells in scRNAseq datasets, facilitating biomarker development, characterization, and drug repurposing. We utilize an in vivo scRNASeq dataset of cells labeled with B-galactosidase activity, a well-established senescence marker. Using SPiDER-B-gal for labeling live cells allows the sequencing of rare senescent populations in vivo. Our random forest classifier trained on SPiDER-B-gal labeled data shows unprecedented performance in independent test data (AUC = 0.967), outperforming traditional methods that use single-gene markers (AUC range 0.472-0.527) or senescence-related gene sets like SenMayo (AUC = 0.729) or CellAge (AUC = 0.645). Trajectory analysis based on ML features recapitulated senescence development across various cell types. Feature importance analysis highlights the role of the IGF signaling pathway in senescence induction and relay, corroborated by cell-cell interaction analysis. We are working towards experimental validation and expanding our model to multiple tissues and organisms to enable large-scale characterization of senescence and help identify drugs that selectively target senescent cells and their expression profiles.

How to Combine Genomics, Deep Phenotyping, and AI to Diagnose Patients with Rare Diseases

Ruçhan Karaman[1,*]

[1]*Genomize, Istanbul, Turkiye*

Presenting Author: ruchan.karaman@gmail.com
**Corresponding Author*

Accurate interpretation of diverse genetic variants remains a pivotal challenge in rare disease diagnosis. Although evidence-based guidelines established by the American College of Medical Genetics and Genomics (ACMG) have enhanced the precision of variant assessment, the practical implementation of these guidelines is quite problematic. The inherent genetic heterogeneity in rare diseases, coupled with the need to integrate information from numerous databases, contributes to this complexity. Therefore, advancements in secondary variant calling, automated variant annotation and prioritization, visualization of variant annotations with the raw data, and a streamlined reporting process are crucial for efficient and robust analysis. To address this problem we collect data from more than 120 different databases to annotate the variants according to ACMG/AMP guidelines and prioritize the variants that could be causative for a patient's clinical representation. This can help clinicians solve more variants of unknown significance (VUS) cases with its extended annotation feature. We validated our method by using real-world Whole exome sequencing (WES) data of 220 patients with the matching pre-diagnostic and phenotypic information in which the causative variants were identified by the clinicians. We identified the causative variants in each patient, achieving 97% success in identifying the relevant variants in this cohort.

Targeting Circadian Clock Components for Therapeutic Advances

Şeref Gül[1,*]

[1]*Institute of Life Sciences and Biotechnology, Bezmialem Vakif University, Istanbul, Türkiye*

Presenting Author: seref.gul@bezmialem.edu.tr
**Corresponding Author*

The circadian clock orchestrates gene expression and metabolic homeostasis, forming a dynamic feedback loop where metabolism influences clock function. Disruptions in core clock genes, such as CLOCK, BMAL1, PER2, and CRY1/2, reveal their pivotal roles in diseases ranging from metabolic syndromes to sleep disorders. Notably, CRY mutations exhibit opposing effects in cancer and glucose regulation, highlighting context-dependent outcomes. Emerging evidence underscores the therapeutic potential of targeting clock proteins to address diabetes, cancer, and aging-related disorders, offering novel strategies for precision medicine through circadian modulation.

Clinical NGS Analyses at the Intersection of Information and Genome Technologies

Mehmet Baysan[1,*]

[1]*Department of Computer Engineering, Istanbul Technical University, 34469 Istanbul, Turkey*

Presenting Author: baysanm@itu.edu.tr
**Corresponding Author*

At BaysanLab, we combine informatics and genome technologies to improve the health services provided in Turkiye and worldwide. Next-generation sequencing (NGS) technologies allow detailed genetic maps of humans at a very low cost and in a short amount of time. In our laboratory, we develop software and online platforms to analyze clinical genomic data using the opportunities offered by information technologies. In this talk, I will briefly overview some of the research projects we have conducted recently. I will start with the Comparative Sequencing Analysis Platform (COSAP), an open-source platform that provides popular sequencing algorithms for SNV, indel, structural variant calling, and their annotations through a user-friendly web interface. Then, I will continue with VCF Observer, a novel web application that provides a user-friendly solution for VCF comparisons via high-level analysis and visualization. In the second part, I will mention our projects for optimizing variant calling performance. I will explain how we evaluate the impact of alternative algorithms and parameters for somatic sequencing and how we can combine alternative algorithms and replicates for better variant detection. I will then mention our class project where we combine practical NGS reproducibility assessment and bioinformatics education. In the third part, I will focus on variant pathogenicity prediction. I will start by describing how we combine and standardize pathogenicity benchmark datasets. Then, I will disclose how we use these datasets to assess the performance of prediction algorithms with a special focus on ACMG pathogenicity guideline categories. In the final part, I will illustrate our efforts on the Turkish Genome Project which aims to reveal the genetic structure of the healthy Turkish Population.

Integrative Computational Approaches Using the Protein Structure-Function Paradigm: Insights into Cerebral Palsy and Obesity

Günseli Bayram Akçapınar[1,*]

[1]*Acıbadem University, Istanbul, Turkiye*

Presenting Author: gunseli.akcapinar@acibadem.edu.tr
**Corresponding Author*

Integrative computational approaches utilizing the protein structure-function paradigm are explored to gain mechanistic insights into two complex disorders, cerebral palsy (CP) and obesity. The Deep-CP Project examines the impact of point mutations in proteins on CP pathogenicity through a comprehensive database of protein variants. Advanced deep-learning architectures, including transformer-based models, are employed to predict variant pathogenicity by integrating protein language and structural features.

The spatial distributions of variants on 3D structures of CP-associated proteins and their functional impacts are analyzed. Pathogenic variants are shown to significantly destabilize protein structures, as indicated by higher ddG values. Statistical analysis confirms clear distinctions between benign, pathogenic, and variants of uncertain significance (VUS), with ddG highlighted as a potent predictor of variant pathogenicity.

In obesity, the impact of pathogenic variants is explored using molecular dynamics simulations, focusing on proteins such as STEAP1B, KCNQ5, and UCP1. Structural predictions are generated using AlphaFold utilizing protein-cofactor, protein-protein interactions, and the effects of these variants on protein function and their roles in the cellular context relevant to obesity are assessed. MD simulations identified one variant of STEAP1B with the largest conformational change upon simulation at ambient and high temperatures with changed cofactor-protein interactions, leading to an altered conformation.

Completing the Loop in Metabolism-Oriented Omics Data Analysis: Data Imputation and Integration Challenges

Ali Çakmak[1,*]

[1]Department of Molecular Biology, Genetics, and Biotechnology, Istanbul Technical University, 34469, Istanbul, Turkey

Presenting Author: ali.cakmak@itu.edu.tr
**Corresponding Author*

In this talk, we discuss two different aspects of multi-omics data integration. The first part focuses on vertical omics data integration that involves the integrated analysis of different types of omics datasets. We introduce a comprehensive metabolism-oriented integrated multi-omics data analysis method which can accommodate major omics datasets including genomics, transcriptomics, proteomics, and metabolomics. Our methodology entails constructing an integrated multi-omic interaction network, which serves as the foundational framework for our analysis. The integrated network covers a wide range of biological interactions, such as expression, translation, transcription factors, post-transcriptional regulation, etc. To understand how omics data changes impact the overall biological system, we utilize information diffusion models to propagate the fold-changes through the interaction network. We demonstrate the usefulness of the presented approach on 5 different cancer datasets.

The second part of the talk presents the challenges of horizontal omics data integration which involves the unification of multiple omics datasets of the same type. Most datasets include measurements for only a very small fraction of the known metabolites. Hence, simply putting together these studies leads to very sparse datasets, which do not lend themselves well to training machine learning models. In this talk, we present two novel approaches for dataset merging and imputation model training: (i) Iterative Similarity-based Merging generates an optimal merge set for each dataset and makes sure that a minimum sparsity threshold is maintained, and (ii) Model-guided Agglomerative Merging combines datasets in pairs to create a single large dataset in an attempt to effectively combine diverse metabolomics datasets while minimizing the likelihood of gaps created by non-overlapping metabolites. We demonstrate promising results from the application of the proposed methods on the entire Metabolomics Workbench datasets, which is the official metabolomics data repository of NIH.

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

ORAL PRESENTATIONS

MetaPepticon: An Automated Pipeline for Anticancer Peptide Prediction from Shotgun Metagenomics Datasets

Ahmet Arıhan Erözden[1,2], Nalan Tavşanlı[1,2], Gamze Demirel,[2,3] Nazmiye Özlem Şanlı,[1] Mahmut Çalışkan,[1] Muzaffer Arıkan[1,*]*

[1]*Biotechnology Division, Department of Biology, Faculty of Science, Istanbul University, 34134, Istanbul, Türkiye*
[2]*Institute of Graduate Studies in Science, Istanbul University, 34116, Istanbul, Türkiye*
[3]*Cancer and Stem Cell Research Center, Faculty of Medicine, Maltepe University, 34857, Istanbul, Türkiye*

Presenting Author: aa.erozden@istanbul.edu.tr
*Corresponding Author: muzafferarikan@istanbul.edu.tr

Since their first discovery, peptides that specifically target and impose cytotoxicity upon cancer cells have provided a promising field of research for therapeutic applications against cancer. The goal to discover these anticancer peptides presents novel research opportunities, and the number of studies on these peptides has been growing at an accelerating pace. Shotgun metagenomics datasets hold great potential as extensive sources for the discovery of anticancer peptides. As these datasets withhold great amounts of data, *in silico* detection tools are required for fast, efficient, and accurate anticancer peptide detection.

Here, we present the MetaPepticon, a bioinformatics pipeline that accepts the raw shotgun metagenomic data as input and outputs anticancer peptide candidates. MetaPepticon utilizes a consensus approach that incorporates multiple anti-cancer peptide prediction tools with distinct algorithms and methodologies, thereby identifying overlapping anticancer peptide candidates with the highest prediction scores across multiple tools. Furthermore, it checks the overall highest scored sequences for potential toxicity and scores any potentially toxic peptides. Ultimately, MetaPepticon provides a selection of peptides that exhibit the highest prediction scores alongside the lowest toxicity scores, providing researchers with optimized candidates for further investigation. Moreover, MetaPepticon provides user with the flexibility to select parameters, such as the number of overlapping predictions or best scored peptides across multiple tools.

MetaPepticon provides a reproducible, flexible, end-to-end standardized workflow for the prediction of anticancer peptide candidates from shotgun metagenomics datasets. By integrating a consensus approach, it enhances prediction consistency and enables the discovery of novel anti-cancer peptide candidates from a variety of environmental microbiome samples.

LARA: A Novel Multimodal Algorithm Achieves Non-Invasive Multicancer Classification

Mehmet Çay[1], Zeynep Erson Omay[2], Bony De Kumar[3], Merve Nur Köroğlu[1], Eyüp Alper Saygılı[1], Murat Günel[2], Kaya Bilguvar[1,2,*]

[1]*Acibadem University School of Medicine, Istanbul, Turkiye*
[2]*Yale University School of Medicine, Neurosurgery*
[3]*Yale Center of Genomic Analysis*

Presenting Author: mehmet.cay@live.acibadem.edu.tr
**Corresponding Author*

Introduction: Early cancer detection and accurate classification are critical to improve patient outcomes in the clinics. Traditional diagnostic methods can be invasive, expensive, and lack accuracy. Here, we present LARA (Logistic Augmented Regression Algorithm), a novel non-invasive diagnostic software that integrates multi-model genomic data to classify various cancers from cell-free DNA.

Methods: Our cohort consisted of 61 individuals: 18 brain cancer patients from Yale Program in Brain Tumor Research, 23 colorectal cancer patients and 20 healthy controls (5 from our institution and 38 from publicly available datasets). LARA examines global DNA methylation patterns and genomic fragmentation profiles from cfDNA samples. Solid data processing and quality control techniques were implemented to create balanced and quality dataset along with identifying and removing inherent batch effects due to sequencing differences. Matrix normalization-based marker selection was utilized to prioritize most informative markers. Machine learning models were developed and integrated using these selected features, to deploy a holistic approach of LARA.

Results: Individual models based on methylation data and fragment size analysis achieved overall accuracies of 94.74% and 88.24%, respectively. By integrating both models, LARA dramatically outperformed the individual models, achieving an overall accuracy of 99.35% and a mean AUC of 0.9996. LARA successfully distinguished different tumors and healthy samples and showed superior classification performance than induvial models.

Discussion: LARA reveals a striking achievement in non-invasive cancer diagnosis, presenting a highly accurate and robust methodology for multicancer detection and classification. By integrating methylation and fragmentation data through a novel computational framework, LARA outperforms state-of-arts models, holding potential to improve patient outcomes and guide personalized treatment strategies.

VCF observer: a user-friendly software tool for preliminary VCF file analysis and comparison

Abdullah Emul[1], Mehmet Arif Ergün[1], Rumeysa Ertürk[1], Mehmet Baysan[1,*]

[1]*Istanbul Technical University, İstanbul, Turkiye*

Presenting Author: asim_emul@hotmail.com
**Corresponding Author: baysanm@itu.edu.tr*

Advancements in DNA sequencing and computing technologies have allowed the development of personalized medicine. There is ongoing growth in the volume of genetic data produced globally, with genetic variant data as the principal basis for actionable knowledge. This data is stored in variant call format (VCF) files [1]. Growing data volumes present challenges in their high-level analysis, while their comparison can help uncover insights [2].

To this end, we developed VCF Observer: a web application that provides a user-friendly solution for VCF file analysis and comparison. VCF Observer enables users to conduct high-level analyses, removing the need for coding skills in the data exploration stage. The application allows users to upload, analyze, and visualize VCF files using a GUI. Common visualizations, including Venn diagrams, clustergrams, and precision–recall plots, are provided to users. A notable feature is the option for dynamic, metadata-based file grouping. Furthermore, the application allows for benchmarking via user-provided validation data.

VCF Observer provides a user-friendly interface and supports commonly used visualizations. It makes VCF file analysis more accessible and assists researchers and clinicians in their studies and practice. VCF Observer is open-source with its source code accessible at https://github.com/MBaysanLab/vcf-observer and it is hosted at https://bioinformatics.itu.edu.tr/vcfo for use by the community.

SABER: Sequencing Analysis Benchmarker–A Comprehensive Tool for Effective Benchmarking

Abdullah Emul[1], Mehmet Arif Ergün[1], Batuhan Eroğlu[2] Özgür Yolcu[3], Mehmet Baysan[1,*]

[1]*Istanbul Technical University, İstanbul, Turkiye*
[2]*Ankara University, Ankara, Turkiye*
[3]*Turkish-German University, İstanbul, Turkiye*

Presenting Author: asim_emul@hotmail.com
**Corresponding Author: baysanm@itu.edu.tr*

Genetic data contains many insights for diagnosing and treating diseases, leading to precision medicine techniques. NGS offers rapid and inexpensive detection of genetic profiles. NGS data must be analyzed through multiple algorithms, also called pipelines, for variant detection. These algorithms are imperfect and pipelines need to be thoroughly evaluated for accuracy since their performance is crucial for the efficient use of NGS technologies [1].

Many community standard reference data sets exist for such assessment and benchmarking processes. However, users are required to navigate complex web pages to access them. Moreover, numerous pipeline output comparison tools report various benchmarking metrics, which can be difficult to use for new users. No tool unifies the myriad resources and tools available for NGS variant detection benchmarking [2].

We present Saber: a unified platform through which the benchmarking of analysis pipelines is facilitated via a standardized interface. We have assembled lists of validation data sets and their metadata, simplifying its access. We have developed standard interfaces for output comparison software tools, making them accessible to inexperienced users. Users only need to select a data set and provide a pipeline output to be benchmarked. Saber greatly simplifies the sequencing analysis benchmarking process, acting as a one-stop shop for NGS pipeline validation and calibration needs.

Efficient Combination of Replicates for Better Somatic Variant Detection in WES

Yunus Emre Cebeci[1], Rumeysa Aslihan Erturk[1], Mehmet Arif Ergun[1], Mehmet Baysan[1,*]

[1]*Department of Computer Engineering, Istanbul Technical University, 34469 Istanbul, Turkey*

Presenting Author: cebeci16@itu.edu.tr
**Corresponding Author: baysanm@itu.edu.tr*

The detection of somatic variants in whole exome sequencing (WES) is essential for identifying genetic mutations associated with diseases such as cancer. However, with the billions of nucleotides in the human genome, even low experimental error rates can significantly affect the accuracy of variant calling. To address this challenge, replicating the sequencing of samples is a frequently used method to minimize the technical errors that occur in a single run. The Sequencing Quality Control Phase 2 (SEQC2) dataset, which contains replicates sequenced with different kits at various centers, offers an opportunity to develop and assess novel replicate-based methods1,2. Leveraging this dataset, we aimed to create methods to improve the reliability of somatic variant detection by reducing false positives and increasing consistency in variant calls.

We analyzed tumor/normal and tumor-only WES replicates from SEQC2 using multiple pipelines with two mappers (bwa, bowtie2) across all analyses. For tumor/normal samples, we employed variant callers including Mutect2, Strelka2, and SomaticSniper, while for tumor-only samples, we utilized Mutect2, VarScan2, VarDict, and LoFreq. Results were validated against declared high-confidence variants in high-confidence regions. Our consensus-based methods significantly improved somatic variant detection accuracy compared to single-pipeline methods. Building upon this foundation, we later trained machine learning models using the replicates, achieving performance comparable to models trained using declared high-confidence variants.

Detailed Evaluation of Structural Variant Detection Algorithms Through a Diverse Set of Benchmark Datasets

Gamze Maden Müftüoğlu[1,2], Fatma Zehra Sarı[3,4], Mehmet Baysan[1,3,*]

[1]*Department of Computer Engineering, Istanbul Technical University, 34485, Türkiye*
[2]*Department of Software Engineering, Istanbul Aydın University, 34295, Türkiye*
[3]*Türkiye Health Data Research and Artificial Intelligence Applications Institute, TUSEB, 34718, Türkiye*
[4]*Institute of Biotechnology, Gebze Technical University, 41400, Türkiye*

Presenting Author: maden21@itu.edu.tr
**Corresponding Author: baysanm@itu.edu.tr*

Structural variants (SVs) are DNA polymorphisms longer than 50 base pairs including insertions, deletions, inversions, and translocations. Accurate SV detection is essential for understanding genetic diversity and disease mechanisms but their size and complexity present significant challenges. Benchmark datasets are used to evaluate SV callers[1], but issues with accessibility and standardization hinder comparisons. We developed a repository combining eight WGS benchmark datasets (NA12878, NA19240, HG002, HG04036, HG02059, HG01352, HG00733, HG00514) and assessed six SV callers: Manta, Delly, Lumpy, Breakdancer, Wham, and GRIDSS. These algorithms follow different approaches: Breakdancer uses paired-end reads to infer SVs, while Manta, Delly, and Wham integrate paired-end and split-read. Lumpy combines paired-end, split-read and read-depth techniques while GRIDSS employs graph-based methods. Unlike SNPs, performance evaluation of SV detection is challenging since the same event can be detected at overlapping but not identical coordinates. We assessed algorithm performance using two detailed SV comparison tools: EvalSVcaller and Survivor, focusing on precision, recall, and F1-score. In EvalSVcaller, Breakdancer and Manta showed low precision and recall, while Delly demonstrated higher precision but struggled with recall. GRIDSS and Lumpy exhibited moderate precision with low recall. In Survivor, Breakdancer and Delly performed poorly while GRIDSS and Lumpy showed weaker results. The NA12878 and HG002 samples generally yielded higher precision and recall across most tools. Specifically, Manta and Wham achieved high precision but low recall, particularly in these high-quality datasets. Results highlight poor overall performance in SV detection which is concerning given the role of SVs in diseases. We observed considerable variability in tool performance, underscoring the need for careful tool selection, improved standardization, and more effective SV detection solutions.

Development of a Web-Based Tool for Identification of Druggable Pockets in Protein Conformational Ensembles and Ensemble Docking

Emre Taha Çevik[1], Onur Serçinoğlu[1,*]

[1]*Gebze Technical University, Engineering Faculty, Bioengineering Department*

Presenting Author: e.cevik2019@gtu.edu.tr
**Corresponding Author: osercinoglu@gtu.edu.tr*

Ensemble docking, which refers to molecular docking simulations targeting multiple conformations of a protein target, is increasingly used in structure-based drug discovery [1]. This method enhances the accuracy of predicting ligand-receptor interactions compared to traditional docking approaches. While several userfriendly tools have been developed for molecular docking, these mostly operate on static protein structures rather protein conformational ensembles. In this project, a novel web-based ensemble docking tool is proposed, integrating binding pocket predictings from protein conformational ensembles and subsequent molecular docking simulations via SMINA [2], which is an Autodock Vina fork [3]. Furthermore, the tool also provides additional functionalities for virtual screening applications [4]. By combining binding pocket prediction algorithms with ensemble docking, this tool offers a unique approach that can significantly advance drug discovery, protein engineering, and structural biology research by exploring the orthosteric and/or allosteric sites detected on the protein structure [1]. An intuitive interface makes the tool accessible to a wider audience. Users can perform druggable pocket predictions on an input of protein conformations, visualize clusters of binding pockets, and choose multiple binding pockets proposed by our workflow, enabling informed selection of conformations for ensemble docking. The utility of the tool is demonstrated by running the workflow on a conformational ensemble of a kinase and its respective active/inactive ligands [5].

Keywords: *Ensemble Docking, Drug Discovery, Allostery, Web Server*

Turkish Genome Project Data Sharing Portal

Ahmed Abunada[1], Mehmet Baysan[1,*]

[1]*Istanbul Technical University, İstanbul, Turkiye*

Presenting Author: a.abunada@msn.com
**Corresponding Author:* *baysanm@itu.edu.tr*

The Turkish Genome Project by TÜSEB aims to sequence 100,000 genomes from Turkish individuals, enhancing understanding of genetic diversity and disease predispositions. Although still in the early stages, the project has sequenced 500 genomes, providing data for research and precision medicine. Whole Genome Sequencing (WGS) produces extensive data, posing challenges for non-technical users.

To address this, we developed the Turkish Genome Data Shard Portal (TGD), an accessible platform that allows the exploration of complex genomic datasets through interactive visualizations and key metrics: https://tgd.tuseb.gov.tr. Users can examine genetic markers and population-specific variations without a technical background.

TGD takes inspiration (data, information etc) from well established bioinformatics platforms and services, one such example would be gnomAD: https://gnomad.broadinstitute.org/

TGD is built with an optimized tech stack for performance and scalability. The backend uses Django, a Python framework with a Model-View-Template architecture for efficient data management. Containerized with Docker, TGD supports seamless deployment across environments, ensuring stability and scalability. Nginx is a reverse proxy, and Gunicorn is the WSGI server that manages incoming requests efficiently. Nginx handles static files and offloads application-level requests to Gunicorn, which uses multiple workers for high concurrency.
The front end is built with HTML, CSS, and vanilla JavaScript, enhanced by jQuery and AJAX for smooth, asynchronous backend interactions. Django's templating system, with reusable blocks and conditional logic, enables responsive user experiences.

The data layer, powered by MySQL, securely stores genomic data, with Django's ORM framework simplifying database interactions. Processed VCF data is organized into structured tables (Chromosomes, Genes, Transcripts, and Variants), delivering near-instant analysis and visualization.

Impact of Computational Parameters on Somatic Variant Calling in Whole Exome Sequencing

Gül Eda Aydemir[1], Mehmet Arif Ergun[1], Abdulkadir Özel[2], Rumeysa Aslıhan Ertürk[1], Mehmet Baysan[1,*]

[1]Faculty of Computer and Informatics, Istanbul Technical University, Sarıyer, 34485, ˙Istanbul, Turkey and
[2]Molecular Biology and Genetics, Istanbul Technical University, Sarıyer, 34469, ˙Istanbul, Turkey

Presenting Author: aydemirg21@itu.edu.tr
**Corresponding author: baysanm@itu.edu.tr*

The complex nature of tumors makes identifying somatic variants challenging, often leading to inconsistent results across variant-calling pipelines. Previous studies underline the need to carefully select tools and parameters to ensure accurate somatic variant detection. This project evaluates the effect of key computational parameters on variant calling in somatic whole exome sequencing data, focusing on mapping algorithms, variant callers, and preprocessing steps. By analyzing samples from five sequencing centers and running multiple configurations, we generated 480 unique variant call sets to assess how parameters like trimming, base recalibration, and duplicate handling impact variant detection performance. Our analysis revealed significant heterogeneity across sample sets, showing considerable variability in precision, recall, and F1-scores depending on parameter configurations. In addition to the variant caller and aligner algorithm; trimming and base recalibration also substantially impacted accuracy. Performing trimming consistently increased recall while base recalibration improved precision, irrespective of other factors. Furthermore, adopting a consensus approach enhanced overall accuracy, yielding much higher F1-scores than individual pipelines alone. The highest consistency was observed with the combination of Bowtie and Mutect, with trimming and base recalibration. This setup also provided the most reliable tumor mutational burden estimates. As expected, regions with low map quality and segmental duplications showed the lowest recall and F1-scores, underscoring the need for tailored computational strategies in complex genomic regions. In this work, we assessed the heterogeneity in somatic variant analysis and highlighted the importance of analysis choices for reliable biomarker discovery to enhance cancer diagnostics.

Enhancing Drug-Target Interaction Prediction via Multimodal Learning and Domain Adaptation

Ardan Yılmaz[1], Volkan Atalay[1,*]

[1]*Middle-Eastern Technical University, Ankara, Turkiye*

Presenting Author: yilmazardan@gmail.com
**Corresponding author: vatalay@metu.edu.tr*

Drug-target interaction (DTI) prediction remains challenging in bioinformatics and cheminformatics due to the scarcity of annotated data in the vast genomic and molecular input space. This can be formulated as a binary classification problem with two inputs: drugs and proteins, thereby suitable for multimodal learning. We employ a state-of-the-art cross-attention mechanism to effectively fuse these modalities, assessing the interactions between features across different modalities. A significant issue in DTI prediction is the distributional difference between training and inference data caused by the lack of annotations in the vast input space. That's why standard random train/test splits, assuming distributional similarity between training and evaluation data, do not reflect a realistic setting and lead to over-optimistic evaluations. We use training and test sets composed of dissimilar drug-protein pairs to better simulate realistic scenarios, introducing a domain shift between source (training) and target (test) data. We employ domain adaptation methods to address this domain shift by training a domain-invariant feature extractor, aligning the source and target distributions alongside the primary classifier. In particular, we employ advanced statistical methods, such as Maximum Mean Discrepancy (MMD) Loss and adversarial training for domain alignment, resulting in a robust feature extractor. Our approach, combining multimodal learning with domain adaptation, exhibits performance on par with the state-of-the-art on widely-used benchmarking datasets, such as BindingDB and BioSnap, for DTI prediction. This demonstrates the effectiveness of our approach in DTI prediction, even in the presence of domain shift between training and test data.

Breast Cancer Research Targetting Programmed Cell Death in the Concept of 3PM Medicine

Elif Kubat Öktem[1,*]

[1]*Istanbul Medeniyet University, Istanbul, Turkiye*

Presenting Author: elifkubat.oktem@medeniyet.edu.tr
**Corresponding author*

Breast cancer is the leading cause of malignant tumor-related deaths worldwide and the most frequently diagnosed malignant tumor in women. Despite advancements in early detection and treatment, it remains crucial to explore new therapeutic strategies. A controlled mechanism of cell suicide, programmed cell death (PCD) is vital to an organism's growth, stability, and balance. A deeper understanding of PCD at the systems biology level could lead to novel treatment approaches for breast cancer. Breast cancer biomarkers linked to PCD are presented in this work, which may pave the way for better treatment approaches. To determine whether genes are up-or down-regulated in breast cancer tissue relative to normal tissue, publicly accessible transcriptome data were used for differential gene expression analyses. There were 244 PCD genes identified among these differentially expressed genes. To investigate network modules that are highly associated with breast cancer, a Weighted Gene Coexpression Network (WGCNA) analysis was conducted. The blue module had the strongest association with breast cancer ($r = 0.78$, $p = e^{-253}$) with 115 PCD genes. The 115 genes were further reduced to 41 using machine learning techniques such as random forest and logistic regression. Nineteen of these biomarkers were associated with apoptosis, ten with autophagy, seven with ferroptosis, four with lysosome-dependent cell death, and one with pyroptosis. Around these biomarkers, a multifactorial regulatory network was built. To determine the prognostic potential of these biomarkers, survival analyses were conducted. Lastly, Ipratropium Bromide was identified as a potential medicine that targets PCD genes for breast cancer by drug repositioning analyses. Further *in vitro* and *in vivo* investigation of these biomarkers and the Ipratropium Bromide would enable the development of 3PM (predictive, preventive, and personalised) medicine for the treatment of breast cancer.

Autoregressive Design of Target-Centric Drug Candidate Molecules with Chemical and Protein Language Models

Atabey Ünlü[1,2], Elif Çevrim[1,2], Tunca Doğan[1,2,*]

[1]*Biological Data Science Lab, Dept. of Computer Engineering, Hacettepe University, 06800, Ankara, Turkey*
[2]*Dept. of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, 06800, Ankara, Turkey*

Presenting Author: atabeyunlu36@gmail.com
**Corresponding author: tuncadogan@hacettepe.edu.tr*

Designing novel molecules is essential in drug development; however, experiments are laborious, costly, and require innovation. Generative modelling with large language models shows promise for optimised molecule design, yet effective targeting of specific proteins with AI-designed molecules remains a critical challenge. This study proposes Prot2Mol, a generative model for designing drug candidate molecules targeting understudied proteins with a few or no experimentally known ligands. The model employs an autoregressive, transformer-based encoder-decoder architecture, representing molecules as tokenised SELFIES and proteins through ProtT5-XL and ESM2 embeddings (Figure 1a). Through cross-attention, Prot2Mol learns specific protein-molecule interactions. Prot2Mol was trained on 418,956 experimental compound-protein pairs for 3,352 proteins and 291,777 compounds from ChEMBL. Targeted molecule generation is achieved by inputting any protein embedding during inference, enabling ligand generation even for proteins absent from the training data. We designed 10,000 molecules for AKT1 and DRD4. Our evaluations displayed strong performance and efficiency on generative benchmarks compared to existing models (Figure 1b). Top AKT1 and DRD4-targeting de novo molecules had average Tanimoto similarities of ~0.320 and ~0.370 to known inhibitors of these proteins, indicating their novelty; and high docking scores (Figure 1e), showing that Prot2Mol effectively modelled compound-protein interactions. Figure 1c displays 4 promising de novo molecules. Cross-attention visualisation of AKT1 and DRD4 with their crystal ligands capivasertib and nemonapride highlighted key binding pocket amino acids (Figure 1d), indicating the model learned critical molecular interactions. These molecules are currently being evaluated for synthesis and in vitro tests on drug-resistant cancer lines. Prot2Mol is open-sourced at https://github.com/HUBioDataLab/Prot2Mol and will be available as a web service.

Phylogeny-Driven Approaches for Variant Effect Prediction and Coevolution

Nurdan Kuru[1], Ogun Adebali[1,2,*]

[1]Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Turkiye
[2]TÜBİTAK Research Institute for Fundamental Sciences, 41470 Gebze, Turkiye

Presenting Author: nurdankuru@sabanciuniv.edu
*Corresponding Author: oadebali@sabanciuniv.edu

Evolutionary conservation is essential for predicting amino acid substitutability, functional loss in proteins, and other bioinformatics challenges. Using multiple sequence alignment (MSA) without phylogenetic context can lead to over-representation of dependent evolutionary events. Here, we introduce three novel, phylogeny-based methods for variant effect prediction and co-evolution, demonstrating the critical role of evolutionary information. Our first focus is on the impact of missense mutations, crucial for diagnosing Mendelian diseases. We developed PHACT, a phylogeny-dependent probabilistic approach that leverages gene-based phylogenetic trees. Starting from the query species, PHACT traverses the tree, recording positive probabilities of change for each amino acid to identify phylogenetically independent substitutions. This approach enables precise predictions by accounting for phylogenetic context and avoiding redundant counts. PHACT's performance, compared with tools in dbNSFP, showed higher AUC and AUPR values, indicating the benefit of expanding analysis beyond MSA alone. Building on PHACT, we developed PHACTboost, a gradient boosting classifier based on PHACT scores, MSA, phylogenetic trees, and ancestral reconstruction-based features. In comparisons across test sets of varying difficulty, PHACTboost outperformed dbNSFP predictors, including recent tools like AlphaMissense, CPT-1, and EVE. Expanding on these insights, we developed PHACE, a co-evolution method utilizing phylogenetic history. PHACE detects parallel substitutions by analyzing probability shifts between nodes, filtering out changes unrelated to co-evolution, and employing a refined MSA to categorize amino acids. In tests on 652 human proteins, PHACE showed superior performance over established tools in AUC, MCC, and F1 score comparisons. In sum, our phylogeny-based approaches advance variant prediction and co-evolution analysis, enhancing our understanding of protein evolution and function.

Multi-omic Single-cell Data Analysis for Biomarker Identification in Viral Pathogenesis

M. Burak Koca[1], F. Erdoğan Sevilgen[2,*]

[1]*Computer Engineering Department, Gebze Technical University, Kocaeli, Türkiye*
[2]*Management Information Systems Department, Fatih Sultan Mehmet Vakıf University, İstanbul, Türkiye*

Presenting Author: b.koca@gtu.edu.tr
**Corresponding author: sevilgen@gtu.edu.tr*

Pathogenesis encompasses the mechanisms that explain the cause, progression, and treatment of diseases. Understanding the pathogenesis of infectious diseases at the cellular level is crucial for developing effective prevention and treatment strategies. Single-cell sequencing technologies enable detailed investigations into the cellular impacts of infections. Single-cell multi-omic technologies facilitate the identification of distinct cell subsets, allowing for a deeper examination of viral pathogenesis across various omic layers. In this study, to understand viral pathogenesis, we identified infection levels of patient cells and explored distinctive biomarkers between different cell types or subsets. In line with this goal, multi-omics data were integrated using the variational graph auto-encoder-based SCPRO-VI method to increase cellular heterogeneity and discover distinct subsets. Infection levels were determined by analyzing the upregulations and downregulations of virus-targeted proteins, sourced from the PHISTO database, through z-score calculations. These values were compared against expected post-infection changes in target proteins and the number of matched changes determined each cell's infection level. Distinctive gene/protein analysis was employed for the subsets with high infection levels to identify biomarkers related to pathogenesis. The method was tested on a CITE-seq dataset of 161,764 cells from 8 HIV-1 patients, sampled pre- and post-vaccination. The highest infection levels were observed in CD4+ T cells, especially within the CTL subset. TCM and TEM subsets also showed high infection levels, while naive CD4+ T cells had mild levels. Among the top 5 distinctive genes identified in the TCM subset, 4 have been reported in the literature to be upregulated at post-HIV-1 infection. These results emphasize the role of highly infected subsets in identifying biomarkers in viral pathogenesis.

