Hellenic Bioinformatics • 16

Main Auditorium, AUTH • Thessalonica, Greece

16-18 May 2025

May 2025

Hellenic Bioinformatics

Greece & Cyprus

http://hbio.info

HBio 16 2025

table of contents

https://www.iscb.org/become-a-member

https://www.iscb.org/become-a-member

conference title & dates

HBioinfo 2025

Conference: Hellenic Bioinformatics 16 • 2025

16-18 May 2025 • AUTH | Thessalonica | Greece

“Bioinformatics and Data Science: From Research to Translation”

conference keywords

artificial intelligence

biodiversity

bioinformatics

bioprospecting

computational biology

data integration

enzyme discovery

functional genomics

functional proteomics

machine learning

metabolic engineering

microbiome data science

pharmacogenomics

phylogeny

precision medicine

predictive analytics

structural genomics

synthetic biology

systems biology

text mining

Google Cloud

Medical Pharmaquality

RGCC

H.bioinfo

MOTIONHELLAS LTD
INNOVATIVE TECHNOLOGY SOLUTIONS

medicair

SYSBIO

appbio
APPLIED BIOINFORMATICS

iGEM Thessaloniki

ARISTOTLE UNIVERSITY OF THESSALONIKI

CENTER FOR BIOLOGICAL COMPUTATION
initiative

RESEARCH COMMITTEE
ARISTOTLE UNIVERSITY OF THESSALONIKI

elixir Greece

ISCB
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY
AFFILIATED CONFERENCE

ARISTOTLE UNIVERSITY OF THESSALONIKI
KNOWLEDGE INNOVATION CULTURE

sponsors

AUTH-BIO

AUTH-CSD

AUTH-RC

AUTH-PHARM

AUTH-MED

In-kind

appbio (graduate program)

CBC initiative

iGEM thessaloniki

sysbiobio.info (consultancy)

Silver

Google Cloud

Bronze

Medicair Bioscience Labs

Medical Pharmaquality

MotionHellas

RGCC SA

welcome messages

organizers: welcome message

Dear Conference Participants,

We are delighted to welcome you to Thessalonica for the 16th Conference of Hellenic Bioinformatics.

This year's conference, entitled "Bioinformatics and Data Science: From Research to Translation", picks up the thread where we last left it in-person during 2018–2019, as the 2020 conference was unfortunately cancelled.

We are especially excited to host 26 renowned speakers from 14 countries. The program features a rich schedule of sessions covering a wide range of topics in the field — Data Science, Research Frontiers and Translation Challenges.

We hope you take full advantage of this unique opportunity to learn from, connect with, and be inspired by our distinguished speakers.

Thank you for joining us and helping to make this conference a truly memorable event.

Enjoy the conference!

The Organizers
(Org Cttee, Local Org Cttee, Program Cttee)

society: welcome message

Dear Colleagues,

The Hellenic Bioinformatics Society warmly welcomes you to our 2025 Annual Conference, held in Thessalonica.

This collective effort between Cyprus and Greece over the past years has shaped our annual conference series into a significant event in the field of bioinformatics — supported by the scientific community, academia and the computational bioscience industry.

H.Bioinfo 16 • 2025 aims to build on this momentum, continuing our efforts to strengthen the bioinformatics sector in both countries and beyond. Our goal is to foster an environment ripe for exchanging ideas, forming collaborations, and exploring funding opportunities across laboratories, nationally and internationally.

This year, after popular demand and the positive response to our 'Bridge' online-only events (HBI4 2021, HBI5 2024), we decided to open up more space for community contributions in an otherwise packed program, to showcase activities across the board.

We wish you all a productive and enjoyable conference experience!

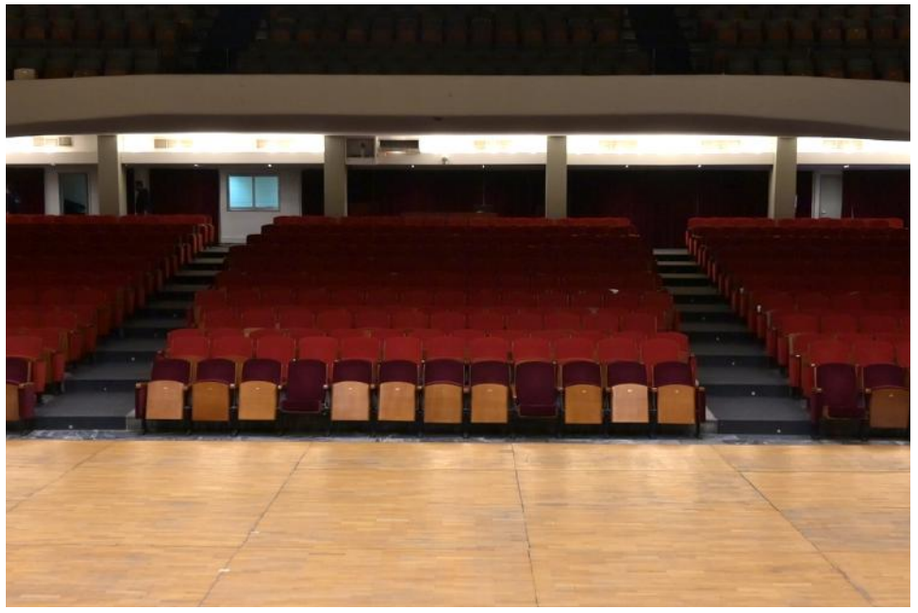
Christos Ouzounis

Chair, on behalf of the Board of Directors
Hellenic Bioinformatics

venue

Aristotle University of Thessalonica

Main Auditorium, *and* lobby



Central University Campus, Thessalonica

voice: (+30) 2310 995210

email: halls@eadp-auth.gr

organizing committee

Michael Aivailiotis, AUTH

Lefteris Angelis, AUTH

Filippos Aravanopoulos, AUTH

Anastasia Chasapi, Vidavo & AUTH [co-chair]

Spyros Gkelis, AUTH

Anna-Bettina Haidich, AUTH

Ilias Kappas, AUTH [chair]

Nikos Laskaris, AUTH

Antigoni Malousi, AUTH

Anastasis Oulas, CING [co-chair]

Spiros Papakostas, IHU

Alexios Polidoros, AUTH

Vasilis Promponas, U Cyprus

George Spyrou, CING

George Tsiamis, UPAT

Ioannis Vizirianakis, AUTH [chair]

local organizing committee

Michael Aivailiotis, AUTH

Christos Bouas, AUTH

Ilias Kappas, AUTH

Christos Kotsopoulos, AUTH

Asimina Kournoutou, AUTH [chair]

Alex Michailidis, AUTH

Despina Neraki, AUTH

Stathis Pateras, AUTH

Antonios Tsanakas, AUTH

and the iGEM Team, AUTH

program committee

Michael Aivailiotis, AUTH, Greece

Grigoris Amoutzias, U Thessaly, Greece

Irene Angelidaki, DTU, Denmark

Eleftherios Angelis, AUTH, Greece

Ismini Baltsavia, U Crete, Greece

Anastasia Chasapi, Vidavo & AUTH, Greece

Rob Finn, EMBL-EBI, UK

Ioannis Iliopoulos, U Crete [co-chair], Greece

Ilias Kappas, AUTH

Nikos Kyrpides, JGI/LBNL [co-chair], USA

Ilias Lagkouvardos, TUM, Germany

Claudine Médigue, CEA-CNRS, France

Folker Meyer, U Duisburg-Essen, Germany

Pericles Mitkas, AUTH, Greece

Anastasis Oulas, CING, Cyprus

Christos Ouzounis, AUTH, Greece & LBNL, USA

Evangelos Pafilis, HCMR, Greece

Eleftherios Panteris, AUTH, Greece

Giorgos Pavlopoulos, BSRC Fleming, Greece

Joan-Carles Pons Mayol, U Balearic Islands, Spain

Vasilis Promponas, U Cyprus [chair], Cyprus

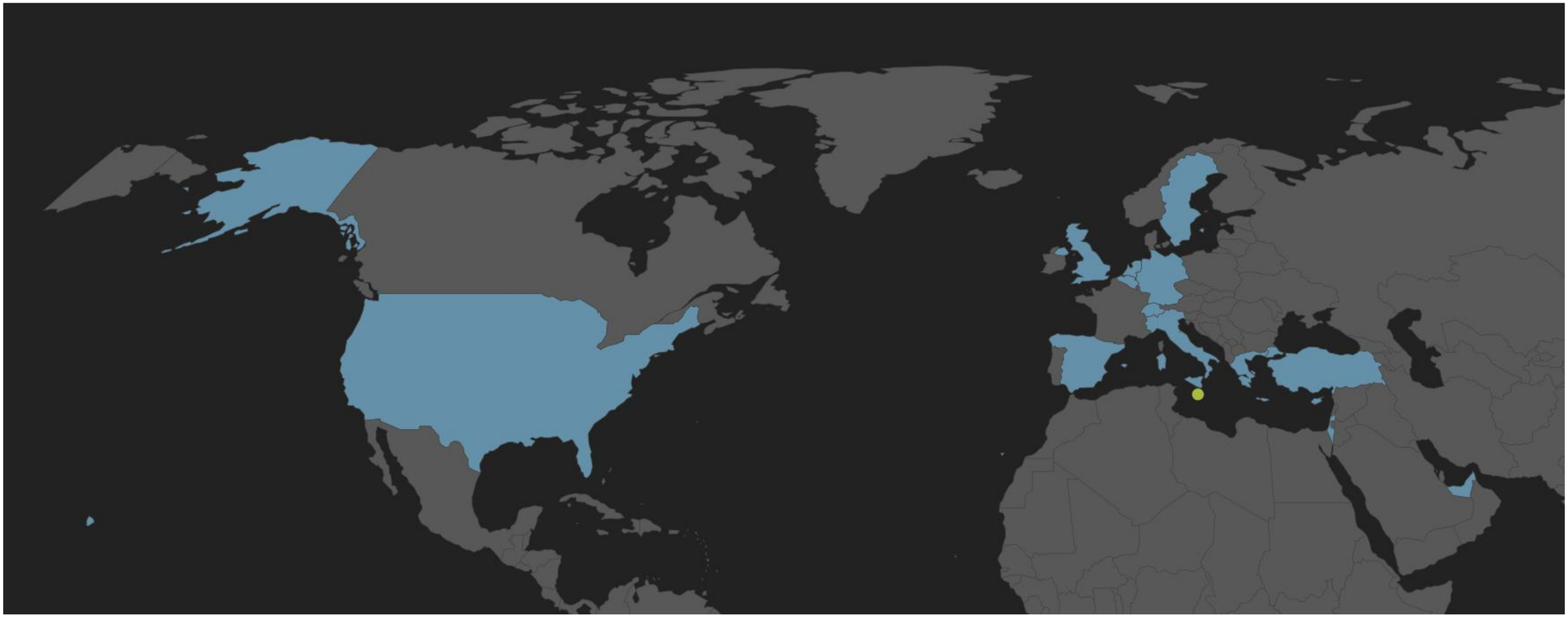
Lynn Schriml, U Maryland, USA

Zacharias Scouras, AUTH, Greece

Stella Tamana, CING, Cyprus

Metaxia Vlassi, NCSR Demokritos, Greece

••



invited speakers

Panagiotis Alexiou, Malta

Miguel Andrade, Germany

Pantelis Angelidis, Greece

Christos Arvanitidis, Spain

Stefano Campanaro, Italy

Vassily Chatzimanikatis, Switzerland

Arthur Declercq, Belgium

Anton Enright, UK

Rob Finn, UK

Milana Frenkel-Morgenstern, Israel

Manuela Helmer-Citterich*, Italy

Ylva Ivarsson, Sweden

Peter Karp*, USA

Eugene Koonin*, USA

Nikos Kyrpides, USA

Christine Orengo*, UK

George Patrinos, Greece

Evangelia Petsalaki, UK

Ioannis Pitas*, Greece

Nataša Pržulj, Abu Dhabi UAE

Ugur Sezerman, Turkey

Nektarios Tavernarakis*, Greece

Laura Treu, Italy

Alfonso Valencia*, Spain

HJG (Harmen) van de Werken, The Netherlands

Ioannis Xenarios*, Switzerland

*keynote lectures

Conference Statistics

Category	Count
Sessions	17
Workshops	2
Poster Session	1
Keynote Talks	8
Invited Talks	18
Short Talks	38
Posters-Only	18
Sponsor Session	1
Principal Investigators	100
Students	200
Participants	300

conference by numbers

17 sessions

2 workshops

1 continuous poster session

keynote talks: 8 — 8

invited talks: 18 — 26

total citations by invited speakers: > 1M!

short talks: 38 — 64

posters-only: 18 — 82

total: 82

silver sponsor session: 1

100 PIs, 200 students, 300 participants

26 speakers from 14 countries

271 authors (210 unique names) from 23 countries

Austria	3
Belgium	11
Cyprus	20
Denmark	3
France	1
Georgia	1
Germany	7
Greece	149
India	1
Israel	1
Italy	14
Malta	1
Netherlands	6
Pakistan	1
Saudi Arabia	1
Slovenia	3
Spain	5
Sweden	1
Switzerland	6
Turkey	1
United Arab Emirates	4
United Kingdom	10
United States	21

pre-conference workshops

may 15

pre-conference workshops

MAY 15 | morning

proteomics: 0900-1300

Brief history of MS-based proteomics

State-of-the-art quantitative MS-based proteomics approaches

Proteome Identification, Quantitation and Characterization – Available bioinformatics platforms for MS-data processing

Demonstration using MaxQuant

Quality control and filtering of proteomics data

Demonstration using Perseus

Comparative proteomics analysis

Functional proteome annotation

GO terms and pathways enrichment analysis

Protein interaction networks

Demonstration using String

Visualization of proteomics data

MAY 15 | evening

machine learning: 1400-1800

RNA-seq gene expression profiles (RPKM, CPM)

Drug response measurements (e.g. IC50, AUC)

Clinical annotations

Variants

Understand data pre-processing steps for transcriptomics data

Train multiple ML models (Elastic Net, KNN, Random Forest, Gradient Boosting Machines, Support Vector Machines) using caret

Evaluate model performance using cross-validation and test sets and several metrics (Confusion matrices, Accuracy, Area Under the ROC curve, etc)

Identify important features

Optionally explore deep learning and AutoML using h2o

program overview

may 16-18

program overview - day I

May 16 | Data Science

session 1: artificial intelligence

keynote: I Pitas » Artificial Intelligence, Biological Intelligence and Life

speaker: P Angelidis

keynote: N Tavernarakis » Nuclear autophagy augments somatic longevity and upholds germline immortality

session 2: structural genomics

keynote: C Orengo » AlphaFold massively expands CATH superfamilies giving insights into protein evolution and functional mechanisms

keynote: M Helmer-Citterich » Uncovering Biological Insights through RNA Sequence and Structure

session 3: other activities

sponsor session: Google Cloud

session 4: predictive analytics

speaker: M Frenkel-Morgenstern

session 5: emerging trends

speaker: M Andrade

keynote: A Valencia » Exploring the Limits and Limitations of Computation in Biology

program overview - day II

May 17 | Research Frontiers

session 6: biodiversity

speaker: C Arvanitidis

session 7: functional genomics

speaker: A Enright

session 8: phylogeny

keynote: E Koonin » *In silico* evolution of globular protein folds from random sequences

session 9: microbiome data science

speakers: N Kyrpides, R Finn

session 10: data integration

speaker: N Pržulj

session 11: systems biology, functional proteomics

speakers: U Sezerman, E Petsalaki

speakers: Y Ivarsson, A Declercq

program overview - day III

May 18 | Translation Challenges

session 12: bioprospecting

speakers: S Campanaro, L Treu

session 13: enzyme discovery

keynote: P Karp » The BioCyc Genome and Metabolic Pathway Web Portal

session 14: metabolic engineering

speaker: V Hatzimanikatis

session 15: pharmacogenomics

keynote: I Xenarios » Health2030 Genome Center@SMOC providing genomic medicine at scale

speaker: G Patrinos

session 16: precision medicine

speaker: H van de Werken

session 17: synthetic biology

speaker: P Alexiou

detailed program

may 16-18

day I • may 16

Day 1-FRI: Data Science (8 speakers + 4 contributions)

11:30-12:00	**Welcome Session**	
12:00-14:00	**Artificial Intelligence**	
12:00-12:40	01 Ioannis Pitas*, Greece	Artificial Intelligence, Biological Intelligence and Life
12:40-13:00	02 Pantelis Angelidis, Greece	Digital Aortic Twin to generate non-invasive Abdominal Aortic Aneurysm rupture risk biomarkers and predictive models
13:00-13:10	03 abstract	Contextualized Matrix Factorization-Based Embeddings Uncover Cell Markers and Housekeeping Genes
13:10-13:20	04 abstract	Machine learning can recover homologous relationships out of “junk” similarity search hits
13:20-14:00	05 Nektarios Tavernarakis*, Greece	Nuclear autophagy augments somatic longevity and upholds germline immortality
14:00-15:00	**Lunch + Poster setup**	
15:00-16:30	**Structural Genomics**	
15:00-15:40	06 Christine Orengo*, UK	AlphaFold massively expands CATH superfamilies giving insights into protein evolution and functional mechanisms
15:40-16:20	07 Manuela Helmer-Citterich*, Italy	Uncovering Biological Insights through RNA Sequence and Structure
16:20-16:30	08 abstract	*De novo* birth of transmembrane yeast proteins from intergenic polyA/T tracts
16:30-17:00	***Google Cloud: silver sponsor session***	
17:00-17:30	**Coffee break + Poster session**	
17:30-18:00	**Predictive Analytics**	
17:30-17:50	09 Milana Frenkel-Morgenstern, Israel	AI-Driven Diagnostics and Liquid Biopsy
17:50-18:00	10 abstract	Machine Learning-Based Staging and Progression Prediction in Multiple Myeloma Using Gene Expression Data
18:00-19:00	**Emerging Trends**	
18:00-18:20	11 Miguel Andrade, Germany	Order within disordered protein sequences: function, structure and evolution
18:20-19:00	12 Alfonso Valencia*, Spain	Exploring the Limits and Limitations of Computation in Biology

day II • may 17

10:00-11:00 **Biodiversity**

10:00-10:20 **13** Christos Arvanitidis, Spain

Bridging Disciplines to Confront Biodiversity Change: The Role of Science, Technology, and Innovation

10:20-10:30 **14** abstract

The use of eDNA metabarcoding for biodiversity monitoring: a comparative study with classical survey methods

10:30-10:35 **15** abstract

Identifying locally adapted forests at different spatial scales, using whole genome sequencing and linear mixed effects modelling

10:35-10:40 **16** abstract

BayPassAcorn: a Genotype–Environment Association analysis pipeline for Pool-Seq datasets

10:40-10:50 **17** abstract

A population genomics analysis of *Pinus nigra* populations with differential exposure to air pollution and climatic change

10:50-11:00 **18** abstract

GenTree SPET pipeline: An automated bioinformatics pipeline for genetic diversity statistics in population genomics

11:00-12:00 **Functional Genomics**

11:00-11:20 **19** Anton Enright, UK

Detection of RNA methylation events in mRNAs and microRNAs using Oxford Nanopore Sequencing

11:20-11:30 **decompression gap**

11:30-11:40 **20** abstract

scRNA-Explorer: an End-user Online Tool for Single Cell RNA-seq Data Analysis Featuring Gene Correlation and Data Filtering

11:40-11:50 **21** abstract

Machine Learning-Based Prediction of Promoter and RBS Sequence Strengths

11:50-12:00 **22** abstract

Transcriptomics analysis to unveil the modulation of ferroptosis biomarkers in the induced erythroid maturation of MEL cells

12:00-12:30 **Coffee Break + Poster session**

12:30-13:30 **Phylogeny**

12:30-13:10 **23** Eugene Koonin*, USA

In silico evolution of globular protein folds from random sequences

13:10-13:20 **24** abstract

PhyloFlask: a software framework for large-scale phylogenetic profile visualization

13:20-13:30 **25** abstract

Climate Change Metagenomic Record Index and Sample Matcher: Leveraging Metagenomic Data for Climate Change Research

13:30-14:30 **Lunch + Poster session**

14:30-15:30 **Microbiome Data Science**

14:30-14:50 **26** Nikos Kyrpides, USA

Microbiome Data Science: Unraveling the Dark Matter of Microbial Diversity

14:50-15:10 **27** Rob Finn, UK

Rapid and consistent genome clustering for navigating bacterial diversity

15:10-15:20 **28** abstract

Quadrupling the protein family space with global metagenomics

15:20-15:30 **29** abstract

Exploring the Genomic and Ecological Distinctions Between Isolates and MAGs

15:30-16:30 **Data Integration**

15:30-16:00 **30** Nataša Pržulj, Abu Dhabi UAE

AI for multi-omics data fusion to personalize medicine

16:00-16:10 **31** abstract

Integration of proteomics and transcriptomics data for the identification of novel biomarkers in mycosis fungoides

16:10-16:20 **32** abstract

Bridging Genotype and Phenotype by Machine Learning for Single-Molecule Multi-Omic Sequencing

16:20-16:30 **33** abstract

Early integration of imaging and omics data for stratifying Mild Cognitive Impairment in Alzheimer's disease

16:30-17:00 **Coffee break + Poster session**

17:00-19:00 **Systems Biology, Functional Proteomics**

17:00-17:20 **34** Ugur Sezerman, Turkey

Personalized Medicine in the era of multiomics data and AI

17:20-17:40 **35** Evangelia Petsalaki, UK

Integrative signalling network analysis to uncover mechanisms of disease

17:40-17:50 **36** abstract

Understanding the changes in cell-to-cell communication patterns during the progression of Multiple Myeloma: A computational approach

17:50-18:00 **37** abstract

A Network Biology Perspective on Multiple Myeloma: Integrating Single-Cell Transcriptomics with Prior Knowledge

18:00-18:10 **decompression gap**

18:10-18:30 **38** Ylva Ivarsson, Sweden

Towards a motif map of the human proteome

18:30-18:50 **39** Arthur Declercq, Belgium

How machine learning unravels the immunopeptidome, from analytics towards biology

18:50-19:00 **40** abstract

Computational Analysis of Fluorescence Recovery After Photobleaching (FRAP) Using FRAPedia: A Simple Graphical Approach

day III • may 18

Day 3-SUN: Translation Challenges (8 speakers + 16 contributions)

11:00-12:00 **Bioprospecting**

11:00-11:20 **41** Stefano Campanaro, Italy

Empowering microbial genome annotation with machine learning: insights from a global metagenomic database of anaerobic species

11:20-11:40 **42** Laura Treu, Italy

Multi-omics and metabolic modeling for understanding microbial cooperation in CO_2-driven methanogenesis

11:40-11:50 **43** abstract

From Genes to Function: The case of kiwifruit cold stress response

11:50-12:00 **44** abstract

Transposable element annotation and centromere analysis in *Brassicaceae*

12:00-13:00 **Enzyme Discovery**

12:00-12:40 **45** Peter Karp*, USA

The BioCyc Genome and Metabolic Pathway Web Portal

12:40-12:50 **46** abstract

PathoMagic: standardizing genome annotation with automated format conversion

12:50-13:00 **47** abstract

Deciphering the Anaerobic Microbiome for Constructing Robust Microbial Consortia to Produce Medium-Chain Fatty Acids

13:00-14:00 **Metabolic Engineering**

13:00-13:20 **48** Vassily Hatzimanikatis, Switzerland

TBA

13:20-13:30 **49** abstract

The impact of biochar on methane yield and microbial community dynamics under increasing organic loading rates in a continuous anaerobic digestion system

13:30-13:40 **50** abstract

Molecular Insights of Soil Microbial Communities in Response to Digestate Fertilization

13:40-13:50 **51** abstract

Comparative analysis of the gut microbiome among the five Greek species of the *Viperidae* family

13:50-14:00 **52** abstract

Predicting Anticancer Drug Efficacy with Machine Learning: A Data-Driven Approach

14:00-15:00 **Lunch + Poster session**

15:00-16:00 **Pharmacogenomics**

15:00-15:40 **53** Ioannis Xenarios*, Switzerland

Health2030 Genome Center@SMOC providing genomic medicine at scale

15:40-15:50 **54** abstract

In silico discovery of novel therapeutic targets that counteract cytokine-induced transcriptomic changes in iPSC-derived β cells associated with type 1 diabetes

15:50-16:00 **55** abstract

Integrative -Omics for Construction of Gene Regulatory Networks towards NKT Cell Identity Understanding

16:00-17:00 **Precision Medicine**

16:00-16:20 **56** George Patrinos, Greece

Translational tools and repositories for translational medicine in the post-genomic era

16:20-16:40 **57** HJG (Harmen) van de Werken, The Netherlands

Unveiling the Complex and Enigmatic Biology of Cancer of Unknown Primary Through Interrogation of the Genomic and Transcriptomic Landscape

16:40-16:50 **58** abstract

Uncertainty-Aware Molecular Subtyping in Cancer Using Conformal Machine Learning Models

16:50–17:00 **59** abstract

From point mutations to Brunner Syndrome: How the point mutation influences electrostatic interactions and performance of monoamine oxidase A enzyme

17:00-17:30 **Coffee break + Poster repatriation**

17:30-18:30 **Synthetic Biology**

17:30-17:50 **60** Panagiotis Alexiou, Malta

Agentomics-ML: an agentic system for automated -omics machine learning model development

17:50-18:00 **61** abstract

In silico antimicrobial peptide discovery in Animal genomes

18:00-18:10 **62** abstract

Exploring Genetic Adaptation and Microbial Dynamics in Engineered Anaerobic Ecosystems via Strain-level Metagenomics

18:10-18:20 **63** abstract

Bio-Electrical Nanonetworks: a New Paradigm of Biological Nanocommunication based on Living Cable Bacteria Filaments

18:20-18:30 **64** abstract

From Design to Validation and vice versa: Bioinformatics as a Catalyst in iGEM Synthetic Biology Projects

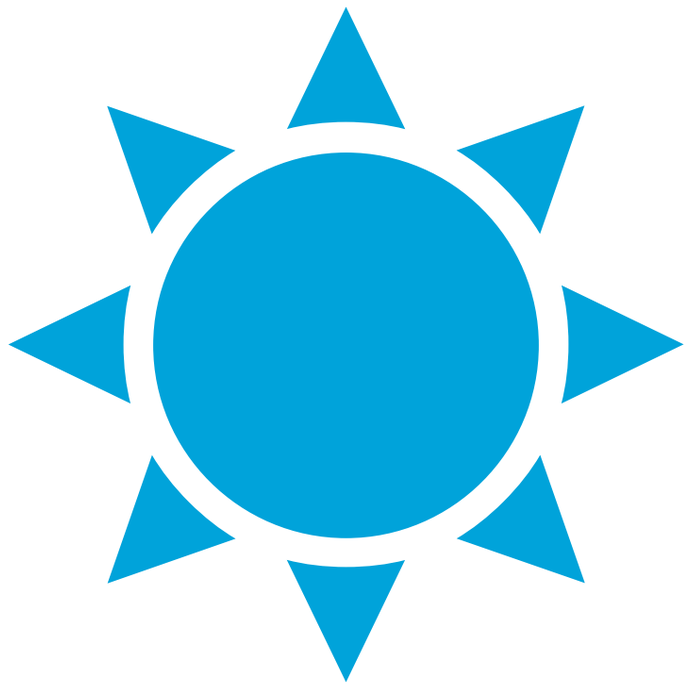
18:30-19:00 **Closing Session**

18:30-18:45 Awards

18:45-19:00 Closing remarks

abstracts

day I • may 16



#01: keynote lecture

Artificial Intelligence, Biological Intelligence and Life

Ioannis Pitas

School of Informatics

Aristotle University of Thessaloniki

Greece

##

#02: invited seminar

Digital Aortic Twin to generate non-invasive Abdominal Aortic Aneurysm rupture risk biomarkers and predictive models

Pantelis Angelidis

University of Western Macedonia, Vidavo

Greece

We develop a Digital Aortic Twin using computational tools and machine learning models to generate non-invasive biomarkers and predictive models for AAA rupture risk and geometric evolution. The Abdominal Aortic Aneurysm (AAA) is a silent and progressive aorta disease that has a high mortality rate and increasing prevalence along aging. The enstaging involves transformation of smooth muscle cells into an aberrant phenotype, with features like secretory phenotype, elevated differentiation marker miR-145, senescence marker SIRT-1, multi-nucleated and aberrant nuclear morphology, and higher levels of DNA damage marker γH2AX, which are potential indicators for pathological premature vascular aging. Most AAAs are asymptomatic until rupture, at which point severe abdominal or back pain, hypotension and a pulsatile abdominal mass occur. Regular screening is crucial to detect AAAs before rupture, as their growth may be rapid or slow / without symptoms. Diagnostic tests include CT/MR Angiography, Duplex Ultrasound, CT scans and clinical examination.

#03: short contribution

Contextualized Matrix Factorization-Based Embeddings Uncover Cell Markers and Housekeeping Genes

Alexandros Xenos, Barcelona Supercomputing Center, Carrer Jordi Girona, 29, 08034 Barcelona, Spain

Noël Malod-Dognin, Computational Biology Department, Mohamed bin Zayed University of Artificial Intelligence, 00000 Abu Dhabi, United Arab Emirates

Marinka Zitnik, Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, MA 02115, Boston, USA; Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Allston, MA 02134, Boston, USA

Nataša Pržulj, Computational Biology Department, Mohamed bin Zayed University of Artificial Intelligence, 00000 Abu Dhabi, United Arab Emirates; Department of Computer Science, University College London, Gower Street, WC1E 6BT, London, United Kingdom; ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

Email: alexandros.xenos@bsc.es

Contextual embeddings are the backbone of modern AI, driving transformative advancements in natural language processing. Unlike traditional word embeddings, contextual embeddings allow the same word to have different representations depending on the context in which it appears (e.g., sentence or passage). Similarly, proteins, much like words, can have different functions depending on the biological contexts in which they are expressed (e.g., cellular or tissue-specific contexts) [1]. In addition, gene expression and consequently, protein function varies significantly between healthy and diseased states [2]. Incorporating biological context into machine learning methods can deepen our understanding of protein function and provide precise, context-specific insights. While contextual embeddings have been applied in biomedicine using attention-based neural network (NN) models [3], they have not yet been generalized for non-NN-based methods.

Non-NN embedding methods, including matrix factorization techniques such as Non-negative Matrix Factorization (NMF), Non-negative Matrix Tri-Factorization (NMTF) and Principal Component Analysis (PCA) have been widely used in biology. Notably, non-negative matrix factorization methods allow for an intuitive and easy interpretation of the resulting matrix factors, making them particularly effective for omics data, where explainability and interpretability are essential [4]. Here, we propose a generic contextualization methodology applicable to any pre-trained embeddings, which we showcase on NMTF-based embeddings. We apply our method to contextualize NMTF-based context-free protein embeddings to generate cell-type-specific protein embeddings leveraging the Tabula Sapiens single-cell transcription atlas [5]. Our results demonstrate that contextualized NMTF embeddings better cluster proteins based on their cell types compared to context-free embeddings derived from standalone NMTF (applied to PPI or scRNA-seq data) and NMTF that jointly integrates scRNA-seq and PPI data.

#04: short contribution

Machine learning can recover homologous relationships out of “junk” similarity search hits

Emilios Tassios, emiliostassios@gmail.com

Jori de Leuw, jorideleuw77@gmail.com

Christoforos Nikolaou, cnikolaou@fleming.gr

Anne Kupczok, anne.kupczok@wur.nl

Nikolaos Vakirlis, nvakirlis@pasteur.gr

email: nvakirlis@pasteur.gr

Elucidating the evolutionary origin of species-specific orphan genes (orphans), lacking homologues outside of a given taxon, is important because orphans frequently underlie unique species traits1. Orphan genes can be the result of sequence divergence beyond recognition, when homologous proteins mutate to a point at which tools that rely on sequence similarity to establish homology can no longer identify them as such. Orphans can also result from other processes, including de novo gene birth from previously noncoding sequences, in which a homologous protein-coding gene truly does not exist2. Accurately distinguishing diverged orphans from the rest would enable us to recover cases of remote homology, get closer estimates of the evolutionary age of protein families, detect cases of rapid divergence and more generally better understand how genetic novelty arises.

Here we propose the hypothesis that orphans resulting from divergence might be recognizable from their patterns of non-statistically significant similarity hits which are usually overlooked. To test this, we simulated diverged orphan protein sequences using as seeds conserved proteins from the Unified Human Gastrointestinal Protein catalogue (UHGP)3 and trained four machine learning classifiers on simple features extracted from the similarity search tool’s DIAMOND output, like total query coverage or maximum bit score, with reversed protein sequences as negative controls. We tested the influence of evolutionary parameters such as simulation tree branch length, indel rate and among-site rate heterogeneity.

#04: short contribution

continued

We found that performance of the models depended on the simulation parameters with those trained on sequences simulated under less divergent scenarios reaching accuracies ~90%, while models trained on mixed parameter datasets had accuracies ~70%. The most important features in the classification were the number of alignments (hits) and the minimum hit E-value, while the overall importance order varied across models. When applying our classifier on a set of ~170,000 eligible orphans from the same UHGP dataset we found that ~30% of them are predicted to be divergent and they are shorter and more disordered than the rest. Our classifiers and pre-processing python scripts are available at https://github.com/emiliostassios/Classification-of-divergent-genes-using-ML and can be readily used as a computationally fast means to obtain a candidate set of diverged orphans from any similarity search output. Studying such orphans across the tree of life will help to disentangle and better understand the different evolutionary processes driving genetic novelty.

REFERENCES

1. Baalsrud, H. T. et al. De Novo Gene Evolution of Antifreeze Glycoproteins in Codfishes Revealed by Whole Genome Sequence Data. Mol. Biol. Evol. 35, 593–606 (2018).
2. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes. Nat. Rev. Genet. 12, 692–702 (2011).
3. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat. Biotechnol. 39, 105–114 (2021).

#05: keynote lecture

Nuclear autophagy augments somatic longevity and upholds germline immortality

Nektarios Tavernarakis

Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas

Medical School, University of Crete, Heraklion, Crete

Greece

Recycling of nuclear material is essential for cellular and organismal homeostasis. Phenotypic alterations in nuclear morphology, as well as, aberrant nuclear activity, dynamics and signaling are universal hallmarks of ageing and age-related pathologies. The crosstalk between autophagy and the nucleus in the context of physiology and pathology remains largely elusive. We investigated this potential association by dissecting the involvement of nuclear membrane components in the autophagic process. We show that selective autophagy of nuclear material is an important determinant of germline immortality and somatic ageing under conditions of stress. Impairment of autophagic recycling of nuclear components diminishes stress resistance, germline immortality and longevity. We find that the Caenorhabditis elegans nuclear envelope anchor protein, ANC-1 and its mammalian orthologue Nesprin 2 are key autophagy regulators that restrict nucleolar size, a common denominator of diverse lifespan extension regimes. We identify and characterize a novel germline immortality assurance mechanism, which involves nucleolar degradation at the most proximal oocyte by ANC-1. Clearance of aberrant germ cells during their differentiation by autophagic cell death requires ANC-1 and LGG-1. Notably, perturbation of this clearance pathway causes tumour-like structures in the C. elegans germline. Similarly, genetic ablation of Nesprin 2 in female mice causes ovarian carcinomas, indicating that the relevant molecular pathways are evolutionarily conserved, across distant phyla. Thus, autophagic recycling of nuclear envelope-associated and nucleolar components is an essential soma longevity and germline immortality mechanism that promotes youthfulness and delays ageing under conditions of stress, by preserving nuclear architecture and preventing nucleolar expansion.

#06: keynote lecture

AlphaFold massively expands CATH superfamilies giving insights into protein evolution and functional mechanisms

Christine Orengo

University College London

UK

##

#07: keynote lecture

Uncovering Biological Insights through RNA Sequence and Structure

Manuela Helmer-Citterich

Dept of Biology

University of Rome Tor Vergata

Italy

##

#08: short contribution

De novo birth of transmembrane yeast proteins from intergenic polyA/T tracts

Nikolaos Vakirlis1* & Timothy Fuqua2,3

1: Hellenic Pasteur Institute, Athens, Greece.
2: Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland
3: Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, Lausanne, Switzerland

email: n.vakirlis@pasteur.gr

Intergenic genomic regions have essential regulatory and structural roles that impose constraints on their sequences, but they also have the potential to evolve into protein-coding genes. De novo gene birth, as this process is called, is now being recognized as a potent force for genetic novelty. In budding yeast, computational predictions have shown that intergenic regions harbor a higher-than-expected propensity to encode transmembrane domains, and this propensity seems to be linked to the high prevalence of predicted transmembrane domains in evolutionarily young de novo genes. However, what accounts for this enriched propensity is not known. Here we show that specific arrangements of polyA/T tracts, which are abundant and enriched in yeast intergenic regions, explain this observation. We provide evidence that polyA/T tracts, which are known to act as Nucleosome Depleted Regions in a regulatory context, have been coopted through de novo gene emergence for the evolution of novel small genes encoding proteins with predicted transmembrane domains. Our findings contribute to our understanding of the process of de novo gene evolution and show how seemingly distinct but interacting levels of functionality can exist within the same genomic loci.

REFERENCES

Vakirlis, N. & Fuqua, T. Intergenic polyA/T tracts explain the propensity of yeast de novo genes to encode transmembrane domains. 2025.03.04.641575 Preprint at https://doi.org/10.1101/2025.03.04.641575 (2025).

Tassios, E., Nikolaou, C. & Vakirlis, N. Intergenic Regions of Saccharomycotina Yeasts are Enriched in Potential to Encode Transmembrane Domains. Molecular Biology and Evolution 40, msad059 (2023).

Vakirlis, N. et al. De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. Nat Commun 11, 781 (2020).

#09: invited seminar

AI-Driven Diagnostics and Liquid Biopsy

Milana Frenkel-Morgenstern

Scojen Institute for Synthetic Biology

8 University Street, Reichman, Herzliya

Reichman University

Israel

AI-driven diagnostics and liquid biopsy are transforming disease detection and monitoring, particularly for cancers and autoimmune conditions. In our study on glioblastoma patients, we integrated AI with advanced molecular analysis of cell-free DNA (cfDNA), circulating tumor DNA (ctDNA), and other biomarkers. This approach enabled earlier, more accurate, and non-invasive disease detection, paving the way for a breakthrough in brain tumor management.

#10: short contribution

Machine Learning-Based Staging and Progression Prediction in Multiple Myeloma Using Gene Expression Data

Nestoras Karathanasis, George M. Spyrou

Bioinformatics Department, The Cyprus Institute of Neurology & Genetics, 6 Iroon Avenue, Ayios Dometios, 2371 Nicosia, Cyprus

email: nestorask@cing.ac.cy

Accurate staging of multiple myeloma (MM) is crucial for optimizing treatment strategies. However, predicting the progression of asymptomatic patients—classified as having monoclonal gammopathy of undetermined significance (MGUS)—to symptomatic MM remains a challenge due to limited data1. This study aimed to develop machine learning models to improve MM staging accuracy and stratify asymptomatic patients based on their risk of progression2.

We utilized gene expression microarray datasets to train machine learning models, incorporating various data transformations and machine learning methods (ElasticNet, Random Forest, Gradient Boosting Machines, Support Vector Machines). For MM staging, models were trained to separate Healthy, MGUS and MM patients, on a single dataset and validated across five independent datasets, with performance assessed using multiclass area under the curve (AUC) metrics. To predict progression in asymptomatic patients, we employed two approaches: (1) training models on a dataset of asymptomatic patients who either progressed or remained stable, and (2) training models on multiple datasets combining asymptomatic and MM samples, then evaluating their ability to distinguish between stable and progressing asymptomatic cases. Feature selection and enrichment analyses were conducted to identify key signaling pathways associated with disease stages and progression.

The MM staging models demonstrated strong performance, with ElasticNet achieving consistent multiclass AUC values of 0.9 across datasets and transformations, indicating robust generalizability. For predicting progression in asymptomatic patients, both modeling approaches yielded similar results, with AUC values exceeding 0.8 across datasets and algorithms (ElasticNet, Boosting, and Support Vector Machines), highlighting their potential in assessing progression risk. Enrichment analyses identified key pathways involved in MM pathogenesis, including PI3K-Akt, MAPK, Wnt, and mTOR3.

#10: short contribution

continued

REFERENCES

1. Cowan, A. et al. Personalised progression prediction in patients with monoclonal gammopathy of undetermined significance or smouldering multiple myeloma (PANGEA): a retrospective, multicohort study. Lancet Haematol. 10, e203–e212 (2023).

2. Karathanasis, N. & Spyrou, G. M. Predicting the Progression from Asymptomatic to Symptomatic Multiple Myeloma and Stage Classification Using Gene Expression Data. Cancers (Basel). 17, (2025).

3. John, L., Krauth, M. T., Podar, K. & Raab, M. S. Pathway-directed therapy in multiple myeloma. Cancers (Basel). 13, 1–19 (2021).

#11: invited seminar

Order within disordered protein sequences: function, structure and evolution

Miguel Andrade

Faculty of Biology

Johannes Gutenberg University in Mainz

Germany

Intrinsically disordered regions in protein sequences contain motifs such as short tandem repeats and other regions of low complexity for example with only one or two amino acids. These regions evolve very fast and are a source of sequence variability. I will present several studies where we assess their evolution, amino acid and codon usage, functional associations and capability to induce structure.

#12: keynote lecture

Exploring the Limits and Limitations of Computation in Biology

Alfonso Valencia

ICREA Prof

Barcelona Supercomputing Center

Spain

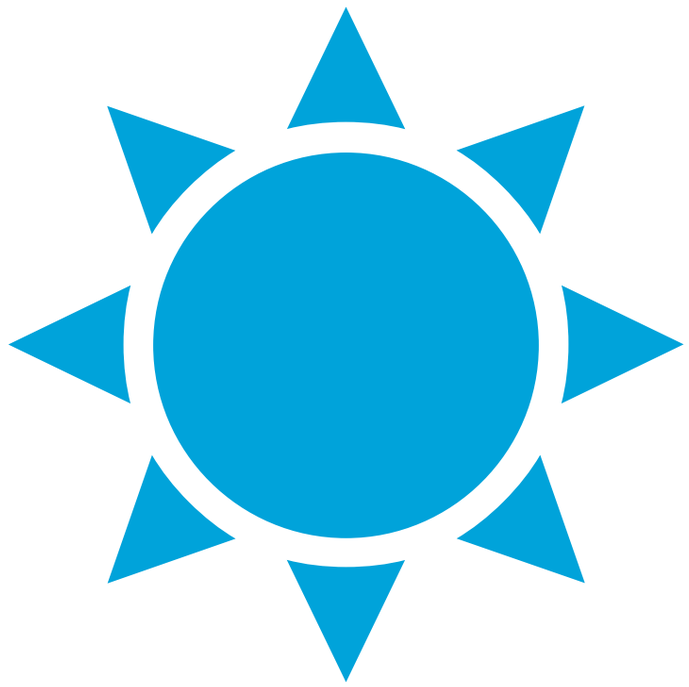
In this talk, I will explore what I consider the three key challenges facing computational biology today: computing capacity, data accessibility, and the ability to simulate biological processes at a mechanistic level.

First, I will discuss the limitations in computing capacity, focusing on recent advances in high-performance computing (HPC) in Europe—including the new European "AI factories." Next, I will address data accessibility, examining the ongoing development of the European Health Data Space and how its implementation in projects like GDI, EUCAIM, and EOSC4Cancer is poised to transform access to biological and health data. Federated systems, in particular, will play a crucial role in enabling secure and efficient data sharing across Europe and at the national level.

The third challenge is how to encode biological systems into models that provide mechanistic interpretations. I will introduce the concept of "digital twins" for cellular systems—dynamic models designed to replicate and predict cell behavior. While these models represent a major advancement, they also face significant technical and scientific hurdles.

Finally, I will examine how Artificial Intelligence (AI) is revolutionizing computational biology. AI is accelerating data analysis, uncovering novel insights, and enhancing predictive modeling. However, these systems are not able by themselves the scientific understanding at the mechanistic level we aim to—a critical issue that I will highlight in this talk.

day II • may 17



#13: invited seminar

Bridging Disciplines to Confront Biodiversity Change: The Role of Science, Technology, and Innovation

Christos Arvanitidis

LifeWatch (ESFRI)

Spain

The drivers of biodiversity change have been well understood for many years. Nations worldwide have committed to numerous conventions, and a substantial body of legislation is currently being implemented across global, European, regional, and national levels. Despite these efforts, significant progress is still needed to reverse the negative trends in biodiversity loss, which profoundly affect ecosystem services and the societal goods and benefits they provide.

Recognizing this urgency, the United Nations has identified science, technology, and innovation as critical levers for delivering evidence-based and technologically advanced solutions to biodiversity challenges. In alignment with this vision, the European Union is working to unlock the full potential of research and innovation through the development of a dynamic ecosystem of Research Infrastructures and the European Open Science Cloud (EOSC). Substantial resources have been dedicated to supporting this initiative.

Addressing complex societal challenges such as biodiversity change requires a holistic approach, which can integrate a wide range of scientific disciplines and engages diverse stakeholders. The most promising environment for this integration lies in the collaborative interfaces between disciplines and domains, often referred to as "trading zones". The Research Infrastructures and the EOSC ecosystem are uniquely positioned to catalyze this process, offering advanced technologies that bridge disciplinary boundaries and accelerate the collective response to biodiversity change.

#14: short contribution

The use of eDNA metabarcoding for biodiversity monitoring: A comparative study with classical survey methods

Styliani Minoudi1,2, Maria Demertzioglou3, Anna Mastrogianni4, Dimitra-Lida Rammou1,5, Dionisios Youlatos5,6, Evangelia Michaloudi3, Alexandros Triantafyllidis1,2, Ioannis Tsiripidis4, Konstantinos Sagonas2,5

1Laboratory of General Biology, Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, GR-54124 Thessaloniki Greece

2Genomics and Epigenomics Translational Research (GENeTres), Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Balkan Center, GR-57001, Thessaloniki, Greece

3Laboratory of Ichthyology, Department of Zoology, School of Biology, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

4Laboratory of Systematic Botany and Phytogeography, Department of Botany, School of Biology, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

5Laboratory of Marine and Terrestrial Animal Diversity, Department of Zoology, School of Biology, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

6International Center for Biodiversity and Primate Conservation, Dali University, China

Email: sminoudi@bio.auth.gr

Biodiversity loss remains a critical global challenge, necessitating reliable biomonitoring strategies. Traditional survey methods, though widely used, often struggle with detecting cryptic species, early life stages and rare or low-density taxa, including endangered species. Environmental DNA (eDNA) metabarcoding has emerged as a promising alternative, offering rapid, non-invasive, high-sensitivity biodiversity assessment. This study evaluates and compares classical biodiversity monitoring methods with eDNA metabarcoding across eight sites in Lake Volvi (GR1220001), encompassing aquatic and terrestrial ecosystems.

#14: short contribution

continued

Seasonal field surveys in 2023 documented the biodiversity of the area using conventional techniques for selected taxonomic groups, including zooplankton, vascular plants, amphibians, reptiles, and wild mammals, while concurrent water and soil samples were collected for eDNA analysis. Molecular assessment of vertebrate taxa was performed by targeting the 12S rRNA gene and the COI for invertebrates. Herein, results concerning animal taxa are presented, while analyses for plant taxa are still in progress. Amplicon libraries were prepared and sequenced on an Illumina NovaSeq platform.

We targeted 2 million reads per sample, ensuring sufficient sequencing depth to ensure the detection of rare species. A custom bioinformatics pipeline was performed, using VSEARCH tool to process amplicon sequence variants (ASVs) for taxonomic identification. To carry out the analysis we used the Aristotle University of Thessaloniki (AUTh) high-performance computing resources. Preliminary results revealed both overlapping and unique species detections between the two approaches, highlighting their complementary strengths. While eDNA metabarcoding successfully identified taxa that remained undetected through classical surveys, some challenges such as reference databases completeness and contamination risks remain. In addition, especially regarding soil samples, traditional methods seem to outperform eDNA methodology. These findings underscore the potential of integrating molecular tools into biodiversity monitoring, advocating for a multi-faceted approach to enhance conservation planning and ecosystem management.