Network-based heterogeneity clustering in primary open-angle glaucoma patients reveals ancestry-specific gene clusters with known and novel genes

Eyyub Unlu[1], Meltem Ece Kars[2], Yuval Itan[2,*]

[1]*Wellcome Sanger Institute, Cambridge, UK.*
[2]*Mount Sinai School of Medicine, New York, USA.*

Presenting Author: eu1@sanger.ac.uk
[*]Corresponding Author: yuval.itan@mssm.edu

Background: Primary open-angle glaucoma (POAG) is a genetically complex neurodegenerative disease that accounts for 74% of all glaucoma cases. While genome-wide association studies (GWAS) have identified over 100 genes associated with POAG, establishing the role of common genetic variants, the contribution of rare genetic variants remains underexplored. Previous studies have shown that POAG prevalence and genetic associations vary across different ancestries. Therefore, investigating the relationship of rare and high-impact variants across different ancestries may provide insights into the missing heritability.

Methods: Whole exomes from three ancestries—African (AFR), European (EU), Hispanic (HIS) — were analysed from the BioMe Biobank. Cases were identified using ICD-10 code (H40.11), and individuals with other eye diseases (H00-H59) were excluded from both case and control groups. Population structure analyses were conducted to generate a genetically matched dataset of cases and controls. The number of cases in each ancestry group were 235 (AFR), 41 (EU), and 219 (HIS), with a fixed case-control ratio of 1:6.5. High-impact rare variants were retained through the following filtration methods: consequence (MisLoF), pathogenicity (CADD, MSC), and frequency (MAF <1%). First, gene-level analyses were conducted using the optimal unified sequence kernel association test (SKAT-O). Afterwards, network-based heterogeneity clustering (NHC) was employed to detect physiologically related gene clusters associated with POAG that might be missed by SKAT-O due to low statistical power resulting from sample size and genetic heterogeneity.

Results: SKAT-O did not identify any significant gene in any ancestry group (AFR, EU, HIS). NHC identified multiple significant ($P < 0.05$) gene clusters in each group. Importantly, five biologically-relevant gene clusters contained POAG-associated genes and did not overlap between groups. One of these significant clusters was in the HIS group and contained NOX1, NOX4, and NOXA1. These genes encode components of NADPH oxidase that have been shown to contribute to oxidative stress in glaucoma. In the future work, final candidate genes will be prioritised by biological relatedness methods.

Leveraging Hypergraph Neural Networks for Drug-Target Interaction Prediction

Elif Çevrim[1,2], Tunca Doğan[1,2,*]

[1]*Biological Data Science Lab, Department of Computer Engineering, Hacettepe University, Ankara, Turkey*
[2]*Department of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, Ankara, Turkey*

Presenting Author: candaselif@gmail.com
**Corresponding author: tuncadogan@hacettepe.edu.tr*

Drug–target interaction (DTI) prediction is essential for advancing drug discovery and repurposing. The growing availability of experimental DTI data enables systematic exploration of previously unknown interactions by constructing biological graphs and conducting link prediction in the context of graph learning. However, DTI prediction remains challenging due to the complex and heterogeneous correlations between drugs and targets. This work addresses the need to capture high-order correlations within heterogeneous DTI networks to enhance prediction performances. To this end, we propose a hypergraph-based framework for DTI prediction, where each vertex represents a drug, and labels are assigned based on their interacting targets, making each target a label category. Each hyperedge represents a specific molecular fragment shared among drugs, encoded as binary digits generated from PubChem fingerprints (Figure 1A). This approach enables the capture of high-order information critical for modelling complex interactions. We trained a model using experimental DTI data from the ChEMBL database and the hypergraph neural network (HGNN) architecture, which effectively aggregate high-order information through vertex-hyperedge-vertex propagation and spectral convolution to predict unknown DTIs (Figure 1B). Our experiments evaluated various hyperedge group sizes, data splits, and learning rates, using classification metrics (accuracy, precision, recall, and F1-score) confirming the framework's effectiveness in accurately identifying drug-target relationships (detailed results will be provided in the presentation). By learning joint biological representations, our model predicts DTIs through molecular fragment associations, offering a powerful drug discovery and repurposing tool. This approach can reveal previously unknown interactions, accelerating the identification of effective therapies and strategically repurposing existing drugs to address therapeutic challenges.

FlowDrug: Target-Specific Drug Candidate Molecule Generation with Latent Flow Matching Models

Melih Gökay Yiğit[1,2], Tunca Doğan[1,3,*]

[1]*Biological Data Science Lab, Dept. of Computer Engineering, Hacettepe University, 06800, Ankara, Turkey*
[2]*Dept. of Computer Engineering, Middle East Technical University, 06800, Ankara, Turkey*
[3]*Dept. of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, 06800, Ankara, Turkey*

Presenting Author: melih.gokay8@icloud.com
**Corresponding author: tuncadogan@hacettepe.edu.tr*

To meet the demand for therapeutic precision in drug development, it is essential to design molecules that bind to target biomolecules with high affinity and selectivity. However, traditional drug discovery methods are often time-consuming and costly. Addressing this challenge, we present FlowDrug, a novel conditional generative model for target-specific drug candidate molecule generation utilising latent Flow Matching (FM). Operating within a latent space constructed by a molecular variational autoencoder (VAE) trained on 4.2 million drug-like compounds from the MOSES, ZINC250K, and ChEMBL datasets (Fig.1A), FlowDrug employs a rectified flow process to iteratively refine latent representations with a transformer-based architecture where self-attention mechanisms capture dependencies within molecular representations, and cross-attention mechanisms (Fig.1B) integrate conditioning protein embeddings from protein language models like ProtT5-XL-U50 and ESM2 (Fig.1C). It leverages classifier-free guidance during sampling to enhance conditional information without explicit classifiers. Moreover, it can be directed through optimisation objectives using classifier guidance to generate molecules satisfying specific properties such as high drug-likeness, synthesizability, or multi-target activity, addressing multiple aspects of drug design in a unified framework. Trained on 663,000 bioactivity records of compound-protein interactions, FlowDrug demonstrates the ability to generate novel, diverse drug-like molecules with high binding affinity and selectivity for target proteins, showcasing its potential to accelerate drug discovery. Future work includes exploring different flow trajectories and constraint optimisation techniques to refine de novo molecules. Utilising advanced generative modelling techniques, FlowDrug represents a significant step forward in the computational design of target-specific small molecules and will be available as an open-access tool soon.

Automated Protein Function Prediction Using Biological Knowledge Graphs and Heterogeneous Graph Transformers

Erva Ulusoy[1,2], Tunca Doğan[1,2,*]

[1]*Biological Data Science Lab, Department of Computer Engineering, Hacettepe University, Ankara, Turkey*
[2]*Department of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, Ankara, Turkey*

Presenting Author: ervaulsy@gmail.com
**Corresponding author: tuncadogan@hacettepe.edu.tr*

Proteins are fundamental to cellular processes, and knowing their functions is essential for understanding complex biological systems. While function prediction methods offer a cost-effective alternative to experimental approaches, most current models rely on single data types—such as protein sequences—limiting their capacity to capture the complexity of protein functions. Knowledge graphs (KG) and geometric deep learning can address these limitations by introducing advanced data structures and frameworks to integrate and process diverse biological data types. In this study, we present "ProtHGT," a novel heterogeneous graph learning model for Gene Ontology (GO)-based protein function prediction (Figure 1A). ProtHGT utilises a comprehensive KG built from 14 diverse data sources, covering numerous biological entity types and their interactions/annotations/similarities. This integrative approach, combined with ProtHGT's heterogeneous graph transformer-based learning architecture, allows for a deeper understanding of protein functions. Through its attention mechanism, ProtHGT captures the unique characteristics of each node and edge type and their roles in determining protein functions, resulting in high performance when predicting molecular functions and cellular components across species (Figure 1B). An ablation study revealed that including diverse data sources in the source KG—especially domains, pathways, and other types of annotations—significantly improves prediction accuracy. Finally, a case study validated ProtHGT's predictions, highlighting its ability to identify biologically relevant protein functions (Figure 1C). These findings emphasise ProtHGT's potential to provide new insights into functional genomics, making it a valuable resource for researchers in bioinformatics. ProtHGT will soon be released as a programmatic tool, complete with its datasets and results, and as a user-friendly web server to ensure accessibility for a broader range of users.

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

POSTER PRESENTATIONS

Paper ID	Screen	Date	Time	Primary Author Name	Paper Title
				Best Poster Presentation Award (Session: Human Genetics)	
123	1	19.Dec	15:25	Dila Nur Çakal	Studying Allele Frequency Trajectories In Anatolia
21	1	19.Dec	15:35	Elif Öz	A Novel Wes Analysis Workflow For Gene Hunting In Primary Ciliary Dyskinesia (Pcd)
78	1	19.Dec	15:45	Atakan Ünlü	A Comparative Study Of Majiq, Fraser, And Dasper For Detecting Aberrant Splicing Events In Alzheimer's Disease Patients
129	1	19.Dec	15:55	Barış Yıldırım	Ccl2 As A Key Mediator Of Neural Invasion In Tumor Progression: An In Silico Study
26	1	20.Dec	15:25	İsra Mavaldi	The Role Of Gut Microbiota In Neurodegenerative Disorders: Targeting The Symptoms Of Alzheimer's Disease And Depression With Probiotics And Prebiotics
108	1	20.Dec	15:35	Kübra Nur Şahin	Investigating Molecular Changes Of Fusiform Gyrus Tissue In Alzheimer's Disease By Transcriptome Analysis
115	1	20.Dec	15:45	Rümeysa Aksu	An Rna-Seq Based Network Approach To Elucidate Molecular Mechanisms Of Asymptomatic Alzheimer's Disease
111	1	20.Dec	15:55	Tuğba Sever	Investigation Of Relationship Between Differential Mirnas And Genes In Parkinson's Disease
45	1	20.Dec	16:05	Kader Saribulak	Prediction Of Metabolite Biomarkers For Alzheimer's Disease Subtypes With An Innovative Algorithm Combining Genome-Scale Metabolic Models With Transcriptome Data
126	1	20.Dec	16:15	Elif Kesekler	Pathogenic Impact Of L284r Mutation In Syngap1 Gene And Its Association With Cerebral Palsy
				Best Poster Presentation Award (Session: Multiomics Approaches)	
116	2	19.Dec	15:25	Mustafa Keleş	Reanalysis Of Transcriptomics With A Focus Of Hox-Tale Genes Reveals Differentially Expressed Genes For Distinct Stages Of Non-Alcoholic Fatty Liver Disease (Nafld), Non-Alcoholic Steatohepatitis (Nash) And Cirrhosis
42	2	19.Dec	15:35	Yasin Kaymaz	An Innovative Approach For Distinguishing Tumor-Associated Macrophage Subtypes Within Single-Cell Rna Sequencing Data
125	2	19.Dec	15:45	Yasin Kaymaz	Correlations Between Immune Cell Compositions In Lung Cancer Tumors And Patient Survival Outcomes
69	2	19.Dec	15:55	Burak Özyürek	Integration Of Multi-Scale Agent-Based Models With Flux Balance Analysis
50	2	19.Dec	16:05	Kübra Tekşen	Evaluation Of The Effects Of Gentamicin On Multidrug Resistant Escherichia Coli Strains Using Proteomic Approaches
51	2	19.Dec	16:15	Kübra Tekşen	Evaluation Of The Effects Of Ampicillin And Ceftazidime Antibiotics On Proteus Mirabilis Using Metabolomic Approaches
67	2	20.Dec	15:25	Ilgim Durmus	Metagenomic Analysis Of Duodenal Bacterial And Fungal Microbiota In Obesity
71	2	20.Dec	15:35	Cyrille Mesue Njume	Multi-Omics Biomarker Analysis Of Alzheimer's Disease
79	2	20.Dec	15:45	Muzaffer Arıkan	Foodprot: Specialized Protein Sequence Databases For Food Metaproteomics
80	2	20.Dec	15:55	Arda Örçen	Comparison Of Different Genome-Scale Metabolic Models Of Pichia Pastoris For Overproduction Strategy Of Drug Precursor Metabolites
83	2	20.Dec	16:05	Aycan Şahin	Metabomics: Metabolism-Oriented Omics Data Integration
84	2	20.Dec	16:15	Ali Berk Alişan	Fibrosis Focused Mirna Profiling Reveals A Common Change In Hsa-Mir-377-3p Expression Under Several Msc Priming Conditions
88	10	19.Dec	15:25	Hilal Kazan	Batch Ordering Significantly Affects Scrna-Seq Data Integration Performance
90	10	19.Dec	15:35	Ferda Abbasoğlu	Imputation Of Dropout Rna Gene Expression Matrix Using Atac Data
92	10	19.Dec	15:45	Mehmet Izmirli	Evaluation Of Physicell Software For Construction Of Multi-Scale, Agent Based Models Of Biological Systems
97	10	19.Dec	15:55	Zeynep Özbiltekin	Deciphering Transcriptional Landscapes Of Mild And Advanced Masld Fibrosis: Insights Into Differential Gene Expression, Pathway Enrichment, And Drug Repurposing Opportunities
103	10	19.Dec	16:05	Gözde Ertürk Zararsiz	Methcomring: A Shiny Web-Tool For Visualization Of Circular Heatmaps To Identify Inter-Laboratory Variation In Ring Studies
107	10	19.Dec	16:15	Gözde Ertürk Zararsiz	Establishing Continuous Reference Intervals In Adult Thyroid Function Tests Using Gamlss Methods
105	10	20.Dec	15:25	Leila Kianmehr	Metabolomics And Transcriptomics Based Multi-Omics Integration Reveals Major Metabolic Pathways And Progression Biomarkers Associated With Autosomal Dominant Polycystic Kidney Disease
122	10	20.Dec	15:35	Tala Al-Rubaye	Single Host-Microbe Metabolic Interactions In The C. Elegans Gut Under Vitamin-Limited Conditions
72	10	20.Dec	15:45	Nura Fitnat Topbas Selcuki	Genetic Insights Into Endometriosis And Adenomyosis In Populations Of Turkish Ancestry

Paper ID	Screen	Date	Time	Primary Author Name	Paper Title
💊 Best Poster Presentation Award (Session: Drug Design and Development)					
4	3	19.Dec	15:25	Buse Meriç Açar	Evaluation Of Effects Of Potential Usp7 Inhibitors On Mdm2-P53 Interaction Through Molecular Docking Studies In Selected Cancers
19	3	19.Dec	15:35	Betul Orucoglu	Identification Of Potential Sars-Cov-2 Inhibitors Among Widely Used, Well-Tolerated Drugs Through Drug Repurposing And In Vitro Approaches
24	3	19.Dec	15:45	Zeynep Şevval Düvenci	Discovery Of Peptide-Based Inhibitor Targeting The Human Il18:Il18rα Complex
27	3	19.Dec	15:55	Mehmet Bakal	Biomedical Embedding Transitions For Enhanced Computational Drug Repositioning
38	3	19.Dec	16:05	Hacer Nur Eksi	Computational Insight Into Molecular Mechanism That Enable Survival Of Mycobacterium Tuberculosis: The Mtra Response Regulator Protein
52	3	19.Dec	16:15	Oktay Vosoughi	Computational Investigation Of Dynamics Of Beta-Arrestin-1 And Beta-Arrestin-2 That Are Co-Mutated In Lung Cancer: A Patient-Centric Approach
55	3	20.Dec	15:25	Zeynep Cinviz	Environment-Dependent Modulation Of Arrestin Dynamics And Activation
68	3	20.Dec	15:35	İremnur Yalçın	Computational Investigation Of The Role Of Ptm's And Mutations On Antibiotic Resistance In Mycobacterium Tuberculosis
96	3	20.Dec	15:45	Ümmü Söylemez	A Novel Prediction Model For Breast Cancer Diagnosis Using Drug-Gene Interaction-Based Pre-Existing Biological Knowledge
110	3	20.Dec	15:55	Dima Amairy	Computational Investigation Of Dynamics Of Circadian Rhythm Regulator, Clock-Bmal-1 Complex, In Dna-Histone Complex
127	3	20.Dec	16:05	Esma Yaz	Computational De-Novo Peptide Design For Gabaa Receptor To Be Used In Epilepsy Disease
128	3	20.Dec	16:15	Bünyamin Şen	Crossbarv2: A Unified Biomedical Knowledge Graph For Heterogeneous Data Representation And Llm-Driven Exploration
16	8	20.Dec	15:45	Berkan Tut	Identification Of Novel Therapeutic Targets In Pancreatic Ductal Adenocarcinoma Through Integrated Transcriptomic Analysis And Molecular Docking Approaches
🧬 Best Poster Presentation Award (Session: Genome Research)					
29	4	19.Dec	15:25	Emre Özzeybek	Species Classification From Short Genomic Reads Using Feedforward Neural Networks
76	4	19.Dec	15:35	Emre Özzeybek	Reproducibility Of Clinical Variant Detection In Next-Generation Sequencing Technologies
40	4	19.Dec	15:45	Büşra Nur Darendeli Kiraz	Centralized Evaluation Of Variant Effect Prediction Tools Using Curated Benchmark Datasets
23	4	19.Dec	15:55	Rumeysa Ertürk	Teaching Ngs To Engineers: Addressing Reproducibility Challenges In Bioinformatics
49	4	19.Dec	16:05	Zeynep Altundal	Comprehensive Evaluation Of Acmg-Based Variant Interpretation Tools
77	4	19.Dec	16:15	Emre Çamlıca	Fastq Or Bam: Alternative Replicate Combining Strategies For Ngs
10	4	20.Dec	15:25	Safa Aydın	Comparative Analysis Of Long-Read Structural Variant Calling Tools: Benchmarking Tools And Different Combinations For Detecting Somatic Structural Variants In Whole-Genome Sequencing Data
11	4	20.Dec	15:35	Elif Güney Tamer	Comparison Of Splice Site Prediction Tools For The Development Of Variant Effect Pipeline For Clinical Practice
12	4	20.Dec	15:45	Kübra Yılmaz	Mutational Signature Detection In Whole Exome Sequencing: Insights From Esophageal Squamous-Cell Carcinoma
13	4	20.Dec	15:55	Şevval Aktürk	Palaeogenomic Investigation Of Host-Pathogen Coevolution: A Computational Workflow For Data Analysis
56	4	20.Dec	16:05	Sema Duruoğlu	Investigating Ahr (Aryl Hydrocarbon Receptor) Signaling In Cancer Cells: An Integrated Bioinformatics And Experimental Framework
41	4	20.Dec	16:15	Doğukan Uğun	Heraclomics: A User-Friendly Local Server App For Scrna-Seq Analysis
57	9	19.Dec	15:25	Sinem Darwish	Medivar: A Next-Generation Sequencing-Based Clinical Variant Analysis And Reporting System
58	9	19.Dec	15:35	Beyza Kurtoğlu	Somatic Mutation Clustering Patterns In Cancers Associated With Uv Exposure
59	9	19.Dec	15:45	Beyza Kurtoğlu	A Comprehensive Rna-Seq Exploration: Variant Calling Analysis Of Transcriptome Data In Sotos Syndrome
61	9	19.Dec	15:55	Burcak Otlu	Analyzing Replication Timing Of Genes Across Different Cell Lines
47	9	19.Dec	16:05	Burcak Otlu	Rapid Construction Of Mutation-Centric Networks Leveraging Long-Range Interaction Data
64	9	19.Dec	16:15	Burcak Otlu	Deciphering Sequence Variations And Splicing Sensitivity: Predictive Analysis Of Psi In Srrm4 Response Groups
98	9	20.Dec	15:25	Elham Tarahomi	Investigation Of The Role Of Neuregulin3b In Synaptic Signaling And Brain Development In Zebrafish
106	9	20.Dec	15:35	Esma Aksel	Identification Of Intron Retention Regions In Breast Cancer By Bioinformatics Approaches
75	5	20.Dec	15:45	Faruk Üstünel	Evolutionary Insights Into The Cry1 Gene And Sleep Phenotypes: Identifying Conserved Regions And Snp Hotspots Across 392 Mammalian Species

Paper ID	Screen	Date	Time	Primary Author Name	Paper Title
				Best Poster Presentation Award (Session: Protein Structure & Function)	
5	5	19.Dec	15:25	Buse Şahin	Alphafold2 And Alphafold3 Leads To Significantly Different Results In Human-Parasite Interaction Prediction
6	5	19.Dec	15:35	Inci Sardag	Computational Modeling Of The Anti-Inflammatory Complexes Of Il37
30	5	19.Dec	15:45	Yağmur Poyraz	A Structural And Modeling Study Of Gklip Lipase For Enhanced Enzyme Activity
34	5	19.Dec	15:55	Betül Gündoğdu	Comparative In Silico Modelling Of Two Novel Eurocin-Like Antimicrobial Peptides From Ascomycota
36	5	19.Dec	16:05	Asli Kutlu	Molecular Dynamics Tool As A Powerful Approach To Prioritize The Structural Impacts Of Scid- And Os-Associated Missense Variants On Rag1
37	5	19.Dec	16:15	Yunus Dilek	Structural Insights Into Obesity-Associated Steap1b Variants Using Molecular Dynamics Simulations
53	5	20.Dec	15:25	Hacı Aslan Onur İşcil	Make Live, Then Kill: A New Structural Biology Based Strategy Against Tuberculosis
54	5	20.Dec	15:35	Ferzan Betül Berber	Mapping Functional Transcription Factor Binding Sites In Breast Cancer
60	5	20.Dec	15:45	Busel Ozcan	Computational Prediction Of Hbv Surface Antigen Structures: Insights For Vaccine Design And Therapeutic Discovery
65	5	20.Dec	15:55	Selahattin Aydoğan	Investigating The Impact Of Pathogenic Bmal1 Snps On The Human Circadian Clock Mechanism
99	8	19.Dec	15:25	Evren Atak	Computational Investigation Of Peptide Modulation Of Pmhc Dynamics
119	8	19.Dec	15:35	Ayca Emanetoglu	Exploring The Substrate Selectivity Of Mfs Proteins Through Structural Bioinformatics Methods
121	8	19.Dec	15:45	Useyma Gülhan	In Silico Analysis Of Latent Transforming Growth Factor B Binding Protein Family Expression Patterns And Prognostic Value In Hepatocellular Carcinoma
124	8	19.Dec	15:55	Zeynep Özer	Development Of A Software For The Analysis Of Atomistic Local Frustration Levels In Protein Structures
82	8	19.Dec	16:05	Levise Tenay	In Silico Characterization Of Htr1d Missense Mutations In Obesity
8	8	19.Dec	16:15	Ahmet Zübeyir Nursoy	Automating Fret Analysis For Enhanced Characterization Of Biomolecular Interactions
				Best Poster Presentation Award (Session: Machine Learning Applications)	
1	6	19.Dec	15:25	Rohan Butani	Evaluating And Comparing Various Machine Learning Driven Prediction Models Of Ulnar Collateral Ligament Reconstruction
14	6	19.Dec	15:35	Emma Qumsiyeh	Mirgedinet: A Comprehensive Examination Of Common Genes In Mirna-Target Interactions And Disease Associations: Insights From A Grouping-Scoring-Modeling Approach
7	6	19.Dec	15:45	Açelya Dalgıç	Machine Learning Based 16s Microbiome Processing Pipeline: A Case Study For Type 1 Diabetes
15	6	19.Dec	15:55	Merve Yarici	Integrating Machine Learning With Gems: A Framework For Advancing Complex Metabolic Simulations
20	6	19.Dec	16:05	Hakan Kahraman	Pan-Cancer Analysis Of Non-Apoptotic Cell Death Mechanisms Using Machine Learning
102	6	19.Dec	16:15	Gokmen Zararsiz	Mlms: Machine-Learning Interface For Metabolomics Data
130	6	20.Dec	15:25	Duha Alioglu	Age-Related Transcriptomic Changes In Mouse Dental Tissue: Insights From Single-Cell Rna Sequencing Using Conventional And Machine Learning Approaches
86	6	20.Dec	15:35	Nur Sebnem Ersoz	Biomarker Identification Via Biological Domain Knowledge Based Feature Grouping And Scoring In Omics Data Analysis
33	6	20.Dec	15:45	Daniel Voskergian	Topic Selection For Text Classification Using Ensemble Topic Modeling With Grouping, Scoring, And Modeling Approach
46	6	20.Dec	15:55	Nurten Bulut	Efficient Gene Expression Analysis Through Simultaneous Scoring In Recursive Cluster Elimination
87	6	20.Dec	16:05	Gülbahar Merve Şilbir	Ppromoter-Fcgr: Deep Learning On Frequencey Choas Game Representation For Prediction Of Dna Promoters
89	6	20.Dec	16:15	Alperen Dalkıran	Molecular Contrastive Learning With Graph Attention Network (Mocl-Gat)
104	7	19.Dec	15:25	Seyit Semih Yiğitarslan	Language Modeling-Based Generative Artificial Intelligence For De Novo Protein Design
22	7	19.Dec	15:35	Maham Khokhar	Improving Efficiency Of The Grouping-Scoring-Modeling Framework Through Statistical Pre-Scoring In Transcriptomic Data Analysis
28	7	19.Dec	15:45	Yasin Inal	The G-S-M Grouping Scoring Modeling Approach
9	7	19.Dec	15:55	Hamza Umut Karakurt	Lunara: A Light-Weight Rshiny Application To Annotate And Network-Based Analyses Of Genomic Variants For Gene Panels
25	7	19.Dec	16:05	Sandeep Pedapudi	A Machine Learning Model To Detect The Symptoms Of Parkinson's Disease Detection Using Spiral And Wave Images
				Best Poster Presentation Award (Session: Artificial Intelligence for Health Informatics)	
94	7	19.Dec	16:15	Pınar Pir	Reconstruction Of Multi-Scale Mathematical Models Of Tumors And Their Microenvironments
35	7	20.Dec	15:25	Umit Akkaya	Histopathological Image Classification With Transfer Learning
101	7	20.Dec	15:35	Serra İlayda Yerlitaş Taştan	Estimation Of Apolipoprotein B Levels Using Machine-Learning Models
112	7	20.Dec	15:45	Barış Can	Estimating Metabolic Differentiation In Diseases With Deep Learning
113	7	20.Dec	15:55	Ceren Yaşar	Predicting Unplanned Hospital Readmissions
118	7	20.Dec	16:05	Hacer Yeter Akıncı	Employing Anatomical Therapeutic Chemical Ontology To Deal With High Dimensionality And Performance Issues In Predictive Clinical Data Analytics
131	7	20.Dec	16:15	Zeynep Özkeserli	Unveiling The Protein Landscape In Cerebral Palsy Through Ai-Based Structural Analysis

Evaluating and Comparing Various Machine Learning Driven Prediction Models of Ulnar Collateral Ligament Reconstruction

Rohan Butani[1], Kedar Chintalapati [2], Anirudh Sridharan [3], Theodore Oltean [4], Arjunn Shastri[1,*]

[1] *University of Washington*
[2] *The Overlake School*
[3,4] *Redmond High School*

Presenting Author: rohanbutani81@gmail.com
**Corresponding Author: arjunnshastri@outlook.com*

Ulnar collateral ligament (UCL) reconstruction, commonly known as Tommy John (TJ) surgery, has become increasingly prevalent among baseball players, particularly pitchers. This paper presents a machine learning (ML) approach to predict the likelihood of TJ surgery in players based on performance metrics determined using public databases. Utilizing a dataset encompassing diverse player profiles and a matched-pair design, we developed various models to predict UCL reconstruction. Deep Neural Decision Forest attained the highest accuracy at 79.2%, displaying a robust and novel capability to predict UCL reconstruction in Major League Baseball pitchers. Diabetes is a chronic disease that requires rigorous management of blood sugar levels to avoid serious complications. Intelligent assistance systems, based on artificial intelligence (AI), have shown promise in personalized diabetes management. This article examines the application of reinforcement learning (RL) as an innovative approach to optimizing treatments, adjusting insulin doses and providing personalized nutritional recommendations. Reinforcement learning is characterized by its ability to learn from continuous interaction with a dynamic environment. In the context of diabetes, this environment is represented by the patient's physiological state, measured by continuous data such as blood glucose, diet and physical activity. Possible actions include adjusting insulin doses, providing dietary recommendations, as well as personalized advice on physical exercise. We propose a model of LR structured around three main components: As the intelligent agent adapts to the patient's physiological responses, it becomes increasingly adept at anticipating insulin needs and preventing imbalances. Looking ahead, we aim to test the intelligent assistant over a sufficiently long period to enable the system to improve its performances and then discuss the results as part of a collaboration between researchers, clinicians and engineers to further improve diabetes management using artificial intelligence technologies.

Using Artificial Intelligence and Reinforcement Learning to Provide Intelligent Assistance to Diabetic Patients

Ali Seridi[1,*], Yamina Bordjiba[1], Riad Bourbia[1]

[1]*Computer Science Department, Labstic Laboratory, 8 May1945 Guelma University*

Presenting Author: seridi.ali@univ-guelma.dz
**Corresponding Author: seridi.ali@univ-guelma.dz*

Diabetes is a chronic disease that requires rigorous management of blood sugar levels to avoid serious complications. Intelligent assistance systems, based on artificial intelligence (AI), have shown promise in personalized diabetes management. This article examines the application of reinforcement learning (RL) as an innovative approach to optimizing treatments, adjusting insulin doses and providing personalized nutritional recommendations. Reinforcement learning is characterized by its ability to learn from continuous interaction with a dynamic environment. In the context of diabetes, this environment is represented by the patient's physiological state, measured by continuous data such as blood glucose, diet and physical activity. Possible actions include adjusting insulin doses, providing dietary recommendations, as well as personalized advice on physical exercise. We propose a model of LR structured around three main components: **States:** representing current blood glucose levels, general health status, treatment history, and the patient's dietary and exercise habits. **Rewards:** aimed at minimizing deviations from optimal blood glucose levels, while reducing the risk of hypoglycemia and hyperglycemia. RL's algorithm maximizes long-term cumulative reward, which corresponds to glycemic stability and overall well-being. As the intelligent agent adapts to the patient's physiological responses, it becomes increasingly adept at anticipating insulin needs and preventing imbalances. Looking ahead, we aim to test the intelligent assistant over a sufficiently long period to enable the system to improve its performances and then discuss the results as part of a collaboration between researchers, clinicians and engineers to further improve diabetes management using artificial intelligence technologies.

Towards Early Diagnosis of Melanoma: Improving the Performance of ResNet50 with SMOTE

Yamina Bordjiba[1,*], Ali Séridi[1], Nabiha Azizi[2], Hayat Farida Merouani[3] and Randa Salmi[1]

[1]*Computer Science Department, Labstic Laboratory, 8 May1945 Guelma University*
[2]*LabGed Laboratory, Badji Mokhtar University, Annaba, Algeria*
[3]*LRI Laboratory, Badji Mokhtar University, Annaba, Algeria,*

Presenting Author: hayet_merouani@yahoo.fr
**Corresponding Author: bordjiba.yamina@univ-guelma.dz*

To improve patients' chances of survival, it is essential to diagnose melanoma at an early stage. Convolutional neural networks (CNNs), specifically the ResNet50 model, have shown promising potential for the automated detection of skin lesions. However, medical datasets often suffer from an imbalance between classes, with the number of melanoma samples being much lower than for other lesion types. This imbalance can bias the results and reduce the model's sensitivity in detecting melanoma. Because of the class imbalance, we chose to use the SMOTE technique to artificially generate examples from the minority class (melanomas). By applying SMOTE to the ISIC2019 dataset, we obtained a balanced dataset on which we trained the ResNet50 model for the binary classification task. The ResNet50 model achieved 70% accuracy with low sensitivity for melanoma detection before the application of SMOTE. However, after applying SMOTE, the model's performance improved considerably, reaching an accuracy of 94.57%, with increased sensitivity for detecting melanoma cases. These results demonstrate the effectiveness of SMOTE in improving the detection capacity of deep learning models in the face of unbalanced data. This study highlights the importance of dealing with data imbalance in automated diagnostic systems and proposes a promising approach combining pre-trained models such as ResNet50 with data augmentation techniques to improve the accuracy and reliability of diagnostic tools in dermatology.

Keywords: *Melanoma, imbalanced medical dataset, ResNet50, SMOTE.*

Evaluation of effects of potential USP7 inhibitors on MDM2-p53 interaction through molecular docking studies in selected cancers

Buse Meriç Açar[1], Engin Ulukaya[2,3,*], Aslı Kutlu[3,4]

[1]*Cancer Biology and Pharmacology, Health Science Institute, Istinye University, Istanbul-Türkiye*
[2]*Clinical Biochemistry, Faculty of Medicine, Istinye University, Istanbul-Türkiye*
[3]*Istinye Molecular Cancer Research Center, Istinye University, Istanbul-Türkiye*
[4]*Molecular Biology and Genetics, Faculty of Engineering and Natural Science, Istinye University, Istanbul-Türkiye*

Presenting Author: 229912810@stu.istinye.edu.tr
**Corresponding Author: eulukaya@istinye.edu.tr*

USP7 is a deubiquitinating enzyme that regulates the p53-MDM2 pathway. Somatic mutations in USP7 stabilize MDM2, causing p53 degradation and helping cancer cells evade death, making it a key target in cancer research. However, the lack of structure-based functional information about the USP7-inhibitor interactions has been a critical gap that has to be fulfilled for the development of effective potent inhibitors toward USP7 in the presence of different mutations. Thus, the main aim of this study, which can be further associated with clinical outcomes, is to evaluate the binding affinity of USP7 inhibitors in the presence of varied missense mutations in USP7 and thus to constitute a deeper knowledge about the selection of possible inhibitors for different scenarios in USP7. We disclosed all missense mutations in the MDM2-p53 binding site and catalytic domain of USP7 by retrieving from the literature for several solid tumors. In total, 181 missense mutations were subjected to further prioritization by using several web-tools. Among all listed ones, the prioritized mutation list is constituted by setting pathogenicity score of mutation as 70% or above for future steps. In this regard, we prioritized 28 missense mutations of USP7. With this prioritized mutation list, we aim to evaluate two things: possible alterations (1) in binding mechanism of p53 and MDM2 to TRAF domain of USP7 and (2) in binding capability of inhibitors having varied modes of actions. To achieve goal 1, we will perform 200 ns MD simulations at 310 K using CHARMM force fields on both native and mutant USP7 conformations. Then, we will evaluate how missense mutations in the TRAF domain affect p53 and MDM2 binding via protein-protein docking studies.

Keywords: *USP7's catalytic domain impact its binding potential, using HADDOCK, HawkDock, and MM/GBSA calculations.*

AlphaFold2 and AlphaFold3 Leads to Significantly Different Results in Human-Parasite Interaction Prediction

Burcu Ozden[1,2], Buse Sahin[1,2], Ezgi Karaca[1,2,*]

[1]*Izmir Biomedicine and Genome Center; Izmir Türkiye*
[2]*Dokuz Eylul University; Izmir, Türkiye*

Presenting Author: burcu.ozden@ibg.edu.tr
**Corresponding Author:* ezgi.karaca@ibg.edu.tr

Parasitic diseases pose a significant global health challenge with substantial socioeconomic impact, especially in developing countries. Combatting these diseases is difficult due to two factors: the resistance developed by parasites to existing drugs and the underfunded research efforts to resolve host-parasite interactions. To address the latter, we focused on 276 human-parasite interaction predictions involving 15 parasitic species known to cause infections in humans. We aimed to validate these interactions using AlphaFold2-Multimer (AF2) and AlphaFold3 (AF3) modeling. Out of the 276 models, only 42 demonstrated a 'trustworthy' confidence score (≥ 0.8) from either AF2 or AF3. Interestingly, among these 42 interactions, only half displayed a high degree of structural similarity. Our results indicate that, within the context of human-parasite interactions, AF2 and AF3 produce distinct models and score distributions, with no significant correlation between their top-ranked predictions. We recognize that predicting host-parasite interactions is particularly challenging due to the lack of strong coevolutionary signals. Nonetheless, the observed discrepancies in scores and models between AF2 and AF3 are noteworthy and highlight the need for further investigation, which will become possible with the release of AF3.