#15: short contribution

Identifying locally adapted forests at different spatial scales, using whole genome sequencing and linear mixed effects modelling

Nikolaos Tourvas, School of Forestry and Natural Environment, AUTh

Simon Jansen, Universität für Bodenkultur Wien

Lars Littmann, Forest Research Institute Baden-Württemberg

Devrim Semizer-Cuming, Forest Research Institute Baden-Württemberg

Colette Blyth, Universität für Bodenkultur Wien

Deborah Marie Leigh, Swiss Federal Research Institute

Christian Rellstab, Swiss Federal Research Institute

Charalambos Neophytou, Forest Research Institute Baden-Württemberg

Filippos Aravanopoulos, School of Forestry and Natural Environment, AUTh

email: aravanop@for.auth.gr

Given the rapid pace of climate change, the question is raised whether forest ecosystems will be able to adapt to changing environmental conditions. Translocating locally adapted forest reproductive material (FRM) to their expected climatic optimum, known as assisted migration/gene flow (Aitken & Whitlock 2013) has been proposed as a promising strategy to ensure their resilience. However, any large-scale geographical transfer of FRM involves risks. For example, the extent of climate change (and hence the distance between source populations and target areas) is uncertain, differences due to local microenvironments may outweigh latitudinal climate differences, and some environmental parameters (e.g. photoperiod) will not change in the future. In addition, outbreeding depression and adaptation to biotic factors such as symbionts or pathogens may pose problems that have not been adequately investigated in large-scale FRM transfer. Consequently, accounting for environmental variation at small spatial scales in studies of local adaptation is crucial, as it has the potential to reveal adapted seed sources, allowing for the translocation of FRM over short geographical distances from arid areas to more humid locations while limiting the above mentioned risks.

#15: short contribution

continued

In this study, the genetic diversity of three white oak species (Quercus petraea, Q. pubescens, Q. robur) with a wide ecological and geographical range was investigated. The sampling strategy included pairs of populations within a short geographical distance (< 10 km) from Central Europe and the Eastern Mediterranean basin. Populations in each pair differed in terms of soil water availability environment (high/low availability). A total of 114 populations were sampled and about 20 DNA samples from each were pooled and underwent whole-genome sequencing. Using linear mixed-effects models, we investigated the presence of allele frequencies correlation with 29 environmental variables associated with drought at the European continent scale. Furthermore, utilizing our paired sampling design, correlations between genomic regions and “arid”/”humid” ecotypes were investigated at small spatial scales. Results are expected to assist in identifying climate-adapted provenances for seed collection and implementation of breeding programs. This research highlights the importance of studying local adaptation in developing effective strategies for managing forest genetic resources in the face of climate change challenges.

REFERENCES

Aitken SN, Whitlock MC (2013) Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu Rev Ecol Evol Syst 44:367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Acknowledgements

This research received funding from Biodiversa and the General Secretariat of Research and Innovation (GSRT), Greece through project “Identifying seed sources for highly adaptable oak forests in a changing climate (ACORN)” and from GSRT through the Aristotle University of Thessaloniki project “Genetic MonCon” (Project. No. 74167).

#16: short contribution

BayPassAcorn: A Genotype–Environment Association analysis pipeline for Pool-Seq datasets

Nikolaos Tourvas, School of Forestry and Natural Environment, AUTh

Simon Jansen, Universität für Bodenkultur Wien

Devrim Semizer-Cuming, Forest Research Institute Baden-Württemberg

Christian Rellstab, Swiss Federal Research Institute

Charalambos Neophytou, Forest Research Institute Baden-Württemberg

Filippos Aravanopoulos, School of Forestry and Natural Environment, AUTh

email: tourvasn@for.auth.gr

Local adaptation is the widespread property of natural populations to exhibit high fitness in their environment. Due to the rapid progress of climate change though, populations may become maladapted to their habitat in the near future. Genotype–Environment Association (GEA) analysis provides a way for uncovering the genetic basis of local adaptation. Hence, it can act as an early warning prediction method for ecosystems at risk (Rellstab et al. 2016). and inform conservation strategies such as assisted migration (Aitken & Whitlock 2013), where populations are translocated to their expected climate optimum.

BayPassAcorn (https://github.com/nikostourvas/BayPassAcorn) is a Snakemake-based pipeline designed to perform GEA analyses using the software BayPass (Gautier 2015) on genomic Pool-Seq data. The utilized software is containerized via Singularity to guarantee reproducibility and portability across diverse computing environments. The pipeline integrates efficient data processing by partitioning large-scale genomic datasets for near-linear multi-core performance scaling and conducts univariate association tests for the user provided environmental data, pinpointing candidate loci that are correlated with environmental clines. Population ecotypes can also be defined to allow for a contrast analysis. Additionally, an option is available to rank-transform non-normally distributed environmental datasets.

#16: short contribution

continued

A standout feature of BayPassAcorn is the ability to apply the window-based Weighted-Z Analysis (Booker et al. 2023) on the BayPass-generated Bayes Factors. This approach aggregates signals from closely linked SNPs within genomic windows, thereby enhancing the power to detect subtle selection signals. The pipeline also generates comprehensive diagnostic visual outputs by employing ggplot2/base R functions. These include p-value histograms/QQ-plots, Manhattan plots and scatter plots of allele frequencies versus environmental values, ensuring robust evaluation of model performance.

Overall, BayPassAcorn offers a scalable, reproducible, and user-friendly solution for GEA studies. By automating data preprocessing, executing advanced statistical analyses, and delivering intuitive diagnostic outputs, it is expected to assist biodiversity researchers' work on elucidating the architecture of local adaptation in natural ecosystems.

REFERENCES

Aitken SN, Whitlock MC (2013) Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu Rev Ecol Evol Syst 44:367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Booker TR, Yeaman S, Whiting JR, Whitlock MC (2024) The WZA : A window-based method for characterizing genotype–environment associations. Molecular Ecology Resources 24:e13768. https://doi.org/10.1111/1755-0998.13768

Gautier M (2015) Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. Genetics 201:1555–1579. https://doi.org/10.1534/genetics.115.181453

Rellstab C, Zoller S, Walthert L, et al (2016) Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. Molecular Ecology 25:5907–5924. https://doi.org/10.1111/mec.13889

Acknowledgements

This research received funding from Biodiversa and the General Secretariat of Research and Innovation (GSRT), Greece through project "Identifying seed sources for highly adaptable oak forests in a changing climate (ACORN)" and from GSRT through the Aristotle University of Thessaloniki project "Genetic MonCon" (Project. No. 74167).This research received partial funding from the General Secretariat of Research and Innovation, Greece, through the Aristotle University of Thessaloniki project "Genetic MonCon" (Project. No. 74167).

#17: short contribution

A population genomics analysis of *Pinus nigra* populations with differential exposure to air pollution and climatic change

Vasiliki Maria Kotina, Aristotle University of Thessaloniki

Nikolaos Tourvas, Aristotle University of Thessaloniki

Fani Lyrou, Aristotle University of Thessaloniki

Filippos A. Aravanopoulos, Aristotle University of Thessaloniki

email: aravanop@for.auth.gr

Forests are invaluable ecosystems that harbor a very large part of terrestrial biodiversity. Hence, it is paramount to investigate the capacity of keystone forest tree species that support ecosystem functions and services to survive under the stress associated to extreme environmental changes, such as the combined effects of air pollution and climate change. Black pine (Pinus nigra L.) is a well-known bioaccumulator and is widely used as a biomarker for monitoring air pollution. This study focused on black pine population genomics by employing the advanced targeted genotyping approach SPET (Single Primer Enrichment Technology), that relies on the sequencing of a region flanking a single primer to detect genomic variation of interest (Farsakoglou et al. 2020, 2024). The research was conducted on natural populations under a paired sampling factorial design using population pairs representing different conditions of air pollution burden (polluted and control), different age cohorts, different altitudinal species natural distribution limits, and different latitudinal distribution (Katsidi et al. 2023). After SPET sequencing, 83108 SNPs were captured. After extensive filtering (minor allele frequency (MAF) filter, biallelic filter, minimum depth filter, quality score filter, missing data filter, paralog filtering), 5616 SNPs remained for downstream analysis. Preliminary results indicate that overall genetic diversity does not differ significantly among populations, although polluted subpopulations tend to exhibit higher heterozygosity. Furthermore, principal component analyses (PCA) were performed, comparing polluted and control subpopulations (within low/high altitudinal distribution limits, and within age cohorts). In all analyses polluted populations displayed a higher number of outliers, whereas control populations assembled more distinct clusters. These outliers may suggest potential loci under selection and are further investigated to unravel the complex pattern of genomic backgrounds and potential genetic control associated to this study.

#17: short contribution

continued

REFERENCES

Farsakoglou AM, I. Scotti, B. Fady & FA Aravanopoulos 2024. Geographical patterns of forest genetic diversity to tackle conservation needs: insights from genomics approaches. In Proc. IUFRO World Congress, Stockholm.

Farsakoglou AM, Scotti I, the GenTree Consortium and FA Aravanopoulos 2020. Genome-wide genetic diversity and differentiation of 12 major European tree species elucidates conservation needs. In: Intern. Sci. Conf.: Genetics to the rescue – managing forests sustainably in a changing environment, 27-31 January, Avignon, France. https://www.canal-u.tv/chaines/au/gentree-scientific-conference-genetics-to-the-rescue.

Katsidi EC, Avramidou EV, Ganopoulos I, Barbas E, Doulis A, Triantafyllou A & FA Aravanopoulos FA 2023. Genetics and epigenetics of Pinus nigra populations with differential exposure to air pollution. Frontiers in Plant Science 14:1139331. https://doi.org/10.3389/fpls.2023.1139331

This research is supported by the research project “Investigation of key genetic and epigenetic processes affecting genome diversity in forest trees exposed to lignite pollutants and climate change-Potential for early warning” (GeneAlert), (PI: Prof. F.A. Aravanopoulos), and it is funded by the Hellenic Foundation for Research and Innovation.

#18: short contribution

GenTree SPET pipeline: An automated bioinformatics pipeline for genetic diversity statistics in population genomics

Anna-Maria Farsakoglou, School of Forestry and Natural Environment, AUTh

Ivan Scotti, Ecologie des Forêts Méditerranéennes (URFM), INRAE, France

Bruno Fady, Ecologie des Forêts Méditerranéennes (URFM), INRAE, France

Filippos Aravanopoulos, School of Forestry and Natural Environment, AUTh

Email: afarsakog@for.auth.gr

Forest trees are ecologically dominant species spanning across different landscape ecosystems, and due to their large genomes they remain mostly unexplored. Recent advanced genomic technologies, such as single primer enrichment technology (SPET), genotyping-by-sequencing (GBS), and exome capture sequencing have allowed us to produce thousands of single nucleotide polymorphism (SNP) datasets, and thus, dive into genetic diversity at a genome-wide scale. However, the size and complexity of these SNP datasets make their analysis quite challenging since they require advanced computational resources and expertise. Here, we introduce the GenTree SPET bioinformatics pipeline, constructed within the H2020 GenTree project, designed for efficient and user-friendly analysis of large SNP datasets (SNP-sets) in population genomics.

The pipeline integrates R and UNIX environments (free and open source), combining R scripts with bash routines for haplotype phasing. We designed and tested the pipeline on five perennial tree species (Abies alba, Pinus cembra, Pinus halepensis, Pinus nigra, and Taxus baccata). The user provides input through a Variant Call Format (.vcf) file and a population identifier (.txt) file. The pipeline can run the same analysis for up to three different SNP-sets: full-SNP-set (containing all the SNPs of the vcf file), single-SNP-set (by selecting the SNP of each contig with the lowest levels of missing data), and haplotypes (by conducting a haplotype phasing analysis, using the fastPhase software (Scheet & Stephens, 2006)). For each SNP-set selected, the pipeline estimates genetic diversity statistics for each population such as: mean and standard deviation (sd) of gene diversity within populations (HS), expected and observed heterozygosity (He, HO), and inbreeding coefficient (FIS) (Goudet, 2005). Also, it estimates genetic differentiation (FST(i)) of one population compared to all the others, mean, sd and deviation from the mean (δ) of θ diversity (Paradis & Schliep, 2019), Site Frequency Spectrum (SFS) parameters, and contemporary effective population size (Ne) (Do et al., 2014). The use of the pipeline requires intermediate R and UNIX programming skills. The outputs are provided in CSV tables and PDF graphs that allow the user to directly perform further downstream analyses, graphical visualization and interpretation. The GenTree SPET pipeline can serve as an accessible, accurate, and efficient tool for analyzing SNP datasets, since it can estimate genetic diversity statistics in an automatic, time- and cost-effective way, for advanced population genomic research.

#18: short contribution

continued

REFERENCES

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re- implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources, 14(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes, 5(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Scheet, P., & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. The American Journal of Human Genetics, 78(4), 629–644. https://doi.org/10.1086/502802

#19: invited seminar

Detection of RNA methylation events in mRNAs and microRNAs using Oxford Nanopore Sequencing

Anton Enright

Department of Pathology

University of Cambridge

UK

Nanopore sequencing technology such as Oxford Nanopore sequencing is one of the few methods available to directly sequence RNA. Non-coding RNAs such as microRNAs are difficult to quantitate with illumina sequencing due to significant biases that occur during PCR amplification and ligation of sequencing adapters on small molecules. Direct sequencing of microRNAs would allow us to develop novel and accurate diagnostic platforms that would have improved accuracy and sensitivity for microRNA detection. Additionally, direct sequencing allows the possibility to detect RNA modifications and methylation events which are now known to be an important part of the human epitranscriptome. Modifications such as m6A and m5C and Pseudouridine can significantly alter RNA biogenesis, structure and function. In this talk I will go through our labs efforts both experimentally and computationally to both detect microRNAs directly using Oxford Nanopore devices and to attempt to detect and resolve individual RNA modifications.

#20: short contribution

scRNA-Explorer: An End-user Online Tool for Single Cell RNA-seq Data Analysis Featuring Gene Correlation and Data Filtering

Ismini Baltsavia, Division of Basic Sciences, University of Crete Medical School, Heraklion 71110, Greece

Anastasis Oulas, Cyprus Institute of Neurology and Genetics, Bioinformatics Department, P.O.Box 23462, 1683 Nicosia, Cyprus

Theodosios Theodosiou, Division of Basic Sciences, University of Crete Medical School, Heraklion 71110, Greece

Matthieu D Lavigne, IMBB, FORTH, 71003 Heraklion, Crete, Greece

Evangelos Andreakos, George Mavrothalassitis, Division of Basic Sciences, University of Crete Medical School, Heraklion 71110, Greece; IMBB, FORTH, 71003 Heraklion, Crete, Greece

Ioannis Iliopoulos, Division of Basic Sciences, University of Crete Medical School, Heraklion 71110, Greece

email: ibaltsavia@gmail.com

scRNA-Explorer is an open-source online tool developed to address critical issues in the downstream analysis of scRNA-sequencing data by providing a simplified user interface. Prior to any kind of analysis, quality assessment of single-cell sequencing is of critical importance. This tool provides an intuitive interface to explore the overall quality of cells under study using various metrics. Visualizing these metrics in interactive plots greatly assists in defining appropriate thresholds and filtering out non-informative cells. The number of UMIs and genes captured per cell, as well as the number of UMIs per gene, represent sequencing depth, while mitochondrial and hemoglobin ratios indicate the presence of dying or unsuitable cells in the assay. These metrics, along with: i) the most highly expressed features across single cells, ii) the correlation between UMIs and genes, with mitochondrial and hemoglobin ratios projected, and iii) the relative contribution of each cell to the total library size for the 300 most highly expressed genes, are plotted before and after filtering, providing a more comprehensive overview of the remaining cells for subsequent analysis. Once unwanted cells have been excluded, a correlation analysis is performed. Using a selected gene as a “bait,” all genes correlated with this bait gene are identified. Correlated genes are then evaluated for significance using a Wilcoxon test to exclude genes with similar expression distributions in cells where the bait gene is expressed versus those where it is not. The pipeline can be applied to all single cells in the assay or to specific cell types. Cell type clustering is highly customizable, allowing users to adjust parameters at multiple steps during cell type annotation. Finally, enriched biological functions associated with correlated, anti-correlated, or all correlated genes are identified, highlighting potential underlying biological mechanisms. scRNA-Explorer can be accessed at https://bioinformatics.med.uoc.gr/shinyapps/app/scrnaexplorer.

#21: short contribution

Machine Learning-Based Prediction of Promoter and RBS Sequence Strengths

Antonios Tsiorvas, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Panagiotis Boutos, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

Michalis Aivaliotis, Laboratory of Biological Chemistry, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

email: atsior@ece.auth.gr

The strengths of both promoters and Ribosome Binding Sites (RBS) play a critical role in genetic engineering and synthetic biology, as they influence gene expression levels. Traditional laboratory methods for quantifying promoter and RBS strengths are often time-consuming, labor-intensive, and resource-intensive, relying on complex experimental assays and bioinformatics analyses. As synthetic biology increasingly intersects with computational tools, machine learning offers a powerful tool, enabling more efficient, scalable, and automated approaches.

To address this challenge, the Dry Lab team of iGEM Thessaloniki 2024, in the framework of “NeuroMusketeer”, developed two machine learning models to predict the strength of prokaryotic genetic elements [1]. The first one is the Promoter Strength Prediction Model [2], a deep learning regression model based on a convolutional neural network (CNN) architecture, enhanced with residual connections. This model analyzes E. coli promoter sequences, using the one-hot-encoding method, in order to detect regulatory motifs and estimate their strength, which is defined by the level of downstream gene expression. The second model is the RBS Strength Prediction Model [3], a support vector machine (SVM) classifier that categorizes RBS sequences into strong, medium, or weak classes, based on RBS sequence analysis, using k-mer encoding.

This presentation will showcase the full development pipeline of both models, including data preprocessing, model architecture and performance evaluation. We will also discuss potential real-world applications and future improvements of the models, emphasizing how such tools can accelerate synthetic biology workflows and reduce experimental overhead.

Keywords: machine learning, synthetic biology, promoter strength, RBS strength

#21: short contribution

continued

REFERENCES

iGEM Thessaloniki 2024 Wiki, Software Page, https://2024.igem.wiki/thessaloniki/software

iGEM Thessaloniki 2024 GitLab Repository, Promoter Strength Prediction Model, https://gitlab.igem.org/2024/software-tools/thessaloniki/-/tree/mai n/Software1_PSP%20Model?ref_type=heads

iGEM Thessaloniki 2024 GitLab Repository, RBS Strength Prediction Model, https://gitlab.igem.org/2024/software-tools/thessaloniki/-/tree/mai n/Software2_RSP%20Model?ref_type=heads

#22: short contribution

Transcriptomics analysis to unveil the modulation of ferroptosis biomarkers in the induced erythroid maturation of murine erythroleukemia (MEL) cells

Aliki Papadimitriou-Tsantarliotou1, Aristeidis Sionakidis2, Andigoni Malousi3, Chrysostomos Avgeros1, Ioannis S. Vizirianakis1,4

1Laboratory of Pharmacology, School of Pharmacy, Aristotle University of Thessaloniki, Greece

2Precision Breast Cancer Institute, Department of Oncology, University of Cambridge, UK

3Laboratory of Biological Chemistry Faculty of Medicine, Aristotle University of Thessaloniki, Greece.

4Department of Health Sciences, School of Life and Health Sciences, University of Nicosia, Cyprus

Keywords: Erythroid maturation, ferroptosis, chemical inducers, transcriptomics, biomarkers.

Erythroid differentiation is a strictly orchestrated process characterized by distinct developmental changes in both the morphology and gene expression profiles of committed erythroid progenitors during terminal maturation (1). The murine erythroleukemia (MEL)-745 cell line is a well-established in vitro model for erythroid differentiation, as these cells are arrested at a proerythroblast stage between Burst Forming Unit (BFU-E) and Colony Forming Unit (CFU-E). Some chemical agents, the so-called chemical inducers such as Hexamethylene Bisacetamide (HMBA), have been shown to effectively induce terminal erythroid maturation of these cells in culture (2). Interestingly, recent bioinformatic analysis suggests a significant connection between the regulation of ferroptosis-related genes and the induced erythroid differentiation in a population of CD34+ cells (3). Moreover, the potential role of ferroptosis in inherited ribosome-related bone marrow disorders has been recently proposed (). To further explore the potential relationship between the erythroid maturation program and the transcriptional regulation of ferroptosis-related genes, MEL cells were induced to differentiate in intro by the addition of HMBA (5 x 10-3M) and the total RNA was harvested at various time points (0, 24, 48, and 72 hours). Each sample underwent RNA sequencing via Illumina technology, while transcriptomic analysis was conducted using the R package DEGseq2 to visualize the statistically significant differences between treated and untreated groups at various time intervals. The results confirmed an increase in the expression level of several ferroptosis biomarkers (e.g. TFR1, DMT1, FP1, SLC3A2) at the transcript level, accompanied by a decrease in the expression of key inhibitors of ferroptosis, such as SLC7A11, FSP1 and FLVCR1. These observations were further validated through quantitative PCR (qPCR) analysis. In conclusion, these data show that ferroptosis related genes may play a crucial role in erythroid differentiation, however, additional research is required to elucidate the precise underlying mechanisms.

#22: short contribution

continued

REFERENCES

1. Ghosh A, Coffin M, West R, Fowler VM. Erythroid differentiation in mouse erythroleukemia cells depends on Tmod3-mediated regulation of actin filament assembly into the erythroblast membrane skeleton. FASEB J. 2022;36(3):e22220. doi: 10.1096/fj.202101011R. PMID: 35195928.

2. Vizirianakis IS, Tsiftsoglou AS. Blockade of murine erythroleukemia cell differentiation by hypomethylating agents causes accumulation of discrete small poly(A)– RNAs hybridized to 3'-end flanking sequences of βmajor globin gene, Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2005;1743(1–2):101-114. doi.org/10.1016/j.bbamcr.2004.09.003.

3. Tsiftsoglou AS, Vizirianakis IS, Strouboulis J. Erythropoiesis: model systems, molecular regulators, and developmental programs. IUBMB Life, 2009;61(8):800-30. doi: 10.1002/iub.226.

4. Liu Q, Lin Z, Yue M, Wu J, Li L, Huang D, Fang Y, Zhang X, Hao T. Identification and validation of ferroptosis related markers in erythrocyte differentiation of umbilical cord blood-derived CD34+ cell by bioinformatic analysis. Front Genet. 2024 Jul 30;15:1365232. doi: 10.3389/fgene.2024.1365232.

5. Papadimitriou-Tsantarliotou A, Avgeros C, Konstantinidou M, Vizirianakis IS. Analyzing the role of ferroptosis in ribosome-related bone marrow failure disorders: From pathophysiology to potential pharmacological exploitation. IUBMB Life, 2024;76(12):1011-1034. doi: 10.1002/iub.2897.

#23: keynote lecture

In silico evolution of globular protein folds from random sequences

Eugene V. Koonin

Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

The origin and evolution of protein folds are among the most challenging, long-standing problems in biology. Although many plausible scenarios of early protein evolution leading to fold nucleation have been proposed, realistic simulation of this process was not feasible because of the lack of efficient approaches for protein structure prediction, a situation that changed with the advent of powerful tools for fast and robust protein structure prediction, such as AlphaFold and ESMFold. I will present a recently developed computational approach for protein fold evolution simulation (PFES) with atomistic details that provide insights into the mechanisms of evolution of globular folds from random amino acid sequences. PFES introduces random mutations in a population of protein sequences, evaluates the effect of mutations on protein structure, and selects a new set of proteins for further evolution. Repeating this process iteratively allows tracking the evolutionary trajectory of a changing protein fold that evolves under selective pressure for protein fold stability, interaction with other proteins, or other features shaping the fitness landscape. PFES was employed to show how globular protein folds could evolve from random amino acid sequences as monomers or in complexes with other proteins. The simulations reproduce the evolution of several simple folds of natural proteins as well as the evolution of a variety of distinct folds not known to exist in nature. Evolution of a stable fold from random sequences, on average, takes about one, and in some simulations, as few as 0.2 amino acid replacement per site, which is comparable to empirical data on protein evolution. Thus, the results of these computational experiments suggest that simple but stable protein folds can evolve relatively easily. These findings could shed light on the enigma of the rapid evolution of protein fold diversity at the earliest stages of life evolution. PFES tracks the complete evolutionary history from simulations that describes intermediate states at the sequence and structure levels and can be used to test a broad variety of hypotheses on protein fold evolution.

https://www.biorxiv.org/content/10.1101/2024.11.10.622830v1

#24: short contribution

PhyloFlask: a software framework for large-scale phylogenetic profile visualization

Alexandros Michailidis, Vasileios S Papagrigoriou, Christos A Ouzounis

BCCB Group, Artificial Intelligence & Information Analysis Laboratory, School of Informatics, Aristotle University of Thessaloniki, Thessalonica, Greece

Phylogenetic profiling is a computational approach for inferring structural and functional properties of genes based on their presence/absence patterns across complete genomes. As genome sequencing continues to expand, large-scale phylogenetic profiling presents significant challenges, especially on the visualization front. Addressing these issues requires scalable methods that efficiently analyze massive datasets with a clean user interface.

We report the development of a robust framework for large-scale phylogenetic profile visualization using Flask, D3, and ECharts. The input is a list of species identifiers derived from BLASTp hits using any arbitrary query dataset and a modified and indexed Reference Proteomes as a target database, using COGENT-like identifiers. This type of large-scale analysis spans tens of thousands of bacterial and archaeal genomes, identifying numerous high-confidence associations, including potentially novel functional predictions for uncharacterized proteins.

This framework enables rapid, scalable visual inference of gene function and evolutionary relationships, facilitating hypothesis generation for experimental validation. The resulting predictions are made accessible through the PhyloFlask interface, offering an intuitive platform for exploring functional associations across diverse lineages.

REFERENCES

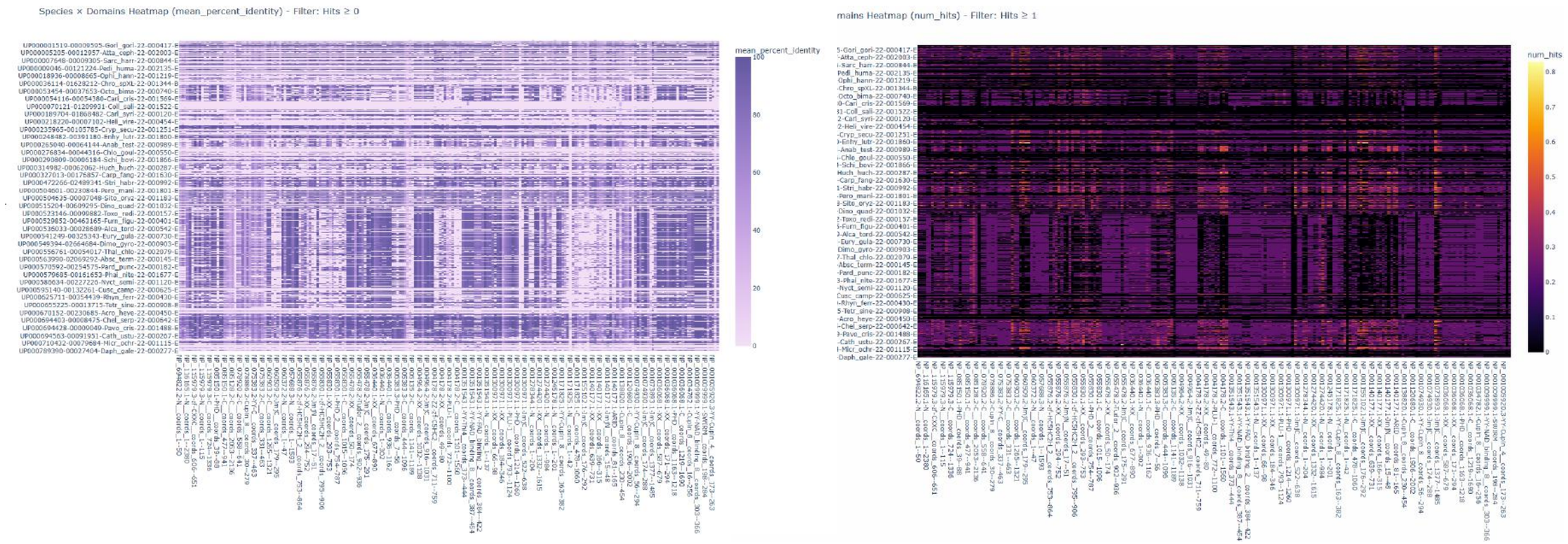
Cohen BA, Mitra RD, Hughes JD & Church GM (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nature Genetics 26(2), 183–186. https://doi.org/10.1038/79896

#24: short contribution

continued

Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, Crandall KA, Deng J, Drew BT, Gazis R, Gude K, Hibbett DS, Katz LA, Laughinghouse HD 4th, McTavish E J, Midford PE, Owen CL, Ree RH, ... & Cranston KA (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. USA 112(41), 12764–12769. https://doi.org/10.1073/pnas.1423041112

Ouzounis CA, Coulson RM, Enright AJ, Kunin V & Pereira-Leal JB (2003) Classification schemes for protein structure and function. Nature Rev. Genetics 4(7), 508–519. https://doi.org/10.1038/nrg1113



#25: short contribution

Climate Change Metagenomic Record Index (CCMRI) and Sample Matcher: Leveraging Metagenomic Data for Climate Change Research

Alexios Loukas1+, Nefeli K. Venetsianou1+, Konstantinos Kalaentzis1, Christina Damianou1, Savvas Paragamian2,1, Vincenzo Lagani3,4, Lars Juhl Jensen5,6, Evangelos Pafilis1

1 Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Greece

2 Department of Biology, University of Crete, Heraklion, Crete, Greece

3 Institute of Chemical Biology, Ilia State University, 3/5, Kakuca Cholokashvili Ave, Tbilisi, Georgia

4 Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology KAUST, Thuwal, Saudi Arabia

5 Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark 6 ZS Discovery, Kongens Lyngby, Denmark

*Correspondence: pafilis@hcmr.gr

Keywords: Climate change, MGnify, metagenomics, microbial ecology, named entity recognition, machine learning, large language models, text classification, similarity metrics, clustering, taxonomic profiles, functional profiles

Climate Change (CC) is reshaping ecosystems globally. Metagenomic data can uncover the impact of CC on microbial communities and the role of the latter in these transformations. However, the vast amount of environmental genomics data complicates finding related studies. The Climate Change Metagenomic Record Index (CCMRI) aims to harvest metagenomic records related to CC and to provide researchers with a pertinent curated database. We are constructing such a database through, initially, the manual curation of all aquatic and terrestrial studies in Mgnify (assessing study titles, descriptions, and linked abstracts for CC clues). To keep future curation efforts scalable, we are developing an automated system that gathers and pinpoints new CC-related studies for a final manual inspection. To this end, text-mining, machine-learning, rule-based classification, and large language model (LLM) methods are being explored. A web platform will offer database access and email notifications upon new CC-studies.

#25: short contribution

continued

In parallel, the Sample Matcher is an algorithm developed to facilitate the comparison and clustering of metagenomic samples based on their taxonomic and functional profiles. Building on MGnify data, Sample Matcher creates a global taxonomic and functional space in which samples can be compared to each other using similarity metrics, like euclidean and cosine distances. Dimensionality reduction techniques refine the analysis and improve clustering accuracy. In the CCMRI context, the Sample Matcher serves as a means to locate climate-change-related studies and samples based on taxonomy and functional data (and not just text-based input). Nonetheless, Sample Matcher is adaptable and could be employed for broader applications.

#26: invited seminar

Microbiome Data Science: Unraveling the Dark Matter of Microbial Diversity

Nikos C. Kyrpides

DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

USA

Microbes represent the most abundant and diverse life forms on Earth, yet the vast majority remain uncultured, uncharacterized, and functionally unknown, a phenomenon often described as microbial dark matter. In this talk, I will explore how large-scale metagenomic sequencing, high-throughput bioinformatics, and open-access data platforms are revolutionizing our ability to explore and understand this hidden microbial universe. Drawing from landmark efforts at the DOE Joint Genome Institute, I will showcase recent advances in microbiome data science, including the creation of global microbial genome catalogs, novel computational tools, and integrative frameworks for identifying previously unknown lineages and functions. I will highlight cutting-edge approaches for mining metagenomes, integrating big data, and building scalable infrastructure to support microbial discovery at planetary scale. Through selected scientific vignettes, I will illustrate how data-driven exploration is transforming our understanding of microbial diversity and complexity, revealing new insights into evolution, ecosystem dynamics, and the metabolic potential of microbial communities. By illuminating the unknown majority, this talk will underscore the pivotal role of microbiome data science in shaping the future of biology, biotechnology, and the global bioeconomy.

#27: invited seminar

Rapid and consistent genome clustering for navigating bacterial diversity

Rob Finn

European Bioinformatics Institute, Cambridge UK

Modern DNA sequence technologies have transformed the way that we can measure global biodiversity, with on-going efforts to sequence all eukaryotic genomes, as well as metagenomics projects producing thousands of metagenome assemble genomes (MAGs) from a single study. However, the sequence archives are incredible bias in the genomes that they contain, for example there are over 500,000 Salmonella enterica genomes alone. As such we are entering an era of millions of bacterial genomes.

MGnify has recently introduced collections of biome specific genome collections, with contain a set of representative species from a given environment, but with large collections of genomes from environment such as the human gut, there are major scaling issues. To address this challenge, we have been developing alignment-free approaches to cluster all bacteria in a matter of hours. In this presentation, I will talk about the technical challenges faced, the particular issues present by MAG incompleteness and how this classification related to classical taxonomy. Finally, I will present the application of the method to MGnify MAG catalogues.

#28: short contribution

Quadrupling the protein family space with global metagenomics

Eleni Aplakidou1,2, Fotis A. Baltoumas1, Maria N. Chasapi1, Ilias Georgakopoulos- Soares3,4, Evangelos Karatzas5, Ioannis Iliopoulos2, Robert D. Finn5, Antonio Pedro Camargo6, Nikos C. Kyrpides6, Georgios A. Pavlopoulos1*

1 Institute for Fundamental Biomedical Research, BSRC "Alexander Fleming", Vari 16672, Greece

2 Division of Basic Sciences, University of Crete Medical School, Heraklion 71110, Greece. 3 Institute for Personalized Medicine, Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, PA, USA

4 Huck Institute of the Life Sciences, Pennsylvania State University, University Park, PA, USA

5 European Bioinformatics Institute, Hinxton, UK

6 DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Corresponding authors: pavlopoulos@fleming.gr

A metagenome encompasses the total genetic material extracted from environmental samples, representing the collective DNA sequences of all microorganisms present. Metagenomics enables the large-scale exploration of this genetic diversity, providing insights into the structure, function, and evolutionary history of microbial communities1. Despite extensive efforts to catalog known proteins, a vast majority of microbial protein space remains uncharacterized. In this study, we employ global metagenomics to systematically uncover novel protein families and provide functional insights into previously unknown sequences2.

Starting with a dataset of 20.7 billion sequences from 40,446 IMG3 metagenomes and 9,540 metatranscriptomes, along with 539 million protein sequences from 167,415 reference genomes, we performed large-scale protein clustering. As a result, we identified 608,258 novel protein families containing 100 or more members, along with an additional 6.5 million families comprising at least 25 members. Notably, none of these families exhibited similarity to known Pfam domains or existing reference genomes. Our findings effectively double the known protein family space for larger families (≥100 members) and quadruple it for smaller families (≥25 members). To assess potential biological relevance, we analyzed taxonomic and functional associations across the newly discovered protein families. Approximately 16.32% of the larger families showed partial alignment with Metagenome-Assembled Genomes (MAGs). Additionally, 15.37% of the families mapped to the NMPFamsDB4 catalog.

#28: short contribution

continued

Beyond sequence-based characterization, AlphaFold25 was used to predict the three-dimensional structures of representative proteins within these families. To systematically identify novel protein folds, we searched against established structural databases, including CATH, PDB, and AlphaFoldDB, enabling us to distinguish previously unknown structures. This analysis revealed numerous novel folds while we provide functional annotation for uncharacterized predicted structures within AlphaFoldDB, bridging the gap between sequence space and functional knowledge.

REFERENCES

Baltoumas, F. A. et al. Exploring microbial functional biodiversity at the protein family level—From metagenomic sequence reads to annotated protein clusters. Front. Bioinforma. 3, 1157956 (2023).

Pavlopoulos, G. A. et al. Unraveling the functional dark matter through global metagenomics. Nature 622, 594–602 (2023).

Chen, I.-M. A. et al. The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res. 51, D723–D732 (2023).

Baltoumas, F. A. et al. NMPFamsDB: a database of novel protein families from microbial metagenomes and metatranscriptomes. Nucleic Acids Res. 52, D502– D512 (2024).

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

#29: short contribution

Exploring the Genomic and Ecological Distinctions Between Isolates and MAGs

Eric Olo Ndela, Antonio Camargo, Gabriele Ghiotto, Dongying Wu, Rekha Seshadri, Natalia Ivanova, Nikos Kyrpides

DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Over three decades of prokaryotic genome sequencing have yielded over 1.5 million microbial genomes in public databases, fueled by a recent influx of metagenome-assembled genomes (MAGs) derived from uncultivated microbes. However, when these genomes are grouped at the species level to evaluate global prokaryotic diversity, we find that only 5.5% of species contain both isolate genomes and MAGs. The majority are represented solely by MAGs (68.2%), while isolates account for a smaller fraction (26.3%).

Strikingly, despite extensive metagenomic sequencing efforts, most cultured species are not recovered through MAGs, suggesting that they are not sufficiently abundant in any habitat to be assembled and binned as a MAG.

Preliminary analyses of the genomic differences between cultivated and uncultivated species revealed that MAGs tend to have smaller genomes, often accompanied by metabolic deficiencies, suggesting distinct ecological adaptations.

These observations, raise the hypothesis that the dominance of uncultivated species in metagenomic samples and their resistance to cultivation, are due to genomic traits indicative of syntrophic or intracellular lifestyles. These traits align with the Black Queen Hypothesis, indicating that metabolic interdependencies not only explain the difficulties in obtaining pure cultures but also contribute to their ecological success.

To explore this hypothesis, we analyzed a dataset of half a million prokaryotic genomes to:

1. Compare genomic properties of isolates and MAGs, identifying key differences that might explain the observed species-level disparity.

2. Examine uncultivated species for metabolic function deficiencies.

3. Assess the global abundance of MAGs versus isolates in natural environments.

The findings of this study could transform our understanding of microbial ecosystems by uncovering genomic dependencies that inform cultivation strategies for MAGs and other novel microbes. These insights have far-reaching implications, from enhancing agro-ecosystems to optimizing engineered microbial communities like bioreactors, ultimately improving our preparedness for global challenges and advancing the reconstruction of the microbial world.

#29: short contribution

continued

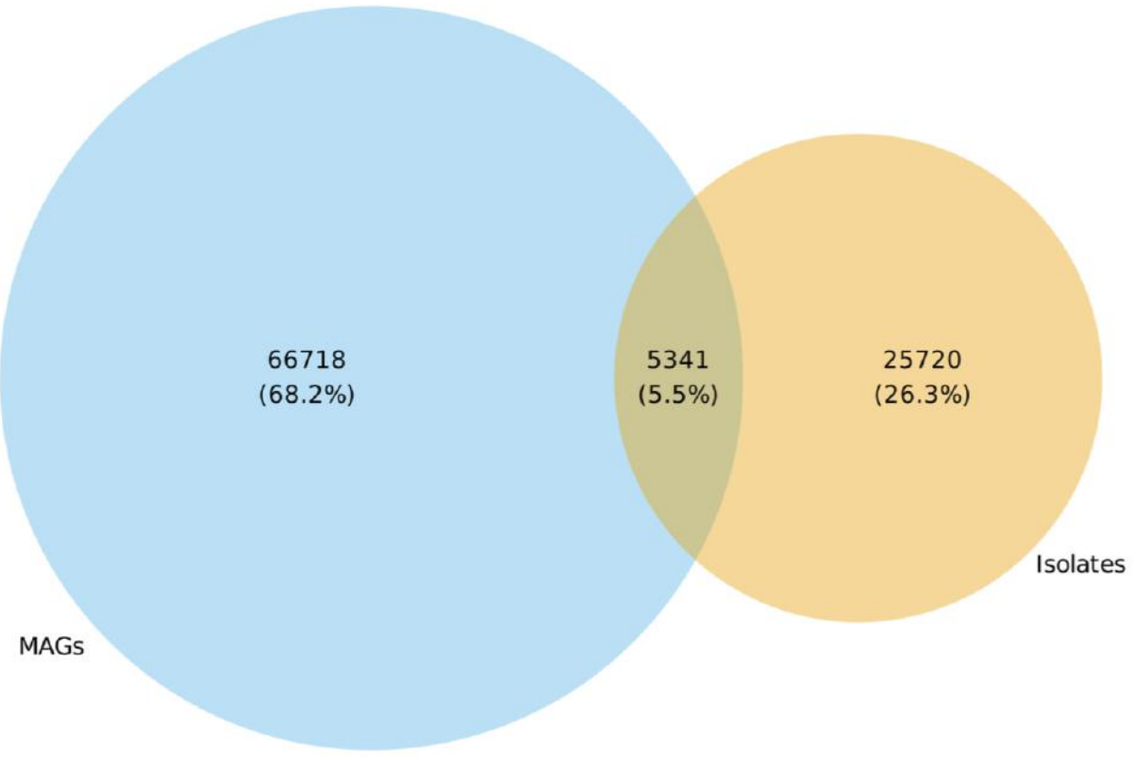


Figure: Venn diagram showing the species level overlap between cultivated (Isolates, yellow) and uncultivated (MAGs, blue) prokaryotic genomes. The dataset comprises 547,467 High-Quality genomes (CheckM2) which gather into 99,779 species level taxa. (Olo Ndela et al. in prep).

#30: invited seminar

AI for multi-omics data fusion to personalize medicine

Nataša Pržulj

1 Professor of Computational Biology, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE

2 ICREA, Pg. Lluis Companys, Barcelona 08010, Spain

3 Department of Computer Science, University College London, London WC1E 6BT, United Kingdom

*E-mail: Natasa.Przulj@mbzuai.ac.ae

Large quantities of multi-omic data are increasingly becoming available. They provide complementary information about cells, tissues and diseases. We need to utilize them to better stratify patients into risk groups, discover new biomarkers and targets, re-purpose known and discover new drugs to personalize medical treatment. This is nontrivial, because of computational intractability of many underlying problems on large interconnected data (networks, or graphs), necessitating the development of new algorithms for finding approximate solutions (heuristics) [1].

We develop versatile artificial intelligence (AI) frameworks for multi-omics data fusion, that also utilize the state-of-the-art network science methods, to address key challenges in precision medicine from time-series, multi-omics data: better stratification of patients, prediction of biomarkers and targets, and re-purposing of drugs, applied to different types of cancer [2,3], Covid-19 [4,5], Parkinson's [6,7] and other diseases. Our new methods stem from graph-regularized non-negative matrix tri-factorization (NMTF), a machine learning (ML) technique for dimensionality reduction, inference, fusion and co-clustering of heterogeneous datasets, coupled with novel graphlet-based network science algorithms. We utilize our new frameworks to for improving the understanding the molecular organization and diseases from the omics data embedding spaces [8,9,10]. Also, we utilize the local network topology to correct for the topological information missed by random walks used in many ML methods [11], and to enable embedding of networks into more linearly separable spaces, allowing for their better mining [12]. The aim is to develop an overreaching framework encompassing all multi-omics data and simplify currently complex and energy inefficient AI methodologies [13].

#30: invited seminar

continued

REFERENCES
[1] Nataša Pržulj, Noel Malod-Dognin: “Network analytics in the age of big data”, Science 353 (6295) 123-124, 2016
[2] Noël Malod-Dognin, Julia Petschnigg, Sam FL Windels, Janez Povh, Harry Hemingway, Robin Ketteler, Nataša Pržulj, “Towards a data-integrated cell,” Nature Communications, 10 (1) 805, 2019
[3] Vladimir Gligorijević, Noël Malod-Dognin, Nataša Pržulj, “Patient-specific data fusion for cancer stratification and personalised treatment,” Biocomputing 2016: Proceedings of the Pacific Symposium, 2016
[4] Alexandros Xenos, Noël Malod-Dognin, Carme Zambrana, Nataša Pržulj, “Integrated data analysis uncovers new COVID-19 related genes and potential drug re-purposing candidates,” International Journal of Molecular Sciences, 24 (2) 1431, 2023
[5] Carme Zambrana, Alexandros Xenos, René Böttcher, Noël Malod-Dognin, Nataša Pržulj, “Network neighbors of viral targets and differentially expressed genes in COVID-19 are drug target candidates,” Scientific Reports, 11 (1) 18985, 2021
[6] Katarina Mihajlović, Gaia Ceddia, Nöel Malod-Dognin, Gabriela Novak, Dimitrios Kyriakis, Alexander Skupin, Nataša Pržulj, “Multi-omics integration of scRNA-seq time series data predicts new intervention points for Parkinson's disease,” Scientific Reports 14 (10983), 2024
[7] Katarina Mihajlović, Noël Malod-Dognin, Corrado Ameli, Alexander Skupin, Nataša Pržulj, “MONFIT: Multi-omics factorization-based integration of time-series data sheds light on Parkinson’s disease,” NAR Molecular Medicine, 1 (4), 2024
[8] Alexandros Xenos, Noël Malod-Dognin, Stevan Milinković, Nataša Pržulj, “Linear functional organization of the omic embedding space,” Bioinformatics, 37 (21) 3839-3847, 2021
[9] Sergio Doria-Belenguer, Alexandros Xenos, Gaia Ceddia, Noël Malod-Dognin, Nataša Pržulj, “A functional analysis of omic network embedding spaces reveals key altered functions in cancer,” Bioinformatics, 39 (5) btad281, 2023
[10] Sergio Doria-Belenguer, Alexandros Xenos, Gaia Ceddia, Nöel Malod-Dognin, Nataša Pržulj, “The axes of biology: a novel axes-based network embedding paradigm to decipher the functional mechanisms of the cell,” Bioinformatics Advances 4(1), 2024
[11] Sam Windels, Noel Malod-Dognin, Nataša Pržulj, “Graphlets correct for the topological information missed by random walks,” arXiv:2405.14194, 2024
[12] Alexandros Xenos, Noel-Malod Dognin, Nataša Pržulj, “Simplifying complex machine learning by linearly separable network embedding spaces,” arXiv:2410.01865, 2024
[13] Nataša Pržulj, Noel Malod-Dognin, “Simplicity within biological complexity,” Bioinformatics Advances, 5(1), vbae164, 2025

#31: short contribution

Integration of proteomics and transcriptomics data for the identification of novel biomarkers in mycosis fungoides

Kotsopoulos Ch.1,2, Eframidou S.1,2, Psatha K.1,2,3, Georgiou E.1, Patsatsi A.4, Aivaliotis M.*1,2,5

1. Laboratory of Biological Chemistry, Department of Medicine, School of Health Sciences, Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece

2. Functional Proteomics and Systems Biology Research Group (FunPATh), Research and Innovation Center (KEDEK AUTH), Thessaloniki, Greece

3. Laboratory of Medical Biology - Genetics, Department of Medicine, School of Health Sciences, AUTH, Thessaloniki, Greece

4. 2nd Department of Dermatology and Venereology, Outpatient Clinic for Cutaneous Lymphomas, Papageorgiou General Hospital, Thessaloniki, Greece

5. Unit of Basic and Translational Research, Special Unit for Biomedical Research and Education, School of Medicine, AUTH, Thessaloniki, Greece

email: aivaliotis@auth.gr

Keywords: Mycosis fungoides, biomarkers, proteomics, transcriptomics

Mycosis fungoides (MF) is the most common type of cutaneous lymphoma and is a rare form of cellular neoplasia, affecting mainly adults. In its early form (eMF) it is localized only in the skin, mimicking benign inflammatory skin diseases (psoriasis, eczema), making its early diagnosis difficult. In some patients who have progressed to late stages of the disease (aMF), the cancer spreads to the blood and lymph nodes, while tumors are also observed. The aim of this work is to find potential biomarkers that will contribute to the correct and early diagnosis of the disease, while it is in its early form. For the analysis, data from two proteomic studies with mass spectrometry were used, involving four groups of individuals: healthy individuals, individuals with psoriasis/eczema and patients with eMF and aMF. Cleaning, filtering, data filling, processing and differential expression analysis were performed using Python. The data integration part as well as a part of the statistical analysis were implemented through Perseus, after correction for batch effect in the two distinct sets with the help of R. In addition, transcriptomic analysis was also performed using R, in order to compare and evaluate the proteomics findings. The bioinformatic comparative analysis showed a clear separation of the four groups from the healthy ones, with a statistically significant differentiation in the relative abundance of their proteins. The proteins unique to the patients were of particular interest, as they are involved in a multitude of pathways related to the control of the transition of mesenchymal to epithelial-type cells (STAT1, CTNNB1), the regulation of epithelial cell adhesion (VSL, JUP, DSP), cell migration, the processing of miRNA molecules (DDX5, SRSF3), etc. In conclusion, the comparative analysis of the protein profile between healthy subjects and patients with psoriasis/eczema, eMF and aMF supports the contribution of the simultaneous identification of the proteome in the distinction between the four study conditions and can be utilized for the more accurate and timely diagnosis of patients with MF.