Computational Modeling of the Anti-Inflammatory Complexes of IL37

Inci Sardag[1], Zeynep Sevval Duvenci[2], Serkan Belkaya[3], Emel Timucin[2,4,*]

[1]*Department of Molecular Biology and Genetics, Bogazici University, Istanbul 34342,Turkey,*
[2]*Department of Biostatistics and Bioinformatics, Acibadem University, Institute of Health Sciences, Istanbul 34752 Turkey*
[3]*Department of Molecular Biology and Genetics, Bilkent University, Ankara 06800,Turkey,*
[4]*Institute of Health Sciences Department of Biostatistics and Bioinformatics, Acibadem University, Istanbul 34752 Turkey*

Presenting Author: inci.sardag@std.bogazici.edu.tr
**Corresponding Author: emel.timucin@acibadem.edu.tr*

Interleukin (IL) 37 is an anti-inflammatory cytokine belonging to the IL1 protein family. Owing to its pivotal role in modulating immune responses, particularly through interfering with the IL18 signaling, elucidating the IL37 complex structures holds substantial therapeutic promise for various autoimmune disorders and cancers. Although the structural homology between IL37 and IL18 suggests a common binding mechanism with the primary members of IL18 signaling, the structures of IL37 complexes have not been experimentally resolved yet. This computational study aims to address this gap through molecular modeling and classical molecular dynamics simulations, revealing the structural underpinnings of its modulatory effects on the IL18 signaling pathway. All IL37 protein-protein complexes, including both receptor dependent and receptor-independent pairs, were modeled using a range of methods from homology modeling to AlphaFold2 multimer predictions. The models that successfully captured experimental features were subjected to molecular dynamics simulations. As positive controls, binary and ternary PDB complexes of IL18 were also included. The comparative look on the IL37 and IL18 complexes revealed a highly dynamic nature for the IL37 complexes. Repeated simulations of IL37-IL18R_ showed altered receptor conformations capable of accommodating IL37 in its dimeric form without clashes, providing a structural basis for the failure of IL18R_ to be recruited to the IL37-IL18R_ complex. Simulations of receptor complexes involving various mature forms of IL37 revealed that the N-terminal loop of IL37 is pivotal in modulating receptor dynamics. Additionally, the glycosyl chains on the primary receptor residue N297 act as a steric block against the IL37's N-terminal loop. The interactions between IL37 and IL18BP were also investigated, and our dynamical models indicated that a homologous binding mode was unlikely, suggesting an alternative mechanism by which IL37 functions as an anti-inflammatory cytokine upon binding to IL18BP. Altogether this study accesses to the structure and dynamics of IL37 complexes, offering molecular insights into IL37's inhibitory function within the IL18 signaling pathway and informing future experimental research.

Keywords: *IL37, IL18 signaling, anti-inflammation, homology modeling, molecular dynamics simulations*

Machine Learning Based 16S Microbiome Processing Pipeline: A Case Study for Type 1 Diabetes

Açelya Dalgıç[1], İdil Yet[1,*]

[1]*Department of Bioinformatics, Institute of Health Sciences, Hacettepe University, Ankara, Turkey 06100*

Presenting Author: aclydlgc69@gmail.com
**Corresponding Author: dilyet@gmail.com*

Type 1 Diabetes (T1D) is a chronic autoimmune disease when the body's immune system mistakenly attacks insulin-producing cells in the pancreas. With the recent progress of microbiome studies, it has become clear that there is a strong connection between the gut microbiome and several diseases. Although T1D is known to be a genetic disease, studies have found a link between the disease and the gut microbiome. In this study, we aimed to investigate the differences in gut microbiota of patients and healthy individuals and to build a pipeline to predict patients using fecal sample sequences with ML models. In previous studies, it has been stated that there are differences between patients and healthy individuals in terms of the abundance of different taxon's and diversity indexes. Although the overall composition of gut microbiota changes between populations, the main differences remain certain. In our study, sequences of fecal samples from three different projects that include different populations and age groups have been processed with QIIME2. A QIIME2 Classifier trained with the SILVA Database is used for taxonomic classification. To overcome population-related differences, catch the fundamental difference between the patient and healthy group, and to reduce the dimension in order to make the chosen models work better, we calculated the diversity indexes of each sample such as alpha and beta diversity. R package vegan is used for diversity calculations. We also incorporated categorical parameters such as age, project, and sex into our models. We applied Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) models in this study. K-Fold Cross-Validation (CV) is applied to each trained model. As a result, we found that the SVM Model had the best performance among all, with 0.94 Average CV Score at K=6. These findings highlight the potential of gut microbiota as biomarkers for T1D and support further exploration in this field.

Automating FRET Analysis for Enhanced Characterization of Biomolecular Interactions

Ahmet Zübeyir Nursoy[1], *Çağdaş Son*[1,*]

[1] *Faculty of Arts and Science, Middle East Technical University,Ankara Turkey*

Presenting Author: *zubeyir.nursoy@metu.edu.tr*
**Corresponding Author:* *cson@metu.edu.tr*

Fluorescence Resonance Energy Transfer (FRET) is a powerful technique to study molecular interactions by detecting energy transfer between donor and acceptor fluorophores. However, challenges such as signal bleed-through and non-uniform noise often affect measurement accuracy, especially in manual analysis. Automating the analysis process can mitigate human error, enhance efficiency, and improve the reproducibility of results. This work aims to develop an automated pipeline for FRET analysis that streamlines image preprocessing, bleed-through correction, cell segmentation, and energy transfer calculations. The goal is to achieve fast, reliable, and user-free analysis while addressing common issues like signal variability and image noise.

Preprocessing: Background subtraction using a moving kernel method to remove noise and enhance image contrast. Bleed-through Correction: Implementing element-wise division and regression-based approaches to reduce fluorophore overlap. **Cell Segmentation:** Using morphological operations and thresholding to accurately identify cell boundaries, even with non-uniform topologies. Also for touching cells, a type of watershed algorithm was harnessed. **Automated FRET Calculations:** Binary transformation and masking techniques to measure energy transfer with high precision. Our automated pipeline provides robust whole-frame FRET analysis, minimizing variability and improving comparability across experiments. Comparative analysis of positive and negative samples demonstrates the effectiveness of the automation in eliminating bleed-through artifacts and enhancing measurement reliability. Automating FRET analysis significantly improves accuracy and efficiency, making it a reliable tool for studying biomolecular interactions. Future directions include the application of fully convolutional network (FCN) and neural networks for novel bio-interaction detection and advanced image classification.

LunaRa: A light-weight RShiny Application to Annotate and Network-Based Analyses of Genomic Variants for Gene Panels

Hamza Umut Karakurt[1,2], Kader Sarıbulak[1,2,*]

[1]*Department of Bioengineering, Gebze Technical University, Kocaeli, Turkey*
[2]*Idea Technology Solutions R&D Center, Sarıyer, Istanbul*

Presenting Author: *hamza_karakurt@windowslive.com*
**Corresponding Author:* *kader.saribulak@gmail.com*

Genomic variant analysis is crucial for understanding disease mechanisms, diagnostics, and identifying potential drug targets. From single-gene disorders to complex diseases involving numerous variants, annotating these variants is a key step. Using annotations such as population frequencies, clinical evidence, and computational predictions from various databases, variants can be interpreted to inform patient diagnosis and prognosis [1]. Tools such as Annovar [2], Ensembl VEP [3], and SnpEff [4] facilitate variant annotation, they often require command-line expertise, which can be a barrier for many users. To address this, we present LunaRa, a user-friendly RShiny-based web application designed for small-scale variant annotation and network analysis, making it ideal for targeted gene panels rather than whole-exome or genome sequencing data. LunaRa integrates data from MyVariant.info [5], HumanNet [6], PharmGKB [7], and DISEASES [8] for now, providing comprehensive variant annotations, pharmacogenomic interactions, proteinprotein interaction networks, and gene-disease associations using VCF files. The app generates annotated variant tables, PPI networks based on genes with variants and their first neighbors, a table and graph based on genes with their associated diseases, and functional enrichment analyses (GO, KEGG). Available as both a desktop application and a web app through RShiny, LunaRa enhances accessibility for researchers and clinicians. LunaRa can be downloaded from https://github.com/hkarakurt8742/LunaRA or can be used https://ideatechnologysolutions.shinyapps.io/LunaRa/ using browser.

Comparative Analysis of Long-Read Structural Variant Calling Tools: Benchmarking Tools and Different Combinations for Detecting Somatic Structural Variants in Whole-Genome Sequencing Data

Safa Aydın[1], Kübra Yılmaz[1,*]

[1]*Department of Bioinformatics, Middle East Technical University, 06800 Ankara, Turkey*

Presenting Author: aydinsafakerem@gmail.com
**Corresponding Author:* kubracelikbas4@gmail.com

Cancer genomes have a complicated landscape of mutations, including large-scale rearrangements known as structural variants (SVs). These SVs can disrupt genes or regulatory elements, playing a critical role in cancer development and progression. Despite their importance, accurate identification of somatic structural variants (SVs) remains a significant bottleneck in cancer genomics. Long-read sequencing technologies hold great promise in SV discovery, and there is an increasing number of efforts to develop new tools to detect them. In this study, we employ eight widely used SV callers on paired tumor and matched normal samples from both the NCI-H2009 lung cancer cell line and the COLO829 melanoma cell line, the latter of which has a well-established somatic SV truth set. Following separate variation detection in both tumor and normal DNA, the VCF merging procedure and a subtraction method were used to identify candidate somatic SVs. Additionally, we explored different combinations of the tools to enhance the accuracy of true somatic SV detection. Our analysis adopts a comprehensive approach, evaluating the performance of each SV caller across a spectrum of variant types and numbers in finding cancer-related somatic SVs. This study, by comparing eight different tools and their combinations, not only reveals the benefits and limitations of various techniques but also establishes a framework for developing more robust SV calling pipelines. Our findings highlight the strengths and weaknesses of current SV calling tools and suggest that combining multiple tools and testing different combinations can significantly enhance the validation of somatic alterations.

Comparison of Splice Site Prediction Tools for the Development of Variant Effect Pipeline for Clinical Practice

Elif Güney Tamer[1], Dr. Büşranur Çavdarlı[2], Doç. Dr. Yeşim Aydın Son[3,*]

[1]*Department of Biological Sciences, Middle East Technical University, 06800,Ankara Turkey*
[2]*Ankara City Hospital, 06800 Ankara, Turkey*
[3]*Department of Medicine, Hacettepe University, 06800, Ankara, Turkey*

Presenting Author: elifguney8@gmail.com
**Corresponding Author:* yesim@metu.edu.tr

RNA splicing is a post-translational modification process in which pre-mRNAs are converted into mature mRNAs. It is a cellular process in that multiple mRNA transcripts can be produced from a single gene which drives protein diversity. There are 3' (acceptor) and 5' (donor) splice sites and branchpoint motifs that regulatory proteins recognize to carry out splicing. Any variants in these regions can disrupt the mechanism and lead to different diseases. It has been reported that up to 62% of all disease-causing point mutations affect RNA splicing. Several tools have been developed to predict splice sites, the splice site motif, and the effect of splice site variants. However, there is a need to translate this knowledge in bioinformatics analysis to clinical practice. In this study, we aimed to compare three deep learning/machine learning-based splice site variant tools to provide a pipeline for the analysis of patient data. We compared Pangolin, SPIP and SpliceAI tools with experimentally validated MFASS dataset. Also, in order to compare with bigger dataset, we repeated our analysis SPIP validation dataset as well. Comparisons are made with confusion matrix, precision-recall, MCC score and F1 score. The variant datasets are separated into three categories: upstream intronic sites, downstream intronic sites and exonic sites. Our results showed that Pangolin predicted better than the other tools in almost all regions with both MFASS data and SPIP data. Performance scores of all three tools on exonic splice sites are low compared to other sites. This may be due to training of all these tools based on canonical splice site regions and motifs. Therefore, as a future study we would like to create a new splice site tool for prediction of exonic splice site variants.

Mutational Signature Detection in Whole Exome Sequencing: Insights from Esophageal Squamous-Cell Carcinoma

Kübra Yılmaz[1], Burcak Otlu[1], Ahmet Acar[1,*]

[1]*Institute of Bioinformatics, Middle East Technical University, 06800, Ankara Turkey,*

Presenting Author: *kubracelikbas4@gmail.com*
**Corresponding Author:* *acara@metu.edu.tr*

Mutational signatures are patterns of mutations associated with biological processes such as DNA damage, repair mechanisms, and environmental exposures. These signatures hold great potential for clinical applications, including early cancer detection, targeted therapies, and predicting treatment responses. While whole genome sequencing (WGS) has successfully identified these signatures and shown promising clinical results, whole exome sequencing (WES) has limited detection power due to the lower number of mutations detected. Considering the affordability and interpretability of WES in clinical settings, there is a need to develop methods to accurately determine mutational signatures from WES data. Our study analyzed 552 esophageal squamous-cell carcinoma WGS samples and corresponding down-sampled WES data. Our results revealed that the detection power of WES was reduced by half for all signature types, including SBS, ID, and DBS. These initial findings suggest that further development, including the application of deep learning models, is necessary to enhance the sensitivity and accuracy of mutational signature detection in WES data, making it a valuable tool for clinical applications.

Palaeogenomic investigation of host-pathogen coevolution: A computational workflow for data analysis

Şevval Aktürk[1], Duygu Deniz Kazancı[2], Emrah Kırdök[3], İdil Yet[1], Füsun Özer[4], Yılmaz Selim Erdal[4], Gülşah Merve Kılınç[1,*]

[1]*Department of Bioinformatics, Hacettepe University, Ankara, Turkey*
[2]*Department of Biological Sciences, Middle East Technical University, Ankara, Turkey*
[3]*Department of Biotechnology, Mersin University, Mersin,Turkey*
[4]*Department of Anthropology, Hacettepe University, Ankara, Turkey*

Presenting Author: sevvalakturk96@gmail.com
**Corresponding Author:* gulsahhdal@gmail.com

Throughout human history, profound changes in diet and lifestyle have shaped human biology, leaving detectable signatures in both the human genome and microbiome. Advances in paleogenomics now enable the reconstruction of time-series genomic and metagenomic datasets. In this study, we aim to analyze ancient holobiomes (n=56 individuals)—encompassing both human and microbial DNA—from three distinct Anatolian populations spanning 7,000 years, from the Late Neolithic to the Medieval period. Here, we reconstruct a computational workflow to analyze human and microbial genomes produced by paired-end sequencing of ancient DNA extracted from teeth and dental calculus. First, the ancient metagenomic tool, aMeta, will be applied to holobiome sequences for *k*-mer and alignment-based taxonomic classification to accurately identify microbial species and authenticate ancient DNA reads. Second, data will be used to assess microbial diversity across individuals and quantify species abundances across periods to uncover changes in microbial composition. Third, to explore potential interactions between genetic background and oral microbiome patterns in ancient individuals with dental caries, linear and logistic regression models will be employed. Finally, probabilistic Bayesian networks will be utilized to investigate the underlying genetic factors linking oral microbiome profiles to the presence of dental caries. This approach enables a detailed, population-specific analysis of microbial community structures and their temporal changes, offering insights into the co-evolution of humans and their microbiomes in response to long-term dietary and lifestyle shifts.

Integrating Machine Learning with GEMs: A Framework for Advancing Complex Metabolic Simulations

Merve Yarıcı[1], Muhammed Erkan Karabekmez[1,2,*]

[1] *Department of Bioengineering, Istanbul Medeniyet University, , Istanbul, Turkey*
[2]*Division of Engineering in Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Cambridge, Massachusetts, USA*

Presenting Author: merveyarici@gmail.com
** Corresponding author:* mkarabekmez@bwh.harvard.edu

Genome-scale metabolic models (GEMs) provide a framework for simulating cellular metabolism, enabling predictions of phenotypic outcomes and facilitating multi-omics data integration. These models are crucial for optimizing bioprocesses, studying microbial communities' metabolism, and identifying drug targets. However, the complexity of metabolic networks, especially in community-level models, creates computational challenges, particularly regarding run-time efficiency and model scalability. As metabolic networks expand, such as in community-level models, the demand for computational resources increases, limiting the practical application of these models. In this study, we propose a new framework that integrates machine learning (ML) with GEMs to address these issues, optimizing both predictive accuracy and computational run-time. Our approach uses ML algorithms to enhance flux prediction, model parameter optimization, and reduce computational burden, particularly for large-scale community GEMs. By training ML models on fluxomics and high-throughput omics data, we identify key reactions and interactions that contribute most to metabolic outputs. Additionally, we apply dimensionality reduction and feature selection to simplify model complexity, enabling faster simulations without compromising accuracy. This framework applies to complex community models, showing significant reductions in run-time while maintaining predictive accuracy compared to traditional constraint-based methods. Our findings show that this approach offers a scalable solution for simulating large and complex metabolic networks, making community-level modeling more practical. Overall, this study demonstrates that integrating ML with GEMs improves both the accuracy and efficiency of metabolic simulations. This advancement has the potential for more effective applications in metabolic engineering, precision medicine, and systems biology, especially for modeling complex metabolic systems. ***Keywords:*** *Genome-scale metabolic models, Machine Learning, Community metabolic models*

Identification of Novel Therapeutic Targets in Pancreatic Ductal Adenocarcinoma Through Integrated Transcriptomic Analysis and Molecular Docking Approaches

Berkan Tut[1], Ebru Sarsılmaz[2,*], Esra Büşra Işık[3]

[1]*Department of Bioengineering, Istanbul Medeniyet University, Istanbul, Turkey*
[2]*Beykoz Institute of Life Sciences and Biotechnology, Bezmialem Vakif University, Istanbul, Turkey*
[3]*Faculty of Health Sciences, The University of Manchester, Manchester, United Kingdom*

Presenting Author: berkantut8@gmail.com
**Corresponding author:* ebruu.sarsilmaz@gmail.com

Pancreatic ductal adenocarcinoma (PDAC) remains one of the deadliest cancers, characterized by its aggressive nature, poor prognosis with limited therapeutic options. This study aims to identify novel therapeutic targets by leveraging publicly available transcriptomic data from multiple GEO datasets (GSE16515, GSE15471, GSE32676, GSE71989). After merging we analyzed a total of 54,675 genes, comparing 114 tumor and 70 normal samples. Differential gene expression analysis using t-tests with Benjamini-Hochberg correction revealed 1073 differentially expressed genes (DEGs), of which 566 were upregulated and 507 were downregulated. Protein-Protein Interaction (PPI) were constructed via STRING and CytoHubba identified key subnetworks, highlighting the top 10 most altered genes. These key genes were further analyzed in the L1000CDS2 to reveal potential drug candidates. Focusing on the top three, molecular docking was performed using AutoDock Vina after predicting optimal binding sites via DogSiteScorer. 3D structures of target proteins and ligands were sourced from AlphaFold Protein Structure Database and DrugBank, respectively. Strong binding affinities were observed, suggesting promising therapeutic interventions for PDAC. These findings could pave the way for novel drug development, offering hope for improving PDAC treatment outcomes.

Identification of Potential SARS-CoV-2 Inhibitors Among Widely Used, Well-Tolerated Drugs Through Drug Repurposing and *In vitro* Approaches

Betül Oruçoğlu[1], İdil Çetin[2], Mehmet Topçul[2], Şeref Gül[1,*]

[1]*Beykoz Institute of Life Science and Biotechnology Institute, Bezmialem Vakif University, , 34540 Istanbul, Turkey*
[2]*Faculty of Science, Istanbul University, 34134 Istanbul, Turkey*

Presenting Author: betul.orucoglu@bezmialem.edu.tr
**Corresponding author:* seref.gul@bezmialem.edu.tr

COVID-19 disease caused by the SARS-CoV-2 virus, became a worldwide pandemic shortly after it was first detected in Wuhan, China in December 2019. According to recent reports, ~675 million people were infected and ~7 million lost their life. COVID-19 pandemic has negatively affected the economic and social life of almost all countries. Despite various strict precautions around the world to stop the spread of the disease, an increasing number of new cases have been reporting and hence the negative effect of the disease persists. SARS-CoV-2 genome consists of genes encoding two polyproteins (ORF1a and ORF1b), four structural proteins and auxiliary proteins. When the viral genome replicates, the polyproteins synthesized from ORFs are cut at appropriate sites by 3C-like protease (3CLpro) and papain-like protease (PLpro), which are also present in these polyproteins, releasing non-structural proteins and continuing the life cycle of the virus. These proteases are very crucial enzymes for the viral life cycle and represent a promising target for antiviral drug discovery due to the absence of similar proteases in humans. Since the drug discovery process is very time-consuming and expensive, drug repurposing of FDA-approved drugs against emergent diseases such as COVID-19 saves time and money. Within the scope of this project, it is aimed to find SARS-CoV-2 inhibitors that are effective at low doses from FDA-approved drugs in widespread use. With the help of computational studies, 27 FDA-approved drugs with high binding energy to the active site of 3CLpro that have potential to be inhibitors of this protease were identified. After analyzing the cytotoxic activity of these drugs using the xCELLigence system, their IC50 values were determined by testing with *in vitro* 3CLpro enzymatic activity. Dual combination trials were conducted to identify drugs with synergistic effects and thus high activity at lower doses. To assess the impact of these drugs on SARS-CoV-2 replication, non-infectious SARS-CoV-2 sub-genomic replicons containing luciferase reporter were used in lung (Calu-3) and intestinal (Caco-2) cell models. Among the tested compounds, amcinonide, eltrombopag, lumacaftor, candesartan, and nelfinavir demonstrated inhibitory activity against 3CLpro *in vitro* at micromolar concentrations. Notably, lumacaftor inhibited SARS-CoV-2 replication in Caco-2 and Calu-3 cells with IC50 values of 964 nM and 458 nM, respectively. Candesartan also showed inhibitory effects, with IC50 values of 714 nM in Caco-2 cells and 1.05 µM in Calu-3 cells. In single-drug trials, amcinonide, pimozide, eltrombopag, and nelfinavir exhibited IC50 values ranging from 2 to 5 µM in both lung and intestinal cells. Furthermore, dual combination experiments showed that amcinonide, pimozide, lumacaftor, and eltrombopag act as potent SARS-CoV-2 inhibitors at nanomolar levels when combined with candesartan at micromolar concentrations. This study identifies a select group of FDA-approved drugs—most notably

lumacaftor, candesartan, and nelfinavir—as potent inhibitors of SARS-CoV-2 3CLpro and viral replication *in vitro*, highlighting their strong potential for repurposing as antiviral therapeutics. Further *in vivo* studies in mice will pave the way for clinical trials, especially benefiting patients suffering from long COVID-19 symptoms.

Pan-cancer Analysis of non-Apoptotic Cell Death Mechanisms Using Machine Learning

Hakan Kahraman[1], Pınar PİR[1,*]

[1]*Department of Bioengineering, Gebze Technical University, 41400, Çayırova, KOCAELİ*

Presenting Author: *hakankahraman@gtu.edu.tr*
**Corresponding author:* *pinarpir@gtu.edu.tr*

This study focuses on the pan-cancer analysis of non-apoptotic cell death mechanisms specifically cuproptosis, ferroptosis, and pyroptosis and their potential implications for cancer diagnosis and prognosis. Cuproptosis, a copper-dependent mechanism, involves cellular oxidative stress induced by dysregulated copper homeostasis, while ferroptosis is iron-dependent and results from lipid peroxidation and oxidative damage to cell membranes. Pyroptosis is an inflammatory programmed cell death, triggered by inflammasome activation, releasing pro-inflammatory cytokines, and contributing to tumor immunity. The objective of this study is to explore the relationship between genes associated with these non-apoptotic cell death pathways and cancer progression by utilizing bioinformatics and machine learning approaches. Gene sets were retrieved from literature, and bulk RNA sequencing data for 15 cancer types from The Cancer Genome Atlas (TCGA) was analyzed. Machine learning algorithms classified patients into subgroups based on their gene expression patterns linked to these pathways. Our results highlights the use of the Stacking Classifier, an ensemble method that improves cancer type classification accuracy by combining the predictions of multiple algorithms. This method enhances overall classification performance by leveraging the strengths of different models. Additionally, a unique gene signature for these cell death pathways was developed through Cox regression analysis. This signature helps stratify patients into risk categories, allowing a deeper understanding of how these pathways affect survival across cancer types. The study also explores the clinical relevance of this gene signature, linking it to cancer prognosis and potentially offering insights for personalized treatment. The findings suggest that integrating non-apoptotic cell death mechanisms into cancer research could provide new avenues for understanding cancer biology and developing targeted therapies.

A Novel WES Analysis Workflow for Gene Hunting in Primary Ciliary Dyskinesia (PCD)

Elif Öz[1], Elif Arik Sever[1], Sercan Güloğlu[2], Ebru Arslan[3], Nur Konyalılar[2], Ela Erdem Eralp[3], Zeynep Seda Uyan[2], Pınar Ata[3], Cansın Saçkesen[2], Yasemin Gökdemir[3], Bülent Karadağ[3], Saniye Girit[4], Serçin Karahüseyinoğlu[2], Hasan Bayram[2], Osman Uğur Sezerman[1,*]

[1] Faculty of Arts and Science, Department of Biostatistics and Bioinformatics Acıbadem University, 34662 Istanbul, Turkey
[2] Department of Cellular and Molecular Medicine Koc University, 34450, Istanbul, Turkey
[3] Faculty of Medicine, Marmara University, 34854. Istanbul, Turkey
[4] Faculty of Medicine, Medeniyet University, 34700, Istanbul, Turkey

Presenting Author: ptbrn7vjl@mozmail.com
**Corresponding author: ugur.sezerman@acibadem.edu.tr*

Primary Ciliary Dyskinesia (PCD) is a rare genetic disorder characterised by impaired cilia function. Despite a low PICADAR score of 3, a team of experienced pediatric pulmonologists diagnosed the patient with PCD. The ciliary beating frequency was recorded at 6.48 ± 1.50 Hz. Whole-exome sequencing (WES) was performed using Illumina NovaSeq 6000 technology. Read quality was assessed using FASTQC, and adapter sequences were removed with Skewer. Alignment of the sequencing data to the hg19 reference genome was performed using the Burrows-Wheeler Aligner (BWA). SAM files were sorted and converted into BAM format using Sambamba, and duplicate reads were marked and filtered using Picard's MarkDuplicates tool. Base quality recalibration was conducted using the GATK BaseRecalibrator and ApplyBQSR tools. Variants were called using GATK HaplotypeCaller and annotated via Ensembl VEP. The pathogenicity of variants was assessed using InterVar following ACMG guidelines. Variant filtering focused on rare variants (MAF ≤ 0.01) present in GnomAD and 1000 Genomes Project databases. Variants associated with PCD were cross-referenced with the NCBI ClinVar database. Enrichment analysis was performed using EnrichR to explore potential involvement in PCD-related pathways.

Improving Efficiency Of The Grouping-Scoring-Modeling Framework Through Statistical Pre-Scoring In Transcriptomic Data Analysis

Maham Khokhar[1], Burcu Bakir-Gungor[2], Malik Yousef[3,*]

[1]*Department of Data Science, Social Sciences Institute, Abdullah Gul University, Kayseri, 38080, Turkey*
[2]*Department of Computer Engineering, Faculty of Engineering, Abdullah Gul University, Kayseri, 38080, Turkey*
[3]*Department of Information Systems, Galilee Digital Health Research Center Zefat Academic College, 13206, Zefat, Israel*

Presenting Author: mahamkhokhar96@gmail.com
**Corresponding author:* malik.yousef@gmail.com

This study enhances transcriptomic data analysis by introducing a Pre-Scoring component into the Grouping-Scoring-Modeling (G-S-M) framework. G-S-M integrates biological knowledge with machine learning by grouping genes based on biological relevance to identify key groups for classification. However, traditional G-S-M requires scoring thousands of gene groups, impacting speed and performance. We hypothesize that statistical pre-filtering will reduce computational demands and improve group selection without sacrificing accuracy. A Pre-Scoring component was integrated into G-S-M, using the limma (Linear Models for Microarray Data) package for differential expression analysis. The Pre-Scoring component applies empirical Bayes methods to prioritize gene groups based on adjusted p-values, reducing the number of groups needing scoring. Nine human gene expression datasets from the Gene Expression Omnibus (GEO) database were analyzed, focusing on disease-related groups. The top 20% of statistically significant groups were selected for further analysis, with redundant gene-disease associations eliminated to reduce data redundancy. Performance was evaluated using a Random Forest classifier with Monte Carlo cross-validation, assessing accuracy, sensitivity, specificity, and AUC. The Pre-Scoring G-S-M model demonstrated improved computational efficiency by reducing the number of gene groups needing intensive scoring. Across nine datasets, the enhanced model achieved comparable metrics to the standard G-S-M approach while using fewer features. For instance, in the GDS1962 dataset, Pre-Scoring G-S-M achieved 94% accuracy with an average of 17.7 genes, compared to the standard G-S-M's 92% accuracy with 81.1 genes, indicating improved precision in feature selection without losing accuracy. The Pre-Scoring component enhances G-S-M efficiency, enabling more focused data analysis. Findings suggest the enhanced framework may improve biomarker discovery and disease classification.

Teaching NGS to Engineers: Addressing Reproducibility Challenges in Bioinformatics

Rumeysa Aslıhan Ertürk[1], Mehmet Baysan[1,*]

[1] *Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey*

Presenting Author: *rumeysa.erturk@itu.edu.tr*
**Corresponding author:* *baysanm@itu.edu.tr*

Next-generation sequencing (NGS) technology has transformed genetics, enabling rapid decoding of genetic information. However, extensive analysis is required because of the complex and errorprone datasets created [1]. This demand usually leads to analytical tools that don't follow rigorous software engineering standards, which causes issues with reproducibility. Reproducibility is crucial in scientific study, yet it is often overlooked in bioinformatics [2]. Ensuring the reliability of analytical tools is crucial as the use of NGS in clinical and scientific applications increases. Training future researchers with strong expertise in computer engineering and genomics can help overcome these problems. To this end, we developed an elective undergraduate bioinformatics course for computer engineering students at Istanbul Technical University. The course covered genomics, statistical analysis, and Python programming, emphasizing reproducibility in NGS analysis. Students analyzed public data from a somatic sequencing benchmark study using 12 analytical pipelines, incorporating tools like BWA, Bowtie2, Mutect, SomaticSniper, and Strelka, and utilized COSAP (COmputational Sequence Analysis Pipeline) [3]. We have gathered 132 variant call files from 11 project groups and conducted a comparative analysis between these files, as well as against a high-confidence variant list. Evaluation of submissions revealed substantial heterogeneity in variant lists, even with consistent input data and procedures. In terms of accuracy, students using Linux-based systems performed better than those using Windows Subsystem for Linux (WSL). Additionally, students who used Docker containers had higher average F1 scores than those who manually installed applications. The study found that variant caller selection had the second-largest effect on pipeline performance after the operating system. While COSAP simplified process management, it was insufficient to provide complete replicability. These findings highlight the importance of using consistent computing practices to ensure NGS analysis accuracy. Including real-world projects in the curriculum helped students better understand the reproducibility problems in bioinformatics.

Discovery of Peptide-based Inhibitor Targeting the Human IL18:IL18Rα Complex

Zeynep Şevval Düvenci[1], Ahmet Can Timuçin[2], Emel Timuçin[1,*]

[1]*Faculty of Engineering and Natural Sciences, Department of Biostatistics and Bioinformatics Acibadem University, Istanbul 34752 Turkey*
[2]*Faculty of Engineering and Natural Sciences Department of Molecular Biology and Genetics Acibadem University, Istanbul 34752 Turkey*

Presenting Author: zeynep.duvenci@live.acibadem.edu.tr
**Corresponding author:* emel.timucin@acibadem.edu.tr

Uncontrolled release of IL18, proinflammatory cytokine that belongs to the IL1 family and plays crucial role in inflammation, is associated with severe inflammatory diseases and therefore therapeutic approaches that block IL18 activity are important for treatment of these diseases. To target the IL18:receptor complex to modulate IL18 activity, binding proteins of human and/or viral origin that bind to IL18 have emerged. In the study of proof of that concept, the efficacy and safety of biological molecules targeting IL18 have been demonstrated in a variety of diseases with high IL18 activity, and directly targeting IL18 has proven to be a valid strategy in clinical models. Despite these positive developments, studies on stopping IL18 activity with different and smaller molecules compared to biologics are still increasing. In that context, two small molecules that block a region on surface of IL18 were found in a recent study. However, to effectively inhibit the IL18:receptor protein-protein binding surface, which has a very large area, compounds that exceed typical drug sizes are needed. Peptide-based compounds can be tested as IL18 inhibitors due to their small intermolecular size with biologics. Peptides are particularly advantageous compared to small molecules, because they can more effectively cover the protein-protein interaction surface than do small molecules and they have relatively low production costs and can more easily penetrate into tissues than do biologics. However, peptides targeting the IL18:receptor complex have not been previously studied. At this point, the aim of this study was to discover peptide-based inhibitors targeting IL18 and its receptor complexes and to evaluate their in vitro efficacy.

Keywords: *interleukin-18, inflammation, peptide, protein-protein interactions, structure-based drug design*

A Machine Learning Model to Detect the Symptoms of Parkinson's Disease Detection Using Spiral and Wave Images

K. Rama Krishna[1], P. Sandeep[1], K. Santhosh Kumar[1], B. KoteswaraRao[1]

[1]*Velagapudi Ramakrishna Siddhartha Engineering College, Andhra Pradesh, India,*

Presenting Author: *sandeepp0603@gmail.com*
**Corresponding author:* *kramakrishna@vrsiddhartha.ac.in*

Parkinson's disease (PD), a progressive neurological illness characterized by motor symptoms such tremors, stiffness, and delayed movements, is frequently identified by clinical evaluations and physical examinations. The inability to diagnose Parkinson's disease (PD) early can postpone the initiation of treatment. Recent studies have shown that certain sketching patterns, especially those in spiral and wave shapes, can highlight mild Parkinson's-related motor deficits. In order to find early indications, this study uses machine learning to examine spiral and wave drawings made by people with and without Parkinson's disease. Spiral and wave drawings of PD patients and healthy controls are included in the dataset. The sharpness of these photos is enhanced using preprocessing methods including noise removal and normalization. As markers of motor disorders, characteristics such as drawing speed, stroke consistency, and amplitude fluctuation are retrieved. Convolutional neural networks (CNNs) are trained with these attributes to categorize drawings as either non-Parkinsonian or Parkinsonian. To guarantee dependability, the model's recall, accuracy, and precision are assessed. This method's excellent accuracy makes it a non-invasive, easily accessible tool for Parkinson's disease early screening, which may lead to an earlier diagnosis and better patient outcomes.