#32: short contribution

Bridging Genotype and Phenotype by Machine Learning for Single-Molecule Multi-Omic Sequencing

Marios Eftychiou (PhD researcher), Laboratory of Integrative Cancer Genomics (VIB-KU Leuven), ON5 Herestraat 49 - box 912, 3000 Leuven

email: marios.eftychiou97@gmail.com

Jonas Demeulemeester (supervisor), Laboratory of Integrative Cancer Genomics (VIB-KU Leuven), ON5 Herestraat 49 - box 912, 3000 Leuven

Alejandro Sifrim (co-supervisor), Laboratory of Multi-omic Integrative Bioifnromatics, ON4 Herestraat 49 - box 652, 3000 Leuven

Recent innovations in long-read sequencing [1] allow to simultaneously obtain long-range sequence, endogenous CpG methylation and base-pair resolution open chromatin information from single DNA molecules (e.g., Fiber-Seq [2]). These approaches hence enable the high-throughput direct linking of genotypes to molecular phenotypes and will boost our understanding of the role of small and complex genetic variation in human health but in complex disease states.

In the current study we focused on developing state-of-the-art supervised deep-learning sequence-to-function architectures for variant effect prediction [3, 4, 5, 6]. Our model, dubbed FiberTron, has a hybrid convolution transformer architecture. FiberTron was trained on Fiber- Seq data (Pacific Biosciences HiFi sequencing) from the B-Lymphoblastic Genome-in-a-Bottle cell line HG002 to predict genome-wide 5mCpG methylation, chromatin accessibility patterns and gene expression profiles (the latter obtained with Oxford Nanopore direct RNA sequencing), with single-base resolution, directly from the DNA sequence alone.

Testing at genomic loci with characteristic accessibility/gene expression patterns (e.g., CTCF binding sites, transcription start sites) showed that FiberTron can accurately predict molecular phenotypes and effectively captures underlying biological signals without showing signs of overfitting.

In summary, FiberTron provides a basis for the accurate prediction of personal genome-wide methylation, chromatin accessibility and gene expression patterns at base resolution only with DNA sequences as input. This capability can be leveraged to investigate the functional effects of genetic aberrations associated with both normal and disease phenotypes.

#32: short contribution

continued

REFERENCES

[1] Logsdon, G.A., Vollger, M.R. and Eichler, E.E., 2020. Long-read human genome sequencing and its applications. Nature Reviews Genetics, 21(10), pp.597-614.

[2] Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S. and Stamatoyannopoulos, J.A., 2020. Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science, 368(6498), pp.1449-1454.

[3] Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A. and Zeitlinger, J., 2021. Base-resolution models of transcription-factor binding reveal soft motif syntax. Nature genetics, 53(3), pp.354-366.

[4] Pampari, A., Shcherbina, A., Nair, S., Schreiber, J., Patel, A., Wang, A., Kundu, S., Shrikumar, A. and Kundaje, A., 2023. Bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. Manuscript in preparation.

[5] Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P. and Kelley, D.R., 2021. Effective gene expression prediction from sequence by integrating long-range interactions. Nature methods, 18(10), pp.1196-1203.

[6] Linder, J., Srivastava, D., Yuan, H., Agarwal, V. and Kelley, D.R., 2025. Predicting RNA-seq coverage from DNA sequence as a unifying model of gene regulation. Nature Genetics, pp.1-13.

#33: short contribution

Early integration of imaging and omics data for stratifying Mild Cognitive Impairment in Alzheimer's disease

Sotiroula Afxenti1, Margarita Zachariou1, Efi Athieniti1, George M. Spyrou1

1Bioinformatics Department, The Cyprus Institute of Neurology and Genetics, Iroon Avenue 6, Nicosia 2371, Cyprus

Email: sotiroulaa@cing.ac.cy

Background:

Mild Cognitive Impairment (MCI) represents a prodromal phase of Alzheimer's disease (AD), marked by subtle yet detectable cognitive deficits. Despite its association with AD, MCI exhibits considerable heterogeneity in clinical symptoms, underlying molecular profiles, and neuroimaging features, and not all individuals with MCI convert to AD. To address this variability, refining MCI classification into meaningful subgroups is crucial for enhancing prognostic accuracy and tailoring interventions. This study focuses on improving MCI subtyping by constructing detailed individual profiles through the integration of imaging and omics datasets, with the aim of supporting precision medicine strategies for reducing AD risk and optimising therapy.

Methods:

Data were retrieved from the Alzheimer's Disease Neuroimaging Initiative (ADNI), including structural MRI imaging, cerebrospinal fluid (CSF) peptidomics/proteomics, and other clinical measures. A multi-modal subtyping strategy was employed to uncover distinct individual clusters within the MCI cohort. Subgroups were characterised using clinical and neuropsychological assessments, established AD biomarkers, and associated neuroimaging and molecular characteristics. Integration analysis for biomarker discovery was conducted to identify subgroup-specific relationships between brain imaging and peptidomics/proteomics-derived features. Individuals within each subgroup were further classified based on their progression trajectory. Longitudinal cognitive function was analysed to compare subgroup trajectories. Finally, a computational drug repurposing analysis was conducted to identify therapeutic candidates targeting subgroup-specific biological mechanisms.

#33: short contribution

continued

Results: 1) Two subtypes emerged within the MCI group through the integration of multi-modal data. 2) The first subgroup was defined by its neuronal hyperplasticity, minimal brain atrophy and cortical thinning, alongside improved clinical scores and an upregulation of peptides/proteins associated with less severe structural changes. 3) The second subgroup exhibited processes commonly deregulated in AD. Additionally, there were pronounced brain atrophy and cortical thinning, executive dysfunction and downregulation of peptides/proteins associated with significant structural changes. 4) Labelling regarding the conversion to AD revealed that fast progressors were primarily found in the second subgroup, while the first subgroup primarily comprised stable individuals. 5) Longitudinal analysis demonstrated a more significant cognitive decline in the second subgroup compared to individuals within the first subtype, consistent with their more pronounced structural and molecular abnormalities. 6) In silico drug repurposing analyses yielded both shared and subtype-specific drug candidates, which align with the distinct biological pathways present in each subgroup.

Conclusions:

This study highlights two distinct MCI subtypes through the integration of imaging and molecular data, each exhibiting unique structural patterns, proteomic profiles, and cognitive trajectories. The findings reinforce the utility of multi-modal profiling for stratifying MCI populations and facilitating precision treatment strategies. The identified subtypes not only enhance our understanding of MCI heterogeneity but also open new possibilities for tailored drug development in the context of AD.

Keywords: Alzheimer’s disease, Mild Cognitive Impairment, multi-modal data, data integration

#34: invited seminar

Personalized Medicine in the era of multiomics data and AI

Ugur Sezerman

Acıbadem University School of Medicine

Istanbul

Türkiye

Advancements in sequencing technologies and other omics technologies and meta data enabled access to different levels of information in establishing a digital twin of an individual. Such a massive amount of data can be analysed efficiently by exploiting recent advancements in Machine Learning algorithms.

In this talk, I will introduce types of data available and what kind of information it can provide about the health status of an individual, then I will go over AI applications using this data and give specific applications in rare diseases, cancer and drug resistance will be explained.

#35: invited seminar

Integrative signalling network analysis to uncover mechanisms of disease

Evangelia Petsalaki

European Bioinformatics Institute, Cambridge UK

Cell signalling describes the processes that happen in the cell in response to changes in its environment and cell communication. Deregulated cell signalling leads to disease including cancer, diabetes and others, therefore understanding its regulation is critical for understanding disease mechanisms and developing therapeutic strategies.

Despite our knowledge that signalling responses happen through complex networks, most signalling research still uses linear pathways as the ground truth. Moreover, signalling responses are highly dependent on context, such as tissue type, genetic background of specific cells or patients etc and therefore these static pathways are not always suitable. There is also a high bias in the literature towards kinases and pathways for which reagents and prior knowledge is readily available. This leaves a huge dark space in our understanding of cell signalling and significantly hinders our ability to understand the principles of its regulation. My group combines mathematical approaches with network inference to study of the principles underpinning the context-specific regulation of human cell signalling in health and disease.

In this talk I will present an overview of my group's research interests and focus on two projects. I will first present our work on melanoma drug resistance where we found that resistant cell state influences the drug response pathway to elicit different kinase activities and phenotype. I will then use Metabolic Dysfunction Associated Liver Disease (MASLD) as a case study to show how using pseudotemporal ordering combined with network analysis we can extract detailed dynamics of the disease progression even from cross sectional data, and identify non-invasive biomarkers, contributing to a more personalised diagnosis and management of this disease.

#36: short contribution

Understanding the changes in cell-to-cell communication patterns during the progression of Multiple Myeloma: A computational approach

Eleni Nicolaidou, Grigoris Georgiou, Anastasis Oulas, George M Spyrou

6 Iroon Avenue, 3271 Ayios Dometios, Nicosia

Cyprus

email: elenin@cing.ac.cy

Single-cell RNA sequencing (scRNAseq) is a cutting-edge technique, that enables the characterization of cellular composition, heterogeneity and gene expression at the resolution of single-cells. This powerful approach provides insights into cell-to-cell communication (CCC), uncovering signaling patterns in both health and disease.

Given that cancer remains one of the leading causes of mortality worldwide, our study focuses on dissecting CCC and its potential perturbations during the progression of Multiple Myeloma (MM). Intercellular communication is mediated by ligand-receptor (LR) pairs, where ligand-expressing sender cells activate receptor-bearing receiver cells, initiating downstream signaling cascades that further regulate downstream gene expression. Global communication patterns and CCC network reconstruction are of utmost importance to visualize cellular interactions, understand disease progression and identify key regulatory mechanisms driving MM pathogenesis and progression.

To achieve this, three publicly available scRNAseq datasets of MM with samples from all the stages of the disease (MGUS, SMM, MM), along with healthy controls, were analysed following basic scRNAseq analytics with Seurat v5 pipeline (1). Stage-specific CCC networks were reconstructed using CellChat (2) and basic network topology analytics were performed in Cytoscape with CytoHubba (3). Node rewiring that calculates the variability of cell nodes and edges in their connections between different stage-specific networks was studied with DyNet, a Cytoscape's plug-in (4). Finally, to give biological significance to our results, follow-up analyses were performed with NicheNet to investigate downstream responses and target genes that may be influenced by LR pairs mediating specific CCCs (5).

#36: short contribution

continued

Throughout this study, we aim to (1) elucidate intercellular communication patterns, (2) detect alterations in healthy and disease states or across progressive stages of MM and (3) identify key players implicated in MM progression. Our findings provide a deeper understanding of MM pathophysiology and may unveil potential therapeutic targets for intervention.

REFERENCES

1. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024 Feb;42(2):293–304.

2. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 Feb 17;12(1):1088.

3. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4(Suppl 4):S11.

4. Goenawan IH, Bryan K, Lynn DJ. DyNet: visualization and analysis of dynamic molecular interaction networks. Bioinforma Oxf Engl. 2016 Sep 1;32(17):2713–5.

5. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020 Feb;17(2):159–62.

#37: short contribution

A Network Biology Perspective on Multiple Myeloma: Integrating Single-Cell Transcriptomics with Prior Knowledge

Grigoris Georgiou1*, Margarita Zachariou1, George M Spyrou1*

1 Bioinformatics Department, The Cyprus Institute of Neurology & Genetics, 6 International Airport Avenue, 2370 Nicosia, Cyprus, P.O.Box 23462, 1683 Nicosia, Cyprus, Electronic address: grigorisg@cing.ac.cy

*Grigoris Georgiou: Tel: +357-22392624; Electronic address: grigorisg@cing.ac.cy

*George Spyrou: Tel: +357-22392852; Electronic address: georges@cing.ac.cy

Background/Aim: Multiple Myeloma (MM) is a plasma cell malignancy that progresses through distinct stages, from asymptomatic Monoclonal Gammopathy of Undetermined Significance (MGUS) and Smoldering Multiple Myeloma (SMM) to active disease. Understanding the molecular mechanisms underlying this progression remains a major challenge in MM research. This study integrates single-cell RNA sequencing (scRNA-seq) analysis with established MM biological knowledge to uncover key molecular regulators of disease progression and identify potential therapeutic targets. Materials and Methods: We utilized publicly available databases to construct a foundational gene network representing MM biology. Single-cell differential expression analysis was performed on an integrated dataset derived from two CD138+ MM scRNA-seq studies (EGAD0001009648 and GSE145977). Combining this analysis with network-based methodologies, we identified key genes and regulatory interactions associated with MM progression. The final validation of these key genes was conducted using a batch-corrected dataset derived from eight bulk RNA-seq studies. This allowed us to identify consistent molecular patterns and potentially shared regulators across different patient cohorts. Results: Our study revealed CCND1, HDAC4, TP53, IRF4, and FGFR3 as critical backbone genes in MM. Further network analysis identified GSK3B and PRKCD as major regulators influencing other genes, while CDKN1B and CCND1 emerged as key regulatory targets. Pathway enrichment and protein-protein interaction (PPI) analyses highlighted essential pathways involved in MM progression. Additionally, we generated a 44-gene signature that includes stage-specific genes such as GSK3B, PRKCD, CEBPB, PRKDC, RELA, and PTEN, providing novel insights into disease mechanisms and potential intervention strategies. Finally, NDNF, DKK1 and CDKN1A were also highlighted through Bulk-RNA differential expression analysis. Conclusion: This study integrates prior MM knowledge with scRNA-seq analytics to identify crucial genes and regulatory networks driving disease progression. These findings offer valuable insights into MM biology and provide a foundation for future research into targeted therapies to disrupt key disease mechanisms.

#37: short contribution

continued

REFERENCES

Saadoune, C., Nouadi, B., Hamdaoui, H., Chegdani, F., Bennis, F., 2022. Multiple Myeloma: Bioinformatic Analysis for Identification of Key Genes and Pathways. Bioinforma. Biol. Insights 16, 11779322221115545. https://doi.org/10.1177/11779322221115545

Xu, Z., Yu, J., Chen, Y., 2024. Hub genes and associated drugs for multiple myeloma with 1q21+: identified by bioinformatic analysis. Hematol. Amst. Neth. 29, 2323890. https://doi.org/10.1080/16078454.2024.2323890

Dong, X., Lu, G., Su, X., Liu, J., Chen, X., Tian, Y., Chang, Y., Wang, L., Wang, W., Zhou, J., 2021. Identification of key miRNA signature and pathways involved in multiple myeloma by integrated bioinformatics analysis. Hematol. Amst. Neth. 26, 976–984. https://doi.org/10.1080/16078454.2021.2003980

Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., Satija, R., 2023. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-01767-y

Lo Surdo, P., Iannuccelli, M., Contino, S., Castagnoli, L., Licata, L., Cesareni, G., Perfetto, L., 2023. SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. Nucleic Acids Res. 51, D631–D637. https://doi.org/10.1093/nar/gkac883

Goenawan, I.H., Bryan, K., Lynn, D.J., 2016. DyNet: visualization and analysis of dynamic molecular interaction networks. Bioinforma. Oxf. Engl. 32, 2713–2715. https://doi.org/10.1093/bioinformatics/btw187

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., von Mering, C., 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 51, D638–D646. https://doi.org/10.1093/nar/gkac1000

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinforma. Oxf. Engl. 25, 1091–1093. https://doi.org/10.1093/bioinformatics/btp101

#38: invited seminar

Towards a motif map of the human proteome

Ylva Ivarsson

Department of Chemistry – BMC

Uppsala University

Sweden

Numerous studies have generated large-scale datasets on human protein-protein interactions. However, many interactions remain undiscovered, particularly low-affinity, conditional, and cell type-specific interactions, which are disproportionately underrepresented. Most of these missing interactions are likely mediated by short linear motifs (SLiMs). Over the past decade, we have developed dedicated experimental methods for large-scale screening of SLiM-based interactions, along with tools and guidelines for data processing. Our optimized proteomic peptide-phage display (ProP-PD) library tiles the disordered regions of the human proteome, enabling the screening of approximately one million overlapping peptides in a single assay. To date, we have released data on over 2,000 SLiM-based interactions, and we have a major new dataset currently in preparation. The amino acid resolution of our interaction data allows for the prediction of functionally important disease mutations and phosphorylation events. Recently, we developed a tailored phage peptidome to pinpoint the effects of disease-associated mutations. By screening a collection of 80 bait proteins, we identified nearly 370 mutation-modulated interactions, with mutations either weakening, enhancing, or creating novel binding interfaces. Collectively, our findings provide novel insights into SLiM-mediated interactions that contribute to shaping the cellular interactome and reveal how disease-associated mutations may perturb or rewire the interactome.

#39: invited seminar

How machine learning unravels the immunopeptidome, from analytics towards biology

Arthur Declercq

VIB-UGent Center for Medical Biotechnology – CompOmics Group

Technologiepark-Zwijnaarde 75 - 9052 Gent

Belgium

The adaptive immune system is marvellous yet highly complex, and functions through the presentation of self-peptides, or epitopes, via Human Leukocyte Antigen (HLA) proteins. The identification of these epitopes through immunopeptidomics offers a promising avenue for advancing novel therapies, including cancer vaccines, TCR-T cell therapies, and pathogen-targeted vaccines. However, the immense search space significantly hampers accurate and efficient epitope identification.

To address this challenge, we have developed a suite of immunopeptidomics-tailored prediction models that not only enhance epitope identification but also extend HLA allele coverage, including for alleles that are typically difficult to detect. Building upon this foundation, we leverage powerful deep learning Graph Neural Networks (GNNs) to represent peptides as molecular structures, enabling the construction of a cutting-edge deep learning epitope predictor, MHC-3PO. This model is capable of generalizing to unseen HLA alleles, accommodating modified epitopes, and revealing biologically relevant binding mechanisms through interpretable saliency maps.

Our integrative approach, combining enriched immunopeptidomics data with advanced GNN-based modelling, shows the power of novel bioinformatics tools to potentially accelerate the discovery of therapeutically relevant epitopes and to broaden the applicability of precision immunotherapies across diverse patient populations.

#40: short contribution

Computational Analysis of Fluorescence Recovery After Photobleaching (FRAP) Using FRAPedia: A Simple Graphical Approach

Alkiviadis Athanasiadis, Veronica Giourieva , George Komis

Department of Botany, School of Biology, Aristotle University of Thessaloniki, 54124 ThessalonikiGreece

email: alkathnik@bio.auth.gr

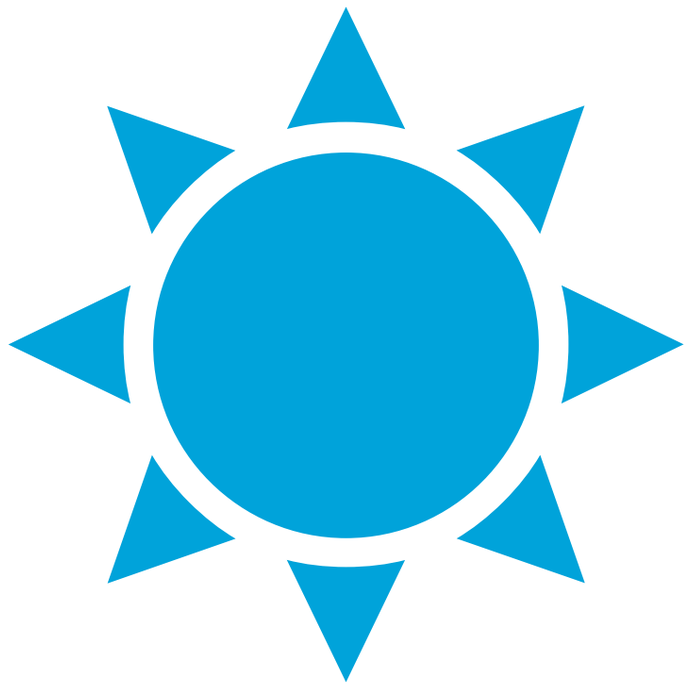
Fluorescence Recovery After Photobleaching (FRAP) is a powerful and increasingly popular live-cell imaging technique used to quantify molecular mobility, protein dynamics, and interactions within single cells. During a FRAP experiment, a specific region of the cell, containing a molecule of interest labelled with a fluorescent marker, is exposed to a brief and strong laser pulse, effectively bleaching the fluorescence in that area. Subsequently, the microscope captures the recovery of fluorescence over time1. This recovery data allows researchers to determine key metrics such as the mobile and immobile fractions of the molecule and the half-time of recovery. More advanced mathematical models allow the calculation of diffusion or kinetic parameters directly from the recovery curve2. Despite the widespread use of FRAP in life sciences and industry, there is a lack of software available online for processing and analysing raw FRAP data. Here, we present FRAPedia, a new open-source, standalone program designed to address this gap. With a simple graphical user interface, FRAPedia allows researchers to easily analyse FRAP data. Developed using Python 3.12 and common scientific libraries, FRAPedia can normalize datasets retrieved from confocal laser scanning microscopy FRAP experiments, generate automated graphs, and calculate key parameters, including mobile/immobile fractions and recovery half-time, based on an exponential curve fitting model. We exemplify and validate FRAPedia using case studies from transgenic Arabidopsis plants expressing appropriately tagged membrane markers corresponding to membrane lipid biosensors or intrinsic membrane proteins. FRAPedia is specifically designed to help researchers in bulk analysis of large FRAP datasets efficiently, reducing the need for complex calculations and enabling faster analysis with just a few clicks.

REFERENCES

1. Giakoumakis N.N. , Rapsomaniki M.A. , Lygerou Z.Analysis of protein kinetics using fluorescence recovery after photobleaching (FRAP). Methods Mol. Biol. 2017; 1563 : 243 – 267.

2. Lorén, N. et al. Fluorescence recovery after photobleaching in material and life sciences: putting theory into practice. Q. Rev. Biophys. 48, 323–387 (2015).

day III • may 18



#41: invited seminar

Empowering microbial genome annotation with machine learning: insights from a global metagenomic database of anaerobic species

Stefano Campanaro

Department of Biology

University of Padova

Padua

Italy

Microbial genomics has witnessed an unprecedented expansion, driven by advances in metagenomics and single-cell sequencing. Yet, our ability to assign functions to newly assembled genomes in a fast and reliable way lags far behind, largely due to the diverse and understudied nature of many environments. We introduced MICROPHERRET, a machine learning-based framework for rapid and accurate functional annotation of microbial genomes, and extended its capabilities to incorporate both standard and less-explored microbial functions. This approach provides a more holistic view of microbial roles in complex ecosystems, with particular focus on anaerobic microbial communities. Drawing on a comprehensive database of prokaryotic and viral genomes derived from over 300 metagenome samples, we illustrate how supervised learning algorithms can predict an extensive range of metabolic and ecological traits, even when genome completeness is below ideal thresholds. We further demonstrate how this framework highlights the significance of specific functional categories and uncovers the selective pressures shaping genome diversity. Overall, our study offers insights into the conceptual, methodological, and practical aspects of using machine learning for microbial genome annotation, ultimately showcasing how these methods can guide future investigations of microbial ecology and drive innovative biotechnological applications.

#42: invited seminar

Multi-omics and metabolic modeling for understanding microbial cooperation in CO_2-driven methanogenesis

Laura Treu

Department of Biology

University of Padova

Padua

Italy

Multi-omics approaches, including metagenomics, metatranscriptomics, and metabolic modeling, provide valuable insights into microbial cooperation in CO_2-driven methanogenesis. In addition, flux balance analysis provides a powerful tool for simulating microbial functions in terms of growth and metabolite exchange.
Three microbial communities, all dominated by hydrogenotrophic archaea (Methanothermobacter wolfeii and Methanothermobacter thermoautotrophicus), were investigated to elucidate their roles in CO_2 conversion. As informed by experimental data, genome-scale metabolic models identified a potential formate cross-feeding from Limnochordia to M. wolfeii, driven by the reductive glycine pathway. Additionally, amino acid exchanges, particularly aspartate released by M. wolfeii, were found to sustain microbial cooperation, with Sphaerobacter thermophilus fermenting aspartate into acetate, which is then converted into formate by Limnochordia. These interactions bolster methanogenesis, highlighting the importance of amino acids in stress resistance and metabolic flexibility. Amino acid supplementation restored growth in communities inhibited by antibiotics, underscoring their role in microbiota stability. Moreover, modeling suggested that Clostridia species, via the Wood-Ljungdahl pathway, exchange acetate with M. thermoautotrophicus through alanine and glutamate. Finally, novel tools have been developed for profiling microbial and viral single-nucleotide variant trajectories. These findings expand the understanding of microbial ecosystem dynamics, highlighting the significance of syntrophic relationships and providing new insights for optimizing CO_2-to-methane conversion in biotechnological applications.

#43: short contribution

From Genes to Function: The case of kiwifruit cold stress response

M. G. Kollaros1, M. Michailidis1, M. Samiotaki2, C. Skodra1, C. Polychroniadou1,3,4, D. Valasiadis1, A. Karamanoli5, C. Bazakos4,6,7, G. Tanou3,4, A. Molassiotis1

1Laboratory of Pomology, Department of Horticulture, Aristotle University of Thessaloniki, 57001 Thessaloniki, Thermi, Greece

2Institute for Bioinnovation, Biomedical Sciences Research Center “Alexander Fleming”, 16672, Vari, Greece

3 Institute of Soil and Water Resources, ELGO-DIMITRA, Thessaloniki-Thermi, 57001, Greece

4Joint Laboratory of Horticulture, Institute of Soil and Water Resources, ELGO-DIMITRA, Thessaloniki-Thermi 57001, Greece

5Laboratory of Agricultural Chemistry, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

6Institute of Plant Breeding and Genetic Resources, ELGO-DIMITRA, Thessaloniki Thermi, 570001, Greece

7Max Planck Institute for Plant Breeding Research, Department of Comparative Development and Genetics, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Corresponding email: amolasio@agro.auth.gr

Kiwifruit (Actinidia chinensis var. deliciosa cv. ‘Hayward’) is a climacteric fruit renowned for its high ethylene sensitivity and elevated respiration rate, which predispose it to rapid softening and senescence, ultimately constraining its postharvest longevity. Although cold storage is a widely adopted strategy to maintain kiwifruit quality and extend shelf-life, prolonged exposure to low temperatures can induce cold-specific physiological disorders, due to its subtropical origin. Cold stress triggers widespread modifications in numerous biochemical pathways by reconfiguring the transcriptional landscape of plant cells. In this study, we employed a holistic multi-omics and functional framework to unravel the mechanisms underpinning cold acclimation in ‘Hayward’ kiwifruit. For this purpose, mature kiwifruits were cold stored (0°C and 95% RH) for different postharvest periods (15 and 90 days). We conducted a comprehensive analysis integrating metabolomic, proteomic, transcriptomic, and whole-genome bisulfite sequencing data from both pericarp and placenta tissues at harvest and postharvest stages. This systems-based investigation revealed distinct cold-responsive signatures across transcriptional, epigenomic, proteomic, and metabolic layers, thereby elucidating the key pathways modulated in response to low-temperature stress. Through the integration of multi-omics datasets with advanced computational bioinformatic models, we identified a number of crucial transcription factors (TFs), including members of the C2H2, GARP-G2, GRAS, HMG and NAC families, that are central to the orchestration of the cold stress response. Functional validation of these TFs further underscored their pivotal roles in mediating cold acclimation. Collectively, our findings deepen the understanding of the molecular mechanisms triggered by low temperatures in kiwifruit and lay the groundwork for the development of biomarkers to guide future breeding programs.

#44: short contribution

Transposable element annotation and centromere analysis in *Brassicaceae*

Anastasia Boutsika, School of Agriculture, Dept. of Field Crops and Ecology, Aristotle University of Thessaloniki, Greece

and School of Health Sciences, Dept. of Molecular Biology & Genetics, Democritus University of Thrace, Greece

Estella Perez-Roman,, Alexandros Bousios, School of Life Sciences, University of Sussex, Brighton, United Kingdom

email: Alex.Bousios@sussex.ac.uk

Despite their conserved function, centromeres undergo rapid sequence evolution, driven by cycles of satellite repeat homogenization and transposable element (TE) invasion (Wlodzimierz et al., 2023a). In Brassicaceae, centrophilic TEs display distinct lineage- specific patterns: ATHILA elements are exclusive to Arabidopsis thaliana, Arabidopsis lyrata contains both ATHILA and ALE (Tsukahara et al., 2025), while other species exhibit CRMs and ALE within centromeres (Bousios et al 2025). This variability raises the question of whether centromere targeting by TEs is evolutionarily conserved or highly dynamic. This study examines the presence and distribution of centrophilic TEs in centromeres across three Brassicaceae genome —Cochlearia excelsa (6 chromosomes) and two Brassica oleracea varieties (A12 and GDDH, each with 9 chromosomes)—to explore whether transposon integration follows conserved patterns or undergoes dynamic evolutionary changes.

TE annotation was performed using state-of-the-art pipelines and downstream analyses for TE lineage refinement. We identified genes in silico using Helixer and predicted centromere boundaries with TRASH for satellite and tandem repeat detection (Wlodzimierz et al., 2023b) and Context Tree Weighting (CTW) to identify regions of DNA compression.

Our analysis identified that the C. excelsa centromeres are defined by a 102 bp satellite repeat, and that of B. oleracea (A12 and GDDH) by a 176 bp satellite repeat. The centromeres of the three species show little sequence homology with the centromeres of A. thaliana and A. lyrata, consistent with rapid satellite sequence evolution between Brassicaceae species. Although centrophilic TEs occupy <10% of the centromeric space in A. thaliana and A. lyrata (Wlodzimierz et al., 2023a; Tsukahara et al., 2025), our results suggest higher levels of TE invasion in Cochlearia excelsa and Brassica oleracea, with 20.89% and around 25% of the centromere space occupied by TEs respectively. In C. excelsa, CRM and ALE elements were predominantly found within centromeres, reinforcing their adaptation for this genomic niche. This was particularly prominent in chromosomes 1 and 3 where alternating blocks of 102 bp arrays and CRM/ALE generated a unique centromere organization. In contrast to C. excelsa, centromeric TEs in Brassica oleracea were only ALE, with CRM and other TE lineages mostly accumulating in pericentromeric heterochromatin. As in the two Arabidopsis species, centromeric TEs are younger than centrophobic TEs (located in the pericentromeres and chromosome arms), reflecting parallel and independent dynamics of recent invasion patterns by ATHILA, ALE and CRM elements across the Brassicaceae.

#44: short contribution

continued

This study sheds further light on the evolutionary plasticity of centromeres and centrophilic TEs in plants. Given the continuous improvement of long-read sequencing technologies that can completely resolve the complex structure of centromeres, we anticipate that our understanding of centromere function and evolution will rapidly advance in the near future across eukaryotes.

REFERENCES

Bousios, A., Kakutani, T., & Henderson, I. R. (2025). Centrophilic retrotransposons of plant genomes. Annual Review of Plant Biology, 76(14), 1–26. https://doi.org/10.1146/annurev-arplant-083123-082220

Tsukahara, S., Bousios, A., Perez-Roman, E., et al. (2025). Centrophilic retrotransposon integration via CENH3 chromatin in Arabidopsis. Nature, 637, 744–759. https://doi.org/10.1038/s41586-024-08319-7

Wlodzimierz, P., Rabanal, F. A., Burns, R., Naish, M., Primetis, E., Scott, A., Mandáková, T., Gorringe, N., Tock, A. J., Holland, D., Fritschi, K., Habring, A., Lanz, C., Patel, C., Schlegel, T., Collenberg, M., Mielke, M., Nordborg, M., Roux, F., Shirsekar, G., … Henderson, I. R. (2023a). Cycles of satellite and transposon evolution in Arabidopsis centromeres. Nature, 618(7965), 557–565. https://doi.org/10.1038/s41586-023-06062- z

Wlodzimierz, P., Hong, M., & Henderson, I. R. (2023b). TRASH: Tandem repeat annotation and structural hierarchy. Bioinformatics, 39(5), btad308. https://doi.org/10.1093/bioinformatics/btad308

#45: keynote lecture

The BioCyc Genome and Metabolic Pathway Web Portal

Peter D. Karp

AI Center, SRI International, Menlo Park CA

USA

BioCyc.org is an extensive web portal containing 20,000 genomes and associated metabolic pathways for microbes, model eukaryotes, and humans. BioCyc provides extensive bioinformatics tools for search and analysis; those tools can also be applied to a company's proprietary genome data.

BioCyc databases combine curated information with data imported from multiple sources, and with computationally inferred data (metabolic pathways, operons, PFam domains, and orthologs). Curated databases receive intensive review and updating by Ph.D. biologists that includes entering new gene functions and metabolic pathways from the experimental literature, defining regulatory relationships, and creating protein complexes. 77 BioCyc databases have been curated from 153,000 publications.

BioCyc genome-related tools include a genome browser, sequence alignment, and extraction of sequence regions. Pathway-related tools include pathway diagrams, a tool for navigating zoomable organism-specific metabolic map diagrams, and a tool for searching for metabolic routes that transform a starting metabolite into a product metabolite. Regulation tools depict operons and regulatory sites, as well as showing full organism regulatory networks. Omics data analysis tools support enrichment analysis and painting of transcriptomics and metabolomics data onto individual pathway diagrams and onto zoomable metabolic map diagrams. The Omics Dashboard tool enables interactive exploration of omics datasets through a hierachy of cellular systems.

This presentation will emphasize new BioCyc tools including a metabolic network explorer tool, our new genome browser, a new comparative genomics tool called the Comparative Genome Dashboard, and extensions to the metabolic network visualization tool to enable display of multi-omics data.

#46: short contribution

PathoMagic: standardizing genome annotation with automated format conversion

Christos Bouas1, Ron Caspi2, Peter D Karp2, Christos A Ouzounis1

1BCCB Group, Artificial Intelligence & Information Analysis Laboratory, School of Informatics, Aristotle University of Thessaloniki, Thessalonica, Greece

2AI Center, SRI International, Menlo Park CA, USA

Genome annotation is the process of identifying and labeling genomic features for structure and function, such as genes and enzymatic properties of gene products. These annotations are often stored across the wider community in tabular formats like CSV for ease of use; however, the lack of standardization in these files creates interoperability challenges with bioinformatics tools such as Pathway Tools, which requires structured input in the PathoLogic format (PF) to generate Pathway/Genome Databases (PGDBs) for BioCyc.org.

To address this issue, we propose a minimum information standard for annotated genomic data by just eliminating redundant columns from a CSV file and establishing essential guidelines for formatting, required properties, and data representation.

Additionally, we present two fast and robust parsers – mi2pf and pf2mi – built using formal methods with flex and yacc. The mi2pf parser converts CSV files into PF format, allowing direct integration with Pathway Tools, while pf2mi performs the reverse operation, converting PF files back into CSV for seamless editing and further analysis before back-conversion.

PathoMagic offers consistency in genome annotation data exchange and automates format conversions, reducing manual errors and improving usability. By streamlining the process of PGDB creation and maintaining compatibility between formats, PathoMagic facilitates efficient genome-scale studies, supporting advances in computational biology and bioinformatics.

#46: short contribution

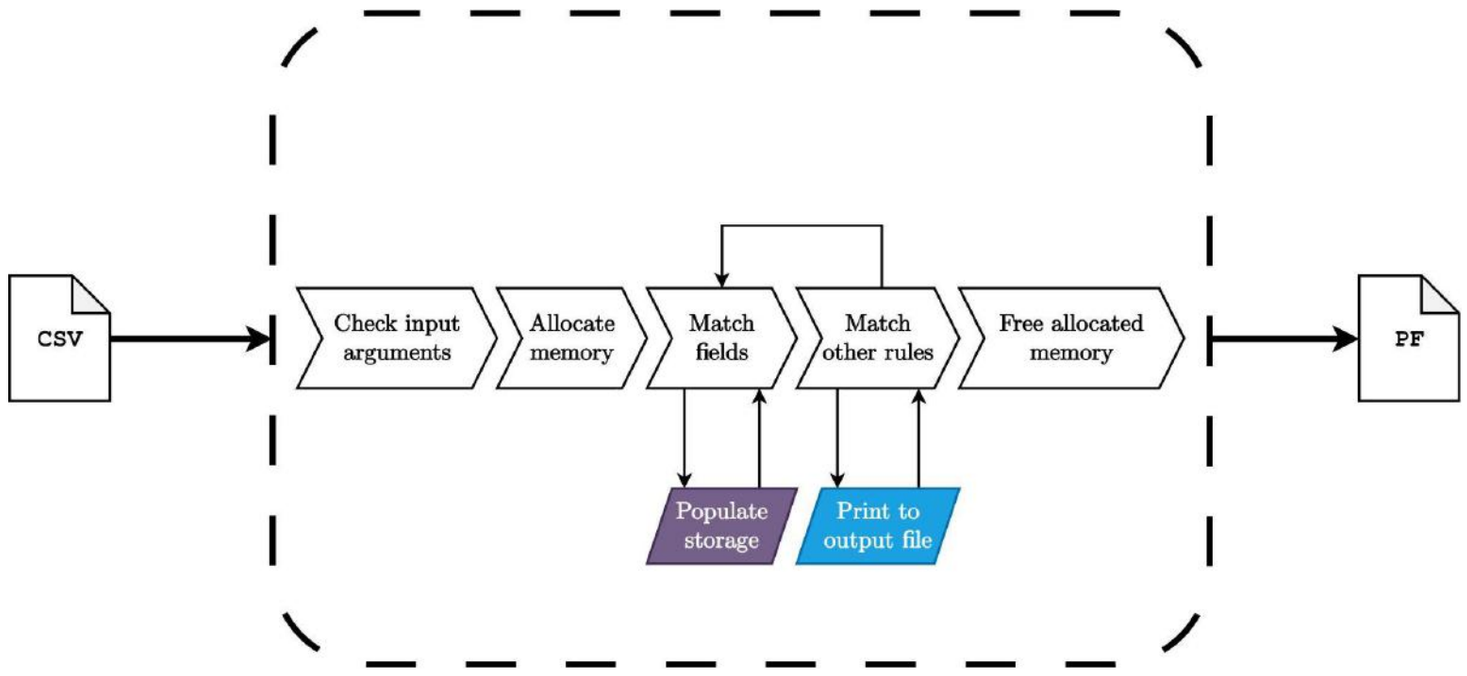
continued

REFERENCES

Baltsavia I, Stamoulos G, Tziavaras K, Bouas C, Katikaridou I, Dermaris A, Kothari A, Iliopoulos I, Caspi R, Karp PD, Kyrpides NC & Ouzounis CA (2024) MjCyc: Rediscovering the pathway-genome landscape of the first sequenced archaeon, Methanocaldococcus (Methanococcus) jannaschii. iScience, 28(1), 111546. https://doi.org/10.1016/j.isci.2024.111546

Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, Latendresse M, Ong WK, Subhraveti P, Caspi R, Fulcher C, Keseler IM & Paley SM (2021) Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Brief. Bioinf. 22(1), 109–126. https://doi.org/10.1093/bib/bbz104

Keseler IM, Skrzypek M, Weerasinghe D, Chen AY, Fulcher C, Li GW, Lemmer KC, Mladinich KM, Chow ED, Sherlock G & Karp PD (2014) Curation accuracy of model organism databases. Database 2014, bau058. https://doi.org/10.1093/database/bau058



#47: short contribution

Deciphering the Anaerobic Microbiome for Constructing Robust Microbial Consortia to Produce Medium-Chain Fatty Acids

Konstantinos Papasakellariou, Irini Angelidaki

Biological Conversions, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DTU, Søltofts Plads 228A, Lyngby 2800, Denmark
email: konpa@kt.dtu.dk

Anaerobic fermentation represents a pivotal sector in biotechnology, contributing significantly to the green transition. Utilizing anaerobic microorganisms is particularly advantageous due to their ability to produce secondary metabolites in substantial quantities, which are extensively employed in industrial applications [1], [2]. An effective approach for producing high-value chemicals, such as medium-chain fatty acids, through microbial chain elongation, is the utilization of mixed- species cultures, comprising microbial species that engage in synergistic interactions. These mixed cultures exhibit superior adaptability to environments characterized by extreme thermodynamic conditions [3]. Microbial chain elongation is a process where an electron donor, like ethanol, elongates the carboxylic chain of a short-chain fatty acid, such as acetic acid, to form medium-chain fatty acids. Medium-chain fatty acids are important chemicals with numerous applications [4]. Currently, they are derived through the catalytic conversion of fatty acids. These fatty acids are mainly produced from coconut or palm oil, which contributes to the devastating effects of tropical deforestation.

This project aims to develop a comprehensive model of the metabolic reactions occurring within mixed-species cultures of anaerobic microorganisms, specifically focusing on taxa involved in microbial chain elongation.

This research will involve the development of a laboratory protocol for the microbial chain elongation of short-chain fatty acids, investigation of microbial consortia interactions, and the generation of community genome-scale metabolic models [5]. The objective is to establish synthetic mixed microbial cultures capable of efficiently producing medium-chain fatty acids. This work will advance the field of microbial community dynamics and contribute to the development of sustainable high-value chemicals.

#47: short contribution

continued

REFERENCES

1. J.L. Rombouts, E.M.M. Kranendonk, A. Regueira, D.G. Weissbrodt, R. Kleerebezem, M.C.M. van Loosdrecht. Biotechnology and Bioengineering 117 (2020) 1281 – 1293.

2. M. Zhou, B. Yan, J.W.C. Wong, Y. Zhang. Bioresource Technology 248 (2018) 68-78

3. R.D. Hoelzle, B. Virdis, D.J. Batstone. Biotechnology and Bioengineering 111 (2014) 2139-2154

4. S. O-Thong, X. Zhu, I. Angelidaki, S. Zhang, G. Luo. Advances in Bioenergy 5 (2020), 63-99

5. J.D. Orth, I. Thiele, B. Ø. Palsson. Nature Biotechnology 28 (2011) 245- 248

#48: invited seminar

Bioinformatics-Driven Translational Research in Metabolic Engineering

Vassily Hatzimanikatis

Laboratory of Systems Biotechnology

Ecole Polytechnique Federale de Lausanne (EPFL)

Lausanne

Switzerland

Recent advances in metabolic engineering have increasingly relied on sophisticated bioinformatics approaches to accelerate design, prediction, and optimization in biological systems. Yet, a translational perspective—connecting computational models to actionable outcomes in medicine and biotechnology—remains a frontier with untapped potential.

Translation involves two core components: analysis and design, which are by definition at the heart of engineering science and practice. In this respect, computational approaches in metabolic engineering aim to analyze biological data to extract actionable insights, upgrade the information content of experimental datasets, and guide the rational design of microbial strains, communities, and therapies.

In this talk, I will present a set of integrative strategies that leverage high-resolution omics data, genome-scale models, and machine learning to inform microbial strain design with real-world clinical or biotechnological constraints. I will highlight how these computational pipelines enable the identification of novel metabolic targets, the analysis of microbiome structure and function from genomic information, and the integration of cell signaling and metabolism. Emphasis will be placed on recent case studies that exemplify the role of systems-level modeling in addressing translational challenges.

#49: short contribution

The impact of biochar on methane yield and microbial community dynamics under increasing organic loading rates in a continuous anaerobic digestion system.

Charitini Nikolaidou1,2, Maria-Chiara Valerin3, Stefano Campanaro3, Panagiotis G. Kougias1

1Soil and Water Resources Institute, Hellenic Agricultural Organization Dimitra, 57001, Thessaloniki, Greece

2University Center of International Programmes of Studies, International Hellenic University, 57001, Thessaloniki, Greece

3Department of Biology, University of Padova, Via U. Bassi 58/b, 35121 Padua, Italy

email: c.nikolaidou@swri.gr

The use of biochar as an additive in anaerobic digestion (AD) systems is a promising yet contentious approach to potentially boost methane yield (Chiappero et al., 2020). The benefits of biochar often emerge under specific conditions, such as high biochar dosages and substrate loadings (Hu et al., 2023). This study aims to identify the optimal biochar dosages in batch systems and then evaluate their effect on both methane yield and microbial community dynamics in continuous systems with increasing organic loading rates (OLR) under realistic scenarios.

In total six concentrations of wood-chip biochar (0, 5, 10, 12.5, 15 and 20 g/L) were screened via Biochemical Methane Potential (BMP) tests under normal substrate loading (3 $gVS*L^{-1}*day^{-1}$). The two most efficient dosages in terms of methane yield were chosen as feedstock additives for continuous operation in Continuous Stirred-Tank Reactors (CSTR). Three mesophilic reactors (R1: control, R2: 10 g biochar/L, R3: 15 g biochar/L) were operated for 110 days across two periods: Period 1 (OLR = 3 $gVS{\cdot}L^{-1}{\cdot}day^{-1}$) and Period 2 (OLR = 4 $gVS{\cdot}L^{-1}{\cdot}day^{-1}$). Throughout operation, pH, VFA concentration, biogas production and composition were monitored. The microbial profile of each reactor for each period was mapped and analyzed using 16S rRNA amplicon sequencing.

All biochar applications improved methane yield in batch tests compared to the control. In contrast, the addition of biochar in the CSTRs had no effect on methane yield, regardless of the OLR, indicating that the retention of biochar within the reactor may be essential for optimizing methane production. As OLR increased, methane yield dropped in all reactors, without any accompanied VFA accumulation or pH drop. The decline was rather the result of CO_2 build-up, as evidenced by the higher CO_2% in the output biogas, most likely due to increased fermentation processes caused by substrate overloading.

While in the initial operational stages the reactors were dominated by Firmicutes phylum, with the increase of the OLR there was an increase alongside to Synergistota, Planctomycetota, and Verrucomicrobiota, likely due to the higher influx of complex cellulosic compounds.

#49: short contribution

continued

REFERENCES

The versatile methanogen *Methanosarcina* was also favored, in response to the higher OLR. The increased prevalence of Synergistota (particularly the genus Syner-01) and *Methanosarcina* suggests a shift towards methanogenesis via syntrophic acetate oxidation (Zhang et al., 2022). This hypothesis is reinforced by the positive correlations found between hydrogenotrophic *Methanoculleus*, *Methanosarcina* and suspected CO_2 producing syntrophic acetate-oxidizing bacteria (SAOBs). Despite their proliferation, hydrogonotrophic methanogens were unable to effectively convert the CO_2 produced by SAOBs into methane due to its abrupt boost at the beginning of period 2.

The discrepancy between batch and continuous feed findings could be attributed to biochar outflow or insufficient retention time, that prevents microbial colonization. Future research is needed to determine the optimal biochar application method in order to maximize its advantageous aspects in realistic long-term anaerobic digestion processes.

REFERENCES

Chiappero, M., Norouzi, O., Hu, M., Demichelis, F., Berruti, F., Di Maria, F., Mašek, O., & Fiore, S. (2020). Review of biochar role as additive in anaerobic digestion processes. *Renewable and Sustainable Energy Reviews*, *131*, 110037. https://doi.org/10.1016/j.rser.2020.110037

Hu, J., Stenchly, K., Gwenzi, W., Wachendorf, M., & Kaetzl, K. (2023). Critical evaluation of biochar effects on methane production and process stability in anaerobic digestion. *Frontiers in Energy Research*, *11*. https://doi.org/10.3389/fenrg.2023.1205818

Zhang, L., Yuan, Y., Zhang, Y., & Liu, Y. (2022). Exploring key factors in anaerobic syntrophic interactions: Biomass activity, microbial community, and morphology. *Bioresource Technology*, *363*, 127852. https://doi.org/10.1016/j.biortech.2022.127852

#50: short contribution

Molecular Insights of Soil Microbial Communities in Response to Digestate Fertilization

Magkdi Mola1,2, Charitini Nikolaidou1,2, Maria-Chiara Valerin3, Stefano Campanaro 3, Ioannis Mylonas4, Vassilis Aschonitis5, Nikolaos Monokrousos2, Panagiotis G. Kougias1

1Waste Management and Bioprocessing Laboratory, Soil and Water Resources Institute, Hellenic Agricultural Organization Dimitra, Thermi, Thessaloniki, 57001, Greece 2University Center of International Programmes of Studies, International Hellenic University, Thessaloniki, 57001, Greece 3Department of Biology, University of Padova, Via U. Bassi 58/b, 35121 Padua, Italy 4Institute of Plant Breeding and Genetic Resources, Hellenic Agricultural Organization Dimitra, Thermi, Thessaloniki, 57001, Greec 5Soil and Water Resources Institute, Hellenic Agricultural Organization Dimitra, Thermi, Thessaloniki, 57001, Greece

Corresponding author: mmola@ihu.edu.gr

Digestate is the solid and liquid byproduct left after organic materials such as food waste, manure, or crop residues are processed in a biogas plant through a procedure called anaerobic digestion. Digestate is increasingly used as a sustainable fertilizer, however, its impact on soil microbial communities remains insufficiently explored, given that it contains a diverse array of microorganisms that could alter soil health. This study investigates the effects of digestate application on soil microbial communities using molecular approaches, comparing its impact with that of conventional fertilizer and a mixed fertilizer composed of conventional and digestate fertilizers. We assessed microbial community composition after one season of maize growth in an agricultural field, focusing on soil bacteria (16S) and fungi (ITS). Our findings indicate that digestate did not negatively affect soil microbial communities, as no significant differences were observed in microbial diversity or relative abundance compared to conventional and mixed fertilizer. This result aligns with a previous mesocosm study, which reported no significant impacts on soil microbial communities attributed to the short-term duration of digestate application (Nikolaidou et al., 2024). These outcomes highlight the resilience and adaptive capacity of native microbial communities, effectively maintaining their balance and structure despite the introduction of exogenous microorganisms (Mola et al., 2024). Going a step further, we investigated the role of specific biomarkers to understand potential shifts in microbial functions. Plots fertilized with conventional fertilizer exhibited a greater number of taxa identified as biomarkers, with most of them being microorganisms involved in nutrient cycling and plant growth (**fig.1**). Moreover, we identified soil nematodes using microscopic analysis to examine their community structure. Our observations revealed that digestate fertilization increased the abundance of bacterivorous nematodes while reducing the population of herbivorous nematodes, indicating that nematodes were more responsive to fertilization compared to soil microorganisms. Overall, these results suggest that digestate can serve as a viable alternative fertilizer, offering a sustainable solution while maintaining or even enhancing soil health, without causing significant disruption to the soil's microbial balance.