Keywords: *Machine Learning, Wave Images, Spiral Images, Image Processing, PD Disease*

The Role of Gut Microbiota in Neurodegenerative Disorders: Targeting the Symptoms of Alzheimer's Disease and Depression with Probiotics and Prebiotics

İsra MAVALDİ[1], Alper YILMAZ[1,*]

[1]*Department of Bioengineering, Yıldız Technical University, 34220 Istanbul, Turkey*

Presenting Author: isra.mavaldi@std.yildiz.edu.tr
**Corresponding Author: alyilmaz@yildiz.edu.tr*

According to the World Health Organization (2023), over 55 million people worldwide suffer from Alzheimer's disease, with more than 60% living in low- and middle-income countries. Annually, there are approximately 10 million new cases.[1] It is characterized as a neurodegenerative disorder; even the main reason behind Alzheimer's disease is unknown, but in all patients, beta-amyloid (Aβ) plaques are detected extracellularly. These plaques form outside brain cells and disrupt cellular function when β-secretase, rather than α-secretase, cleaves the transmembrane amyloid precursor protein (APP).[2] On the other hand, depression is thought to be a risk factor for Alzheimer's disease; in other words, behavioral and psychological symptoms that appear in depression, including mood disturbances, appetite changes, and sleep disruptions, often appear early in AD progression. Even though the main reason behind depression is still unknown, in all patients low serotonin levels were detected.[3], [4].

In this study, we aimed to reveal possible metabolites produced by gut microbiota through probiotics and prebiotics that can reduce amyloid plaque formation, increase the degradation of existing plaques, and alleviate depression symptoms by enhancing serotonin levels. We combined datasets containing information about inhibitors of beta-secretase, microbial metabolites produced either directly or as part of its metabolic activities, and metabolites that are decreased in depressed patients. These datasets were retrieved from various databases, including Brenda[5], GMMAD2[6], GutMGene[7], and a published research article[8]. In this study, we propose a novel strategy that combines probiotic and prebiotic interventions. We identify a specific probiotic capable of producing neuroprotective metabolites and a beta-secretase inhibitor. This approach could lead to valuable therapeutic breakthroughs, potentially transforming how we treat neurodegenerative diseases. All datasets are analyzed using RStudio.

Keywords: *Alzheimer's Disease, Depression, Gut Microbiota, Psychobiotics, Amyloid Plaques*

Biomedical Embedding Transitions for Enhanced Computational Drug Repositioning

Betül Erkantarcı[1], Gökhan Bakal[1,*]

[1]*Department of Computer Engineering, Faculty of Engineering, Abdullah Gul University*

Presenting Author: betul.erkantarci@agu.edu.tr
**Corresponding Author: gokhan.bakal@agu.edu.tr*

The high cost and lengthy timelines in traditional drug discovery underscore the critical need for innovative and efficient approaches. Extensive clinical trials, rigorous regulatory approvals, and substantial financial investments are the main challenges in bringing new drugs into the market. Conversely, computational methods offer a promising avenue for mitigating these challenges and accelerating drug discovery. Drug repositioning (repurposing existing drugs for new therapeutic indications) represents a particularly effective strategy. This approach leverages the already-established safety profiles of approved drugs, significantly reducing development time, costs, and risks associated with traditional drug discovery. This research utilizes a novel computational method for drug repositioning, employing a translational entity embedding-based neural network model. The model is trained on the comprehensive biomedical knowledge graph provided by the Semantic Medline Database, learning to effectively represent the complex relationships between biomedical entities such as drugs, diseases, genes, and proteins. The model's performance is rigorously validated using repoDB, a widely recognized gold-standard dataset for evaluating drug repositioning methods. Technically, the model learns to minimize the vector distance between semantically related entities. This distance-based approach allows for the prediction of potential drug-disease associations. Shorter distances between entity embeddings indicate a higher likelihood of a therapeutic relationship. The presented study offers a computationally efficient and powerful tool for exploring drug candidates, which expedites drug discovery and leads to new treatments for various diseases. For instance, the model yielded a potential treatment relation between Rolipram and Edema. The results show the model's potential to significantly accelerate the drug development process and improve the efficiency of drug repositioning efforts.

G-S-M: A Comprehensive Framework for Integrative Feature Selection in Omics Data Analysis and Beyond

Malik Yousef [1,*], Jens Allmer [2], Yasin İnal [3], and Burcu Bakir Gungor [3]

[1] *Department of Information Systems, Zefat Academic College, Zefat, Israel*
[2] *Medical Informatics and Bioinformatics, Institute for Measurement Engineering and Sensor Technology, Hochschule Ruhr West, University of Applied Sciences, Mülheim an der Ruhr, Germany*
[3] *Department of Computer Engineering, Abdullah Gul University, Kayseri, Türkiye*

Presenting Author: bakirburcu@gmail.com
**Corresponding Author: malik.yousef@gmail.com*

The treatment of human diseases is a major research question in many fields related to medicine. It has become clear that patient stratification is of utmost importance so that patients receive the best possible treatment. Bio/disease markers are critical to achieve stratification. Markers can come from many different sources such as genomics, transcriptomics, and proteomics. Establishing markers from such measurements often involves data analysis, machine learning, and feature selection. Traditional feature selection techniques often rely on the estimation of individual feature importance or significance by assigning a score to each feature, disregarding the inter-feature relationships. In contrast, the G-S-M (grouping scoring modeling) approach considers a group of features as a set that is organized based on prior knowledge. This approach takes into account the interdependence among features, providing a more meaningful evaluation of feature relevance and utility. Prior knowledge can encompass much compiled information such as microRNA-target interactions and protein-protein interactions. Here we present a new tool called G-S-M that presents the generalization of our previous works such as maTE, CogNet, and PriPath. The G-S-M tool combines machine learning and prior knowledge to group and score features based on their association with a binary-labeled target such as control and disease. This approach is unique in that computational and domain knowledge is utilized concurrently. Embedded feature selection, repeatedly employing machine learning during the selection process results in the identification of the most discriminative groups. Furthermore, the G-S-M tool allows for a more holistic understanding of the underlying mechanisms of a given system to be achieved through the combination of machine learning and prior domain knowledge, which can lead to new insights and discoveries. The implementation of the G-S-M workflow is freely available for download from our GitHub repository: https://github.com/malikyousef/The-G-S-M-Grouping-Scoring-Modeling-Approach . With this generalized approach we aim to make the feature selection approach available to a broader audience and hope it will be employed in medical practice. An example of such an approach is the TextNetTopics that is based on the G-S-M approach. TextNetTopics uses Latent Dirichlet Allocation (LDA) to detect topics of words, where those topics serve as groups. In the future, we aim to extend the approach to enable the incorporation of multiple lines of evidence for biomarker detection and patient stratification via combining multi-omics data.

Species Classification from Short Genomic Reads using Feedforward Neural Networks

Emre Özzeybek[1,2,3], Aybar C. Acar[3,*]

[1]*Izmir Biomedicine and Genome Center, Dokuz Eylul University, Izmir, Türkiye*
[2]*Department of Biomedicine and Health Technologies, Dokuz Eylul University, Izmir International Biomedicine and Genome Institute, Izmir, Türkiye*
[3]*Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, Ankara, Türkiye*

Presenting Author: emre.ozzeybek@gmail.com
**Corresponding Author: acacar@metu.edu.tr*

With the cost of Next Generation Sequencing technologies in decline, the need for fast and efficient classification of genomic findings has become of utmost importance. Due to the output length limitations of most Second Generation Sequencing techniques, it is important that we are able to classify short reads of DNA. In this research, we trained a basic Artificial Neural Network model, built using Keras, with three hidden layers on short reads(ranging from 50 to 500 base pairs) taken from two species' reference genomes. We selected Escherichia Coli and Saccharomyces Cerevisiae for their short and well-studied reference genomes. Their taxonomic difference makes them ideal candidates for ascertaining the viability of using neural networks for species classification using short read data. Our model yielded promising results classifying these short reads with classification accuracies ranging between 80% and 91% corresponding to differing hyperparameters and read lengths. Notably, accuracy increased with the read length, with the model reaching its highest performance on 500 base pair reads. In future applications, this model could be adapted to classify multiple species, even in genomic regions where comprehensive DNA barcode data is unavailable or limited. Additionally, by leveraging neural networks, this method has the flexibility to be applied to non-barcode regions of the genome, an area where traditional methods have limited application. Integration of more sophisticated deep learning methods can possibly improve the results.

A Structural and Modeling Study of *Gklip* Lipase for Enhanced Enzyme Activity

Yağmur Poyraz[1], Ahmet Tülek[2], Nurettin Paçal[2], F. İnci Özdemir[3], Gözde Şükür[3], Deniz Yıldırım[4], Şebnem Eşsiz[1,5,*]

[1]*Kadir Has University, Postgraduate Education Institute, Computational Sciences and Engineering, Istanbul, Türkiye.*
[2] *Postgraduate Education Institute, Department of Bioengineering and Sciences, Iğdır University, Iğdır, Türkiye.*
[3] *Faculty of Science, Department of Molecular Biology and Genetics, Gebze Technical University, , Kocaeli, Türkiye.*
[4] *Faculty of Ceyhan Engineering, Department of Chemical Engineering,Çukurova University, Adana, Türkiye*
[5]*Kadir Has University, Faculty of Natural Sciences and Engineering, Department of Molecular Biology and Genetics, Istanbul, Türkiye.*

Presenting Author: 20151709002@stu.khas.edu.tr
**Corresponding Author: sebnem.gokhan@khas.edu.tr*

Lipases are hydrolase class enzymes that catalyze the hydrolysis of fats. Particularly in biodiesel applications, lipases are used to obtain biodiesel from vegetable oils or animal fats by catalyzing the transesterification reaction of fats. However, there are some problems with the use of lipase in biodiesel production. Firstly, reaction efficiency is a significant problem, especially in transesterifying crude oils with high free fatty acid content. Also, lipases may not be fast enough in terms of reaction speed. Additionally, these enzymes are sensitive to environmental conditions; inappropriate temperature or pH conditions can negatively affect enzyme activity, while stability issues also exist. Thus, researchers are continuously exploring innovative methods and technologies to enhance the performance of lipases and improve process efficiency. Enzyme mutation and lipase immobilization are vital areas of focus in this endeavor. This study aims to improve our understanding of the structural and dynamic properties of *Geobacillus kaustophilus* thermophilic lipase and use mutations to improve its activity, substrate selectivity, and thermal stability. For finding the ideal mutations, the *Gklip* lipase was modeled by MODELLER, and docking studies were conducted to assess ligand selectivity of both wild-type and mutated structures with different-length carbon chain fatty acids. Then, 100 ns of molecular dynamic simulations were performed for all protein-ligand complexes. The last 30 ns of the simulations were used for MM/GBS analysis to calculate the binding free energy of fatty acid chains. Out of the 50 mutations, the top 5 were taken to experimental studies for testing the stability and activity changes in *Geobacillus kaustophilus* thermophilic lipase. This research was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant Number 123Z977. We extend our gratitude to TUBITAK for their support.

Topic Selection for Text Classification Using Ensemble Topic Modeling with Grouping, Scoring, and Modeling Approach

Daniel Voskergian[1], Burcu Bakir-Gungor[2], Malik Yousef[3,*]

[1]*Computer Engineering Department Al-Quds University, Jerusalem, Palestine*
[2]*Department of Computer Engineering, Abdullah Gul University, Kayseri, Turkey*
[3]*Information System Department Zefat Academic College, Zefat, Israel*

Presenting Author: burcu.gungor@agu.edu.tr
**Corresponding Author: malik.yousef@gmail.com*

TextNetTopic is a recently developed approach that performs text classification-based topics (a topic is a group of terms or words) extracted from an LDA Topic Modeling as features rather than individual words. Following this approach enables TextNetTopic to fulfill dimensionality reduction while preserving and embedding more thematic and semantic information into the text document representations. In this article, we introduced a novel approach, ENTM-TS, an advancement of TextNetTopics. ENTM-TS integrates multiple topic models using the Grouping, Scoring, and Modeling approach, thereby mitigating the performance variability introduced by employing individual topic modeling methods within TextNetTopics. We conducted our comprehensive evaluation utilizing two datasets: the Drug-Induced Liver Injury textual dataset from the CAMDA community and the WOS-5736 dataset. The experimental results show that the performance of ENTM-TS surpasses or aligns with the optimal outcomes obtained from individual topic models across the two datasets, establishing it as a robust and effective enhancement in text classification tasks.

Comparative in Silico Modelling of Two Novel Eurocin-like Antimicrobial Peptides from Ascomycota

Betul Gundogdu[1], Busel Ozcan[1], Yunus Emre Dilek[1], Gunseli Bayram Akcapinar[1,*]

[1]*Department of Medical Biotechnology, Institute of Health Sciences, Acibadem University*

Presenting Author: Betul.Gundogdu@live.acibadem.edu.tr
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

Defensins are small, cationic peptides with conserved CYS motif which confer stability and functional versatility. They are primarily recognized for their role in innate immune response and essentially all living organisms produce defensin-like peptides(DLPs) to protect themselves against pathogens and external threats. Defensins possess a diverse spectrum of biological activities, such as antimicrobial, antiviral, and antifungal properties. Importantly, DLPs also have the ability to target cancer cells. As a result, DLPs have emerged as a promising class of bioactive molecules for therapeutic applications.

Limited structural studies have offered insights into the many roles played by DLPs. However, given the pronounced structural conservation of these peptides, a deeper structure-function analysis is needed to elucidate their mechanisms. Therefore, computational comparative structural analysis of different DLPs may reveal unique structure-activity relationships.

In this study, previously uncharacterized fungal DLPs mined from NCBI database. Two candidates from Aspergillus udagawae and Hyaloscypha hepaticicola selected according to their structural and physicochemical properties. These proteins are expressed by our group in the methylotrophic yeast, Pichia pastoris and were shown to undergo pH-dependent oligomerization, which suggests a potential mechanism for their activity regulation(unpublished results). Monomeric DLPs were modelled using AlphaFold and I-TASSER to ensure accuracy and robustness, while oligomeric forms were modelled with AlphaFold Multimer to explore protein-protein interactions. ConSurf server was used to decipher the evolutionary conservation of key DLP motifs. Additionally, a comparative analysis of structural properties such as surface electrostatics&hydrophobicity identified potential residues involved in oligomerization. This study advances our understanding of DLP oligomerization, potentially guiding the development of novel therapeutics.

Histopathological Image Classification with Transfer Learning

Ümit Murat Akkaya[1], Habil Kalkan[1,*]

[1]Gebze Technical University, Faculty of Engineering, Computer Engineering Department, Gebze, Kocaeli, Türkiye

Presenting Author: umit.akkaya@gtu.edu.tr
**Corresponding Author: hkalkan@gtu.edu.tr*

Accurate and rapid classification of histopathological cancer images is essential for enhancing diagnostic support through computer-aided systems, particularly when only a limited dataset is available. This study explores a deep learning-based approach to cancer detection, leveraging the DenseNet121 architecture to classify tissue image patches as cancerous or normal. Using transfer learning, the model was trained on a dataset of 100 normal and 70 cancerous histopathological image patch samples, achieving improved accuracy despite data constraints. Transfer learning proved significantly more effective than training a model from scratch, with an accuracy of 0.91 compared to 0.76. Additionally, data augmentation techniques and the "fit one cycle" method were employed to further optimize learning and enhance model performance. This proposed approach demonstrates strong potential for reliable cancer detection in histopathological images, even with limited sample sizes.

Keywords: *Image Classification, Deep Learning, Convolutional Neural Networks, Transfer Learning*

Molecular dynamics tool as a powerful approach to prioritize the structural impacts of SCID- and OS-associated missense variants on RAG1

Aslı Kutlu[1], Sinem Fırtına[2], Begüm Işıkgil[3,*]

[1]*Molecular Biology and Genetics, Faculty of Engineering and Natural Science, Istinye University, Istanbul-Türkiye*
[2] *Medical Genetics, Medical School, Istanbul-Cerrahpaşa University, Istanbul-Türkiye*
[3] *Medical Biology and Genetics, Institute of Graduate Education, Istinye University, Istanbul-Türkiye*

Presenting Author: asli.kutlu@istinye.edu.tr
**Corresponding Author: begumisikgil@gmail.com*

The immune system is a defence mechanism of cells and organs that protect the body from foreign substances (pathogens). More than one element of this mechanism works in harmony with each other. The dysfunction of any elements in the system is defined as an 'immune disease'. Several genes play a role in the differentiation process that starts immediately after the formation of B and T cells, which are the most important elements of the system, such as RAG1 and RAG2 genes. Severe combined immunodeficiency and Omenn syndrome result from defects in the differentiation of B and T cells controlled by RAG1/RAG2. In this project, we aimed to investigate the structural effects of 37 different missense mutations on the RAG1 protein. We investigated whether the variants associated with SCID or OS clinics cause structural changes at different levels using molecular dynamics simulations. Out of 37 missense variants reported in RAG1, we run 100 ns classical MD simulations with 13 variants. Here, we concluded that SCID-related variants cause much more significant and sharp structural changes than OS-associated variants. We did not observe up to 100% structural loss and complete disintegration of the protein molecule in any of the variants we analysed in the trajectories collected over 100 ns. However, the effects of the variants were often more than local, affecting regions far from their positions. In particular, significant functional changes with large increases in flexibility values were found for residues in the heptamer and nanomer binding regions. The theoretical knowledge gained within the framework of this project can be used for the prioritisation of variants and the evaluation of their structural effects. The other major finding was that, although some of the mutations were similar to other mutations in the same clinical category, each mutation had its own unique structural change.

Keywords: *RAG1, Molecular Dynamics Simulations, Severe combined immunodeficiency, Omenn Syndrome*

Structural Insights into Obesity-Associated STEAP1B Variants Using Molecular Dynamics Simulations

Yunus Emre Dilek[1], Levise Tenay[2,3], Onur Emre Onat[2], Günseli Bayram Akçapinar[1,*]

[1] Department of Medical Biotechnology, Institute of Health Sciences, Acibadem University, Istanbul, Turkey
[2]Department of Molecular Biology, Institute of Life Sciences and Biotechnology, Bezmialem University, Turkey
[3] Department of Biotechnology, Institute of Health Sciences, Bezmialem University, Turkey

Presenting Author: yunusemredilek@hotmail.com
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

Obesity affects over half a billion individuals worldwide and is influenced by both genetic and environmental factors. In our previous study involving a Turkish cohort, we identified the Six-Transmembrane Epithelial Antigen Protein 1B (STEAP1B) gene as associated with obesity, uncovering seven mutations through family segregation analysis. The STEAP protein family, known for its roles in metal ion transport and redox reactions, is crucial in cellular metabolism and cancer progression. Given STEAP1B's structural similarities to STEAP1 and the limited understanding of its function and membrane assembly, we aimed to predict key molecular interactions and conformational changes related to its pathogenic role in obesity.

In this study, we applied structural bioinformatics and molecular dynamics (MD) simulations—the first time for STEAP1B—to investigate its structural and functional characteristics. Using AlphaFold-based modeling and MD simulations, we examined the wild-type (WT) STEAP1B protein and five missense variants. Analysis of MD trajectories for the apo form of membrane-embedded STEAP1B monomers revealed that variants with single-nucleotide polymorphisms (SNPs), Q163R and L257P exhibited slight structural alterations, suggesting potential functional implications in obesity. We further modeled and simulated the WT and these two variants in homotrimeric structures bound to heme and flavin adenine dinucleotide (FAD) ligands, providing initial insights into possible membrane-embedded assembly forms of STEAP1B. The results of our study provide new insights into the homotrimeric structure of STEAP1B and its obesity-associated variants, particularly Q163R and L257P.

Computational Insight into Molecular Mechanism that Enable Survival of Mycobacterium Tuberculosis: The MtrA Response Regulator Protein

Hacer Nur Eksi[1], Ozge Sensoy[1,2,*]

[1] *Graduate School of Engineering and Natural Sciences, Istanbul Medipol University*
[2]*Regenerative and Restorative Medicine Research Center (REMER), Research Institute for Health Sciences and Technologies (SABITA), Istanbul Medipol University, Istanbul, Turkey*

Presenting Author: hacer.eksi@std.medipol.edu.tr
**Corresponding Author: osensoy@medipol.edu.tr*

Tuberculosis (TB), which is caused by bacterium Mycobacterium tuberculosis (MT), remains a leading global health challenge. TB is splitted into two categories: the latent and active phase. In the former, bacterium stays in a non-replicative state within granulomas, which are associated with hypoxia, acidic pH, and limited carbon sources. The bacterium survive in these challenging microenvironments by utilizing two-component systems (TCS). Among them is the MtrB/MtrA, which is essential for acquiring intrinsic resistance. Stress factors are detected by membrane-bound histidine kinase, MtrB, which phosphorylates MtrA response regulator, thus activating it. Consequently, expression of several virulence-related genes, including those involved in cell growth and mycolic acid synthesis, are regulated. Recent studies have established MtrA as a crucial drug target; however, there has been no therapeutics that can be used to modulate the function of the protein. We are motivated by the urgent need for developing effective therapeutics that can combat TB. This, on the other hand, necessitates holistic mechanistic understading of molecular mechanism that enables survivial of the bacterium in challenging environments. To this end, we examined structure and dynamics of MtrA at acidic and physiological pH and demonstrated that both H143 and H187 act as pH sensors by accepting protons at acidic pH. Eventually, the opening of the N-domain with respect to the C-domain is regulated, thus assisting adaption to active-like conformation, which is necessary for binding DNA. We also studied the impact of acetylation, which occurs at the K110 residue, and showed that domain opening was prohibited. These findings, which are in correspondance with experimental studies, provide, for the first time, a platform for drug development studies that target MtrA in its dormant and active states, thus offering novel strategies for combating MT.

Centralized Evaluation of Variant Effect Prediction Tools Using Curated Benchmark Datasets

Busra Nur Darendeli Kiraz[1], Zeynep Betul Altundal[2], Rumeysa Erturk[2], Alper Yilmaz[3], Mehmet Baysan[2,*]

[1] *Yildiz Technical University, Department of Bioengineering, Istanbul, Turkey*
[2] *Istanbul Technical University, Department of Computer Engineering, Turkey*
[3] *Yildiz Technical University, Faculty of Arts and Sciences, Department of Molecular Biology and Genetics, Istanbul, Turkey*

Presenting Author: bndarendeli@gmail.com
**Corresponding Author: baysanm@itu.edu.tr*

The rapid development in Next-Generation Sequencing (NGS) technologies has made genomic profiles increasingly accessible. Despite the accelerated production of genetic data, associating the identified genetic variants with diseases remains a significant challenge. Numerous computational tools have been developed for variant effect prediction. The performance evaluations of existing computational tools are often hindered by concerns related to circularity and bias, which can undermine their reliability1. To enable unbiased assessments of these tools, benchmark datasets are required. However, there is currently no single platform offering easy access to datasets that include such benchmark comparisons. This study aims to facilitate access to these datasets by centralizing them into a unified platform. The datasets include clinically relevant and leading genome-wide variant databases such as ClinVar, HGMD, and ClinGen, as well as germline datasets compiled from UniProt, the CHARGE sequencing project, and VariBench. Additionally, datasets containing somatic variants associated with childhood cancers, as determined by expert consensus, were incorporated. By consolidating these curated datasets, we evaluated the performance of variant effect prediction tools using different approaches. In our analyses, BayesDel outperformed other predictors for somatic variants in general, while REVEL showed better performance for germline variants. The results suggest that selected benchmark dataset has a substantial impact on predictor performance and detailed evaluation of existing methods through heterogeneous benchmarks can significantly improve the performance of variant pathogenicity prediction.

HERACLOMICS: A User-Friendly Local Server App for scRNA-Seq Analysis

Doğukan Uğun[1,*]

[1]*Yildiz Technical University, Department of Bioengineering, Istanbul, Turkey*

Presenting Author: dogukan.ugun@std.yildiz.edu.tr
**Corresponding Author*

HERACLOMICS is a local server designed to empower users without R programming skills to perform comprehensive single-cell RNA-seq analysis. Featuring a tutorial-style interface, HERACLOMICS guides users through each stage of preprocessing, analysis, and visualization, with customizable quality control thresholds, adjustable dimensionality reduction, and clustering parameters supported by 2D and 3D UMAP and t-SNE visualizations. Flexible differential expression analysis between user-defined cell groups, coupled with gene set enrichment analysis, aids in identifying biologically impacted pathways. Multimodal data visualization is available for CITE-seq and 10X Genomics Multiome data, while a save-state function preserves progress to avoid repeating computational steps. HERACLOMICS also provides advanced tools for further analysis, such as cluster labeling based on marker genes, gene expression visualization across clusters, and identification of gene co-expression networks to reveal functional relationships. Gene regulatory networks are explored through transcription factors and target gene interactions, while trajectory analysis uncovers differentiation pathways using RNA velocity and pseudotime. Cell communication is examined through inferred ligand-receptor interactions, and differential expression analysis identifies genes unique to selected cell groups. These features make HERACLOMICS a powerful and accessible tool for single-cell RNA-seq analysis, enabling users to complete a full workflow within 1–2 hours.

An Innovative Approach for Distinguishing Tumor-Associated Macrophage Subtypes within Single-Cell RNA Sequencing Data

Duygu Keremitçi[1], Yasin Kaymaz[1,*]

[1]*Ege University, Faculty of Engineering, Bioengineering Department, Izmir, Türkiye*

Presenting Author: duygukeremitci@gmail.com
**Corresponding Author: yasinkaymaz@gmail.com*

Identifying cell subtypes in single-cell RNA sequencing data is challenging, especially for macrophages, due to transcriptomic oscillations and noise. This issue is crucial for tumor-associated macrophages (TAMs) in lung cancer, where distinguishing between pro-inflammatory M1 and pro-tumorigenic M2 states is essential. Using the lung cancer single-cell atlas (LuCA), which includes over 1.3 million cells from 318 donors across 29 studies, we developed a bioinformatics strategy to classify TAMs into seven subtypes: interferon-primed, immune regulatory, inflammatory cytokine-enriched, lipid-associated, pro-angiogenic, resident tissue macrophages, and proliferating TAMs. We calculated AUCell scores for each subtype in a pool of 179,689 macrophages, selecting the top 250 cells per subtype. Using these, we created a classifier, HieRFIT, to label the remaining macrophages using predictive models at branching points of cell differentiation trajectories, enabling further analysis in a larger cohort. Our main goal was to examine TAM subtype compositions across tissues and their associations with disease status. We found that lipid-associated TAMs significantly increased in advanced tumors compared to early-stage or healthy tissues ($p<0.01$), while RTMs decreased as the disease progressed. This trend persisted across UICC stages but lessened in metastatic tumors. We also identified genes uniquely expressed by lipid-associated TAMs, discovering upregulated genes such as PLA2G7, CCL18, APOE, GPNMB, and LIPA, along with novel markers including LGMN, PLTP, A2M, GM2A, PLD3, CTSZ, and CPM associated with lipid metabolism and phagocytosis. Our findings suggest that macrophages may convert to lipid-associated subtypes as tumor burden increases, or that these macrophages drive disease progression. This bioinformatics approach provides valuable insights into TAM roles in lung cancer and offers a framework for studying other immune cell types in single-cell data.

Prediction of Metabolite Biomarkers for Alzheimer's Disease Subtypes with an Innovative Algorithm Combining Genome-Scale Metabolic Models with Transcriptome Data

Kader Sarıbulak[1], Ecehan Abdik[1], Tunahan Çakır[1,*]

[1]*Department of Bioengineering, Gebze Technical University*

Presenting Author: kader.saribulak@gmail.com
**Corresponding Author: tcakir@gtu.edu.tr*

Alzheimer's Disease (AD) is the most common neurodegenerative disorder with cognitive decline and functional impairments that progress rapidly[1]. AD starts with subtle molecular changes in the brain years before clinical symptoms, therefore early diagnosis through novel biomarkers is essential for effective treatment[2-4]. Large cohort studies have revealed AD's heterogeneity, suggesting that studying metabolic and pathological subtypes may lead to a better characterization of molecular perturbations in the patients[5,6]. A recent study identified 5 different AD subtypes based on transcriptomic profiles[5]. Identifying biomarkers for these subtypes could enable early diagnosis and development of targeted treatments. In this study, the TAMBOOR[7] (TrAnscriptome-based Metabolite Biomarkers by On-Off Reactions) algorithm, a constraint-based modeling approach, was used for personalized and subtype-specific biomarker predictions for AD. TAMBOOR compares control and disease states in terms of the number of reactions that must be active to secrete each metabolite, given gene expression level changes. Previous research has demonstrated TAMBOOR's effectiveness in predicting biomarkers for Parkinson's disease[7]. RNA-seq data from the Religious Orders Study and Memory and Aging Project[8] (ROSMAP), including 396 AD and 164 healthy samples, was analysed in this study. Transcriptome data of each sample was mapped on a generic human genome-scale metabolic model (Human-GEM, v1.18)[9] with 2,889 genes, 8,456 metabolites, and 12,995 reactions to identify fold-changes of gene expression-based reaction scores. Then, the TAMBOOR algorithm was applied to predict potential biomarkers for each patient. The algorithm's predictive power was evaluated using a curated list of AD biomarkers. From these biomarkers, L-lactate, glutamate, and taurine appeared in one-third of the overall AD patients, but were predicted in over half of the patients when the samples were divided into subtypes.

Keywords: *Alzheimer's Disease, Genome-scale Metabolic Network, Transcriptome, Biomarker, Disease subtypes*

Acknowledgements: *This project is supported by TUSEB under Project No 2023-A3-02.*

Efficient Gene Expression Analysis through Simultaneous Scoring in Recursive Cluster Elimination

Nurten Bulut[1], Burcu Bakir-Gungor[1], Bahjat F. Qaqish[2], Malik Yousef[3,4,*]

[1] *Department of Computer Engineering, Abdullah Gül University, Kayseri, Türkiye.*
[2] *Department of Biostatistics, University of North Carolina at Chapel Hill, Chape Hill, NC, USA.*
[3] *Department of Information Systems, Zefat Academic College, Zefat, Israel*
[4] *Galilee Digital Health Research Center, Zefat Academic College, Zefat, Israel.*

Presenting Author: nurten.bulut@agu.edu.tr
**Corresponding Author: malik.yousef@gmail.com*

The analysis of gene expression data presents challenges due to limited sample sizes and high data complexity. Dimensionality reduction is essential to address these issues, and feature selection (FS) methods are commonly employed for this purpose. Support Vector Machines Recursive Cluster Elimination (SVM-RCE) is a technique designed for this purpose, utilizing K-means clustering to identify gene clusters.

In this method, the expression levels of genes within each cluster are included in the sub-data for that cluster, while the sample class labels are preserved. Subsequently, each cluster is assigned a score using internal cross-validation and SVM, and clusters with low scores are eliminated. This iterative process continues until a predefined number of clusters remains.
In this study, the existing method is refined to enable simultaneous scoring of cluster centers. Scores are calculated for each cluster by applying either a linear SVM or Random Forest (RF), with the absolute values of the coefficients from the linear SVM or the feature weights from the RF classifier used to assign scores.

The revised methodology was applied to 17 transcriptomic datasets from the Gene Expression Omnibus (GEO) database. Findings indicate that while the performance of this new method is comparable to the original SVM-RCE technique, it reduces computation time by approximately 80%. This improvement highlights the potential of refined computational strategies for the efficient analysis of high-dimensional biological data.

Keywords: *Feature Selection, Clustering, gene expression data analysis, transcriptomics*

Rapid Construction of Mutation-Centric Networks Leveraging Long-Range Interaction Data

Ramal Hüseynov[1], Burçak Otlu[1,*]

[1]*Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey*

Presenting Author: ramal.huseynov@metu.edu.tr
**Corresponding Author: burcako@metu.edu.tr*

Somatic mutations are essential for the transformation of normal cells into cancerous cells. Mutations can be categorized as driver mutations, which confer a growth advantage to cancer cells, and passenger mutations, which do not contribute to tumorigenesis but occur alongside driver mutations. To distinguish between driver and passenger mutations, we propose a novel graph-based approach that constructs a mutation-centric network leveraging long-range interaction data. Our method efficiently constructs this network by representing genomic intervals as nodes and their interactions as edges. By iteratively expanding the network from a seed mutation, we utilize long-range interactions and capture overlaps between these genomic intervals. This method employs positive and negative indexing for interacting intervals, with the seed mutation serving as the graph's root at index zero. Indices of overlapping intervals are stored at each index if there are any. This approach enables the quantification of a mutation's influence, the identification of complex interaction patterns such as the number of cycles in the graph, and the assessment of proximity to known driver genes or any gene set of interest. By providing a comprehensive view of a mutation's impact on the genomic landscape, we aim to improve the identification of driver mutations and advance our understanding of cancer biology.

Keywords: *Somatic mutations, Driver mutations, Network, Long-range interactions*

Comprehensive Evaluation of ACMG-based Variant Interpretation Tools

Zeynep Betul Altundal[1], Busra Nur Darendeli Kiraz[2], Rumeysa Aslihan Erturk[1], Mehmet Baysan[1,*]

[1]*Istanbul Technical University, Department of Computer Engineering, Istanbul, Turkey*
[2]*Yildiz Technical University, Department of Bioengineering, Istanbul, Turkey*

Presenting Author: zeynepbetulaltundal@gmail.com
**Corresponding Author: baysanm@itu.edu.tr*

Next-generation sequencing (NGS) technologies have revolutionized genetics by enabling the rapid generation of large-scale genetic data. However, the exponential increase in sequence data has introduced significant challenges in interpreting genetic variants, particularly in determining their pathogenicity. Accurate interpretation of genetic variants is essential for diagnosing genetic disorders and guiding clinical decisions.

The interpretation of genetic variants varied widely between organizations, causing inconsistencies in clinical practices. To address these inconsistencies, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) established guidelines in 2015. These guidelines classify genetic variants into five pathogenicity categories: Pathogenic, Likely Pathogenic, Uncertain Significance (VUS), Likely Benign, and Benign [1]. This pathogenicity assessment relies on 28 criteria, incorporating diverse data sources, including population data, computational predictions, functional studies, and segregation data.

Several computational tools have been developed to implement the ACMG/AMP guidelines for variant classification. This study evaluates these tools using benchmark datasets from reputable sources such as ClinVar, the ClinGen Evidence Repository, Oxford High Mobility Group (HMG), and pediatric cancer data from the New England Journal of Medicine (NEJM). The tools were assessed for concordance with expert classifications and their computational efficiency. This analysis provides an in-depth comparison of current ACMG-based variant interpretation tools across multiple datasets.