#50: short contribution

continued

Keywords:

Digestate, Alpha diversity, Soil microbial communities, Sustainability, Soil nematodes

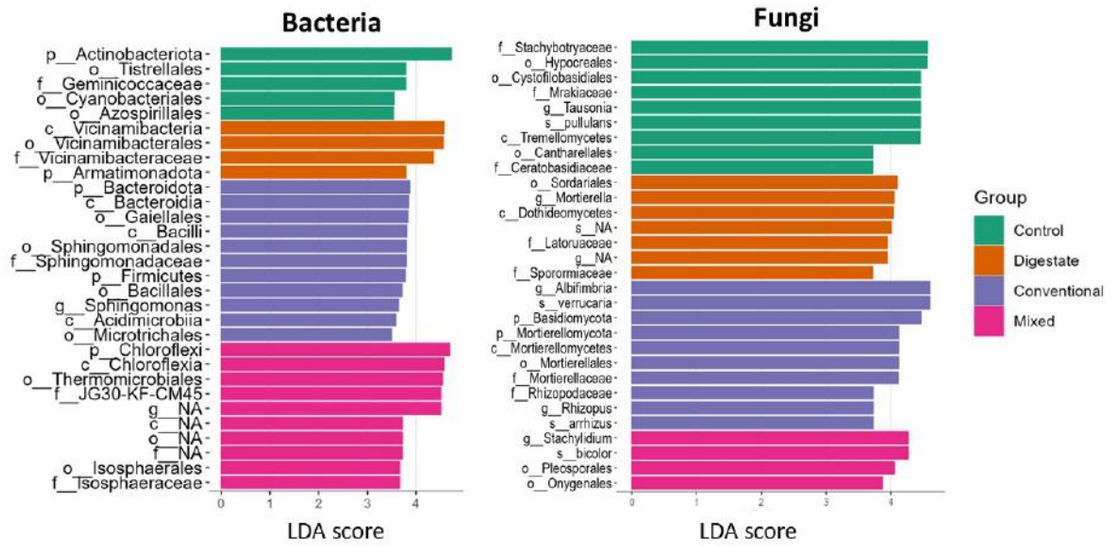


Fig. 1. Linear discriminant analysis effect size (LEfSe) analysis identifying differences in abundant fungal and bacterial taxa between treatments. The bar plots represent the significantly differential taxa between treatments based on the effect size (LDA score > 3).

REFERENCES

Mola, M., Kougias, P.G., Statiris, E., Papadopoulou, P., Malamis, S., Monokrousos, N., 2024. Short-term effect of reclaimed water irrigation on soil health, plant growth and the composition of soil microbial communities. Science of The Total Environment 949, 175107. https://doi.org/10.1016/j.scitotenv.2024.175107

Nikolaidou, C., Mola, M., Papakostas, S., Aschonitis, V.G., Monokrousos, N., Kougias, P.G., 2024. The effect of anaerobic digestate as an organic soil fertilizer on the diversity and structure of the indigenous soil microbial and nematode communities. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-024-32850-9

#51: short contribution

Comparative analysis of the gut microbiome among the five Greek species of the *Viperidae* family

Vasileios Balaktsis1, Thomas Daftsios1, Maria-Marily Christou1, Magkdi Mola2, Spiros Papakostas2, Konstantinos Sagonas1

1Department of Zoology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

2Department of Science and Technology, School of Science and Technology, International Hellenic University, Thessaloniki, Greece

The gastrointestinal tract of vertebrates hosts diverse microbial communities that play an essential role in various physiological processes and may even influence the host's evolutionary potential. However, research on gut microbiota has primarily focused on humans and model organisms leaving a significant knowledge gap regarding its role in wild animal populations, including snakes, and its role in local adaptation. Vipers are widely distributed worldwide, with Greece being home to five species, among which *Vipera ammodytes* is the most common. The venom of the *Viperidae* family is predominantly hemotoxic, containing toxins that severely impact blood coagulation, making their bites not only painful, but also potentially life-threatening. Studying the gut microbiota of these species is not only relevant from ecological and evolutionary standpoints but also because it may uncover potential links between microbial communities and venom composition. To investigate this relationship, we collected intestinal samples from several individuals representing all five different species of *Viperidae* found in Greece (7 gut samples from *Vipera ammodytes*, 8 from *Macrovipera schweizeri*, 6 from *Montivipera xanthina*, 3 from *Vipera berus bosniensis* and 3 from *Vipera graeca*). We amplified the microbial 16S rRNA gene and sequenced it using the MinION sequencing platform. Here, we present the initial characterization and comparative analysis of those microbial community profiles.

#52: short contribution

Predicting Anticancer Drug Efficacy with Machine Learning: A Data-Driven Approach

E. Pateras1, I. Vizirianakis2, G. Tzimagiorgis3, A. Malousi3

1Interdepartmental Postgraduate Program in Precision Medicine, School of Medicine, A.U.Th. 2Department of Pharmacy, A.U.Th. 3Laboratory of Biological Chemistry, School of Medicine, A.U.Th

Email: epater@auth.gr

Introduction: Predicting the pharmacological response to anticancer drugs remains a significant challenge due to the heterogeneity of cancer and the complexity of the underlying molecular mechanisms. Machine learning (M), as a field of artificial intelligence, employs computational algorithms to generate predictive models, uncovering patterns and insights that may not be discernible through conventional analytical methods. As a result, ML not only enhances pharmacological response prediction but also facilitates the discovery of biologically relevant insights (1,2). This study aims to develop ML methodologies for predicting the pharmacological response of anticancer drugs based on the gene expression profiles of cancer cell lines. Furthermore, the study compares the predictive performance of features selected through computational techniques with those derived from biological knowledge. The proposed methodology is implemented using two anticancer agents: Cetuximab and Uprosertib.

Materials and Methods: The half-maximal inhibitory concentration (IC50) measurements of the cancer cell lines were obtained from the GDSC resource (3) and the gene expression data was retrieved using the PharmacoGX pharmacogenomics package (4). In total, 783 cancer cell lines were analyzed for Cetuximab and 820 for Uprosertib. Feature selection was implemented using the Recursive Feature Elimination (RFE) algorithm, while the biological pathway information related to the Cetuximab and Uprosertib targets was obtained from KEGG via the R package KEGGREST (5). The ErbB and MAPK pathways were investigated for the Cetuximab target mechanism, while for Uprosertib, the investigated pathways were PI3K-Akt, mTOR, and Insulin.

Results: For Cetuximab, a predictive model was developed using a Linear Regression algorithm following computational feature reduction, achieving R2 = 0.982 and RMSE = 0.146. The regression models trained on gene features that were derived from biological pathways exhibited lower predictive performance. Specifically,the model based on the ErbB signaling pathway achieved R2 = 0.24 and RMSE = 0.96, while the model based on the MAPK signaling pathway yielded R2 = -0.23 and RMSE = 1.2, both using Linear Regression. For Uprosertib, the model developed using computational feature reduction and a Linear Regression algorithm demonstrated high predictive accuracy (R2 = 0.984, RMSE = 0.090). The regression models trained on gene features from specific biological pathways using a Support Vector Regression (SVR) algorithm showed more modest performance: the PI3K-Akt pathway model achieved R2 = 0.121 and RMSE = 0.675, the mTOR pathway model obtained R2 = 0.134 and RMSE = 0.669, and the Insulin signaling pathway model resulted in R2 = 0.098 and RMSE = 0.682.

continued

Conclusion: Feature selection using computational techniques resulted in models with significantly better performance compared to the biologically-informed feature sets that were derived by the drug target pathways. Interestingly, the genes selected by computational methods included significant known genes directly related to the biological target pathways of each drug, such as EGF, AREG, MAPK9, and PTEN for Cetuximab, and PTEN, FOXO3, IL7R, and EGF for Uprosertib, suggesting that computational methods could be informative in deciphering the complex interactions of the underlying molecular mechanisms.

REFERENCES

Rebala G, Ravi A, Churiwala S. Machine Learning Definition and Basics. An Introduction to Machine Learning [Internet]. 2019 [cited 2025 Feb 8];1–17. Available from: https://link.springer.com/chapter/10.1007/978-3-030-15729-6_1

KORDING KP, BENJAMIN AS, FARHOODI R, GLASER JI. The Roles of Machine Learning in Biomedical Science. Frontiers of Engineering [Internet]. 2018 Jan 22 [cited 2025 Feb 8]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK481619/

Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res [Internet]. 2013 Jan 1 [cited 2025 Feb 10];41(Database issue). Available from: https://pubmed.ncbi.nlm.nih.gov/23180760/

Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. Bioinformatics [Internet].

2016 Apr 15 [cited 2025 Feb 8];32(8):1244–6. Available from: https://dx.doi.org/10.1093/bioinformatics/btv723

5. Package “KEGGREST” Title Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). 2025 [cited 2025 Feb 8]; Available from: https://www.kegg.jp/kegg/rest/

#53: keynote lecture

Health2030 Genome Center@SMOC providing genomic medicine at scale

Ioannis Xenarios

Faculty of Biology and Medicine. Centre for Integrative Genomics (CIG)

Lausanne

Switzerland

The Health2030 Genome Center, part of the Swiss Multi-Omics Center, offers cutting-edge technological platforms and analytical expertise to advance genomic medicine. In this presentation, I will discuss the key requirements for establishing and operating a center capable of delivering rapid, fully accredited genomic services to hospitals across a wide range of diseases. I will illustrate these capabilities through a series of specific use cases and highlight the associated research challenges encountered in their implementation.

#54: short contribution

In silico discovery of novel therapeutic targets that counteract cytokine-induced transcriptomic changes in iPSC-derived β cells associated with type 1 diabetes

Theodora Papadopoulou, Chiara Vinci, Hadis Shakeri, Florian Szymczak, Yue Tong, Izzet Mehmet Akçay, Decio Eizirik, Miriam Cnop

ULB Center for Diabetes Research, Université Libre de Bruxelles, Brussels, Belgium

Corresponding author email: Miriam.cnop@ulb.be

Keywords: Type 1 diabetes, Cytokines, Inflammation, Induced pluripotent stem cells, pancreatic β cells, RNA sequencing, Differential expression, Drug repurposing

Introduction:

Type 1 diabetes (T1D) is a chronic autoimmune disease, affecting 8.4 million people worldwide, in which the immune system targets pancreatic β cells1. Destruction of β cells in T1D results from the combined effects of inflammation and autoimmunity. The cytokines interferon-α (IFNα), interferon-γ (IFNγ) and interleukin-1β (IL1β) are mediators of β cell death in the pathogenesis of T1D2. Human induced pluripotent stem cells (iPSCs) differentiated into β-like cells represent an exciting model to investigate disease mechanisms and identify novel therapeutic targets. We aimed to identify potential therapeutic targets for β cell preservation in T1D by evaluating the similarities between β cells from T1D patients and iPSC-β cells exposed to inflammatory stress and crossing them against publicly available chemical compound signatures.

Data & Methods

We charted by RNA-seq the transcriptome of magnetic-activated cell sorting (MACS)-purified iPSC-derived β-like cells exposed to IFNα or IFNγ+IL1β for 24 hours (n=6 for both). We compared the transcriptomic signatures to the transcriptomes of β cells from T1D donors3 using RedRibbon rank-rank hypergeometric overlap4. We crossed overlapping genes against the transcriptomes of chemical-exposed human cell lines, available in the Connectivity Map database5, to identify compounds potentially reverting the gene signature of β cell failure in T1D.

#54: short contribution

continued

Results

IFNγ+IL1β and IFNα exposure of MACS-purified iPSC-derived β cells resulted in 5693 and 500 differentially expressed genes, respectively (FDR<0.05). In both conditions, inflammatory pathways were upregulated and translation-related pathways downregulated. In the comparison of transcriptomes of β cells from T1D patients and cytokine-exposed iPSC-β cells, RedRibbon showed a significant intersection of 2159 upregulated genes for IFNα and 1254 for IFNγ+IL1β. Enrichment analysis of the commonly upregulated genes showed interferon responses and lymphoid-non-lymphoid interactions for both. The top 150 commonly upregulated genes were submitted to the Connectivity Map to identify drugs that may potentially reverse the gene signature. This pointed to bromodomain inhibitors (tau score: IFNα -89, IFNγ+IL1β -91). Bromodomain inhibitors have previously been shown to reduce the proinflammatory effects of these cytokines in human islets and EndoC-βH1 cells in vitro.

Discussion

The iPSC-β cell transcriptomic responses to IFNα and IFNγ+IL1β recapitulate those of β cells from people with T1D. Mining these signatures identifies known and novel therapeutic targets, including bromodomain inhibitors. The anti-inflammatory properties of these compounds could protect β cells from inflammation in T1D.

REFERENCES

1. <IDF_Atlas_10th_Edition_2021.pdf>.
2. Wachlin, et al., J Autoimmun, 2003. 20(4): p. 303-12
3. Russell, et al., Diabetes, 2019. 68(5): p. 988-1001
4. Piron, et al., Life Science Alliance, 2023: 10.26508/lsa.202302203
5. Subramanian, et al. Cell, 2017. 171(6):1437–1452

#55: short contribution

Integrative -Omics for Construction of Gene Regulatory Networks towards NKT Cell Identity Understanding

Dafni Anagnostopoulou, Arianna Ntatsopoulou, Dimitris Konstantopoulos, Mihalis Verykokakis

Institute for Fundamental Biomedical Research BSRC Alexander Fleming

email: verykokakis@fleming.gr

Natural killer T (NKT) cells are an oligoclonal, relatively homogeneous T cell population that acquire life-long effector functions during a step-wise thymic differentiation process, without prior encounter with foreign antigen1. Their thymic maturation involves transitioning from progenitor to mature cells and culminates in the differentiation of three NKT cell subsets, with distinct functional characteristics. Although there are minor variations among NKT subsets, complex interactions between transcriptional and epigenetic regulators control their differentiation. We used a multi-omics bioinformatics framework to decipher these regulatory mechanisms, combining transcriptomic and epigenomic data at the single-cell level.

Cell identity is defined by the function of lineage-determining transcription factors (TFs), which control unique gene expression programs, by directly binding to accessible cis-regulatory elements at the genomic loci of their target genes. Therefore, TF expression/activity and chromatin architecture contribute to the wiring of gene regulatory networks (GRNs), which are essential to orchestrate cell fate decisions in a cell-intrinsic manner and are maintained throughout life. Here, we integrated single-cell transcriptomic and epigenomic data2, with the aim to reconstruct dynamic GRNs and reveal regulatory mechanisms that guide cell differentiation.

To begin to untangle the transcriptional programs involved in NKT cell development, we purified thymic NKT cells through flow cytometry cell sorting and performed single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq). Cell clustering based on gene expression profiling revealed greater heterogeneity within the NKT cell population than previously anticipated, uncovering several distinct subpopulations. To explore developmental relationships between cell types, we generated pseudotime trajectories3, which allowed identification of transient cell states and highly dynamic expression patterns, particularly at branching points. In parallel, our scATAC-seq analysis4 identified distinct chromatin accessibility landscapes that characterize specific NKT subtypes and uncovered accessible TF binding motifs. Integration of scRNA-seq and scATAC-seq data, revealed TF regulators that influence NKT cell differentiation, providing a thorough understanding of the interaction between TFs and chromatin.

#55: short contribution

continued

We are currently expanding these analyses by integrating accessible TF binding motifs with the expression and epigenetic patterns of the corresponding transcription factors and their putative target genes in order to construct robust GRNs that control differentiation of NKT cells. Our goal is to project these GRNs across pseudotime, to identify key regulons that are dynamically regulated throughout differentiation, especially those that play vital roles at developmental branch points and are thus lineage-specific. By highlighting important TFs and cis-regulatory elements that determine cell fate decisions, our methodology will provide an in-depth comprehension of the regulatory mechanisms influencing NKT lineage commitment and diversification.

Our approach provides a framework for analyzing GRNs during NKT cell development. With implications for immunology, developmental biology, bioinformatics, and possible therapeutic applications that target NKT-mediated immune responses, these findings advance our understanding of cell identity.

REFERENCES

[1] M. Verykokakis, et al., Immunological Reviews 261, 177-197(2014).

[2] P. Badia-i-Mompel et al., Nature Reviews Genetics 24, 739–754 (2023).

[3] K. Street, et al., BMC Genomics 19, 477 (2018).

[4] J.M. Granja, et al., Nature Genetics 53, 403–411 (2021).

#56: invited seminar

Translational tools and repositories for translational medicine in the post-genomic era

George P. Patrinos

I Director General and Director of Research, Hellenic Pasteur Institute, Athens, Greece; 2 Professor; University of Patras School of Health Sciences, Department of Pharmacy, Patras, Greece; 3 Adjunct Full Professor; United Arab Emirates University, College of Medicine and Health Sciences, Department of Genetics and Genomics, Al-Ain, United Arab Emirates; 4 Adjunct Faculty; Erasmus MC, Faculty of Medicine and Health Sciences, Department of Pathology, Clinical Bioinformatics Unit, Rotterdam, the Netherlands. Email: gpatrinos@upatras.gr; general.director@pasteur.gr

In the post-genomic era, the rapid evolution of high-throughput genotyping technologies and the increased pace of production of genetic research data, are continually prompting the development of appropriate informatics tools, systems and databases as we attempt to cope with the flood of incoming genetic information. Alongside new technologies that serve to enhance data connectivity, emerging information systems should contribute to the creation of a powerful knowledge environment for genotype-to-phenotype information in the context of translational medicine. In the area of pharmacogenomics and personalized medicine, it has become evident that database applications providing important information on the occurrence and consequences of gene variants involved in pharmacokinetics, pharmacodynamics, drug efficacy and drug toxicity, will become an integral tool for researchers and medical practitioners alike. At the same time, two fundamental issues are inextricably linked to current developments, namely data sharing and data protection. In this lecture, the impact of high throughput and next generation sequencing technology and its impact on pharmacogenomics research and clinical implementation of genomic medicine will be addressed. In addition, advances and challenges in the field of pharmacogenomics information systems will be discussed, which in turn prompted the development of an integrated electronic ‘pharmacogenomics assistant’. The system is designed to provide personalized drug recommendations based on linked genotype-to-phenotype pharmacogenomics data, as well as to support biomedical researchers in the identification of pharmacogenomic related gene variants.

#57: invited seminar

Unveiling the Complex and Enigmatic Biology of Cancer of Unknown Primary Through Interrogation of the Genomic and Transcriptomic Landscape

Harmen J. G. van de Werken

Department of Immunology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands

h.vandewerken@erasmusmc.nl

Introduction: Cancer of Unknown Primary (CUP) is a challenging and burdensome diagnosis, defined by the inability to identify its tissue of origin. CUP tumors typically exhibit high self-renewal, migratory capacity, plasticity, and evasion of the immune system. However, the biological mechanisms responsible for these behavioral patterns in CUP remain to be further understood and are essential for improving diagnostic accuracy and developing effective therapies. Using in-depth bioinformatics methods, we examine the whole-genome and transcriptome landscapes of 283 CUP patients. Furthermore, we compared the CUP genomic and transcriptomic profiles of metastatic tumors with known primary sites, such as metastatic Non-Small Cell Lung Cancer (mNSCLC), to identify and understand biological differences unique to CUP.

Materials and Methods: Whole Genome sequencing (WGS; N = 283) and RNA Sequencing (RNA-seq; N = 79) data from CUP patients, as well as WGS (N = 615) and RNA-seq (N = 304) data from mNSCLC samples, were obtained through the Hartwig Medical Foundation. The HMF WGS workflows provided data to analyze genomic features, predict mutational signatures, and assess cancer drivers using the ratio of non-synonymous to synonymous substitutions (dN/dS). Somatic copy number alterations were assessed using GISTIC 2.0. Tissue-of-Origin predictions and potential therapeutic targets were identified using the CUPPA algorithm and iClusion database, respectively. Dimensional reduction analysis and hierarchical clustering were applied to identify distinct clusters, while proliferation, immune, and fibrotic signatures were mapped to assign dominant functional groups.

Results: The genomic and transcriptomic profiles of CUP revealed substantial heterogeneity among patients. Mutational signatures identified four subgroups associated with smoking, APOBEC activity, UV radiation, and DNA damage repair pathways. Smoking- and UV-related signatures were strongly associated with predicted lung and skin tumors, respectively. The frequently mutated genes included cell-cycle regulators (e.g., TP53, RB1, and CDKN2A) and immune-related genes (e.g., HLA-I). Notably, 66% of patients exhibited potentially actionable genetic alterations, and 58% of CUP samples could be assigned to a potential Tissue-of-Origin. Interestingly, the whole transcriptome data revealed three distinct tumor microenvironment CUP subgroups, characterized by less immune cell infiltration and higher expression of transcription factors MYC and E2F targets, indicating a higher proliferation rate.

Discussion: WGS analysis revealed highly diverse genomic alterations in CUP, similar to those found in other metastatic tumors. Additionally, our findings support the clinical applicability of WGS to indicate the Tissue-of-Origin and offer tumor-agnostic therapeutic strategies. Transcriptomic data suggested that immune escape and rapid proliferation contribute to the aggressive behavior of CUP tumors.

#58: short contribution

Uncertainty-Aware Molecular Subtyping in Cancer Using Conformal Machine Learning Models

Christina Papangelou1, Konstantinos Kyriakidis2, Pantelis Natsiavas3, Ioanna Chouvarda1, Andigoni Malousi1,4

1School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece,
2Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA

3Institute of Applied Biosciences, Center for Research and Technology Hellas, Thessaloniki, Greece,
4GENeTres Research Group, Center for Interdisciplinary Research and Innovation, Thessaloniki, Greece

email: andigoni@auth.gr

Machine learning (ML) and genomic medicine could potentially transform personalized healthcare by improving disease diagnosis, risk stratification, and treatment planning. However, ML models' stochasticity, domain shifts, and data heterogeneity raise concerns about prediction reliability, making rigorous uncertainty quantification essential. Conformal Prediction (CP)1,2 provides a robust distribution-free framework for assessing uncertainty3, enhancing trust in AI-driven genomic applications under user-defined error rates.

This study applies CP in the molecular subtyping of diffuse large B-cell lymphoma (DLBCL)using transcriptomicdata.DLBCLisahighlyheterogeneousmalignancy classified into three subtypes based on the transcriptional profiles of the patients (activated B-cell like (ABC), germinal center B-cell like (GCB), and molecular high-grade B-cell lymphoma (MHG)). Despite their shared molecular profiles, the differences between subtypes are critical for prognosis and treatment selection, especially with targeted therapies. To address the uncertainty of the predictions, we applied an inductive CP model on 1,311 formalin-fixed, paraffin-embedded biopsies (GEO Data series: GSE181063), after feature selection and statistical validation of the independent and identically distributed (i.i.d.) assumption.

The CP-augmented XGBoost model reduced misclassification rates from 16.25% to 4.8% at a 95% confidence level. It flagged 35.6% of cases as uncertain, preventing 12 out of 17 misclassifications made by the base model, and provided singleton predictions for 30.2% of previously unclassified patients. Results suggest that MHG has a distinct transcriptional signature, while further analysis is needed to distinguish GCB and ABC profiles. Despite reducing false predictions, the CP model was more conservative, yielding fewer singleton predictions (482 in total: {MHG} = 32, {GCB} = 295, {ABC} = 155).

#58: short contribution

continued

In a generalizability test on an external 789 RNA sample dataset (GEO Data series: GSE117556) with distribution shift, the CP model achieved 96.6% empirical coverage, flagging 112 out of 129 misclassified samples as uncertain, reducing failure risk by 80% under distribution shifts. This confirms the ability of the conformalized models to generalize to different distributions, emphasizing the robustness in handling data variability and the need to integrate CP frameworks in genomic medicine for safer predictions especially when critical in clinical decision making.