Evaluation of the Effects of Gentamicin on Multidrug Resistant *Escherichia coli* Strains Using Proteomic Approaches

Kübra Teksen[1], Melis Sardan Ekiz[2], Selinay Ozel[3], Ceylan Polat[4], Gulsah Merve Kilinc[5], Ergin Murat Altuner[1], İdil Yet[5,*]

[1] *Kastamonu University, Faculty of Science, Biology Department, Kastamonu, Türkiye*
[2] *Hacettepe University, Advanced Technologies Application and Research Center (HUNITEK), Ankara, Türkiye*
[3] *Hacettepe University, Faculty of Science, Department of Chemistry, Ankara, Türkiye*
[4] *Hacettepe University, Faculty of Medicine, Department of Medical Microbiology Ankara, Türkiye*
[5] *Hacettepe University, Institute of Health Sciences, Bioinformatics Department, Ankara, Türkiye*

Presenting Author: kubrateksenn@gmail.com
**Corresponding Author: idilyet@gmail.com*

The misuse of antimicrobials has rapidly spread antimicrobial resistance (AMR) in microorganisms, posing a global health threat to effective infection treatments. Understanding how microorganisms respond to antibiotics is crucial for developing new strategies and drug discovery. In this study, *E. coli* strains with different resistance profiles against gentamicin, which is an aminoglycoside antibiotic that is primarily used to treat serious infections caused by Gram- bacteria, were examined by MS-based proteomics methods to understand the changes induced by the antibiotic. Proteomic analysis of E. coli strains was performed using a UPLC connected to timsTOF-MS/MS system mass spectrometry followed by a data processing step using MaxQuant and Perseus. In total, 2080 proteins were identified and 340 proteins were found to be significant. The differentially expressed proteins were classified by Gene Ontology analysis based on their location, biological processes, and molecular functions. It was observed that the majority of these proteins play a key role in intracellular anatomical structures, including the cytoplasm, membrane and cytosol, which are crucial for maintaining cellular organization, facilitating transport, and enabling communication between cellular compartments. Notably, gentamicin was found to impact primary cellular processes, including the cell cycle, cell division, and metabolic regulation, particularly influencing protein synthesis and the cellular response to stress. This suggests gentamicin not only alters protein expression but may also disrupt essential cellular functions, potentially causing to cytotoxic effects. In terms of molecular functions, the analysis revealed that the differentially expressed proteins possess catalytic activity, highlighting their roles as enzymes that facilitate biochemical reactions within the cell. This implies that many of the affected proteins are vital for metabolic pathways and play a significant role in the overall cellular response to gentamicin. Furthermore, some proteins showed binding capabilities, including nucleic acid, ion, and lipid binding, highlighting their diverse functions in cellular processes.

Keywords: *Antimicrobial Resistance, Proteomics, E. coli, timsTOF-MS/MS, Bioinformatics*

Evaluation of the Effects of Ampicillin and Ceftazidime Antibiotics on Proteus mirabilis Using Metabolomic Approaches

Kübra Teksen[1], İdil Yet[2], Ergin Murat Altuner[1,*]

[1] *Kastamonu University, Faculty of Science, Biology Department, Kastamonu, Türkiye*
[2] *Hacettepe University, Institute of Health Sciences, Bioinformatics Department, Ankara, Türkiye*

Presenting Author: kubrateksenn@gmail.com
**Corresponding Author: ergin.murat.altuner@gmail.com*

Antibiotic resistance is a growing public health crisis that occurs when bacteria evolve and become resistant to the drugs designed to kill them. As a consequence, common infections can become harder to treat, leading to increased morbidity, mortality, and healthcare costs. Analyzing the metabolic activities of microorganisms is essential in combating antibiotic resistance. By understanding how bacteria adapt their metabolism to antibiotics, researchers can develop more effective treatment strategies, ultimately reducing the impact of resistant infections on public health. Proteus mirabilis is a critical opportunistic pathogen associated with urinary tract infections, especially in women. The rise of resistant strains complicates treatment and poses a serious challenge in clinical settings. This study aims to investigate the development of antibiotic resistance in P. mirabilis when exposed to sub-inhibitory concentrations of ampicillin and to explore whether such resistance leads to cross-resistance to other antibiotics. Using the disk diffusion method, resistant strains were generated through gradual exposure to sub-inhibitory ampicillin concentrations. It was found that ampicillin-resistant strains exhibited cross-resistance to ceftazidime. Subsequently, it was compared the metabolomic profiles of these resistant strains with sensitive control strains. For metabolic profiling, each sample was analyzed by GC-MS. The metabolomic data, supported by statistical and bioinformatic analyses, enhance our understanding of metabolic pathways, elucidate metabolic changes, and clarify interactions among metabolites. It was concluded that during the development of antibiotic resistance, P. mirabilis passages exhibited significant alterations in vitamin B6 metabolism, unsaturated fatty acid biosynthesis, and tyrosine metabolism pathways. These pathways interact to regulate vitamin B6 metabolism and support the survival and resistance of P. mirabilis in response to antibiotics.

Keywords: *Antimicrobial Resistance, Metabolomics, P. mirabilis, GC-MS, Bioinformatics*

Computational Investigation of Dynamics of beta-Arrestin-1 and beta-Arrestin-2 that are co-mutated in lung cancer: A patient-centric approach

Oktay Vosoughi[1], Özge Şensoy[1,*],Zeynep Cinviz[1],Barış Süzek[2,*]

[1] *Istanbul Medipol University, İstanbul, Türkiye*
[2] *Muğla Sıtkı Kocaman University, Muğla, Türkiye*

Presenting Author: oktay.vosoughi@gmail.com
**Corresponding Authosr: osensoy@medipol.edu.tr , barissuzek@mu.edu.tr*

G protein-coupled receptors (GPCRs) play an important role in cellular communication. Precise and timely regulation of GPCR-mediated signaling is critical for maintaining homeostasis in the cell. When an external signaling molecule binds to a GPCR, the message is conveyed to the cytoplasm by the G protein, whereas the termination of signaling pathway is accomplished by Arrestins, which have been shown to initiate alternative signaling pathways, besides their role in desensitization.

GPCRs are involved in many physiological/pathological processes, hence mutations in this family cause serious diseases like cancer. In addition to GPCRs, different Arrestin subtypes have also been implicated in modulation of pathways that are associated with cancer. Interestingly, the contribution of subtypes to the disease might be different. β-Arrestin-2 negatively regulates lung cancer progression; however, overexpression of β-Arrestin-1 causes a progressive disease in patients having EGFR inhibitor therapy. Apart from expression levels, co-mutation of different Arrestin subtypes in the same patient might also cause cancer. These findings suggest that simultaneous targeting of different Arrestin subtypes may provide effective outcomes in cancer treatments. This requires holistic understanding of the impact of mutations on the i) dynamics of Arrestin, and ii) emergence of potentially druggable binding pockets.

With this motivation, we followed a two-step approach. Accordingly, we searched for mutations pertaining to β- Arrestin-1 and β-Arrestin-2 seen in the same patient diagnosed with lung cancer using TCGA and COSMIC databases. We identified two mutations, namely, T224N and T268I in β-Arrestin-1 and β-Arrestin-2, respectively. We performed molecular dynamics simulations in both water and membrane environment using these mutants, and showed that the mutants displayed restricted dynamics compared to wild type protein, hence resembling an inactive state.

Make Live, Then Kill: A New Structural Biology Based Strategy Against Tuberculosis

Hacı Aslan Onur İşcil[1,2], Saliha Ece Acuner [1,2,*]

[1] *Istanbul Medeniyet University, Departmant of Bioengineering, Istanbul, Turkey*
[2] *Istanbul Medeniyet University, Science and Advanced Technologies Research Center (BILTAM), 34700, Istanbul, Turkey*

Presenting Author: iscil.onur98@hotmail.com
**Corresponding Author: ece.ozbabacan@medeniyet.edu.tr*

Tuberculosis, caused by the bacterium *Mycobacterium tuberculosis*, is a highly lethal disease, with approximately one-quarter of the world's population infected by this bacterium. When comparing infection and mortality rates, the reason high mortality is not observed among those infected is that the bacterium remains dormant in many individuals. Tuberculosis only exerts its deadly effects when the host's immune system weakens, at which point it transitions from a dormant to an active state. One of the key mechanisms controlling the bacterium's transition from dormancy to an active state is the breakdown of inorganic pyrophosphate by the bacterium's PPX2 enzyme. As with many bacteria, the control of inorganic pyrophosphate levels influences tuberculosis's metabolic reactions and drug resistance, thus regulating the disease's dormant-active cycle. When active, the PPX2 enzyme breaks down inorganic pyrophosphate; it becomes inactive in acidic conditions (pH 4-4.5), such as the macrophage environment where tuberculosis remains dormant.

In this study, the structural dynamics of the PPX2 enzyme were examined under different pH conditions using accelerated molecular dynamics simulations, revealing that the enzyme shifts to a closed form in acidic pH conditions. It is suggested that, in this form, the enzyme becomes inactive or dormant, leading to the accumulation of inorganic pyrophosphate in the cell. Understanding this mechanism at the molecular level could enable pathogens that lie dormant, such as tuberculosis, to be activated while the host's immune system is strong, allowing for antibiotic targeting. This approach could make it possible to kill the tuberculosis bacteria with the support of the immune system.

Mapping Functional Transcription Factor Binding Sites in Breast Cancer

F. B. Berber[1,2], G. Turan[2,3], G. Korkmaz[2,4,*]

[1] *Koç University, Graduate School of Science and Engineering, İstanbul, Türkiye*
[2] *KUTTAM Research Center for Translational Medicine, İstanbul, Türkiye*
[3] *Koç University, Graduate School of Health Sciences, İstanbul, Türkiye*
[4] *Koç University, School of Medicine, İstanbul, Türkiye*

Presenting Author: fberber20@ku.edu.tr
**Corresponding Author*

Breast cancer (BC), a leading cause of cancer-related deaths among women, presents significant heterogeneity, making it challenging to develop effective treatments. Understanding BC's molecular subtypes, particularly the luminal subtype, is essential given the diverse genomic alterations observed in these cancers. Key transcription factors (TFs) such as estrogen receptor alpha (ERα), FOXA1, and GATA3 are central to luminal BC progression. ERα, activated by estrogen, drives tumor growth, with alterations in ERα often leading to therapy resistance. TFs like FOXA1 and GATA3 play supportive roles, helping to maintain luminal cell identity and enhance ERα function, thus contributing to tumor development and progression.

This study aims to construct a detailed map of functional TF binding sites and their roles in BC, with a focus on altered genomic regions—particularly those with copy number gains on chromosome 1q, which have been implicated in BC progression and metastasis. To achieve this, we conducted ChIP-seq analysis to identify ERα, FOXA1, and GATA3 binding sites in luminal BC cell lines (MCF7, T47D, and ZR75-1), analyzing a total of 36 publicly available ChIP-seq samples and performing motif analysis to pinpoint direct binding sites of these TFs.

To further investigate the functional impact of these binding sites, we generated a CRISPR-based library using the CRISPR-DO algorithm, followed by functional genetic screening in MCF7 and T47D cell lines, along with a control triple-negative breast cancer (TNBC) cell line, MDA-MB-231, which lacks these TFs. By examining enriched and depleted functional sites identified through CRISPR screening, this research seeks to uncover molecular drivers of breast cancer. This study provides critical insights into the transcriptional regulation of breast cancer, setting a foundation for future research on targeted therapies based on these essential transcription factors.

Environment-dependent Modulation of Arrestin Dynamics and Activation

Zeynep Cinviz[1], Özge Şensoy[1], Giulia Morra[1,*]

[1]*Istanbul Medipol University, İstanbul, Türkiye*

Presenting Author: zeynep.cinviz@std.medipol.edu.tr
**Corresponding Author: giulia.morra@scitec.cnr.it*

The discovery that Arrestins, besides their roles as terminators of G protein-coupled receptor (GPCR)-mediated signaling, also activate certain signaling pathways has been a breakthrough, and promoted development of biased ligands that enable coupling of the receptor to a specific Arrestin subtype. Towards this end, either i) the orthosteric site or ii) allosteric regions like cytosolic or membrane-exposed regions of receptor are targeted, both of which has its own limitation. Hereby, we propose an alternative strategy, where we propose to stabilize a certain conformation of β-Arrestin, which is required for binding to a specific receptor, by means of small molecules. As our reference system, we chose Arrestin-biased ligand, carvedilol,-bound β2-adrenergic receptor (β2AR) in complex with β-Arr2 and aimed to stabilize conformation of Arrestin in that complex by small molecules. To enable observation of rearrangements associated with Arrestin activation, we performed coarse grained simulations using Martini force-field, and optimized parameters accordingly. Since Arrestin shuttles between cytosol and the membrane, we performed simulations in both environments to examine possible conformations. We clustered conformations adapted by β-Arr2 and identified possible binding pockets, in comparison to β-Arr1 to increase the possibility of finding molecules specific for β-Arr2. We show that β-Arr2 transiently samples active-like conformations both in cytosol and at the membrane, whereas b β-Arr1 requires interaction with the membrane. Interestingly, we demonstrate that β-Arr2 samples small domain-rotation angles without any need of C-tail detachment as long as the tail doesn`t interact with the lariat loop as opposed to what has been proposed in the field.

Investigating AhR (Aryl Hydrocarbon Receptor) Signaling in Cancer Cells: An Integrated Bioinformatics and Experimental Framework

Sema Duruoğlu[1,2], Bihter Muratoğlu[2,3], Cemil Can Eylem[4], Emirhan Nemutlu[4], Cansu Özdemir[2,3], İdil Yet[1,*]

[1]*Institute of Health Sciences, Department of Bioinformatics, Hacettepe University, Ankara, Türkiye*
[2]*Institute of Health Sciences, Department of Stem Cell Sciences, Hacettepe University, Ankara, Türkiye*
[3]*Center for Stem Cell Research and Development (PEDI-STEM), Hacettepe University, Ankara, Türkiye*
[4]*Faculty of Pharmacy, Department of Analytical Chemistry, Hacettepe University, Ankara, Türkiye*

Presenting Author: duruoglusema@gmail.com
**Corresponding Author: idilyet@gmail.com*

AhR signaling has both tumor promoting and suppressing roles in cancer cells. Given the complexity of cancer biology, examining AhR is crucial for contributing to cancer research. This study investigated the effects of AhR activation on cancer cell metabolism and gene expression. The effects of AhR activation on cancer cells were investigated using the HeLa cell line with short-term (24 hours) and long-term (120 hours) treatments. 6-Formylindolo(3,2-b)carbazole (FICZ) at 1 μM was used as an AhR agonist in all analyses. RT -qPCR assessed the effects of 120h AhR activation on genes CCNB1 (cell cycle/cancer), GLUT1 and GLUT4 (glucose), and CYP1A1 and CYP1B1 (xenobiotic). Metabolomic analysis was conducted using GC-MS, and metabolites were annotated using MS-DIAL. A fold change threshold of >1.2 and a p-value of <0.05 were used for statistical significance. 120h exposure to FICZ increased the expression of CYP1A1 and CYP1B1, which are downstream activators of the AhR signaling. The increase in CCNB1 expression following FICZ treatment suggests that this compound may promote cell cycle progression. Elevated GLUT1 and GLUT4 expression indicated metabolic modulation, possibly to meet higher energy demands. To validate, intracellular metabolomic analysis was conducted, which revealed different metabolite levels following FICZ treatment. At 24h, 13 metabolites increased and 16 decreased, with seven significant results. At 120h, 23 metabolites increased and 11 decreased, with three significant results. PCA revealed that the 24h and 120h treatments exhibited 85.9% and 94.5% variance, respectively, thereby indicating a separation between the FICZ and control groups. These results show that activation of AhR alters cell proliferation dynamics and gene expression related to xenobiotic metabolism and energy uptake, and highly impacts cellular metabolism. This study provides a basis for further research into AhR signaling as a potential therapeutic target in cancer.

MediVar: A Next-Generation Sequencing-Based Clinical Variant Analysis and Reporting System

Sinem Darwish[1], Basak Abak Masud[1,2], Ahmed H. Ibrahim[1,2], Busra Nur D. Kiraz[1,3], Muhsin Elmas[1], Baris E. Suzek[4], Akif Ayaz[1,*]

[1]*Istanbul Medipol University, Genetic Diseases Assessment Center, Istanbul, Turkey*
[2]*Mugla Sitki Kocman University, Graduate School Of Science And Engineering, Bioinformatics Graduate Program, Mugla, Turkey*
[3]*Yıldız Technical University, Department of Bioengineering, Istanbul, Turkey*
[4]*Mugla Sitki Kocman University, Faculty of Engineering, Department of Computer Engineering, Mugla, Turkey*

Presenting Author: siinemselvi@gmail.com
**Corresponding Author: akcayaz26@gmail.com*

Next-generation sequencing (NGS) is a significant diagnostic tool in clinics. Identifying disease-associated variants and not missing candidate variants is critical. Recently, we focused on this essential need and developed a variant analysis program for use in the clinic. MediVar is a system designed in collaboration with genetic experts, focused on expediting the NGS-based diagnosis, enabling clinicians to provide more accurate treatments in patient follow-up and treatment, and facilitating the generation of diagnostic reports.

MediVar leverages ReactJS for a dynamic, modern frontend, NodeJS for efficient backend operations, and Redis for fast in-memory caching. MySQL, supported by Bash and Python scripting, ensures robust and reliable data storage and management.

MediVar allows clinicians to enlist patient phenotypes using HPO's standard terminologies and easily view variants associated with phenotypes, filter variants in various criteria, choose reportable variants, and generate custom reports. MediVar follows ACMG guidelines for variant classification in analysis and reporting processes while leveraging clinical databases. MediVar has query and storage functions to enable data reuse to Access relevant variants, phenotypes, or diagnoses retrospectively. MediVar enables the examination of variant reads with ease through IGV integration.

MediVar has been used in the clinical diagnosis of over 10,000 patients until now, including approximately 3,700 WES, 2,400 targeted single-gene, and 7,300 genetic panel cases. MediVar reduced the analysis times per patient and improved the genetic diagnosis process. Furthermore, the retrospective genetic data in MediVar based on over 10,000 patients, provides increased accuracy and effectiveness in diagnosis. MediVar will continue to be updated based on clinicians' requests and technological advancements in the NGS market.

Somatic Mutation Clustering Patterns in Cancers Associated with UV Exposure

Beyza Kurtoğlu[1], Marat Kazanov[1,*]

[1]*Sabancı University, Istanbul, Turkey*

Presenting Author: beyza.kurtoglu@sabanciuniv.edu
**Corresponding Author: marat.kazanov@sabanciuniv.edu*

Ultraviolet (UV) radiation is a known carcinogen linked to the development of various skin cancers, including melanoma, basal cell carcinoma, and squamous cell carcinoma. UV exposure induces characteristic somatic mutations, predominantly C>T transitions at dipyrimidine sites. However, the extent to which UV-induced mutations form clusters remains largely unexplored. This knowledge could provide valuable insights into UV-induced mutational processes and reveal mechanisms underlying tumor progression and adaptation. In this study, we systematically characterized somatic mutation clusters across multiple UV-associated cancer genomes by applying bioinformatics approaches to high-throughput sequencing data. We examined the distribution, frequency, and functional impact of these clusters, identifying regions with significantly elevated mutation densities. We assessed the relationship between mutation clusters and genomic features, including gene density, chromatin accessibility, and replication timing. Our findings uncovered distinct patterns of mutation clustering in UV-driven cancers, offering new insights into the mutational landscape shaped by UV exposure.

A Comprehensive RNA-Seq Exploration: Variant Calling Analysis of Transcriptome Data in Sotos Syndrome

Beyza Kurtoğlu[1], Onur Serçinoğlu[1], Pınar Pir[1,*]

[1]*Gebze Technical University, Department of Bioengineering, Kocaeli, Türkiye*

Presenting Author: b.kurtogglu@gmail.com
**Corresponding Author: pinarpir@gtu.edu.tr*

Sotos syndrome is the most prevalent of the "overgrowth with intellectual disability disorders" with a diagnosis made in about one out of every 10,000 infants and NSD1 is the primary gene associated with Sotos syndrome.

Complexity of transcriptome is not fully explored by current analysis methods and valuable information remains hindered. To use the unexplored wealth of transcriptome, we developed a transcriptome-based variant calling pipeline and used it to analyse 10 patient samples. In the first part of pre-processing, quality control, adapter trimming and alignment were performed. Next, we continued to pre-processing by using Picard and performed variant calling analysis with GATK. After variant calling, VCF files were created for the ten individuals and a single VCF file was produced with shared variants across all patients using bcftools. Lastly, variant prioritization was done and variants were interpreted based on the VEP results, as well as their associated phenotypes, diseases, and traits found in ClinVar.

All the discovered variants were located on genes associated with inborn genetic diseases. The identified variant in BCL6 gene was previously submitted to the ClinVar as a height-associated variant. This may indicate a potential link between BCL6 gene and tall stature in Sotos syndrome. MED12L, NIBAN1 and CTSS genes were thoroughly investigated and a novel variant in CTSS gene was discovered. We identified rare variants, particularly focusing on those with MAF<0.01 and possible links to genetic diseases. Identifying such variants is crucial for better understanding the mechanisms and characteristics of Sotos syndrome. Further studies are planned to explore variants' impact on gene regulation and protein function.

Keywords: *Sotos syndrome, transcriptome-based variant calling, RNA sequencing, rare variants, inborn genetic diseases*

Computational Prediction of HBV Surface Antigen Structures: Insights for Vaccine Design and Therapeutic Discovery

Busel Ozcan[1], Betul Gundogdu[1], Gunseli Bayram Akcapinar[1,*]

[1]*Department of Medical Biotechnology, Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey*

Presenting Author: buselzcan@yahoo.com
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

Hepatitis, a form of liver damage prevalent worldwide, including in Turkey, can be caused by both viral and non-viral agents. A significant viral cause is the Hepatitis B virus (HBV), a double-stranded DNA virus from the Hepadnaviridae family. Despite the availability of safe hepatitis B vaccines for all age groups, cases of incomplete vaccine efficacy and immune escape variants still persist. Although vaccines based on the surface antigens of the Hepatitis B virus are widely available, the three-dimensional (3D) structures of these proteins are still lacking.

In this study, local, global, and multiple sequence alignment algorithms were applied for protein sequence analyses, prediction of consensus modeling candidates, and selection of immune escape variants. Homologous sequences for HBV S and M proteins were identified using BLASTp against the non-redundant protein database, and if available, structures were retrieved from the Protein Data Bank (PDB). These served as templates for 3D modeling of the Hepatitis B surface antigen (HBsAg). Candidate proteins were modeled using AlphaFold and I-TASSER. Protein sequence and structural features were calculated using various computational biology tools for both models. The Kyte-Doolittle scale was applied to assess protein hydrophobicity, and the Grand Average of Hydropathy (GRAVY) values were calculated. Additionally, the hydrophobic moments (surface polarities) of the proteins were determined using the 3D Hydrophobic Moment Vector Calculator (3D-HM). Surface electrostatics were calculated with PyMOL. A comparative analysis of the structures and structural features as well as quality metrics were performed to gain a deeper understanding of the immune escape variants.

These findings enhance our understanding of the structural properties of HBsAg and provide valuable insights into the mechanisms of immune escape in HBV, which may contribute to the development of more effective vaccines and therapeutic strategies.

Keywords: *Molecular Modeling, AlphaFold, S protein, M protein, Hepatitis B, vaccine*

Analyzing Replication Timing of Genes Across Different Cell Lines

Abdullah Tercan[1], Burçak Otlu[1,*]

[1]*Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey*

Presenting Author: burcako@metu.edu.tr
**Corresponding Author*

DNA replication, while essential for genetic inheritance, is also a significant source of mutations. This process is highly orchestrated, with different genomic regions replicating at specific times during the S phase of the cell cycle. The replication timing is intimately linked to mutation rates, both in germline and somatic cells. Certain mutational signatures, such as those associated with UV exposure and tobacco smoking, exhibit a preference for late-replicating regions, while others, like those linked to DNA repair deficiency and alcohol consumption, tend to occur in early- replicating regions. This specificity highlights the complex relationship between replication timing and the mutation rate. Replication timing of genomic regions can be used to explain the accumulation of mutations. In this study, we quantitatively analyze the replication timing of protein- coding genes across multiple cell lines using SigProfilerTopography. We assign higher scores to genes that replicate earlier. By investigating the mutation burden of these genes in various cancer types, we aim to develop a model that can explain the accumulation of mutations based on their replication timing preferences and the specific mutational signatures involved. This research has the potential to provide valuable insights into the mechanisms underlying mutagenesis. It can serve as a resource for future studies focusing on mutation rate, genome evolution, and cancer development.

Keywords: *DNA Replication Timing*

Integrating Mutation Analyses with Molecular Dynamics Simulations for Investigating Protein Structure and Dynamics

Oktay Vosoughi[1]; Özge Şensoy[1,*]; Zeynep Cinviz[1]; Barış Süzek[2,*]

[1] *Istanbul Medipol University, İstanbul, Türkiye*
[2] *Muğla Sıtkı Kocaman University, Muğla, Türkiye*

Presenting Author: oktay.vosoughi@gmail.com
**Corresponding Authors: osensoy@medipol.edu.tr, barissuzek@mu.edu.tr*

G protein-coupled receptors (GPCRs) play an important role in cellular communication. Precise and timely regulation of GPCR-mediated signaling is critical for maintaining homeostasis in the cell. When an external signaling molecule binds to a GPCR, the message is conveyed to the cytoplasm by the G protein, whereas the termination of signaling pathway is accomplished by Arrestins, which have been shown to initiate alternative signaling pathways, besides their role in desensitization.

GPCRs are involved in many physiological/pathological processes, hence mutations in this family cause serious diseases like cancer. In addition to GPCRs, different Arrestin subtypes have also been implicated in modulation of pathways that are associated with cancer. Interestingly, the contribution of subtypes to the disease might be different. beta-Arrestin-2 negatively regulates lung cancer progression; however, overexpression of beta-Arrestin-1 causes a progressive disease in patients having EGFR inhibitor therapy. Apart from expression levels, co-mutation of different Arrestin subtypes in the same patient might also cause cancer. These findings suggest that simultaneous targeting of different Arrestin subtypes may provide effective outcomes in cancer treatments. This requires holistic understanding of the impact of mutations on the i) dynamics of Arrestin, and ii) emergence of potentially druggable binding pockets.

With this motivation, we followed a two-step approach. Accordingly, we searched for mutations pertaining to beta-Arrestin-1 and beta-Arrestin-2 seen in the same patient diagnosed with lung cancer using TCGA and COSMIC databases. We identified two mutations, namely, T224N and T268I in beta-Arrestin-1 and beta-Arrestin-2, respectively. We performed molecular dynamics simulations in both water and membrane environment using these mutants, and showed that the mutants displayed restricted dynamics compared to wild type protein, hence resembling an inactive state.

Deciphering Sequence Variations and Splicing Sensitivity: Predictive Analysis of PSI in SRRM4 Response Groups

Ümit Sude Böler[1], Burçak Otlu[1,*]

[1]*Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey*

Presenting Author: sude.boler@metu.edu.tr
**Corresponding Author: burcako@metu.edu.tr*

Microexons, short exonic sequences between 3 and 42 nucleotides, are essential for regulating protein function, particularly in the nervous system. Their abnormal inclusion has been associated with diseases like autism and cancer. This study explores how sequence variations in the upstream intron, microexon variant, and downstream intron affect splicing efficiency under the regulation of SRRM4, a splicing factor crucial for neural tissue specificity. Leveraging data from a Massively Parallel Alternative Splicing Assay (MaPSy) with over 17,500 variants, we assessed splicing sensitivity by measuring Percent Spliced In (PSI) across four SRRM4 expression conditions. We examined key sequence features, including exon and intron length, UGC mutations, and splice site strength, using them to build predictive models for splicing outcomes. Our findings show that deep learning models, particularly those using Conv1D and LSTM layers, outperform traditional methods, with the best model explaining up to 99% of PSI variance. Motif analysis further revealed specific sequence motifs likely to influence PSI, adding to our understanding of SRRM4 responsiveness. In future work, we aim to expand motif discovery to identify significant k-mers across different SRRM4 conditions, refine LSTM models, and explore additional deep learning architectures. To address potential overfitting, we plan to perform rigorous hyperparameter tuning, apply cross-validation, and incorporate regularization techniques to improve model robustness. These combined efforts deepen our understanding of splicing regulation's molecular underpinnings and offer potential for therapeutic interventions in splicing- related neurological disorders.

Keywords: *Microexon, Massively Parallel Splicing assays*

Investigating the Impact of Pathogenic BMAL1 SNPs on the Human Circadian Clock Mechanism

Selahattin Aydoğan[1,#], Hursima İzgiş[1,#], Mahmut Çalışkan[1], Şeref Gül[2,*]

[1]*Department of Biology Biotechnology Division, Istanbul University, Istanbul, Türkiye;*
[2]*Institute of Life Sciences and Biotechnology, Bezmialem Vakif University, Istanbul, Türkiye*
[#]*Equal contribution*

Presenting Author: selahattin.aydogan@bezmialem.edu.tr
**Corresponding Author: seref.gul@bezmialem.edu.tr*

The circadian clock is an intrinsic biological system that establishes approximately 24-hour physiological and behavioral rhythms, regulating them in response to environmental cues. Core clock genes, including *Bmal1*, *Clock*, *Cry1/2*, and *Per1/2/3*, drive this mechanism. The transcription factors BMAL1 and CLOCK form a dimer that binds to E-box regions on promoters of clock-controlled genes, initiating the transcription of downstream effectors. As CRYs and PERs are synthesized, they dimerize and translocate to the nucleus, where they inhibit BMAL1/CLOCK-mediated transactivation, forming a feedback loop that regulates approximately 40% of cellular mRNA synthesis. While most clock genes have paralogs (e.g., *Cry1/Cry2* and *Per1/2/3*), *Bmal1* is unique in lacking a paralog, making it particularly susceptible to mutations that alter its function and potentially impair the circadian clock mechanism.

This study aims to examine the effects of single nucleotide polymorphisms (SNPs) in the *BMAL1* gene on the circadian clock in humans. From a pool of 3,000 SNPs in the Ensembl database, we identified 11 missense mutations classified as "pathogenic" by in silico prediction tools: Gly41Arg, Ala154Val, Arg166Gln, Arg216Gln, Arg218Trp, Arg238Gln, Ala357Thr, Val440Gly, Glu501Lys, Ser511Leu, and Ser513Leu. These mutations were introduced into *BMAL1* cDNA via site-directed mutagenesis.

Using a luciferase reporter system, we found that several mutations (Arg166Gln, Arg216Gln, Arg218Trp, Glu501Lys, and Ser513Leu) significantly altered BMAL1/CLOCK transactivation. Subsequent analysis of these SNPs in a repressor assay demonstrated that Arg218Trp and Glu501Lys mutations markedly affected CRY1's suppression of BMAL1/CLOCK transactivation. To assess the impact of these mutations on BMAL1 stability, a cycloheximide (CHX) chase assay was performed, revealing that Glu501Lys altered the protein's half-life. Additionally, immunoprecipitation assays indicated that Glu501Lys and Ser513Leu significantly affected the BMAL1-CLOCK interaction.

Based on these findings, we plan to conduct phenotypic analyses of the Arg218Trp, Glu501Lys, and Ser513Leu mutations using a real-time bioluminescence rescue assay in *Bmal1* knockout cells. This approach will help elucidate how these mutations affect circadian rhythm at the cellular level, offering insights into the mechanistic disruptions in the clock system due to specific *BMAL1* SNPs.

Metagenomic Analysis of Duodenal Bacterial and Fungal Microbiota in Obesity

Ilgım Durmuş[1], Ali Durmuş[2] , Günseli Bayram Akçapınar[1,*], Güldal Süyen[1]

[1]*Acibadem University, Istanbul, Turkiye*
[2]*Avrasya Hospital, Istanbul, Turkiye*

Presenting Author: ilgim.durmus@gmail.com
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

Background: Obesity is a preventable disease resulting from genetic, environmental, and lifestyle factors. While microbiota has been implicated as a contributor to obesity, most studies have focused on fecal samples, potentially overlooking the small intestine's role in food digestion and nutrient absorption. Additionally, these studies predominantly focus on bacteria, with fungi receiving less attention in the context of obesity. This study aims to examine the differences in bacterial and fungal microbiota composition in the duodenal tissue of obese and normal-weight individuals using metagenomics.

Methods: Five obese patients (BMI >35 kg/m^2) and five normal-weight control patients (BMI <25 kg/m^2) were included in this study. After (DNA) extraction from duodenal tissue samples, 16S rRNA gene sequencing was performed to identify bacteria, and ITS sequencing was used for fungi. Bioinformatic analysis was conducted using QIIME2 to assign taxonomic classifications to the bacterial and fungal sequences.

Results: In bacterial communities, the most abundant phyla were Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria. Streptococcus was the most common at the genus level, followed by Lactobacillus and Prevotella. For fungi, two phyla were identified: Ascomycota and Basidiomycota. Candida was more prevalent at the genus level, with species such as Candida glabrata, Candida dubliniensis, and Candida tropicalis observed exclusively in the control group. In contrast, Malassezia was observed only in obese patients.

Conclusion: This study revealed differences in the duodenal bacterial and fungal microbiota composition between obese and normal-weight individuals, highlighting the potential role of duodenal microbiota in obesity. These findings suggest that specific microbial taxa may be associated with obesity. Future studies with larger sample sizes and diverse populations are needed further to elucidate the relationship between obesity and gut microbiota.

Computational Investigation of the Role of PTM's and Mutations on Antibiotic Resistance in Mycobacterium Tuberculosis

İremnur Yalçın[1], Obaida Tayara[1], Hasnae Harmi[1], Özge Şensoy[1,*]

[1]*İstanbul Medipol University, Istanbul, Turkiye*

Presenting Author: iremnur.yalcin@std.medipol.edu.tr
**Corresponding Author: osensoy@medipol.edu.tr*

Antibiotic resistance in Mycobacterium tuberculosis (MTB) is a growing challenge, particularly due to resistance mechanisms within key regulatory proteins involved in the Tuberculosis pathology. This study investigates EmbR, a transcription factor critical in ethambutol resistance, by analyzing the effects of various mutations and post-translational modifications (PTMs) using molecular dynamics (MD) simulations. Our approach focuses on wild-type EmbR and modified forms, including phosphorylation at T57, T109, and T384, along with mutations P49A, R230W, and P243S. The simulation outcomes reveal distinct structural and functional changes: phosphorylation at T384 under hypoxic conditions enhances flexibility under hypoxic conditions, potentially increasing DNA-binding affinity, also T57 and T109 modifications contribute to localized stability shifts that may impact EmbR's regulatory role. P49A and R230W exhibit unique destabilizing effects, with P49A causing increased flexibility in critical binding regions and R230W leading to altered structural dynamics that impair protein stability. Whereas the P243S mutation, particularly in the bacterial transcriptional activation (BTA) domain, disrupts structural integrity and affects domain-specific stability, which could interfere with gene regulation in antibiotic resistance pathways. With help of RMSD, RMSF, and DCCM analyses, our findings reveal how the interplay of PTMs and mutations collectively shapes EmbR's role in resistance mechanisms, underscoring novel therapeutic targets for combating resistant MTB strains.

Keywords: *Mycobacterium tuberculosis, EmbR protein, ethambutol resistance, post-translational modifications, mutations, molecular dynamics simulations, structural analysis.*

Integration of Multi-scale Agent-based Models with Flux Balance Analysis

Burak Özyürek[1], Ecehan Abdik[1], Tunahan Çakır[1,*], Pınar Pir[1]

[1]*Department of Bioengineering, Gebze Technical University, 41400, Kocaeli, Turkiye*

Presenting Author: ozyurek_burak_41@hotmail.com
**Corresponding Author: tcakir@gmail.com*

Multiscale frameworks enable a deeper understanding of complex biological phenomena with heterogeneous behaviors, such as cancer, offering valuable insights for developing precision therapeutic strategies. Integrating Flux Balance Analysis (FBA) with Agent-Based Modeling (ABM) in computational cancer modeling has gained recent importance as it enables a more comprehensive simulation of tumor microenvironments, capturing both metabolic changes and cell-cell interactions. This approach provides critical insights into tumor growth dynamics, metabolic adaptations, and therapeutic responses, supporting the development of precision oncology strategies by modeling cancer's complex, heterogeneous behavior at both single-cellular and multiscale levels.

Agent-based models such as PhysiCell are capable of computing diffusion of chemicals and physical interactions between cells, but when it comes to intracellular metabolism, they assume that the agents representing the cells are metabolically inactive, i.e. they do not allow the representation of metabolic activities. Moreover, existing plugins developed to bridge this gap are not considered adequate for the representation of multiscale systems and do not offer the efficient performance to run large-scale models.

Considering this, we developed the Flux Balance Analysis (FBA) module for Physicell, a computational ABM development framework, using the GLPK library, a linear programming package. This module can solve linear optimization problems in very short times such as a few milliseconds thanks to the optimizations we perform at the code level. In this way, in addition to intercellular chemical and physical interactions, the intracellular metabolic response of agents to changes in the microenvironment can be simulated, and complex interactions that drive the heterogeneity in biological systems can be represented mathematically to unravel the dynamic behavior of the system in response to external stimulants.

Multi-omics Biomarker Analysis of Alzheimer's Disease

Cyrille Mesue Njume[1], Ali Cakmak[1,*]

[1]*Department of Molecular Biology, Genetics, and Biotechnology, Istanbul Technical University, 34469, Istanbul, Turkey*

Presenting Author: cyrillemesue@gmail.com
**Corresponding Author: ali.cakmak@itu.edu.tr*

Alzheimer's disease (AD), a primary cause of dementia, involves complex genetic, proteomic, and metabolic disruptions. This study investigates potential biomarkers and therapeutic targets through a multi-omics analysis of AD, integrating microRNA, metabolomics, and transcriptomics data from ROSMAP dataset and a GEO dataset (GSE53697). Our statistical and machine learning-based (ML) analysis identified key differentiating miRNAs, including hsa-miR-132-3p and hsa-miR-129-5p, which target neuroprotective and regulatory genes such as SIRT1, CDKN1, and FOXO1 (for hsa-miR-132-3p) and SOX4 and KEAP1 (for hsa-miR-129-5p). Additionally, hsa-miR-517c-3p and hsa-miR-519a-3p regulate PTK2B and PTEN, respectively, underscoring pathways linked to neuroinflammation and synaptic resilience in AD. Metabolomic analysis revealed myo-inositol as the top metabolite (which plays a crucial role in neuronal signaling), as well as N-acetylputrescine, homocarnosine, 7-hydroxy-cholesterol, and others. Transcriptomics data revealed enriched pathways related to neurotransmission, such as synaptic vesicle exocytosis, the synaptic vesicle cycle, and signal release from the synapse. Using SHAP, LIME, ML feature importance scores, and t-tests, we determined omics-specific feature importance by calculating a mean ranking across these methods. This approach identified key miRNAs, genes, and metabolites with robust classification power for distinguishing AD from control subjects. Clustering analyses (PCA, t-SNE, UMAP) enabled us to examine the differentiating power of the identified features across each omics layer (metabolomics, transcriptomics, and microRNA), while pathway enrichment further corroborated the involvement of these features in neurodegenerative pathways. Our integrative approach highlights specific miRNAs, genes, and metabolites as potential AD biomarkers and therapeutic targets, presenting a comprehensive molecular landscape of Alzheimer's disease.

Genetic Insights into Endometriosis and Adenomyosis in Populations of Turkish Ancestry

Nura Fitnat Topbas Selcuki[1,2], Feyza Nur Tuncer[3], Sevcan Aydın[3], Pinar Yalcin Bahat[4], Christian Becker[1], Engin Oral[5], Krina Zondervan[1,6], Nilufer Rahmioglu[1,6,*]

[1]*Oxford Endometriosis CaRe Centre, Nuffield Department of Women's and Reproductive Health, John Radcliffe Hospital, University of Oxford, Oxford, UK.*
[2]*Department of Obstetrics and Gynaecology, Istanbul Sisli Hamidiye Etfal Training and Research Hospital, University of Health Sciences Turkey, Istanbul, Türkiye.*
[3]*Department of Genetics, Aziz Sancar Institute of Experimental Medicine, Istanbul University, Istanbul,Türkiye.*
[4]*Department of Obstetrics and Gynaecology, Istanbul Kanuni Sultan Suleyman Training and Research Hospital, University of Health Sciences Turkey, Istanbul, Türkiye.*
[5] *Department of Obstetrics and Gynaecology, Biruni University Faculty of Medicine, Istanbul, Türkiye.*
[6] *Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.*

Presenting Author: nura.topbasselcuki@wrh.ox.ac.uk
**Corresponding Author: nilufer.rahmioglu@wrh.ox.ac.uk*

Introduction: Endometriosis and adenomyosis are distinct yet related gynecological conditions characterized by ectopic endometrial-like tissue. The genetic differences between these conditions remain largely unexplored, with most research focusing on endometriosis in European populations. Endometriosis GWAS has identified 42 significant loci, while emerging adenomyosis studies indicate condition-specific loci. This study explores the genetic architecture of adenomyosis in relation to endometriosis within Turkish ancestry populations.
Materials and Methods: Data were sourced from the COHERE Initiative in Northern Cyprus (7,646 women; 114 endometriosis cases and 670 controls) and the TROX study in Türkiye, comprising 241 endometriosis, 262 adenomyosis patients, and 236 controls. Genotyping used the Infinium Global Screening Array-24 BeadChip, with imputation via the TOPMed reference panel, and GWAS through SNPTEST.
Results: Two GWAS were conducted: one included 259 adenomyosis patients (88% pathologically confirmed) versus 871 controls, and the other included 345 endometriosis patients versus 871 controls. The adenomyosis GWAS revealed 24 nominally significant loci ($p<5x10^{-5}$), none overlapping with the 42 endometriosis loci. Top two SNPs ($p<5x10^{-7}$) were rs2745309, an eQTL for a novel transcript antisense to PAX7, involved in muscle stem cell proliferation and differentiation, and rs116908616, an eQTL for transcript RP11-666E17.1, expressed in both uterus and vagina. This is the first GWAS on adenomyosis and endometriosis in Turkish ancestry populations.
Conclusion: This study enhances understanding of the genetic differences between endometriosis and adenomyosis in Turkish populations, revealing unique genetic loci associated with each condition. By expanding the genetic focus to non-European populations, these findings contribute to a more inclusive genetic profile and may inform future research into personalized treatments and diagnostics for these conditions.
Keywords: *endometriosis, adenomyosis, genome-wide association study*

Evolutionary Insights into the CRY1 Gene and Sleep Phenotypes: Identifying Conserved Regions and SNP Hotspots across 392 Mammalian Species

Faruk Üstünel[1,2], Onur Emre Onat[1,2,*]

[1]*Department of Drug Discovery and Development, Institute of Health Sciences, Bezmialem Vakıf University, Istanbul, Türkiye*
[2]*Beykoz Institute of Life Sciences and Biotechnology, Bezmialem Vakıf University, İstanbul, Türkiye*

Presenting Author: faruk.ustunel@bezmialem.edu.tr
**Corresponding Author: onur.onat@bezmialem.edu.tr*

The CRY1 gene, a key component of the circadian clock, plays a vital role in regulating sleep and physiological rhythms across mammals. However, the genetic links between CRY1 and sleep phenotypes remain incompletely understood. We analyzed 4,211 positions within CRY1 across 392 mammalian genomes to investigate evolutionary conservation and SNP hotspots associated with diurnal (211 species), nocturnal (131 species), cathemeral (38 species), and crepuscular (12 species) sleep behaviors. Phenotypes were assigned through an extensive literature review and biodiversity database searches. Using Cactus alignments and calculating evolutionary conservation scores such as PhyloP and PhastCons, we identified 873 highly conserved positions with PhastCons scores above 0.9 across all phenotypes. Specific analyses revealed 538 positions with PhyloP scores above 2 in diurnal species and 1,075 such positions in nocturnal species, highlighting functionally relevant SNPs that may underlie phenotypic diversity. Machine learning approaches were then applied to detect associations between genotype and phenotype, uncovering SNPs in conserved regions potentially linked to behavioral adaptation. Our findings reveal SNP hotspots in conserved CRY1 regions, suggesting genetic variations that may drive differences in circadian behavior and sleep regulation. This research provides a foundation for further exploration of CRY1's role in sleep evolution and highlights specific targets for understanding sleep-related genetic diversity across mammalian species.

Keywords: *Circadian rhythms, Sleep genetics, Evolutionary biology, Mammalian genomics*

Reproducibility of Clinical Variant Detection in Next-Generation Sequencing Technologies

Emre Özzeybek[1,2,3], Mehmet Baysan[3,4,5], Uğur Özbek[1,3,*]

[1]*Izmir Biomedicine and Genome Center, Dokuz Eylul University, Izmir, Türkiye*
[2]*Department of Biomedicine and Health Technologies, Dokuz Eylul University, Izmir International Biomedicine and Genome Institute, Izmir, Türkiye*
[3]*Rare and Undiagnosed Platform (RUDIP), Izmir Biomedicine and Genome Center, Izmir, Türkiye*
[4]*Health Institutes of Türkiye, Istanbul, Türkiye*
[5]*Department of Computer Engineering, Istanbul Technical University, Istanbul, Türkiye*

Presenting Author: emre.ozzeybek@gmail.com
**Corresponding Author: ugur.ozbek@ibg.edu.tr*

Next-generation sequencing (NGS) technologies have revolutionized clinical genetics by allowing rapid and comprehensive analysis of genetic data, which is essential for diagnosing rare genetic diseases. Currently, in Türkiye, clinical genetic testing practices prioritize clinical exome sequencing (CES) and whole exome sequencing (WES) as their primary comprehensive clinical diagnostic tests, selected for their cost-effectiveness. As a broader test, WGS is expected to replace WES in practice as sequencing costs diminish. The reproducibility of clinical variant detection is crucial, considering the rapid adoption of sequencing technologies in disease diagnosis.

Utilizing the resources of the RareBoost project, which aims to advance rare disease research and innovation, we have reanalyzed raw data of previously undiagnosed rare disease patients. From these analyses, we identified six patients whose samples had previously undergone WES tests but needed resequencing which can be used to assess the reproducibility of detected variants. Four of these patients had raw data from outdated WES but definite causative variants could not be detected, therefore we performed WGS for further investigation. We repeated WES and compared variants for the remaining two patients since raw data was unavailable.

We observed a limited overlap between old and new analyses on exon regions. The documentation for the old samples was not detailed enough to support a detailed reproducibility assessment. Moreover, the differences between WES kits and the selection of different reference genomes further complicated the analyses. Our results demonstrate the importance of detailed record-keeping for clinical sequencing analyses to establish reproducibility for NGS-based diagnosis in rare diseases. These comparisons would also help to understand the marginal gain through WGS compared to WES.

This project was funded by the Horizon 2020 ERAChair Project RareBoost (Project ID:952346).

FASTQ or BAM: Alternative Replicate Combining Strategies for NGS

Emre Çamlıca[1], Mehmet Arif Ergün[1], Mehmet Baysan[1,*]

[1]*Istanbul Technical University, Istanbul, Turkiye*

Presenting Author: camlica21@itu.edu.tr
**Corresponding Author: baysanm@itu.edu.tr*

Next-generation sequencing (NGS) has revolutionized genomics by providing unprecedented throughput and accuracy for variant detection, making it a cornerstone of modern biomedical research. The ability to identify genetic variations quickly and reliably has significant implications for understanding diseases and advancing genomic studies. In order to benefit thoroughly from the NGS data, however, effective algorithms and computational resources are required[1]. Two common methods of merging technical replicates of same samples are; concatenating the raw reads in FASTQ files or combining separately processed analysis ready BAM files. GATK recommends against the former method by arguing that the base recalibration is done per lane (or library preparation) and concatenating raw reads prevents distinguishing these tags in technical replicates, therefore hampering the base recalibration step.

The aim of this study is to compare effect of technical replicate merging strategies on variant calling performance. For this purpose, we analyzed WGS samples with known ground truth variants from GIAB consortium. We used 32 FASTQ pairs of sample HG002 which are from two different lanes with sequencing depths varying from 0.66x to 8x. All scenarios are tested on BWA + HaplotypeCaller pipeline with Comparative Sequencing Analysis Platform (COSAP) [3]. Pipelines are run on UHeM, which is a high performance computing center, to tackle the high computational needs of WGS variant calling.

Our analysis highlights a significant distinction between merging strategies: pipelines where technical replicates were combined at the FASTQ stage outperformed those combined at the BAM stage. These findings underscore the importance of merging strategy in optimizing variant detection accuracy, particularly in scenarios involving multiple sequencing lanes or technical replicates.

A Comparative Study of Majiq, Fraser, and Dasper for Detecting Aberrant Splicing Events in Alzheimer's Disease Patients

Atakan Ünlü[1], Tunahan Cakir[1,*]

[1]*Department of Bioengineering, Gebze Technical University, 41400, Kocaeli, Turkiye*

Presenting Author: a.unlu2023@gtu.edu.tr
**Corresponding Author: tcakir@gmail.com*

Alternative splicing (AS) is a crucial mechanism in gene regulation and plays a significant role in the development and progression of complex diseases, including Alzheimer's Disease (AD), the most common neurodegenerative disorder worldwide. Detecting AS events is critical for elucidating the molecular mechanisms underlying AD. RNA sequencing (RNA-seq) data provide a comprehensive view of AS, enabling researchers to uncover splicing alterations associated with disease pathology. However, selecting the most suitable splicing analysis tools is vital, as variability in performance and output can significantly influence research outcomes. In this study, we followed a personalized approach to compare three AS detection tools—Majiq, Fraser, and Dasper—using RNA-seq data from 20 AD patients obtained from the Mount Sinai Brain Bank (MSBB) consortium. Each tool employs distinct methodologies for detecting AS events. Our comparison focused on runtime efficiency and the statistical and functional characteristics of identified genes reported by each tool. Our results show that the tools differ in capturing aberrant splice junctions in AD-associated genes and in runtime efficiency. Differences were also observed across the patients in terms of AD-associated genes within the same bioinformatic tool setting.

Keywords: *Alzheimer's Disease, alternative splicing, RNA-seq, aberrant splicing*

FoodProt: Specialized Protein Sequence Databases for Food Metaproteomics

Gül Sena Ada[1,#], Sanem Coşkun[2,#], Muzaffer Arıkan[3,*]

[1]*Department of Molecular Biotechnology, Faculty of Science, Turkish-German University, 34820, Istanbul, Türkiye*
[2]*Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Bursa Technical University, 16310, Bursa, Türkiye*
[3]*Biotechnology Division, Department of Biology, Faculty of Science, Istanbul University, 34134, Istanbul, Türkiye*
[#]*Equal contribution*

Presenting Author: senaada3@gmail.com
**Corresponding Author: muzafferarikan@istanbul.edu.tr*

Food microbiomes are complex ecosystems that influence the quality, safety and nutritional value of foods, thus playing a vital role in human health. In recent years, advances and the increasing use of meta-omics technologies have provided valuable opportunities to deepen understanding of the structure and function of food microbiomes. Among these methods, metaproteomics analyses provide in-depth information about the functional characteristics of food microbiomes, the roles proteins play in food systems, and microbial interactions. The protein identification in metaproteomics typically relies on a reference protein sequence database. Therefore, building a reference protein sequence database that is thorough yet manageable in size is essential for effective protein identification. As a result, there is an increasing demand for specialized protein sequence databases to enhance performance in metaproteomics research.

Here, we introduce FoodProt, a collection of specialized protein sequence databases tailored for food metaproteomics. FoodProt covers a diverse range of food categories, including fermented beverages, dairy, fish products and probiotics and currently offers ready-to-use reference protein sequence databases for 14 distinct food categories. These databases include a total of 5.8 million unique predicted protein sequences derived and curated from over 10,000 prokaryotic and eukaryotic genomes. The utility of FoodProt was demonstrated through the analysis of three food metaproteomics datasets from various food categories, as well as comparisons with other database construction strategies.

FoodProt serves as a comprehensive and valuable resource for food microbiome studies utilizing metaproteomics analyses.

Comparison of different Genome-Scale Metabolic Models of Pichia pastoris for overproduction strategy of drug precursor metabolites

Arda Örçen[1], Günseli Bayram Akçapınar[1]

[1]*Acıbadem Mehmet Ali Aydınlar Üniversitesi, Institute of Health Sciences, Medical Biotechnology*

Presenting Author: orcenarda@gmail.com
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

Pichia pastoris has emerged as a powerful host for metabolic engineering, attracting significant interest in the production of pharmaceutical precursor metabolites. Overproduction of Tyrosine, a vital amino acid serving as a main precursor to a diversity of pharmaceuticals synthesized through the phenylpropanoid pathway, is highly desired in pharmaceutical applications. To obtain optimal tyrosine overproduction, a logically tailored selection of base metabolic routes utilizing various models is necessary because Pichia pastoris lacks the complete phenylpropanoid pathway and the quality of the metabolic models currently available is unknown. This work examined and compared many Pichia pastoris genome-scale metabolic models (GSMMs). This study examines different genome-scale metabolic models (GSMMs) of Pichia pastoris to attain maximal tyrosine overproduction. We use MATLAB's "COBRA-Toolbox" to perform constraint-based modeling and examine the benefits and drawbacks of several GSMMs. Particular attention was given to how precisely the models represent tyrosine biosynthesis pathways. The models' capabilities to boost tyrosine production through metabolic flux redistribution, pathway improvements, and targeted genetic modifications were evaluated. Moreover, the models' capacity to adapt to increased tyrosine rates while minimizing the generation of byproducts was investigated in an attempt to provide recommendations for potential approaches to production scaling up in manufacturing environments. The comparative analysis identified the model that enables tyrosine overproduction in Pichia pastoris with minimal knockouts or genetic insertions, offering predictions and valuable insights on optimizing the phenylpropanoid pathway. These findings have significant implications for the biomanufacturing of pharmaceutical precursor metabolites, facilitating more sustainable and cost-effective production methods, and opening avenues for greener approaches inmanufacturing.

In Silico Characterization of HTR1D Missense Mutations in Obesity

Levise Tenay[1,2], Onur Emre Onat[1,2,*]

[1]*Bezmialem Vakif University, Institute of Health Sciences, Department of Biotechnology*
[2]*Bezmialem Vakif University, Institute of Life Sciences and Biotechnology, Department of Molecular Biology*

Presenting Author: levise.tenay@bezmialem.edu.tr
**Corresponding Author: onur.onat@bezmialem.edu.tr*

Obesity is a complex disease marked by substantial genetic heterogeneity. In a recent study, scientists demonstrated that the serotonin 2C receptor (HTR2C) is involved in the regulation of human appetite, weight, and behavior (He et al.,2022). In this study, we evaluated serotonin receptor variants in exome sequencing data from 3,853 individuals from Turkish families. Among these genes, we observed that variants in the 5-Hydroxytryptamine Receptor 1D (HTR1D) gene are associated with obesity. We prioritized 8 mutations based on allele frequencies and evolutionary conservation scores. HTR1D is one of the serotonin receptor subtypes that bind the neurotransmitter serotonin and belongs to the G protein-coupled receptor (GPCR) family, which plays a crucial role in transmitting signals related to mood, pain, and appetite. To assess the potential impact of mutations, we conducted molecular docking simulations to investigate how these mutations could affect serotonin (5-HT) binding affinity and receptor interactions. Computational analyses predicted structural and functional alterations by evaluating protein stability, flexibility, and dynamic behavior. Our findings revealed that two mutations, M92I and L307P, induced significant conformational alterations and reduced protein stability, while the R136G mutation enhanced serotonin binding affinity compared to the wild-type receptor. Upon analyzing the specific locations of these mutations, we discovered that the R136G and L307P mutations are located near the G-alpha protein binding site of the HTR1D receptor. Since HTR1D is coupled to GPCR proteins and mediates inhibitory neurotransmission by inhibiting adenylate cyclase activity, these mutations may directly impact receptor function and modify its role in neurotransmission. Given that serotonin has an established role in appetite control and energy metabolism, a mutation in this receptor could contribute to overeating or altered energy expenditure that promotes obesity.

Keywords: *HTR1D, missense mutations, obesity, serotonin receptor, in silico analysis.*

MetabOmics: Metabolism-Oriented Omics Data Integration

Aycan Şahin[1], Utku Sabri Kaya[1], Ali Çakmak[1,*]

[1]*Istanbul Technical University, Faculty of Computer and Informatics Engineering, Department of Computer Engineering*

Presenting Author: sahinay21@itu.edu.tr
**Corresponding Author: ali.cakmak@itu.edu.tr*

The integration of multi-omics data is crucial to understand the complexity of diseases and uncovering insights that single-omics approaches cannot reveal. The phenotype of diseases often reflects metabolic changes, with some pathways boosted and others reduced, collectively explaining disease etiology. In this paper, we introduce a comprehensive metabolism-oriented integrated multi-omics data analysis method which can accommodate omics datasets including genomics, transcriptomics, proteomics, and metabolomics. Our approach builds an integrated multi-omic network covering key biological interactions like expression, translation, and miRNA regulation. We map genes, proteins, and metabolites onto a network, quantify their alterations using fold-changes, and propagate these changes with information diffusion models to capture cascading effects that providing a holistic view of the biological interactions. After propagation, we update the bounds of metabolic reactions based on the propagated measurements, reflecting dynamic metabolic processes. For individual-specific analysis, we apply a personalized objective function for flux variability analysis. Finally, we apply an extended Metabolitics algorithm to compute reaction and pathway differential scores. These scores offer a detailed assessment of the metabolic reactions and pathways, highlighting key changes and offering insights into the underlying metabolic processes. We applied the proposed algorithm to five cancers: breast, kidney (stage III and IV), colon, pancreas, and prostate. Using K-fold cross-validation, we evaluated models based on metabolomics data (Metabolitics), transcriptomics data (deltaFBA), and our integrated approach (MetabOmics). The average F1 scores without integration were 0.781 and 0.713, while integration achieved 0.869. Our findings demonstrate that integrated analysis significantly enhances classification performance compared to using metabolomics or transcriptomics data alone.

Fibrosis focused miRNA profiling reveals a common change in hsa-miR-377-3p expression under several MSC priming conditions

Ali Berk Alişan[1] , Özge Burcu Şahan[1], Mustafa Keleş[1] , Ayşen Günel-Özcan[1,*]

[1]*Department of Stem Cell Sciences, Graduate School of Health Sciences , Center for Stem Cell Research and Development, Hacettepe University , Ankara, Türkiye*

Presenting Author: aliberk.alisan@hotmail.com
**Corresponding Author: aysen.ozcan@hacettepe.edu.tr*

Background: Fibrosis is characterized by excessive extracellular matrix deposition, leading to impaired tissue function in several chronic diseases. Mesenchymal stem cells (MSCs) can relieve fibrosis by secreting several factors, such as miRNAs. Priming strategies may empower MSCs' antifibrotic effects. In this study, we aimed to assess antifibrotic power of MSC priming strategies by using fibrosis focused miRNA expression profiles. For this aim, MSCs from two different source were induced with three diffent factors and their miRNA expression profile were compared.

Methods: Human bone marrow-derived MSCs (BM-MSCs) and umbilical cord-derived MSCs (UC-MSCs) were induced with melatonin (5 µM, 24h), TNF-α (10 ng/mL, 48h), or OSM (2 ng/mL, 24h). RNA was isolated, and qRT-PCR was performed using a 384-well fibrosis miRNA panel. Differential miRNA expression was compared between each primed group against control group propagated in the standard MSC medium. miRNAs were scored as fibrotic (-1) or anti-fibrotic (1), multiplied by their fold regulations, and an anti-fibrotic score was calculated.

Results: Analyses showed hsa-miR-377-3p was modulated by all three conditions. Fibrosis scores were -1.6 for melatonin, 8.51 for OSM, and 61.78 for TNF-α priming. Pathway analyses of six miRNAs commonly regulated by OSM and TNF-α highlighted the TGF-β signaling pathway in both KEGG and Reactome databases (p values 5.60E-05 and 1.80E-05, respectively).

Conclusion: Our findings underscore the anti-fibrotic potential of Melatonin, TNF-α, and OSM in MSC therapies, with hsa-miR-377-3p emerging as a key regulator. The fibrosis scoring system and pathway analysis may provide insights into fibrosis mechanisms, aiding in developing new anti-fibrotic strategies.

Biomarker Identification via Biological Domain Knowledge Based Feature Grouping and Scoring in Omics Data Analysis

Nur Sebnem Ersoz[1], Burcu Bakir-Gungor[2,*], Malik Yousef [3,4]

[1]*Department of Bioengineering, Graduate School of Engineering and Science, Abdullah Gul University, Kayseri, Turkey*
[2]*Department of Computer Engineering, Faculty of Engineering, Abdullah Gul University, Kayseri, Turkey*
[3]*Galilee Digital Health Research Center (GDH), Zefat Academic College, Israel*
[4]*Department of Information Systems, Zefat Academic College, Zefat, Israel*

Presenting Author: ersoz.nursebnem@gmail.com
**Corresponding Author: burcu.gungor@agu.edu.tr*

Numerous studies have utilized various computational feature selection (CFS) methods while analyzing omics datasets. These methods primarily rely on statistical techniques to individually rank the features and neglect the biological domain-knowledge. However, integrative gene selection approaches incorporate biological domain-knowledge from external resources during gene expression data analysis. These approaches generate ranked groups of genes based on both biological background information, such as gene ontology, pathways, and biological networks obtained from external resources. A recent survey noted that GO and KEGG resources are prominently employed as external knowledgebases for integrative gene selection.

Recently, our group introduced the Grouping–Scoring–Modeling (G-S-M) approach to select groups of features (i.e., genes) rather than evaluating features individually. The feature groups can be either generated by utilizing pre-existing domain knowledge stored in a biological database or through a fully data-driven approach using statistical measures. The G-S-M idea has been explored in the development of various computational tools, including CogNet, maTE, PriPath. These tools leverage external biological data from diverse sources such as KEGG pathways, miRNA-gene targets. Along this line, GeNetOntology (Ersoz, Bakir-Gungor, and Yousef 2023) utilizes Gene Ontology (GO) as external biological information to enhance its classification performance by selecting the most relevant genes from transcriptomic datasets. Hence, scientists can identify genes linked to diseases under study and discover biomarkers that can assist disease diagnosis and targeted treatment strategies. We have identified glioma-associated genes and gene ontology groups which can be used as biomarkers. Hence, we hope to enlighten the main molecular mechanisms behind glioma development and progression. These findings indicate potential of new drug targets and clinical early disease diagnosis.

Keywords: *biomarker identification, machine learning, biological domain-knowledge based feature selection, feature grouping, glioblastoma*.

pPromoter-FCGR: Deep Learning on Frequencey Choas Game Representation for Prediction of DNA Promoters

Gülbahar Merve Şılbır[1,*]

[1]*Department of Biostatistics and Medical Informatics (Bioinformatics), Karadeniz Technical University,Trabzon, Turkey*

Presenting Author: gmerve.cakmak@gmail.com
**Corresponding Author*

Promoters bind RNA polymerase to start transcription. Promoters regulate gene expression and protein production in specific cells, influencing genetic mutations and disease development. Understanding and identifying them is crucial for genetic regulation and gene therapy. The present study proposes the development of a novel model for the prediction of promoters and the generation of a feature vector comprising DNA sequences with an image representation. This model development study introduces a methodological innovation to the existing literature on this topic. The dataset employed in the study was derived from the E. coli K-12 genome sequence data, which was obtained from the RegulonDB database. The 100-1000 bp promoter DNA sequences were divided into 81 bp core sequences based on nucleosome and linker DNA properties. The 100-1000 bp DNA sequences from the dataset were divided into 81 bp core sequences according to nucleosome and linker DNA properties. Non-promoter DNA sequences were created from random 81 bp regions. Similar sequences were removed with CD-HIT. This process yielded 3382 promoter sequences and 3382 non-promoter sequences. This study used k-mer 5 and k-mer 6 images of both promoter and non-promoter sequences created with Frequency Chaos Game Representation (FCGR) in a promoter classifier model. To predict if DNA sequences created with FCGR are promoters, pre-trained models (Resnet50, VGG-16, InceptionV3) were used. Also, a deep learning model based on LeNet-5 was constructed. The performance of the developed models was evaluated using metrics including accuracy, sensitivity, specificity, MCC and AUC. This assessed the precision and reliability of the promoter prediction. The study revealed that the Le-Net5 model outperformed pre-training models, and the results were comparable to existing literature. The FCGR-based promoter prediction approach is recommended for further studies.

Keywords: Promoters, frequency chaos game representation, pre-training model, transfer learning, deep learning

Batch Ordering Significantly Affects scRNA-seq Data Integration Performance

Hilal Kazan[1], Aissa Houdjedj[1,*]

[1]*Antalya Bilim University, Antalya, Turkiye*

Presenting Author: hilal.kazan@antalya.edu.tr
**Corresponding Author: aissa.houdjedj@antalya.edu.tr*

The complexity and the scale of single-cell RNA-seq (scRNA-seq) experiments is growing rapidly. A current challenge is the integration of multiple scRNA-seq with the potential to provide new insights by leveraging the larger volume of data. An important step for achieving successful data integration is to correct for batch effects -- technical artifacts due to differences in equipment, sequencing technologies and capture times.

Several methods have been proposed for batch effect correction in scRNA-seq data integration. An important subset of the current methods often rely on pairwise batch integration and therefore require a batch ordering strategy to integrate more than two batches. Here, we perform comprehensive experiments with two popular methods Seurat (Butler et al, 2018) and Scanorama (Hie et al, 2019) on both simulated and real datasets with 13 diverse metrics to show that different batch orderings can affect the performance significantly. Interestingly, our analysis of both simulated and real datasets reveals multiple instances where prioritizing the integration of the most dissimilar batch pairs over the most similar ones enhances performance, challenging the conventional practice of ordering batches by increasing similarity. Finally, we propose a novel ordering strategy and show that it gives superior results compared to the default option for Seurat and Scanorama. Our findings underscore the need to reconsider batch ordering practices in scRNA-seq data integration and have implications for many of the current integration tools.

Molecular Contrastive Learning with Graph Attention Network (MoCL-GAT)

Alperen Dalkıran[1,*]

[1]*The University of Edinburgh, Edinburgh, UK*

Presenting Author: dalkiran@ceng.metu.edu.tr
**Corresponding Author*

This study presents a novel self-supervised molecular representation framework that leverages a Graph Attention Network (GAT) called Molecular Contrastive Learning with Graph Attention Network (MoCL-GAT) to improve drug-target interaction (DTI) and molecular property prediction. With many proteins and potential drug compounds lacking extensive bioactivity data, creating effective models for DTI becomes challenging, especially for novel or understudied proteins. Our approach addresses this limitation using self-supervised and transfer learning techniques, allowing the model to generate molecular graph embeddings without extensive labeled data. In MoCL-GAT, we generate transferable molecular graph embeddings using a dual self-supervised approach that combines local structural detail and global molecular descriptors. Specifically, we incorporate K-hop subgraph sampling and attribute masking to capture detailed local structures, while molecular descriptor prediction provides high-level molecular features, creating representations that are both detailed and generalizable (Figure 1). We used 1.9 million compounds obtained from the ChEMBL database to train the model. Using a scaffold-based split, we observed an AUROC of 0.934 on blood-brain barrier penetration (BBBP) and 0.749 on drug side effect (SIDER) prediction tasks, which outperforms all existing methods. MoCL-GAT also achieved an RMSE of 0.570 on aqueous solubility (ESOL) and 1.818 on the Free Solvation Database (FreeSolv) tasks, ranking as the top performer on these regression benchmarks. MoCL-GAT also obtained results comparable to the state-of-the-art for the remaining tasks. These results highlight the potential of integrating GATs with self-supervised and transfer learning, enabling improved molecular representation and prediction capabilities. By learning a generalized model on unlabeled compounds and then fine-tuning for DTI with limited bioactivity data, MoCL-GAT effectively addresses the scarcity of data.

Imputation of Dropout RNA Gene Expression Matrix Using ATAC Data

Ferda Abbasoğlu[1,2], Alper Yılmaz[3], Nizamettin Aydın[4,*]

[1]*Department of Computer Engineering, Yildiz Technical University, Istanbul, Türkiye*
[2] *Department of Computer Engineering, Gebze Technical University, Kocaeli, Türkiye*
[3] *Department of Molecular Biology And Genetics, Yildiz Technical University, Istanbul, Türkiye*
[4] *Department of Computer Engineering, Istanbul Technical University, Istanbul, Türkiye*

Presenting Author: ferdaabbasoglu@gtu.edu.tr
**Corresponding Author: nizamettin.aydin@gmail.com*

The gene expression matrix is obtained from RNA-Seq data. With the gene expression matrix, it is possible to identify which gene is expressed in which cell. As a result of the analysis using the expression matrix, biological results such as cell type identification, cluster detection, gene regulation analysis can be obtained. However, the gene expression matrix can contain a large number of zero values. As a result of the excess of zero values, the success of the analysis obtained from gene expression decreases. Zero values may indicate that the specified gene is not actually expressed in that cell, or they may be the result of misreading. Data that appears zero as a result of a misread is called dropout data. In order to fill the dropout data, imputation methods have been developed in the literature. These imputation methods usually use the RNA itself or another omics data. One of these methods is deep learning based autoencoders. In this study, both the autoencoder structure was utilized, and ATAC data, which is known to be less commonly used in imputation. For both RNA and ATAC data, the SNARE-Seq dataset was used, which contains values for genes from the same cell. RNA-Seq data was translated into gene expression matrix and ATAC-Seq data was translated into gene activity matrix using the information obtained from RNA. Common cells and common genes were identified and the number of features in the two matrices was determined as 811 and the number of samples as 8055. The gene expression matrix obtained from RNA-Seq was masked to a certain extent, as in the studies seen in the literature. A certain percentage of non-zero values were masked as 1 to measure how close the imputation result could get to the original value. The masking percentage is set between 10% and 50%. The matrix formed by combining the two matrices was trained with the help of the autoencoder model. The ARI between the imputed matrix and the missing-valued matrix was 0.37 and the NMI was 0.61.

Evaluation of PhysiCell Software for Construction of Multi-Scale, Agent Based Models of Biological Systems

Mehmet İzmirli[1], Pınar Pir[1,*]

[1]*Gebze Techical University, Department of Bioengineering, Kocaeli, Türkiye*

Presenting Author: m.izmirli2023@gtu.edu.tr
**Corresponding Author: pinarpir@gtu.edu.tr*

Agent-based modeling (ABM) represents individuals within a population as distinct agents in a continuous environment. By defining each agent's characteristics, and interactions, ABMs capture heterogeneity and emergent behaviors making them valuable tools for modeling diverse cell types and their interactions in different health and disease scenarios, such as development, evolution, infections, cancer, immune surveillance, and more. PhysiCell, an open-source software platform, enables detailed ABMs of multicellular systems, supporting simulations of cell cycle dynamics, cell-cell interactions, secretion and uptake of soluble factors, and other processes. We reviewed research articles that used PhysiCell for developing ABMs, using keywords like "physicell", "the model", and "the simulation" in the Google Scholar database, focusing on the articles citing the original PhysiCell paper by Ghaffarizadeh et al. Since PhysiCell's release in 2018, 33 studies have developed AMBs using this framework. 11 of these models used simple cell definitions, capturing the emergent behaviors from basic cell and environment interactions, while others built complex models with multiple cell types, signaling molecules, and intricate intracellular and extracellular dynamics. Of the 33 studies, 26 were related to cancer, while the rest explored other biological systems such as organism development, healing, SARS-CoV-2 infection, and more. Some researchers used ABMs alone for inferences and conclusions, while others combined ABMs with wet-lab experiments to verify findings, calibrate models, and gain additional insights with less time and resource investment.

Our work indicated that while ABMs are frequently used in tumor modeling, most biological systems remain underrepresented in this field. Due to the technical skills and computational power required, ABMs potential is hindered. We believe user-friendly tools and affordable cloud computing could enhance ABM accessibility and popularity.

Reconstruction of multi-scale mathematical models of tumors and their microenvironments

Burak Özyürek[1], Ecehan Abdik[1], Mehmet İzmirli[1], Barış Altunkara[1], Simge Şengül Babal[1], Cemal Yıldız[1], Emre Taylan Duman[1], Tunahan Çakır[1], Pınar Pir[1,*]

[1]*Gebze Techical University, Department of Bioengineering, Kocaeli, Türkiye*

Presenting Author: b.ozyurek2021@gtu.edu.tr
**Corresponding Author: pinarpir@gtu.edu.tr*

Cancer has an increasing global prevalence and has drastic economic and social effects. Discovery of cancer drugs is a long and expensive process, the pharmaceutical industry highlights the importance of mathematical modelling in accelerating the process and reducing the costs. Reconstructed models are time dependent, multi-dimensional, hybrid models to represent lung, liver, brain and lymph nodes. The extracellular matrix (ECM) is represented in continuum while the individual cells are represented as agents, hence each cell in the tissue has its own unique profile. Single cell RNA data (scRNA) from heathy and cancer tissues is used to identify the cell types in each tissue, the ligand-receptor pairs expressed by the cells that mediate intracellular signalling, the cell growth rates and the highly expressed enzymes of the active reactions in their metabolisms. The models are constructed in BioFVM / PhysiCell modelling and simulation environment, a new module is developed and integrated to the environment to be able to use flux balance analysis (FBA) in simulations. Cell type specific genome-scale metabolic models were constructed for each tissue and used in FBA to represent the metabolic activities of the cells. The models are then simulated to observe the effects of mutations that change the cell cycle rate, nutrient availability and immune profile on tumor growth. Both primary and metastatic tumor growth were simulated in lung, liver, brain and lymph nodes tissues. Our results demonstrated that ABMs provide an accurate representation of tumor dynamics and a flexible test environment for prognosis and treatment strategies in cancer.

Keywords: *tumor, metastasis, lung, liver, brain, scRNA, mathematical modelling, ABM*

A Novel Prediction Model for Breast Cancer Diagnosis Using Drug-Gene Interaction-based Pre-Existing Biological Knowledge

Ümmü Gülsüm Söylemez[1], Burcu Bakir-Gungor[2], Fei Zou[3], J.Stephen Marron[4],Malik Yousef[5,]*

[1]Department of Software Engineering, Faculty of Engineering, Muş Alparslan University, Muş, Turkey
[2]Department of Computer Engineering,Faculty of Engineering,Abdullah Gul University, Kayseri,Turkey
[3]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[4]Department of Statistics and Operations Research, University of North Carolina,Chapel Hill, NC, USA
[5]Department of Information Systems, Zefat Academic College, Zefat 13206, Israel

Presenting Author: og.uzut@alparslan.edu.tr
**Corresponding Author: malik.yousef@gmail.com*

A drug-gene interaction refers to the relationship between a drug and a genetic variation that may influence a patient's response to drug therapy. Understanding these interactions is crucial for personalized medicine, as it allows for the customization of treatment plans based on an individual's genetic makeup. Previous research has introduced various machine learning models to analyze this interaction based on biological knowledge, highlighting the potential of these methods to improve drug efficacy and reduce adverse effects.

Breast cancer is the second most common cancer globally and the leading cause of cancer-related deaths among women. Diagnosing breast cancer and interpreting test results require expert human knowledge, making it a complex process. In order to assist these tasks, recently machine learning techniques which offer promising avenues for enhancing diagnostic accuracy and treatment plans, have been developed. Along this line, here we aim to develop a novel prediction algorithm based on the Grouping-Scoring-Modeling (G-S-M) methodology. Our approach comprises the following three main components.

The first component (i.e., grouping) involves utilizing pre-existing biological knowledge about drug-gene interactions. In this phase, each drug is associated with specific interacting partner genes, leading to the formation of drug groups. This step is crucial for categorizing the complex interactions in a structured manner that can be analyzed effectively.

The second component (i.e., scoring) involves evaluating each group based on their ability to accurately distinguish between breast cancer and non-breast cancer cases. We applied statistical techniques to assess the performance of each group, ranking them according to their scores. This ranking helps to prioritize the most relevant drug-gene interactions for further analysis.

The final component (i.e., modeling) employs a random forest model to classify the data based on the rankings obtained from the scoring phase. Random forest models are known for their robustness

and their ability to handle complex datasets, making them ideal for this problem. By integrating the rankings into the model, we aim to enhance its predictive accuracy and reliability.

We assessed the performance of our method using various metrics, including accuracy, precision, recall, and F1-score. The results indicate that our novel approach demonstrates significant success across all performance measures, outperforming traditional methods. Our experimental findings suggest that the developed G-S-M methodology could be a valuable tool for early detection and treatment planning of breast cancer, ultimately improving patient outcomes.

Deciphering Transcriptional Landscapes of Mild and Advanced MASLD Fibrosis: Insights into Differential Gene Expression, Pathway Enrichment, and Drug Repurposing Opportunities

Ece Janet Dinç[1], Zeynep Özbiltekin[1], R. Useyma Gülhan[1], Arda Baran Nişan[1], Esra Nalbat[1], Deniz Cansen Kahraman[1,*]

[1]*Cancer Systems Biology Laboratory, Department of Health Informatics, Graduate School of Informatics, METU, Ankara 06800, Turkiye*

Presenting Author: zeynepozbiltekin@outlook.com
**Corresponding Author: cansen@metu.edu.tr*

Metabolic-associated steatotic liver disease (MASLD), previously termed non-alcoholic fatty liver disease (NAFLD), encompasses conditions from simple steatosis to steatohepatitis (MASH) and is a leading cause of liver fibrosis. Despite progress in treatments targeting metabolic dysfunction, drugs addressing liver inflammation remain urgently needed. Here we identified pivotal genes and pathways between mild and advanced stages of liver fibrosis in patients with MASLD and elucidated specific target(s) and drugs for repurposing. Four datasets obtained from GEO (GSE193066, GSE193080, GSE49541, GSE135251) were analyzed using the GEO2R tool to determine the differentially expressed genes (DEGs) between mild (F0-1) (n=209) and advanced (F3-F4) (n=166) fibrosis. 78 DEGs were identified between the two groups ($p \leq 0.05$). Heatmap was created using Pandas, Matplotlib, and Seaborn Python libraries and clustering analysis demonstrated high similarity across all datasets. Enrichr was used to explore pathways/ontologies associated with these DEGs based on KEGG 2021 Human and GO Cellular Component 2021, respectively. ECM-receptor interaction, focal adhesion, cell adhesion molecules, and PI3K/Akt signaling were enriched in advanced MASLD fibrosis. Protein–protein interaction (PPI) network construction using STRING revealed that COL1A1, PDGFRA, LUM, EPCAM, LAMA2, SOX9, and THBS2 were central genes/proteins implicated in disease progression. Drug Repurposing Hub and DGI database were used to identify potential agents against the top seven targets. COL1A1 and LAMA2, targeted by ocriplasmin (a proteolytic enzyme for treating symptomatic vitreomacular adhesion), and PDGFRA, targeted by TAK-593 (tyrosine kinase inhibitor), stand out as promising candidates for novel therapeutic approaches in MASLD. Our study provides valuable insights into the molecular mechanisms underlying MASLD fibrosis progression and highlights promising therapeutic targets and repurposable drugs for its treatment.

Investigation of the Role of Neuregulin3b in Synaptic Signaling and Brain Development in Zebrafish

Elham Tarahomi[1,*], Nuray Söğünmez Erdoğan

[1]Kadir Has University, Istanbul, Turkiye

Presenting Author: elham.tarahomi@stu.khas.edu.tr
**Corresponding Author*

SCO-spondin (Sspo) is essential for central nervous system development, neural growth, repair, and synaptic formation, with strong hippocampal and cortical expression. While its role in neuronal development and the extracellular matrix is partly known, its interaction network, including neuregulin signaling, is under investigation. Neuregulin3b (Nrg3b), an understudied protein, may play a key role in these processes, using zebrafish models, multi modal sequencing, and in situ hybridization (ISH) to identify interactions and expression patterns. A multi-modal approach analyzed available whole-genome, -exome, and RNA-Seq datasets from wild-type and Sspo-hypomorphic mutants (MatZyg and Het) across 5dpf, 15dpf, and 30dpf to detect genetic variations and their impact on temporal RNA expression. Principal component and differential gene expression (DEG) analysis were used for temporal expression profiling. The DEGs were evaluated in network and enrichment analyses to determine the biological functions. In-situ hybridization (ISH) and whole-mount in-situ hybridization (WISH) validated the findings. We identified 16 variants coinciding with the age showing the highest variation in PC1. Temporal co-expression analysis revealed Nrg3b, involved in synaptic signaling pathways, including synapse organization and synaptic transmission, is central to the interaction network. Comparative analysis found Nrg3b expression in MatZyg ~2-folds higher than the Het (Cohen's d: 0.832 (5dpf), 0.621 (15dpf), 0.499 (30dpf)) emphasizing its early-stage role when Sspo is low. WISH on 5dpf zebrafish embryos confirmed brain-specific Nrg3b expression. The analysis identified Nrg3b as a key gene in synaptic signaling and brain development, with higher expression in Sspo MatZyg during early development, highlighting its role under low Sspo conditions. These findings highlight the need for cell-ablation and regeneration studies to explore Nrg3b's role in neurodevelopment and plasticity.

Keywords: *Neuregulin3b, SCO-spondin, synaptic signaling, zebrafish model, multi-modal sequencing, genetic variations, neurodevelopment*.

Computational Investigation of Peptide Modulation of pMHC Dynamics

Evren Atak[1], Onur Serçinoğlu[1,*]

[1] *Department of Bioengineering, Faculty of Engineering, Gebze Technical University, Gebze, Türkiye*

Presenting Author: e.atak2022@gtu.edu.tr
**Corresponding Author: osercinoglu@gtu.edu.tr*

The Major Histocompatibility Complex (MHC) in vertebrates plays a crucial role in modulating adaptive immune responses. The MHC typically binds short peptides derived from either self or foreign proteins, and present these as antigens to the T-cell receptors (TCRs) of Antigen Presenting Cells (APCs). This interaction is a major determinant of the outcome of the cytotoxic T cells response. One often-overlooked aspect of TCR-pMHC interaction is the structural dynamics of the interacting complexes. In particular, there is a need to better understand the mechanisms of how peptide ligands modulate the MHC dynamics. This work investigates this phenomenon using molecular modeling approaches. Using the Immune Epitope Database and Analysis Resource (IEDB), 5 HLA supertypes alleles with the highest amount of high-affinity peptide binders were studied. Peptide-loaded MHC (pMHC) structures of these peptide-supertype pairs were modeled using the PANDORA package. After that, pairwise residue interaction-energy calculations were performed for 1000 structures of HLA-A and HLA-B subtypes, and for 607 structures of HLA-C using gRINN. The resulting interaction energy profiles were subjected to a clustering analysis. Finally, molecular dynamics simulations of representatives from each cluster were conducted. A comparison of global dynamics of these representatives indicated that cluster representatives show distinct behaviors in terms of protein dynamics. In conclusion, our findings provide critical insights into the relationship between peptide ligands and pMHC dynamics, and pave the path for predictive models of pMHC dynamics, which can be applied in fields such as basic immunology research, cancer vaccine development, and autoimmune disease treatment.

Keywords: *Molecular Dynamics, Peptide Binding Affinity, MH*

Estimation of Apolipoprotein B Levels Using Machine-Learning Models

Serra Ilayda Yerlitaş Taştan[1,2], Federica Fogacci[3], Gözde Ertürk Zararsız[1,2], Halef Okan Doğan[4,5], Andrea Balbinot[3], Marina Giovannini[3], Gökmen Zararsız[1,2,5,*] ,Arrigo Francesco Giuseppe Cicero[3,6]

[1]*Department of Biostatistics, Erciyes University School of Medicine, Kayseri, Türkiye,*
[2]*Drug Application and Research Center (ERFARMA), Erciyes University, Kayseri, Türkiye*
[3]*Hypertension and Cardiovascular Risk Research Unit, Alma Mater Studiorum University of Bologna, Bologna, Italy*
[4]*Department of Biochemistry, Sivas Cumhuriyet University, School of Medicine, Sivas, Türkiye*
[5]*Erciyes Teknopark, Hematainer Biotechnology and Health Products Inc, Kayseri, Türkiye*
[6]*Cardiovascular Medicine Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy*

Presenting Author: serrayerlitas@erciyes.edu.tr
**Corresponding Author: gokmenzararsiz@erciyes.edu.tr*

Apolipoprotein B-100 (Apo B) is an important biomarker for cardiovascular risk assessment and management of lipid-lowering therapy. Therefore, it is very important for diagnosing and predicting cardiovascular disease. While the immunoturbidimetric method is the traditional approach for measuring Apo B, this technique has limitations, leading to the development of 15 estimation formulas. In this study, we used data from 12,376 participants aged 18 years and older from the NHANES database (between 2007-2016) to train Apo B with machine learning (ML) algorithms. Two scenarios were used for ML training: in the first scenario, the input variables TC, HDL-C, and TG, and in the second scenario, non-HDL-C and LDL-C were used as inputs. Apo B levels were measured directly and predicted using 11 ML algorithms: SVM, GBM, kNN, MARS, BRNN, XGBoost, Random Forest and Linear, Ridge, Lasso, and Elastic-Net regression models. Model performances were compared with existing formulas. Linear regression analysis was performed, and residual error plots were generated for each prediction method and direct measurement. Performance metrics such as root mean square error (RMSE), concordance, and mean absolute deviations were calculated. The results demonstrated that all ML algorithms outperformed the existing formulas, achieving lower RMSE and higher concordance. Among the ML algorithms, the BRNN model showed the most accurate Apo B predictions in Scenario 1, with an RMSE of 8.351 and concordance of 0.671. Similarly, in Scenario 2, BRNN again delivered superior performance, achieving the lowest RMSE of 8.424 and the highest concordance of 0.667. Among the formulas, Hermans Model-1 had the lowest RMSE of 9.000 and concordance of 0.648, followed by Hwang Model-1 (RMSE of 9.701, concordance of 0.639). To our knowledge, this is the first study to predict Apo B levels with ML algorithms. The results showed that ML algorithms can predict Apo B levels with better accuracy than existing formula.

Keywords: *Apolipoprotein B; Estimation; Formulas; Machine-Learning*

MLMS: Machine-Learning Interface for Metabolomics Data

Ahu Cephe[1,2], Gözde Ertürk Zararsız[2,3], Vahap Eldem[4], Halef Okan Doğan[5,6], Gökmen Zararsız[2,3,6,*]

[1]*Erciyes University Rectorate, Institutional Data Management and Analytics Unit, Kayseri, Türkiye*
[2]*Erciyes University, Drug Application and Research Center (ERFARMA), Erciyes University, Kayseri, Türkiye*
[3] *Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri, Türkiye*
[4]*Istanbul University, Faculty of Science, Department of Biology, Istanbul, Türkiye*
[5]*Cumhuriyet University, Faculty of Medicine, Department of Biochemistry, Sivas, Türkiye*
[6]*Erciyes Teknopark, Hematainer Biotechnology and Health Products Inc., Kayseri, Türkiye*

Presenting Author: ahucephe.87@gmail.com
**Corresponding Author: gokmenzararsiz@erciyes.edu.tr*

Analyzing metabolomics data provides insights into the discovery of potential new biochemical pathways and the interactions of molecules within these pathways, thereby contributing to a deeper understanding of diseases at the molecular level. Various packages and platforms exist for metabolomics data analysis; however, there is a lack of comprehensive tools specifically designed for classification analyses using machine-learning algorithms. The aim of this study is to introduce MLMS, a comprehensive R package developed to facilitate the application of machine-learning methods for classifying metabolomics data. MLMS offers tools for tasks such as disease prediction, identifying key features, and sorting features based on their importance in model prediction. Additionally, it provides pre-processing functionalities, including handling missing values, filtering, normalization, and transformation of raw metabolomics data. The package supports the implementation of more than a hundred machine-learning algorithms available in the literature for building classification models. It also includes features for parameter optimization, model performance evaluation, model comparison, and prediction of class labels for test datasets. Designed with the S4 coding system, MLMS ensures compatibility with existing metabolomics data analysis packages. Unit root tests were conducted using the testthat R package. MLMS is user-friendly, open-source, and freely available, making it one of the most comprehensive machine-learning tools in metabolomics. To support researchers with limited coding expertise, a web-based interface has also been developed. This interface allows users to perform analyses via a graphical menu system, with MLMS handling computations in the background. The source code for both the MLMS R package and its web interface can be accessed at https://github.com/gokmenzararsiz/MLMS.

Keywords: *MLMS, metabolomics, classification, R package, web-tool*

MethComRing: A Shiny Web-Tool for Visualization of Circular Heatmaps to Identify Inter-Laboratory Variation in Ring Studies

Gözde Ertürk Zararsiz[1,2], Dilek Taylı[4], Gökmen Zararsız[1,2,3], Gabi Kastenmüller[5,*]

[1]*Department of Biostatistics, Erciyes University School of Medicine, 38039 Kayseri, Türkiye,*
[2]*Drug Application and Reseach Center (ERFARMA), Erciyes University, 38280 Kayseri, Türkiye,*
[3]*Erciyes Teknopark, Hematainer Biotechnology and Health Products Inc, Kayseri, Türkiye*
[4]*Department of Computer Engineering, Abdullah Gül University Faculty of Engineering, 38080 Kayseri, Türkiye,*
[5]*Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany*

Presenting Author: gozdeerturk@erciyes.edu.tr
**Corresponding Author: gabi.kastenmuller@helmholtz-munich.de*

Ring studies are critical for assessing the consistency, reliability, and reproducibility of analytical methods. These studies aim to standardize methods and examine inter-laboratory variations by testing whether specific analytical methods produce similar results across different laboratories. In this study, a web-tool was developed to analyze analytical differences and provide visual reports for metabolomics and lipidomics data. Each laboratory is given a unique letter label from A to Z, and standard Ring study protocols were applied. Data were generated according to three simulation scenarios; systematic differences between laboratories were analysed by Deming regression, Passing-Bablok and OLS method. Measurement errors were estimated from the repeated measurements, and confidence intervals were calculated with jackknife, bootstrap and nested bootstrap methods. Minimum, median, and maximum values were evaluated through linear regression analyses, and relative differences were determined using regression coefficients. The results were visualized using circular heatmap plots. Simulation scenarios revealed systematic differences in metabolite concentrations between laboratories. Absolute and relative differences were calculated and minimum, median and maximum values for each variable were evaluated separately. The web tool included comprehensive descriptions of statistical methods, facilitating users' data analyses. This study provided a practical approach for validating analysis results, such as metabolomics and lipidomics data, between laboratories employing similar measurement methods. By identifying systematic differences, the feasibility of common analyses was ensured, and the developed web tool supported validation processes with fast and visual reporting.

Keywords: *Interlaboratory Comparison, Relative difference, Deming regression, systematic difference, method comparison*

Language Modeling-Based Generative Artificial Intelligence for De Novo Protein Design

Seyit Semih Yiğitarslan[1,2], Tunca Doğan[1,2,*]

[1]*Biological Data Science Lab, Department of Computer Engineering, Hacettepe University, Ankara, Turkey*
[2]*Department of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, Ankara, Turkey*

Presenting Author: ssyigitarslan@gmail.com
**Corresponding Author: tuncadogan@hacettepe.edu.tr*

The project aims to design de novo protein sequences using generative artificial intelligence techniques. In the first stage, a comprehensive train/validation dataset containing protein sequences, domains, functions, and structural information was created by combining data from various biological databases. The main use-case of the project is to design de novo DNA methyltransferases – DNMT (EC 2.1.1.37). For this, we first fine-tuned ProGen2, a novel protein language model (pLM), using natural DNMT protein sequences to serve as a baseline model. This fine-tuning process enables the model to capture the structural and functional patterns of DNMT proteins. The next stage will involve the development of our primary protein design model by transferring the weights of the fine-tuned Progen2 to a new pLM, which utilises function tags obtained from the Gene Ontology (GO) terms' embeddings (by the node2vec algorithm) along with tokenised sequences of nearly 300,000 proteins with GO annotations from the UniProtKB/Swiss-Prot database (Figure 1). This approach will allow the design of new functional sequences using the trained model, controlled by the given function tags (GO embeddings). We will employ generative metrics (novelty and validity) and classification metrics (precision, recall, F1-score and accuracy) to assess models. The analysis of the predicted structures (by AlphaFold) of our de novo sequences will allow us to evaluate the performance of the model with docking and molecular dynamics simulations and to compare them with both the baseline generative model (fine-tuned Progen2) and the natural DNMT proteins. The detailed examination of the model's architecture, training, and testing processes will provide valuable insights into how autoregressive generative networks can effectively generate samples of specific protein families like DNMT. This study will contribute new knowledge regarding the potential of AI-assisted de novo protein design.

Metabolomics and Transcriptomics Based Multi-Omics Integration Reveals Major Metabolic Pathways and Progression Biomarkers Associated with Autosomal Dominant Polycystic Kidney Disease

Gökmen Zararsız[1,2,3], Leila Kianmehr[1,2,*], Ahu Cephe[4], Necla Kochan[5], Gözde Ertürk Zararsız[1,2], Aleyna Erakcaoglu[1,2], Alparslan Demiray[6], Salih Güntuğ Özaytürk[6], Onur Şenol[7], Vahap Eldem[8], Serpil Taheri[10,11], Halef Okan Doğan[3,9], İsmail Kocyigit[6]

[1]*Department of Biostatistics, Erciyes University School of Medicine, Kayseri, Türkiye*
[2]*Drug Application and Research Center (ERFARMA), Erciyes University, Kayseri, Türkiye*
[3]*Erciyes Teknopark, Hematainer Biotechnology and Health Products Inc, Kayseri, Türkiye*
[4]*Erciyes University Rectorate, Institutional Data Management and Analytics Unit, Kayseri, Türkiye*
[5]*Department of Mathematics, Izmir Economy University, İzmir, Türkiye*
[6]*Department of Nephrology, School of Medicine, Erciyes University, Kayseri, Türkiye*
[7]*Department of Analytical Chemistry, Faculty of Pharmacy, Ataturk University, Erzurum, Türkiye*
[8]*Department of Biology, Faculty of Science, Istanbul University, Istanbul, Türkiye*
[9]*Department of Biochemistry, Sivas Cumhuriyet University, School of Medicine, Sivas, Türkiye*
[10]*Department of Medical Biology, Medical Faculty, Erciyes University, Kayseri, Turkey*
[11]*Betul-Ziya Eren Genome and Stem Cell Center, Erciyes University, Kayseri, Turkey*

Presenting Author: gokmenzararsiz@erciyes.edu.tr
**Corresponding Author: kianmehr96@gmail.com*

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a hereditary kidney disorder affecting over 12.5 million individuals worldwide, often progressing to end-stage renal disease requiring dialysis or transplantation. Despite advancements, ADPKD pregression biomarkers remains unexplored. This study aims to identify key biomarkers and pathways influencing ADPKD progression through a multi-omics approach. Plasma samples from 254 ADPKD patients aged 30-70 were analyzed. Transcriptomics was performed using RNA-sequencing and untargeted metabolomics for metabolite profiling. Patients were categorized by Mayo classification (slow vs. rapid progression) and hypertension status. Differentially expressed genes (DEGs) were subjected to Gene Ontology (GO), KEGG pathway, and protein-protein interaction (PPI) analyses. Hub genes were identified using CYTOHUBBA and MCODE in Cytoscape. DEG analysis revealed 1,665 DEGs for Mayo classification and 48 DEGs for hypertension groups. Pathway analysis revealed cardiomyopathy pathways. Hub gene analysis identified 15 critical genes, including CDK1. Key GO terms of hub genes involved cell cycle, chromosome segragation, mitotic spindle checkpoint. Metabolomics identified 16 significant metabolites, enriched in biotin metabolism, linoleic acid metabolism, and fatty acid biosynthesis. Integrated multi-omics analysis highlighted overlapping pathways, such as glycerolipid metabolism, linoleic acid metabolism, lipid metabolism, and drug metabolism (FDR<0.05). This study demonstrates that multi-omics analysis provides valuable insights into ADPKD progression mechanisms, offering potential biomarkers and therapeutic targets for personalized care and improved patient outcomes.

***Keywords**: ADPKD, biomarker, metabolomics, multi-omics analysis, transcriptomics*

Identification of Intron Retention Regions in Breast Cancer by Bioinformatics Approaches

Esma Gamze Aksel[1], Vahap Eldem[2], Gökmen Zararsız[3,4,5,*]

[1]*Erciyes University, Faculty of Veterinary, Department of Genetic, Kayseri, Türkiye*
[2]*İstanbul University, Faculty of Science, Department of Biology, Division of Zoology, İstanbul, Türkiye*
[3]*Drug Application and Research Center (ERFARMA), Erciyes University, Kayseri, Türkiye*
[4]*Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri, Türkiye*
[5]*Erciyes Teknopark, Hematainer Biotechnology and Health Products Inc, Kayseri, Türkiye*

Presenting Author: gamzeilgar@erciyes.edu.tr
**Corresponding Author: gokmenzararsiz@erciyes.edu.tr*

This study aimed to evaluate intron retention in publicly available breast cancer (Triple Negative Breast Cancer, TNBC) samples using IRFinder, IRFinder-S, İREAD, and Whippet algorithms. The data were subjected to quality control. The reference genome indexed and aligned with in STAR 2.7.10b algorithm. Differential expression analyses in IRFinder, and IRFinder-S were determined by DEseq2, edgeR in IREAD and Delta PSI results in the Whippet. The genes remaining in the intersection set of the genes passing the threshold values of the algorithms in the Venny 2.1.0 application were analyzed regarding the functions and pathways they are related to in ShinyGO 0.80 and gProfiler g:GOSt functional profiling web tools. According to the results obtained, in the intersection set of IRFinder, IRFinder-S, and IREAD algorithms in terms of intron involvement, 47 (0.4%) genes out of 10395 gene regions were increased and 3 (0.3%) of 1184 genes were decreased. According to the results of pathway analysis of upstream intron retention genes, especially ENSG00000074181 (NOTCH3), ENSG00000142949 (PTPRF), ENSG00000104946 (TBC1D17), ENSG00000180900 (SCRIB), ENSG00000143614 (GATAD2B), ENSG00000182492 (BGN) were found to be retained in breast cancer and general cancer pathways. According to the Cohen Kappa analyses it was determined that the most compatible results were found in the IRFinder and IRFinder-S algorithms, the agreement of the IREAD algorithm was low with both algorithms, and the Whippet algorithm was incompatible by not showing a common intersection gene region. In this direction, it can be suggested that researchers should consider the genes in the intersection regions of IRFinder, IRFinder-S, and IREAD algorithms. It is recommended that researchers verify the regions of intron retention to understand the mechanism of TNBC.

***Keywords**: Intron Retention, IRFinder, IRFinder-S, IREAD, Whippet*

Establishing Continuous Reference Intervals in Adult Thyroid Function Tests Using GAMLSS Methods

Funda İpekten[1,2], Gözde Ertürk Zararsız[1,2], Halef Okan Doğan[4], Çiğdem Karakükçü[5], Gökmen Zararsız[1,2,6,*]

[1]*Department of Biostatistics, Faculty of Medicine, Erciyes University, Kayseri, Turkey*
[2]*Drug Application and Research Center, Erciyes University, Kayseri, Turkey*
[3]*Department of Biostatistics and Medical Informatics, Faculty of Medicine, Adiyaman University, Adiyaman, Turkey*
[4]*Department of Biocmehistry, Faculty of Medicine, Cumhuriyet University, Sivas, Turkey*
[5]*Department of Biocmehistry, Faculty of Medicine, Erciyes University, Kayseri, Turkey*
[6]*Erciyes Teknopark, Hematainer Biotechnology and Health Products Inc, Kayseri, Turkey*

Presenting Author: gozdeerturk9@gmail.com
**Corresponding Author: gokmenzararsiz@erciyes.edu.tr*

Background: Traditional reference intervals are often partitioned into discrete reference intervals by major covariates such as age and sex. However, in the context of thyroid function tests, discrete reference intervals can oversimplify the complex relationship between analyte concentrations and age, potentially leading to less accurate clinical assessments. Calculating continuous reference intervals can provide a more precise understanding of thyroid function variations across age groups, improving diagnostic accuracy and personalized patient care.
Methods: Thyroid function test results were collected from Cumhuriyet University Hospital for five years (between 2017 and 2022). Generalized Additive Models for Location, Scale, and Shape (GAMLSS) were used to establish continuous reference intervals in thyroid function tests for adults. The LMS, LMST, and LMSP methods were implemented for each gender and measurement. Distribution parameters were estimated using the maximum penalized likelihood method with the RS algorithm and Fisher scoring method. Generalized Akaike information criterion (GAIC) was used for model comparison.
Results: Continuous reference intervals were established for three thyroid function parameters. The requiring sex-specific reference curves were fT3 and fT4. Continuous reference intervals assessed by GAMLSS analysis more accurately represented the relationship between thyroid function tests and age.
Conclusions: Continuous reference intervals provide a more accurate estimation of the age-related dynamic changes in thyroid function test reference values, enhancing the interpretation of laboratory test results and supporting better clinical decision-making in adults. Establishing continuous reference intervals for thyroid function tests helps improve early diagnosis and treatment processes by more accurately reflecting the biological differences between individuals and the changes that may occur over time.

Keywords: *GAMLSS, thyroid, continuous reference interval, partitioned*

Investigating Molecular Changes of Fusiform Gyrus Tissue in Alzheimer's Disease by Transcriptome Analysis

Kübra Nur Şahin[1], Ecehan Abdik[2], Tunahan Çakır[2], Hatice Büşra Lüleci[2,*]

[1]*Yıldız Technical University, Molecular Biology and Genetics Department, Istanbul, Turkey*
[2]*Gebze Technical University, Bioengineering Department, Kocaeli, Turkey*

Presenting Author: kkubranur.sahinn@gmail.com
**Corresponding Author: hbsrknk@gmail.com*

The fusiform gyrus (FG), part of the occipital and temporal lobes, plays a critical role in visual and language processing. In severe stage of Alzheimer's Disease (AD), individuals may experience symptoms such as an inability to recognize family members. High tau accumulation in the FG is an important factor in AD progression especially for Early-Onset Alzheimer's Disease (EOAD) patients. In this study, we aimed to investigate the potential molecular changes in the FG tissue for AD patients by analyzing transcriptome data. Differentially expressed genes (DEGs) analysis was conducted using the GSE125583 dataset including 219 AD and 70 control samples, retrieved from Gene Expression Omnibus. DESeq2 package from R programming was used to determine DEGs ($p < 0.001$) and 2,480 significant genes were identified. Enrichment analysis was conducted for DEGs using clusterProfiler, R package. Enriched terms highlighted tau protein binding, fatty acid binding, neuron spine structure, axon guidance and also female-specific processes, such as oxytocin signaling, ovarian steroidogenesis, and sex differentiation. Other studies also indicate that females with AD experience greater language decline, show faster neurodegeneration than males, and have higher phosphorylated tau levels in EOAD. Network-based analysis was also applied with DEGs to reveal molecular mechanisms based on miRNA-gene, chemical-gene, and disease-gene interactions by NetworkAnalyst. EOAD related key miRNAs, female-specific chemicals, and white matter-related diseases were highlighted in the generated networks. Overall, our results suggest that molecular abnormalities in FG tissue might play an important role in AD pathology through sex-specific tau accumulation. Moreover, findings also demonstrated that FG can be evaluated as a key tissue for understanding the molecular mechanisms of EOAD and white matter-related neurological diseases.

Keywords: *Alzheimer's Disease, transcriptome data, DEGs, network, the fusiform gyrus*

Computational Investigation of Dynamics of Circadian Rhythm Regulator, CLOCK-BMAL-1 Complex, in DNA-Histone Complex

Dima Amairy[1], Hanife Pekel[1], Şeref Gül[2] & Özge Şensoy[1,*]

[1]*School of Engineering and Natural Sciences, Department of Biomedical Engineering, Istanbul Medipol University, Beykoz, Istanbul*
[2]*Life Sciences and Biotechnology Institute, Bezmialem Vakif University, Beykoz, Istanbul*

Presenting Author: dima.amairy@std.medipol.edu.tr
**Corresponding Author: osensoy@medipol.edu.tr*

The mammalian circadian rhythm is regulated by the CLOCK-BMAL-1 complex, which drives rhythmic gene expression. This complex interacts with DNA and histones, influencing transcriptional activity. Mutations within BMAL-1 protein can alter these interactions, leading to dynamic and structural shifts, mainly in the transcriptional output, thus affecting circadian regulation.

This study employs coarse-grained molecular dynamics simulations to explore the structural and dynamic behaviors of the CLOCK-BMAL-1 complex, analyzing both wild-type and mutated systems in the presence of DNA and histones. Two specific BMAL-1 mutations—R173Q, which is associated with enhanced transcriptional activity, and R223Q, which inhibits it—are examined to understand their impact on the dynamics of the systems. Simulation results demonstrated that the R173Q mutation reduced interaction between histones and DNA, thereby enhancing DNA accessibility and stabilizing the complex in a conformation that favors transcription. This alteration in the basic Helix-Loop-Helix (bHLH) domain likely improves the DNA binding efficiency of the complex, enabling a transcriptionally active state. In contrast, the R223Q mutation increases interactions between histones and the DNA, which may limit DNA accessibility and affect the structural flexibility of the bHLH domain. These increased histone contacts may disrupt the optimal orientation of the CLOCK-BMAL-1 complex, reducing its ability to bind DNA efficiently and leading to decreased transcriptional activity.

These findings provide new insights into the structural and dynamic mechanisms by which specific BMAL-1 mutations modulate transcriptional regulation. The results of this study provide a holistic understanding of molecular mechanisms on how these mutations influence circadian rhythm regulation, thus offering a foundation for future research that focuses on development of effective therapeutics with an emphasis on treatment of circadian-related disorders.

Investigation of Relationship Between Differential miRNAs and Genes in Parkinson's Disease

Tuğba Sever[1], Hatice Büşra Lüleci[2], Tunahan Çakır[2], Ecehan Abdik[2,*]

[1]*Yıldız Technical University, Molecular Biology and Genetics Department, Istanbul, Turkey*
[2]*Gebze Technical University, Bioengineering Department, Kocaeli, Turkey*

Presenting Author: svrtugba@gmail.com
**Corresponding Author: ecehanabdik@gtu.edu.tr*

Parkinson's disease (PD) is the most common neurodegenerative movement disorder. PD is marked by the gradual loss of dopaminergic neurons in the substantia nigra (SN), along with alpha-synuclein accumulation. Mitochondrial dysfunction and oxidative stress are the main contributors to dopaminergic neuron loss. Changes in microRNAs (miRNAs), are associated with PD and influence the accumulation of toxic proteins like alpha-synuclein. Since miRNAs can cross the blood-brain barrier, they hold potential as therapeutic agents and biomarkers for early PD diagnosis. This study evaluated the regulatory impact of miRNAs in peripheral blood on neurodegenerative processes in the SN. Differential expressed genes (DEGs) analysis was performed for SN samples of GSE114517 and GSE136666 datasets that are publicly available at Gene Expression Omnibus. Using DESeq2, 273 genes were identified as DEGs in both datasets ($p < 0.05$). Additionally, an analysis of differentially expressed miRNAs (DEmiRNAs) was conducted on the GSE16658 dataset including samples from peripheral blood mononuclear cells (PBMCs) of PD patients. 40 DEmiRNAs were identified using the "limma" package ($p<0.05$). Enrichment analysis was applied for 273 DEGs and 40 DEmiRNAs separately. Pathways related to cellular stress response, protein folding, and ATPase regulation were revealed as significantly enriched in DEGs-based analysis. Neurodegenerative disease pathways were found to be enriched in DEmiRNA-based analysis. Through NetworkAnalyst, an online tool, protein-protein interaction (PPI) networks specific to the SN were created using DEGs. Interactions between the hub proteins of the created PPI (55 proteins) and DEmiRNAs were screened with the multiMiR, R package. As a result, it was revealed that 31 of the 40 DEmiRNAs interacted with at least one of the hub proteins. In conclusion, the findings suggest that PBMC miRNAs could play a regulatory role on mRNAs in the SN.

Keywords: *Parkinson's disease, transcriptome data, miRNA, DEGs, network*

Estimating Metabolic Differentiation in Diseases with Deep Learning

Barış Can[1], Ali Cakmak[1] ,Sadi Çelik[1,*]

[1] *İstanbul Technical University, Istanbul, Turkiye*

Presenting Author: canb22@itu.edu.tr
**Corresponding Author: sadi.cel@gmail.com*

Changes in metabolic events are particularly important in the identification of many diseases, especially cancer. In the literature, there are many algorithms that quantitatively estimate the activity changes of metabolic pathways based on the abundance changes of metabolites to detect diseases. In the Metabolitics algorithm, Flux Variability Analysis (FVA) is employed to calculate pathway scores. FVA involves solving an optimization problem with linear programming solvers like CPLEX. The calculation of the reaction differentiation scores for a dataset takes approximately 22 minutes. In this study, we propose to calculate pathway differentiation scores by predicting them with neural network-based architectures. This way, metabolic differentiation scores could be calculated within seconds rather than minutes using pre-trained models in an efficient manner. To this end, we developed two architectures. Firstly, a Multiout Deep Neural Network (MODNN), which is a fully connected network, was designed to avoid overfitting by using batch normalization and a dropout level. Secondly, we transform metabolite data into an image format by utilizing two different methods and then employ a pre-trained ResNet18 model. One of the image conversion methods is based on data duplication. The other one is based on the pathway-metabolite relationships in the metabolic network, which creates images for each individual with a size of the number of pathways times the number of metabolites. The approaches are tested on the above-mentioned dataset with 10-fold cross-validation to ensure robust results. When the dataset size increased, the performance of the model did not change significantly. However, it is observed that the robustness has increased. The best result was obtained using ResNet18 and Recon3D-based transformation, which boosted the performance of CNN-based architecture with an average MSE of 0.064.

Predicting Unplanned Hospital Readmissions

C. Yasar[1], H. Akinci[1], M. Y. Ulukus[1], F. Akdogan[1], Z. Akdogan[1], A. Cakmak[1,*]

[1]*Istanbul Technical University, İstanbul, Turkiye*

Presenting Author: yasarcereen@gmail.com
**Corresponding Author: ali.cakmak@itu.edu.tr*

Unplanned hospital readmissions impose economic and health burdens, resulting in high costs, increased mortality rates, and diminished patient quality of life. In this study, we propose a comprehensive decision support system that integrates machine learning models to monitor readmission risk for multiple high-burden diseases, including myocardial infarction, chronic obstructive pulmonary disease, heart failure, diabetes mellitus, and stroke. Traditional readmission models rely on limited variables and arbitrary time thresholds, failing to account for individual patient differences and the dynamic nature of health data. To address these limitations, our system utilizes personalized models that predict readmission risk by analyzing patient-specific data from initial admission through discharge. Our methodology leverages real-life hospital data with 957,356 hospital visit records of 18,775 patients, where readmission cases are labeled based on patient records. Our dataset is highly imbalanced, with readmission cases representing 3.16% of data. To overcome this issue, we employ several data balancing techniques such as undersampling, oversampling, and SMOTE variations. This way, we could ensure that the models do not overfit the majority class, and instead, generalize well to predict rare readmission events. Furthermore, the dataset contains missing values, which we handle with imputation methods such as K-Nearest Neighbors and MissForest. We employ commonly used machine learning algorithms, ranging from logistic regression and decision trees to ensemble models. We ensure model quality through cross-validation and parameter tuning on validation datasets. Our models achieve approximately 92% F1 scores. This framework enables personalized readmission predictions for healthcare systems. Ultimately, our work aims to enhance patient outcomes by empowering clinicians with reliable, data-driven discharge decisions.

An RNA-Seq Based Network Approach to Elucidate Molecular Mechanisms of Asymptomatic Alzheimer's Disease

Rümeysa Aksu[1], Hatice Büşra Lüleci[1], Tunahan Çakır[1,*]

[1]*Department of Bioengineering, Gebze Technical University, Kocaeli, Turkiye*

Presenting Author: r.aksu2022@gtu.edu.tr
**Corresponding Author: tcakir@gtu.edu.tr*

Alzheimer's disease (AD) causes major challenges for both patients and their families due to its extremely challenging management and treatment. Asymptomatic Alzheimer's disease (AsymAD) is a preclinical stage of AD identified by amyloid plaques and neurofibrillary tangles in cognitively normal individuals and offers essential understanding for early diagnosis and treatment of AD. The goal of this research is to uncover molecular insights into AsymAD, which will lead to the identification of new candidate biomarkers and molecular mechanisms for early detection and intervention of AD. Highlighting the focus on gene expression analysis, two different RNA sequencing (RNA-seq) datasets from ROSMAP (Religious Orders Study and Memory and Aging Project) and MSBB (Mount Sinai Brain Bank) cohorts were investigated. The individuals in the datasets were grouped into AD and AsymAD based on clinical and neuropathological criteria. Applied methods here identified 995 differentially expressed genes (DEGs), 1398 differentially expressed transcripts (DETs), and 293 differentially used transcripts (DUTs) between AD and AsymAD samples from ROSMAP and 320 DEGs, 2450 DETs, and 541 DUTs from MSBB. Subsequently, the significant genes from these three analyses were mapped onto a human protein-protein interaction (PPI) network, revealing subnetworks associated with AsymAD. The results were interpreted through enrichment analysis and compared with the predefined lists of AD-related and Learning-Memory-Cognition-related genes, resulting in the discovery of biomarker candidate genes. The candidate genes are gabarapl2, pacsin2, fbxo2, and rasgrf2 from the differential analyses and rhog, prnp, rasgrf1, hspa2, and camk4 from the subnetwork analyses. This study introduces an innovative approach beyond the standard focus on DEGs by integrating DEGs, DETs, and DUTs in the analyses, pointing out the first comprehensive insights into the molecular mechanisms of AsymAD.

Keywords: *Asymptomatic Alzheimer's Disease, Differential Gene Expression, Differential Transcript Expression, Differential Transcript Usage, Subnetwork Analysis*

Reanalysis of Transcriptomics with a focus of HOX-TALE genes reveals differen ally expressed genes for Dis nct Stages of Non-Alcoholic Fa y Liver Disease (NAFLD), Non-Alcoholic Steatohepas (NASH) and cirrhosis

Mustafa Keleş[1], Aysen Gunel-Ozcan[1,*]

[1]*Hacettepe University, Dept. of Stem Cell Research, Ankara, Turkiye*

Presenting Author: mkeles27@gmail.com
**Corresponding Author*

HOX proteins are homebox-domain containing proteins that are expressed in spa otemporal mul collinearity during embryonic development and TALE proteins are canonical cofactors of HOX proteins to ensure DNA binding sequence specificity. Both HOX and TALE proteins also have non canonical roles in adult diseases, such as fibrosis and cancer. Non-alcoholic fa y liver disease (NAFLD) is the condi on characterized by increased intrahepa c triglyceride content without overconsump on of alcohol or any other liver injury e ology. NAFLD can further proceed to Non-Alcoholic Steatohepa s (NASH), which is characterized by inflamma on and fibrosis of the liver ssue and increased liver enzyme levels in serum. Further progression of NASH can result in cirrhosis and liver failure. Although there are known differen ally expressed genes which may partake in the progression of NAFLD to NASH to cirrhosis, the underlying gene c mechanism needs to be studied further. In this study, we have reanalyzed liver biopsy transcriptomic data with a focus on 63 HOX-TALE genes, including coding and non-coding RNAs, from mul ple datasets on Gene Expression Omnibus (GEO) database (GSE48452, GSE66676, GSE135251) with healthy liver, NAFLD and NASH with dis nct pathologically confirmed grades of fibrosis. We have compared mul ple normaliza on methods (TPM, MoR and TMM) for raw RNAseq count data, as well as mul ple tools for reanalysis (GREIN and BEAVR). Within comparisons of healthy vs NAFLD (4 differen ally expressed), healthy vs NASH (10 differen ally expressed) NAFLD vs NASH (9 differen ally expressed), mul ple differen ally expressed genes were found. Gene set enrichment analysis of differen ally expressed genes also revealed some oncogenes such as MEIS1 and some genes with previously associated func ons in different organ fibroses such as PKNOX2. In-vitro experiments will be performed to further confirm these findings.

Employing Anatomical Therapeutic Chemical Ontology to Deal with High Dimensionality and Performance Issues in Predictive Clinical Data Analytics

Hacer Yeter Akıncı[1], Ceren Yaşar[1], Mehmet Yasin Ulukuş[1], Fatih Akdoğan[2], Zeyneb Akdoğan[3], Ali Cakmak[1,*]

[1]*Istanbul Technical University, İstanbul Turkiye*
[2]*Sincan Eğitim ve Araştırma Hastanesi, Ankara Turkiye*
[3]*Üsküdar Üniversitesi, İstanbul, Turkiye*

Presenting Author: hacerakinci07@gmail.com
**Corresponding Author: ali.cakmak@itu.edu.tr*

The increasing volume of health data poses great challenges, such as data abnormalities and missing or erroneously created values, which undermine the accuracy of predictive models and decision support systems. While previous research has focused on addressing errors in clinical data, less attention has been paid to balancing computational efficiency with data accuracy. In this study, we address this gap by examining diagnostic data errors in 18,772 readmitted patients, diagnosed with diabetes and heart failure, using a dataset of over 950,000 records encompassing laboratory tests, prescriptions, and medication data. By applying the Anatomical Therapeutic Chemical (ATC) ontology for feature aggregation, we reduced the dimensionality of medication-related features from 11,000 to 50-100, resulting in a 97% reduction in memory usage and a 92% decrease in training time, while maintaining the model performance with almost no change. Random Forest and XGBoost models demonstrated comparable or better F1 scores, with XGBoost achieving up to 98.89% for diabetes and 83.98% for heart failure when higher levels in ATC ontology are employed. These findings emphasize the potential of the ATC hierarchy to enhance both the computational efficiency and the predictive reliability of clinical data analysis systems. Further studies may explore dynamic ATC level selection tailored to specific datasets and problem definitions.

Exploring the Substrate Selectivity of MFS Proteins through Structural Bioinformatics Methods

Ayça Emanetoğlu[1], Gülçin Deniz Yavuz[1], Onur Serçinoğlu[1,*]

[1]*Gebze Technical University, Faculty of Engineering, Bioengineering*

Presenting Author: a.emanetoglu2019@gtu.edu.tr
**Corresponding Author: osercinoglu@gtu.edu.tr*

The Major Facilitator Superfamily (MFS) proteins are essential carrier proteins that facilitate the exchange of chemicals across the cell membrane. Prediction of the substrates of these carriers is vital for understanding drug resistance in cancer and infectious diseases, as well as for drug development. In this work, the substrate selectivity of MFS proteins has been investigated using structural bioinformatics methods. Interactions between MFS-type carriers and substrates have been characterized through molecular docking simulations using MFS structures obtained from the AlphaFold2 Structure Database. Substrates of each of these proteins have been obtained from the ChEBI (Chemical Entities of Biological Interest) database. Only those MFS proteins with at least ten substrates identified from the ChEBI were included. Finally, molecular docking simulations between each MFS transporter and their substrates as well as non-substrates (substrates of other MFS transporters) were conducted using the AutoDock Vina molecular docking program by taking into account the most druggable pocket within the MFS internal cavity. Statistically-significant differences between docking scores of have been obtained for only two MFS transporters, further highlighting the difficulty in prediction of substrates for MFS proteins. The obtained data could serve as a basis for the development of more accurate structure-based prediction methods for MFS substrates.

Keywords: *MFS proteins, Major Facilitator Superfamily, substrate selectivity, bioinformatics, molecular docking simulations*

In Silico Analysis of Latent Transforming Growth Factor β Binding Protein Family Expression Patterns and Prognostic Value in Hepatocellular Carcinoma

R. Useyma Gülhan[1], Esra Nalbat[1,*]

[1]*Cancer Systems Biology Laboratory, Department of Health Informatics, Graduate School of Informatics, METU, Ankara 06800, Türkiye*

Presenting Author: useymag@gmail.com
**Corresponding Author: oesra@metu.edu.tr*

The latent transforming growth factor β binding protein (LTBP) family, comprising LTBP-1, -2, -3, and -4, modulates transforming growth factor-β (TGF-β), a vital regulator of cell growth and tissue remodeling, and plays key roles in the extracellular matrix (ECM). Dysregulated TGF-β signaling and altered LTBP expression are implicated in various cancers, including hepatocellular carcinoma (HCC). Evidence shows altered LTBP expression in chronic liver diseases, i.e. hepatocellular carcinoma. Our analysis showed distinct RNA and protein expression patterns of LTBP family members between tumor and normal tissues. The clinical data from UALCAN indicated higher LTBP expression in Caucasians and females, with no significant association with age. Protein-protein interactions (PPI) analysis from STRING revealed that all LTBP members interacted with key proteins involved in the TGF-β signaling pathway and ECM-receptor interactions, the most enriched KEGG pathways. Kaplan–Meier survival curves showed a worse prognosis for patients with elevated LTBP1 levels in HCC patients. GEPIA revealed that LTBP1 expression is associated with a more aggressive pathological stage. Pearson correlation analysis from UALCAN resulted in LTBP1 positively correlated with LASS6, PPPICB, FUT11, CDCA7L, and OSMR. Single-cell RNA sequencing data showed elevated LTBP1 expression across liver cell types, including hepatocytes, cholangiocytes, fibroblasts, and endothelial cells. These cell types are integral to ECM structure and maintenance, suggesting that LTBP1's role in ECM organization and TGF-β activation may be cell-type specific. The immune cell-specific analysis identified prominent LTBP1 expression in neutrophils, T cells, and B cells, suggesting roles in immune modulation. These findings highlight LTBP1 as a potential prognostic biomarker in HCC, linking its expression to tumor aggressiveness, immune interactions, and ECM-TGF-β pathway regulation.

Single Host-Microbe Metabolic Interactions in *C. elegans* Gut Under Vitamin-Limited Conditions

Tala Al-Rubaye [1,*]

[1]*Kadir Has University, Istanbul, Turkiye*

Presenting Author: tala.alrubaye@stu.khas.edu.tr
**Corresponding Author*

Background: Monocolonization in *Caenorhabditis elegans* allows controlled exploration of microbial effects on host metabolism. Nutrient-restricted media isolate microbial contributions by emphasizing metabolic dependencies. BacArena simulates nutrient exchange and interdependencies, highlighting metabolism's role in host-microbe interactions and therapeutic potential (Bauer et al., 2017).

Aim: To characterize metabolic exchanges in monocolonized *C. elegans* intestinal environments under vitamin-limited conditions, focusing on reciprocal impacts on host metabolism and bacterial growth.

Methods: Simulations of *C. elegans* monocolonized with 47 bacterial strains were conducted using BacArena in defined in silico arenas representing the gut lumen and host mucosal epithelial cells. Studies have shown that vitamins influence guy microbiota composition and have effects on gastrointestinal health (Pham et al., 2021). Hence, essential vitamins were systematically removed to assess nutrient flow and metabolic interactions. Bacterial growth, metabolic exchanges, and contributions to host metabolism were tracked over time. Matrix validation ensured accuracy, and data analyses identified patterns supporting host and bacterial growth under nutrient-limited conditions.

Expected Outcomes: The study will assess metabolic exchanges between *C. elegans* and bacterial strains under vitamin-limited conditions, focusing on host metabolism and bacterial biomass production. Strains producing essential metabolites are expected to sustain host metabolism, while others may lead to reduced fitness. Similarly, some strains may benefit from host exchanges more than others. These results will clarify metabolic interdependencies, highlighting bacterial contributions to nutrient availability and resilience, offering insights for microbiome-based therapies.

Keywords: C. elegans*, host-microbiome interactions, metabolic modeling, BacArena, vitamin-limited conditions*

Studying allele frequency trajectories in Anatolia

Dila Nur Çakal[1], Özgür Taşkent[2], Kıvılcım Başak Vural[2], Duygu Deniz Kazancı[2], Hande Çubukcu[1], Ayça Doğu[2], Damla Kaptan[2], Yılmaz Selim Erdal[3], Füsun Özer[3], Mehmet Somel[2], Gülşah Merve Kılınç[1,*]

[1]*Department of Bioinformatics, Hacettepe University, Ankara, Turkey*
[2]*Department of Biological Sciences, Middle East Technical University, Ankara, Turkey*
[3]*Department of Anthropology, Hacettepe University, Ankara, Turkey*

Presenting Author: dlnrckl@gmail.com
**Corresponding Author: gulsahhdal@gmail.com*

Background/aim: Genetic architecture behind complex traits in humans has been shaped under complex evolutionary processes. Ancient DNA allows studying these processes by directly investigating the allele frequency trajectories over time. Here, we aim to explore allele frequency changes over time in Anatolia for a set of 36 distinct complex traits. Materials and methods: A total of 34 ancient and 16 modern human genomes from Anatolia were analyzed, focusing on allele frequency dynamics across three distinct time periods: the Neolithic (c.8,000 years BP), Late Chalcolithic (c.6,000 years BP), and present day. Using a maximum likelihood-based approach, frequency shifts were identified for all genetic variants. After determining the allele frequencies, the variants associated with phenotypes of interest were compared with neutral variants to identify those showing significant differentiation. Frequency changes were observed across all three periods, our focus was on 19 variants that showed frequency shifts within the range of 0.05 to 0.15 and specifically increased from Late Chalcolithic period to the modern era. Forward-time neutral simulation scenarios were performed 1,000 times using the SLiM program, with initial frequencies set within the range of 0.05 to 0.15. Then, summary statistics, including Tajima's D, heterozygosity, and Hstatistics, were calculated for both the observed variants and the simulation results. Results: The results indicate that 19 genetic variants associated with phenotypes such as age at menarche, ADHD, body height, HDL cholesterol, LDL cholesterol, schizophrenia, type 2 diabetes, and vitamin D measurement showed an increase in frequency from the Late Chalcolithic period to the present day in Anatolia. PCA analysis of these summary statistics suggested differentiation between variants showing strong frequency shifts and both simulated and empirical neutral variants. These findings raise the possibility that some of the studied traits, related to metabolism, physical characteristics, and age-related health conditions, might have been subject to positive selection in Anatolia, although this would require confirmation by further analyses.

Keywords: *Allele frequencies, selection, population genetics, ancient DNA*

Development of a Software for the Analysis of Atomistic Local Frustration Levels in Protein Structures

Zeynep Özer[1], Onur Serçinoğlu[2,*]

[1]*Gebze Technical University, Faculty of Engineering, Department of Bioengineering*
[2]*Gebze Technical University, Faculty of Engineering, Department of Bioengineering*

Presenting Author: z.ozer2020@gtu.edu.tr
**Corresponding Author: osercinoglu@gtu.edu.tr*

Grasping protein structure is important because it helps appreciate the immense variability and versatility of proteins. Understanding the interplay between local frustration levels and conformational changes in protein structures varies significantly across different force fields. A key challenge in atomic-scale local frustration analysis is the lack of accessible software. In order to overcome this constraint, the goal of this study is to create a simple computational tool that can be used to compute atomistic local frustration levels. This tool builds upon an existing open-source package known as gRINN, which computes pairwise amino acid energetic interactions in MD simulation methods. The tool is specifically designed to simplify the calculation of atomistic local frustration levels within protein structure datasets, utilizing a user-friendly command-line interface. This tool enhances the analysis of protein structural ensembles. Structures of Endolysin, Cold shock protein CspB, and Ribonuclease HI proteins, encompassing mutations were obtained. Furthermore, a detailed analysis was performed on the effects of each position on protein stability due to mutations in the three proteins that were chosen. This analysis includes determining the impact of mutations on the overall stability of the proteins. Connections between the local frustration expected by the generated tool and the overall conservation level of each position were examined. This analysis aimed to compare patterns in highly and less conserved positions, revealing the significance of evolutionary conservation in these proteins.

Keywords: *Structural biology, atomistic local frustration analysis*

Correlations Between Immune Cell Compositions In Lung Cancer Tumors And Patient Survival Outcomes

Özlem Tuna[1], Yasin Kaymaz[1,*]

[1] *Ege University, Faculty of Engineering, Bioengineering Department, İzmir, Türkiye*

Presenting Author: yasinkaymaz@gmail.com
**Corresponding Author: yasinkaymaz@gmail.com*

The composition of immune cells in tumor tissue is crucial for predicting treatment response and overall survival. Current histopathological methods lack precise metrics for clinical outcomes. With high-throughput transcriptomics data, we developed a bioinformatic scheme to estimate immune cell infiltrations and their abundances. We propose a risk score metric to estimate patient prognostic outcomes. For this purpose, we reanalyzed tissue-based RNA sequencing data using a deconvolution approach by incorporating single-cell RNA sequencing datasets. The objective was to correlate immune cell compositions in tumor tissues with lung cancer patient survival rates. Data from 21 studies (298 patients, 505 samples) were utilized to identify gene expression sets for the tumor microenvironment and immune cell subgroups. Deconvolution methods predicted cell compositions for RNA sequencing data of 787 lung cancer patients from TCGA, with gene expression profiles specific to 44 cell types. Cell composition was determined by calculating tumor contents as a percentage of cell ratios. Patients were categorized into 'low' and 'high' content groups based on median values. Kaplan-Meier analysis examined survival rates between these groups to explore potential prognostic relationships. Our results indicate significant heterogeneity in immune cell compositions among patients. Significantly, both adenocarcinoma and squamous cell carcinoma subtypes, $CD8^+$ effector memory T cells and dividing B cells, were consistently associated with favorable prognosis (p=0.018 and p=0.0054, respectively), while neutrophils (p=0.0079), cDC1 (p=0.0036), and $CD8^+$ activated T cells (p=0.00072) were linked to unfavorable outcomes. Signature gene expressions of these tumor-infiltrated immune cells were incorporated into a risk score estimation metric and correlated with overall patient survival.

Pathogenic Impact of L284R Mutation in SYNGAP1 Gene and its Association with Cerebral Palsy

Elif Kesekler[1], Zeynep Özkeserli[2],Yunus Emre Dilek[2] Gunseli Bayram Akcapinar[2,*]

[1]*Department of Molecular Biotechnology, Faculty of Science, Turkish-German University, Istanbul, Turkey*
[2]*Department of Medical Biotechnology, Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey*

Presenting Author: keseklere319@gmail.com
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

SYNGAP1 is a synaptic Ras GTPase-activating protein that is very important during neurodevelopmental processes. It participates in synaptic plasticity and motor coordination. Recently, variants in this gene were implicated in neurodevelopmental disorders and some authors suggested its possible contribution to the occurrence of cerebral palsy. This study investigated a patient diagnosed with CP through whole-exome sequencing, identifying the L284R mutation in the SYNGAP1 gene. Our analyses classified this variant as pathogenic. To evaluate its molecular consequences, we utilized in silico tools including PyMOL, FoldX, and I-Mutant. The L284R mutation, which substitutes a nonpolar leucine with a positively charged arginine, demonstrated destabilizing effects on protein structure with a calculated ΔΔG of -1.81 kcal/mol. The mutation also altered the epitope structure (KRYYCELCLDDMLYA), potentially disrupting interactions necessary for motor function. Additionally, narrowing of the tunnel within the C2 domain (from 1.6 Å to 1.2 Å) was observed, potentially affecting ligand binding and synaptic stability. These findings suggest that the L284R mutation may contribute to synaptic and motor dysfunction. Further studies are necessary to confirm its association with CP and to explore therapeutic strategies.

Computational De-Novo Peptide Design for Gabaa Receptor To Be Used In Epilepsy Disease

Esma Yaz[1], Sefer Baday[1] ,Özge Şensoy[2], Özgür Yılmaz[3,*]

[1] *Istanbul Technical University, Istanbul, Turkiye*
[2]*Medipol University, İstanbul, Turkiye*
[3]*TÜBİTAK, Ankara, Turkiye*

Presenting Author: yaz24@itu.edu.tr
**Corresponding Author: yilmaz.ozgur@tubitak.gov.tr*

The primary neurotransmitter of the central nervous system is γ-Aminobutyric acid (GABA), which plays a vital role in coordinating brain activity. The central nervous system's inhibitory neurotransmitter, GABA, plays a vital role in brain activity. It plays a role in the transmission of neurotransmission in disorders of the brain. One of the most common targets of drugs for treating various neuropsychiatric disorders, such as anxiety and epilepsy, is GABAAR. Studies have shown that the presence of certain genes that are involved in the development and maintenance of epilepsy is linked to the presence of the main subunits γ2β2α1β2α1 of the GABAAR family mostly beeing on synaptic. The various subunits of the GABAAR family have been studied in order to understand their structure, localization, and expression. However, it is not clear how they interact with one another to form different isoforms but the binding site on the receptor. As GABA adheres to the receptor, the domains embedded in the membrane open and Cl- ions flows into the cell, but since synaptic GABA receptors show phasic inhibition, it is important for the channel to remain open during an epilepsy seizure. In this study, it was aimed to design a peptide that would help GABA maintain its effect while bound to the protein. The behavior of the receptor in the presence of peptides of different lengths that cross the brain-blood barrier suitable for the target region was calculated using molecular computing methods. The behavior of the receptor in the presence of peptides of different lengths that cross the brain-blood barrier appropriate to the target site was calculated using molecular computing methods. The findings obtained by computational methods were examined by energy calculations and protein-peptide residue interaction analysis of the peptides that remained bound to the benzodiazepine binding site of the GABAA receptor during MD simulations

CROssBARv2: A Unified Biomedical Knowledge Graph for Heterogeneous Data Representation and LLM-Driven Exploration

Bünyamin Şen[1,2], Erva Ulusoy[1,2], Melih Darcan[1], Mert Ergün[1], Elif Çevrim[1,2], Atabey Ünlü[1,2], Tunca Doğan[1,2,*]

[1]*Biological Data Science Lab, Dept. of Computer & AI Engineering, Hacettepe University, Ankara, Turkey*
[2]*Dept. of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, Ankara, Turkey*

Presenting Author: bjk.990.benim@gmail.com
**Corresponding Author: tuncadogan@hacettepe.edu.tr*

Developing effective therapeutics against prevalent diseases requires a deep understanding of molecular, genetic, and cellular factors involved in disease development/progression. However, such knowledge is dispersed across different databases, publications, and ontologies, making collecting, integrating and analysing biological data a major challenge. Here, we present CROssBARv2, an extended and improved version of our previous work (https://crossbar.kansil.org/), a heterogeneous biological knowledge graph (KG) based system to facilitate systems biology and drug discovery/repurposing. CROssBARv2 collects large-scale biological data from 32 data sources and stores them in a Neo4j graph database. CROssBARv2 consists of 2,709,502 nodes and 12,688,124 relationships between 14 node types (i.e., protein, gene, organism, domain, biological process, molecular function, cellular component, drug, compound, disease, pathway, phenotype, EC number, and side effect). On top of that, we developed a GraphQL API and a large language model interface to convert users' natural language-based queries into Neo4j's Cypher query language back and forth to access information within the KG and answer specific scientific questions without LLM hallucinations, mainly to facilitate the usage of the resource. To evaluate the capability of CROssBAR-LLM to generate scientific knowledge, we constructed several benchmark datasets and compared various open- and closed-source LLMs. Our biological true-false question benchmark revealed that CROssBAR displays a significantly improved accuracy, almost without any hallucinations in answering these scientific questions compared to base LLMs (directly asking these questions to the LLM). CROssBARv2 (https://crossbarv2.hubiodatalab.com/) is expected to contribute to life sciences research considering (i) the discovery of disease mechanisms at the molecular level and (ii) the development of effective personalised therapeutic strategies.

CCL2 as a Key Mediator of Neural Invasion in Tumor Progression: An In Silico Study

Barış Yıldırım[1], Esra Nalbat[1,*]

[1]*Cancer Systems Biology Laboratory, Department of Health Informatics, Graduate School of Informatics, METU, Ankara 06800, Türkiye*

Presenting Author: brsyldrm07@gmail.com
**Corresponding Author: oesra@metu.edu.tr*

Perineural invasion (PNI), the infiltration of cancer cells into nerves, is linked to increased metastasis and poor survival outcomes. Peripheral nerves, a unique component of the tumor microenvironment (TME), interact with cancer cells through signaling pathways that drive cancer progression through PNI. CCL2, a potent chemokine released by nerves, recruits monocytes and macrophages to inflammation sites and has been implicated in cancer progression. This study investigated the expression pattern and prognostic value of CCL2 in various cancer types. Pan-cancer analysis revealed that CCL2 expression showed widespread dysregulation between tumors and normal tissues. The detailed analysis of CCL2 expression showed that it is upregulated in PNI-related cancer types, including pancreatic ductal adenocarcinoma, biliary tract tumors, ovarian serous cystadenocarcinoma, glioma, prostate, colorectal, and head and neck cancers, while it is downregulated in other types of cancer. Survival analyses using Kaplan-Meier plots linked CCL2 expression with poor prognosis in PNI-associated cancers, especially glioblastoma. Our single-cell RNA transcript expression of CCL2 was revealed as enriched in macrophages, monocyte, eosinophil, neutrophil, peripheral blood mononuclear cells, and B cells based on immune cell specificity using single-cell data from human protein atlas. PNI is known to be highly active in pancreatic cancer and is associated with the invasiveness and metastasis of the disease. Our data revealed that enhanced CCL2 expression in ductal cells, secretory cells, pancreatic endocrine cells, fibroblast, smooth muscle cells, and exocrine glandular cells from HPA single cell types results. As a conclusion, CCL2 expression exhibited significant dysregulation across PNI-related cancers and is associated with immune cell recruitment, poor prognosis, and enhanced expression in key cell types involved in cancer progression.

Age-Related Transcriptomic Changes in Mouse Dental Tissue: Insights from Single-Cell RNA Sequencing Using Conventional and Machine Learning Approaches

Duha Alioglu[1], Miguel Thomas[1], Louis Casteilla[1], Christophe Guissard[1], Emmanuelle Arnuad[1], Marie-Laure Renoud[1], Olivier Deny[1], Philippe Kemoun[1], Valérie Planat-Benard[1], Paul Monsarrat[1], Marielle Ousset[1,*]

[1]RESTORE - Geroscience and rejuvenation research center, INCERE 31100 Toulouse, France

Presenting Author: duhaalioglu@gmail.com
**Corresponding Author: marielle.ousset@inserm.fr*

Periodontitis, a disease affecting the teeth-supporting tissues, affects one in two adults over the age of 50, significantly reducing patients' quality of life. The primary risk factor for this condition is age, which increases the likelihood of periodontitis and decreases the regenerative capacity of tissues. Mesenchymal stroma cells (MSCs) are phenotypically plastic, found in all tissues, and are critical for maintaining tissue homeostasis and regenerative capacity. We hypothesize that loss of regenerative capacities of gingiva with age involves changes in MSCs heterogeneity that preclude tissue regeneration of periodontal defects. Single-cell RNA sequencing (scRNA-seq) provides a powerful way to explore the complex transcriptional landscape of MSCs, revealing details that bulk RNA sequencing often misses. Despite its benefits, scRNA-seq data can be difficult to interpret because of issues like high variability, cell type diversity, and noise that can hide important signals. To address these challenges, we applied a comparative approach that combined conventional scRNA-seq methodologies with machine-learning techniques. Specifically, we trained XGBoost models along with explanations techniques to identify and rank genes associated with aging, identifying differences compared to conventional differential analyses. Additionally, dimensionality reduction methods such as UMAP and PCA were used to visualize age-related transcriptomic patterns and highlight key genes responsible for differentiating among age groups. These findings could potentially inform regenerative strategies and enhance our understanding of periodontal health in aging populations.

Unveiling the Protein Landscape in Cerebral Palsy Through AI-Based Structural Analysis

Zeynep Özkeserli[1], Ayça Yiğit[2], Özlem Doğan Akgün[1], Kaya Bilguvar[1], Uğur Özbek[2], Gunseli Bayram Akcapinar[1,*]

[1]*Acıbadem Üniversitesi, İstanbul Turkiye*
[2]*Izmir Biomedicine and Genome Center, İzmir Turkiye*

Presenting Author: Zeynep.ozkeserli@live.acibadem.edu.tr
**Corresponding Author: gunseli.akcapinar@acibadem.edu.tr*

Cerebral palsy (CP) is a group of non-progressive neurological disorders causing impaired movement and posture due to developmental anomalies or brain damage in early life. Affecting 2–3 per 1,000 live births globally, CP displays significant clinical heterogeneity and frequent neurological comorbidities, complicating diagnosis and management. Despite advances in identifying CP-associated genes, gene-phenotype relationships remain poorly understood, posing challenges for therapy.

We employed computational methodologies to analyze the structural and functional landscapes of CP-related proteins. A dataset of CP-associated protein sequences was curated, and their three-dimensional structures were obtained from the Protein Data Bank (PDB) and predicted via AlphaFold2. Structural properties, including physicochemical parameters, were analyzed using ProtParam. To explore features influencing protein interactions and functions, we performed electrostatic potential mapping using APBS with PyMOL to visualize electrostatic surfaces affecting interactions and binding affinity. Hydrophobicity mapping identified hydrophobic and hydrophilic regions on protein surfaces, highlighting potential interaction interfaces. Functional insights were gained through protein-protein interaction mapping with STRING and Gene Ontology (GO) enrichment analysis to identify overrepresented biological processes and molecular functions. Our analysis revealed pronounced heterogeneity in structural and functional attributes of CP-associated proteins. Notably, higher conservation of hydrophobic residues suggested implications for protein stability and interaction networks critical in CP pathogenesis. Integration of AI-driven structural predictions with functional annotations provided deeper insights into CP's molecular mechanisms.

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

HIBIT 2024
The International Symposium on Health Informatics and Bioinformatics

17TH INTERNATIONAL SYMPOSIUM ON HEALTH INFORMATICS AND BIOINFORMATICS

Bezmialem Vakıf University

Erich Frank Conference Hall,
Fatih / İSTANBUL / TÜRKİYE

18 - 20 December 2024