Overall, our findings highlight the potential of the conformalized model to improve ML in multi-class molecular subtyping, advancing uncertainty-informed applications in real-world practice.

REFERENCES

Vovk, V., Gammerman, A., and Shafer, G. (2005). Algorithmic learning in a random world, 29. Springer.

Shafer, G., and Vovk, V. (2008). A tutorial on conformal prediction. J. Mach. Learn. Res. 9. doi:10.48550/arXiv.0706.3188.

Angelopoulos, A. N., and Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. doi:10.48550/arXiv.2107.07511

Sha, C., Barrans, S., Cucco, F., Bentley, M. A., Care, M. A., Cummin, T., et al. (2019). Molecular high-grade b-cell lymphoma: defining a poor-risk group that requires different approaches to therapy. J. Clin. Oncol. 37, 202–212. doi:10.1200/jco.18.01314.

#59: short contribution

From point mutations to Brunner Syndrome: How the point mutation influences electrostatic interactions and performance of monoamine oxidase A enzyme

Martina Rajić, Janez Mavri, Jernej Stare

National Institute of Chemistry, Ljubljana, Slovenia

Enzymes play a vital role in essential biological processes, and even minor disruptions in their function can lead to various disorders or diseases. One potential cause of such disruptions lies in genetic mutations that alter the amino acid sequence of the enzyme. Although a single amino acid change may appear insignificant, it can profoundly influence enzyme function—especially when charged residues are involved—by modifying interactions within the active site. These mutations can also affect local structural stability and protein folding, thereby influencing catalytic efficiency.

This study investigates clinically relevant point mutations in the monoamine oxidase A (MAO-A) enzyme, a key player in regulating serotonin levels within the central nervous system. These mutations have been associated with neurodevelopmental disorders, including aggression and intellectual disability. Using a multiscale molecular simulation approach, including the empirical valence bond (EVB) method, we analyze the catalytic step of serotonin degradation. Our results show that mutations significantly increase the activation energy barrier by few kcal/mol, leading to significant reduction in the reaction rate. This reduction is functionally comparable to a complete loss of enzymatic activity.

Additionally, our analysis indicates that these mutations disrupt electrostatic stabilization of the transition state, highlighting the importance of preorganized electrostatics in enzymatic catalysis. This work provides valuable insights into the molecular basis of neuropsychiatric disorders and demonstrates how computational techniques may assist in predicting disease-related enzyme dysfunctions based on genetic information.

REFERENCES

H. G. Brunner, et al. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science 1993, 262, 578 – 580.

T. M. A., et al. New insights into Brunner syndrome and potential for targeted therapy. Clin Genet 2016, 89, 120 – 127.

A. Prah, D. Pregeljc, J. Stare, J. Mavri. Brunner syndrome caused by point mutation explained by multiscale simulation of enzyme reaction. Scientific Reports 2022, 12, 21889.

M. Rajić, and J. Stare. Investigation of Electrostatic Effects on Enzyme Catalysis: Insights from Computational Simulations of Monoamine Oxidase A Pathological Variants Leading to the Brunner Syndrome. Journal of Chemical Information and Modeling, 2025.

#60: invited seminar

Agentomics-ML: an agentic system for automated -omics machine learning model development

Panagiotis Alexiou

University of Malta

Recent developments in large language models allow for the development of LLM-based agentic systems that autonomously pursue a complex goal that requires reasoning. Such agents can be augmented with various programmatic tools such as code execution and data retrieval, furthering their capabilities beyond token generation and expanding their knowledge base beyond training data.

We showcase the development of Agentomics-ML, an autonomous agentic system that can produce ready-to-run machine learning models for classifying arbitrary -omics datasets. This system can explore the provided dataset and problem domain, plan necessary steps to reach a sufficient classifier, execute these steps, adjust the plan based on the output of used tools and encountered errors, and iterate when necessary.

We measure the generalization of models produced by our system on various classification datasets from the field of computational biology and compare against other agentic systems and existing state-of-the-art classifiers that are published in peer-reviewed journals.

#61: short contribution

In silico antimicrobial peptide discovery in Animal genomes

N. Pyrgelis 1, N. Vakirlis 2 1,2Hellenic Pasteur Institute, Athens

Greece

email: n.pirgelis@pasteur.gr

Antibiotic resistance, a pressing public health issue, urges for the discovery of novel antimicrobial agents. Antimicrobial peptides (AMPs), ubiquitous small proteins that are part of the innate immune response of all organisms are a promising source of novel potential therapeutics. In eukaryotes they are usually expressed as precursor peptides and they remain inactive until enzymatic cleavage removes their signal and propeptide domains, leading to a mature and functional peptide1. AMPs can also be encrypted, hidden within larger proteins, that can be released through protein degradation2. The increasing availability of genomes and transcriptomes coming from the animal kingdom, provides an untapped resource of sequence space for the discovery of putative novel AMPs3. Even though in silico identification with the use of machine learning approaches has made significant progress over the last years, most studies focus on prokaryotes4 and no dedicated pipeline exists capable of mining entire complex eukaryotic genomes.

In this work we present a custom pipeline, that screens eukaryotic genomes and transcriptomes and isolates small peptides which are then in silico cleaved, removing their respective signal peptides and propeptides. At the same time, we scan the complete predicted proteome to predict encrypted AMPs. The resulting “mature” and encrypted peptides are analysed by multiple state-of-the-art classifiers, which label them as AMPs or non-AMPS. The final list of candidate-AMPs is characterised by multiple confidence score levels, determined by the consensus among the different classifiers and their presence in both genomic and transcriptomic data. Using the sponge Ephydatia muelleri as a case study, our results demonstrate the pipeline’s capability to identify numerous candidates and its scalability to thousands of species.

#61: short contribution

continued

REFERENCES

1. Bin Hafeez, A., Jiang, X., Bergen, P. J. & Zhu, Y. Antimicrobial Peptides: An Update on Classifications and Databases. Int. J. Mol. Sci. 22, 11691 (2021).
2. Torres, M. D. T. et al. Mining for encrypted peptide antibiotics in the human proteome. Nat. Biomed. Eng. 6, 67–75 (2022).
3. Hotaling, S., Kelley, J. L. & Frandsen, P. B. Toward a genome sequence for every animal: Where are we now? Proc. Natl. Acad. Sci. 118, e2109019118 (2021).
4. Santos-Júnior, C. D. et al. Discovery of antimicrobial peptides in the global microbiome with machine learning. Cell 187, 3761-3778.e16 (2024).

#62: short contribution

Exploring Genetic Adaptation and Microbial Dynamics in Engineered Anaerobic Ecosystems via Strain-level Metagenomics

Ghiotto G.1, Xirostylidou A.2, Gaspari M.2, Kougias P.G.2, Campanaro S.1, Treu L.1

1Department of Biology, University of Padua, via U. Bassi 58/b, 35131 Padova, Italy

2Soil and Water Resources Institute, Hellenic Agricultural Organisation Dimitra, Thermi, Thessaloniki 57001, Greece

E-mail: stefano.campanaro@unipd.it

Genetic heterogeneity is inherent in all microbial populations, with sympatric cells of the same species often exhibiting single nucleotide variants (SNVs) that influence phenotypic traits, including metabolic efficiency. Variant frequencies fluctuate over time due to genetic drift or selection processes driven by specific environmental pressures, shaping microbiome evolution. However, the precise evolutionary dynamics of strain-level differences under environmental stress remain poorly understood. Here, we present a first-of-its- kind study tracking the adaptive evolution of an anaerobic, carbon-fixing microbiota in a controlled engineered ecosystem designed for carbon dioxide bioconversion into methane. Genetic heterogeneity was accessed over a one-year continuous adaptive laboratory evolution, examining responses to variations in gaseous substrate flow rate (Stage I) and to carbon and electron source deprivation (Stage III). Genome-resolved metagenomics characterized the microbiome of two thermophilic anaerobic reactors filled with plastic or carbon particles. A total of 330 medium-to-high quality MAGs were recovered, comprising 17 archaeal and 313 bacterial genomes. Using strain-resolved metagenomics combined with an ad hoc variant calling and phasing approach, we mapped SNVs trajectories and observed that the two dominant microbes, Methanothermobacter thermautotrophicus and Methanothermobacter marburgensis, maintained distinct sweeping haplotypes over time, likely due to niche-specific metabolic roles. At this population genetic level, strain dynamics were considerably more intricate than those described by the neutral theory of molecular evolution1. Mapping SNVs onto the genes and calculating the dN/dS revealed selective pressure on genes associated with the hydrogenotrophic methanogenesis pathway, particularly mtr and mcr. Furthermore, for the first time, by integrating population genetic statistics with three-dimensional peptide reconstruction, we identified mer and mcrB as potential drivers of archaeal strain competition.

continued

Structural analysis leveraging AlphaFold2 predictions indicated that amino acid substitutions in M. thermautotrophicus reduced steric hindrance in conserved binding clefts of McrB, potentially enhancing substrate docking3. Conversely, the hydrophilic-to-hydrophobic shift amino acid observed in M. marburgensis has the potential to alter the tetrameric state of Mer. Previous studies indicate that the tetramer interface in Methanothermobacter species has lower hydrophobicity than M. kandleri4, which affects protein complex assembly. Therefore, the observed switch from Glu52 to Val52, corresponding to Val54 in M. kandleri4, enhances the region's hydrophobic character, likely promoting Mer tetramerization. These alterations, supported by moderate Grantham distance metrics5, may have conferred a phenotypic advantage, driving shifts in strain dominance during Stages I and III. Our findings structurally demonstrate how environmental perturbations in bioconversion systems induce selection, leading to the fixation of specific haplotypes characterized by their unique genetic makeup when compared to other strains within the populations. This study advances our understanding of microbial adaptation and paves the way for targeted engineering of microbial communities to enhance bioconversion efficiency, with significant implications for sustainable energy and carbon management in anaerobic systems.

REFERENCES

Kimura M. Evolutionary Rate at the Molecular Level. Nature. 1968;217(5129):624-626. doi:10.1038/217624a0

Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

Ermler U. On the mechanism of methyl-coenzyme M reductase. Dalton Trans. 2005;(21):3451-3458. doi:10.1039/B506697B

Shima S, Warkentin E, Grabarse W, et al. Structure of Coenzyme F420 Dependent Methylenetetrahydromethanopterin Reductase from Two Methanogenic Archaea. J Mol Biol. 2000;300(4):935-950. doi:10.1006/jmbi.2000.3909

Grantham R. Amino Acid Difference Formula to Help Explain Protein Evolution. Science. 1974;185(4154):862-864. doi:10.1126/science.185.4154.862

#63: short contribution

Bio-Electrical Nanonetworks: a New Paradigm of Biological Nanocommunication based on Living Cable Bacteria Filaments

Konstantinos Kantelis1, Vassilios Asteriou1, Aliki Papadimitriou2, Anthi Petrou2, Christos A. Ouzounis1, Eleftherios Aggelis1, Ioannis Vizirianakis2, Petros Nicopolitidis1, Georgios I. Papadimitriou1

1Department of Informatics, 2School of Pharmacy

Aristotle University of Thessaloniki, Greece

Biological nanonetworks offer a unique communication paradigm in environments where biocompatibility, low complexity and energy availability are the primary concerns. However, they suffer from low bit rates, high propagation delay and high bit error rate. Bio-Electric Nanonetworks, a totally new class of biological nanonetworks is introduced, that is based on living wires made primarily of Cable Bacteria. Information is transmitted through the living wires in the form of electrical signal, and consequently, the aforementioned transmission characteristics are expected to be on par with the ones of electronic networks. Therefore, the new class of biological nanonetworks introduced in this work is expected to achieve a dramatic improvement of the transmission characteristics, in relation to the classic biological nanonetworks.

REFERENCES

Atkinson JT, Chavez MS, Niman CM & El-Naggar MY (2023) Living electronics: A catalogue of engineered living electronic components. Microbial biotechnology 16(3), 507–533. https://doi.org/10.1111/1751-7915.14171

Kantelis K, Asteriou V, Papadimitriou A, Petrou A, Ouzounis CA, Angelis L, S. Vizirianakis IS, Nicopolitidis P & Papadimitriou GI (2023) Bio-Electrical Nanonetworks: A New Paradigm of Biological Nanocommunication based on Living Cable Bacteria Filaments. In Proceedings of the 10th ACM International Conference on Nanoscale Computing and Communication (NANOCOM '23). Association for Computing Machinery, 148–149. https://doi.org/10.1145/3576781.3608737

Scilipoti S, Koren K, Risgaard-Petersen N, Schramm A & Nielsen LP (2021) Oxygen consumption of individual cable bacteria. Science advances 7(7), eabe1870. https://doi.org/10.1126/sciadv.abe1870

#64: short contribution

From Design to Validation and vice versa: Bioinformatics as a Catalyst in iGEM Synthetic Biology Projects

Errika Sylai*1,2, Margarita Papastergiou*1,2, Michalis Aivaliotis1,3,4

iGEM Thessaloniki, Thessaloniki, Greece

Department of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

Laboratory of Biological Chemistry, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece

Functional Proteomics and Systems Biology (FunPATh), Center for Interdisciplinary Research and Innovation (CIRI-AUTH), GR-57001 Thessaloniki, Greece

*These authors contributed equally to this work

The integration of bioinformatics and computational biology methods into synthetic biology has significantly enhanced the design and optimization of engineered biological systems. This work explores the pivotal role of computational biology across the design-build-test-learn (DBTL) cycle, with a focus on its application in the International Genetically Engineered Machine (iGEM) competition. Members of the dry lab team, in close cooperation with the wet lab team, play a critical role in the project by focusing on modeling, as well as software and hardware development, enabling integration of computational and experimental approaches. We elaborate on the use of computational tools in key synthetic biology workflows, including sequence design, codon optimization, molecular modeling, metabolic pathway simulation, and multi-omics analysis. We highlight platforms such as Benchling, SnapGene, Geneious, and genome-scale metabolic models that have supported construct development and in silico validation, through selected iGEM case studies. Our findings demonstrate how bioinformatics enhances reproducibility, scalability, and system-level insight in synthetic biology projects. By integrating computational and experimental approaches, iGEM teams are able to engineer more robust and innovative biological solutions. This presentation underscores the essential role of bioinformatics in synthetic biology and supports its continued integration into educational and research initiatives.

Keywords: Synthetic Biology, Bioinformatics, iGEM competition, Design-Build-Test-Learn (DBTL), Metabolic Modeling, in silico simulation

poster-only contributions

Certain abstracts appear here as poster presentations due to author preference or considerations related to program balance, representation, and thematic coherence. This designation does not reflect on the quality or merit of the work.

poster #01

Optimizing plant metabarcoding workflows for biodiversity monitoring

1 Nikos Pechlivanis, 2. Sofoklis Keisaris, 3. Fotis Psomopoulos, 4. Panagiotis Madesis

Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thermi, 57001, Thessaloniki, Greece

email: fpsom@certh.gr

Metabarcoding leverages high-throughput sequencing (HTS) technologies to identify plant species from environmental samples. However, challenges such as sequence quality discrepancies, incomplete taxon representation, and contamination can hinder accurate species identification1. For instance, the lack of standardized methods in DNA barcoding, including variations in DNA preservation, extraction techniques, marker and primer selection, and PCR protocols, contributes to these issues. Additionally, the incompleteness and accuracy of molecular reference databases can lead to mismatches between conventional morphological identification and barcode-based identification, further complicating precise species determination.

Enhanced reference databases and defined protocols for sampling, DNA isolation, sequencing, and bioinformatic analyses are essential for advancing taxonomic resolution and conservation efforts1. Currently researchers use open-source computational tools for raw sequence processing, like DADA2, Deblur, and VSEARCH for noise filtration and sequence clustering. These tools are often implemented into workflows and pipelines such as QIIME2, MOTHUR, and ampliseq. However, none of these have been generated with plant metabarcoding analysis in mind, and their usability and filter setting recommendations for plant metabarcoding have never been analyzed.

The objectives of this study are to improve plant metabarcoding through the optimization of reference databases, assessment of available bioinformatics tools, and formulation of best practices for sequence processing and analysis. A range of bioinformatics tools and pipelines will be evaluated for precision and efficacy, ensuring alignment with FAIR data principles. Ultimately, the study aims to provide a standardized and optimized protocol for plant metabarcoding that is reproducible and interoperable across different platforms and datasets, as well as, standardized methodologies and optimal practice guidelines for plant metabarcoding. These developments will facilitate biodiversity monitoring activities and improve the accuracy of taxonomic identification in ecological research.

REFERENCES

1. Kolter, A. & Gemeinholzer, B. Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. Genome 64, 265–298 (2021)

poster #02

On Early Detection of Prostate Cancer Through Machine Learning

Sadia Zafar

affiliation?

Men in the later years are more likely to develop prostate cancer, which is a deadly and heterogeneous disease. Prostate gland cells in cancer can be either benign or metastatic depending on whether they have spread from the original site, and because they have gone out of control, many factors cause prostate cancer including genetics, that can contribute to begun and continued. Receptors are used for the disease classification. Our goal of study is to identify genes that are differentially expressed under normal and tumors conditions. These genes can be useful as biomarkers or training data for a classifier predicting unknown data. To this end, we used two datasets with cancer and normal prostate cancer samples. After quality control, mapping, feature counting, and differential expression analysis, genes were extracted with a p-value greater than 1 and log2FC less than 1 or greater than -1 Network analysis was performed using Cytoscape and String, and functional enrichment and pathway analysis using DAVID. This was done to identify key pathways and interactions associated with our genes. Subsequently, the expression values of the selected genes were extracted from the feature count files. The files were divided into training and testing sessions for subsequent classifier training and testing. SVM, Random Forest.KNN, LR, and MLP classifiers were used through the application of various ML algorithms, including SVM Linear and LR, robust insights into classification effectiveness are gained With an accuracy, precision,recall, F1 score, specificity, and sensitivity of 1.00, KNN showed impeccable results on all parameters. This shows that there were no false positives nor false negatives in the KNN's classification of all normal and tumor samples. Given its immaculate performance, KNN appears to be a good fit for this datasets. However, further studies with larger samples are required for promising results to validate the effectiveness of the classification. In addition, the identified genes required further investigation as they may serve as biomarkers for the diagnosis of PCa.

poster #03

Systems Bioinformatics and Advanced Machine Learning for Cardiovascular disease – The PoCCardio Project

Anastasis Oulas & George Spyrou

Bioinformatics Department, The Cyprus Institute of Neurology & Genetics, 6 Iroon Avenue, 2371 Ayios Dometios, Nicosia, Cyprus

email: anastasioso@cing.ac.cy

The PoCCardio Project focusses on personalised medicine by using an advanced point-of-Care tool for stratified treatment in high-risk cardiovascular (CVD) patients. We are heading one of the work packages that entails the development of systems bioinformatics (SB) and advanced machine learning (ML) methodologies for obtaining proof of concept for the usability of a newly optimized analytic software tool that is expected to be deployed within the project. The software is anticipated to provide:

1) insights and real-time information on diagnostic and clinical data for CVD.

2) Optimized ML model for disease severity prediction and response to treatment for CVD and patients with extremely high risk for myocardial infraction (MI).

3) Feedback for a highly informative panel of biomarkers. A flow chart of the PoCCardio project is shown in Figure 1.

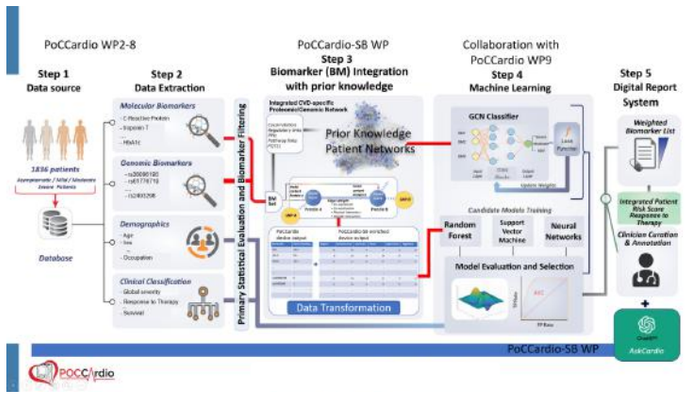


Figure 1. PoCCardio project flow chart describing the major WP with emphasis on the PoCCardio-SB WP which entails the Systems Bioinformatics (SB) and Advanced Machine Learning (ML) methodologies.

poster #04

Bult RNA *in silico* Drug Repurposing for pre-symptomatic and symptomatic Multiple Myeloma

Kyriaki Savva1, Marilena Bourdakou1, Dimitris Stellas2, Jerome Zoidakis3 and George M. Spyrou1

1.Bioinformatics Department, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus

2. Institute of Chemical Biology, National Hellenic Research Foundation (NHRF), Athens, Greece

3.Biotechnology Division, Biomedical Research Foundation, Academy of Athens, Athens, Greece

email: kyriakis@cing.ac.cy

Multiple Myeloma (MM) is a complex B-cell cancer marked by the uncontrolled growth of plasma cells in the bone marrow. Despite notable advancements in treatment options, MM continues to be an incurable cancer, leading to the majority of patients succumbing as the disease becomes resistant to therapy. MM arises from a pre-cancerous condition known as monoclonal gammopathy of unknown significance (MGUS), which can progress to either smoldering (asymptomatic) multiple myeloma (SMM) or MM. De novo drug discovery, focused on finding new treatments for specific diseases, has become both time-consuming and expensive. As a result, drug repurposing, which involves identifying existing drugs for new diseases, can be used. To tackle this crucial and still unmet medical challenge, we performed a stage-specific transcriptomic analysis of multiple myeloma (MM) progression, identifying key differentially expressed genes (DEGs) across mGUS, sMM, and MM stages. Gene Ontology analysis revealed dynamic changes in immune-related biological processes, highlighting a shift from heightened immune activity in mGUS to immune dysregulation in MM. Using multiple in silico drug repurposing tools and an internal scoring scheme that allowed for additional filtering and prioritisation, we identified stage-specific candidate drugs, leading to 25 candidate repurposed drugs for mGUS, 23 for sMM, and 66 for MM. These include geldanamycin and roscovitine (mGUS), olprinone and lamotrigine (sMM), and radicicol and entinostat (MM). Structural similarity assessments indicated minimal overlap with ongoing clinical trial drugs, suggesting structural novelty. GO enrichment based on drug targets uncovered both common and stage-specific biological processes, such as oxidative stress response, kinase activity regulation, and epigenetic modification. Drug synergy analyses revealed multiple promising combinations, including lenalidomide + retinoic acid and erlotinib + bortezomib, supported by consistent synergy across ZIP, Bliss, HSA, and Loewe models. These findings highlight novel therapeutic opportunities and support a precision medicine approach in MM.

Keywords: Multiple Myeloma; progression; computational drug repurposing; drug synergies

poster #05

Identification of ROH and HRR islands across the genome of European sheep breeds

Spyridon Vasileiadis1*, Arianna Manunza2, Johanna Ramirez-Diaz2, Paolo Cozzi2, Alessandra Stella2, Spyridon Papakostas3, Alexandros Triantafyllidis1,4, Valentina Tsartsianidou1,4

1Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece
2 Institute of Agricultural Biology and Biotechnology, National Research Council 20133 Milan, Italy 3International Hellenic University, 57001, Thermi, Thessaloniki, Greece
4 Genomics and Epigenomics Translational Research (GENeTres), Center for Interdisciplinary Research and Innovation, Balkan Center, 57001, Thessaloniki, Greece
*Corresponding author

Runs of Homozygosity (ROH) are continuous homozygous segments in an individual's genome, typically resulting from inbreeding or population bottlenecks (Ceballos et al., 2018). ROH can be categorized by length, with long ROH indicating recent inbreeding and shorter ROH reflecting historical demographic events. Within ROH, certain genomic regions may show recurrent homozygosity across multiple individuals, forming ROH-enriched islands (Bosse et al., 2014). That may arise from selection, genetic drift or evolutionary processes, aiding studies on local adaptation and deleterious mutation accumulation. Heterozygosity Rich Regions (HRR) are genomic regions with consistently high heterozygosity across a population, indicating loci under balancing selection, recombination hotspots or introgressed regions from hybridization events. Studies on ROH and HRR islands provide insights into demographic history, inbreeding, and selection (Szpiech et al., 2019). In the present study, 11 sheep breeds originating from northern (Babolna Tetra), central (Rouge du Rusillon, Hortobagy Racka), southern Europe (including Greek native populations) (Turcana, Rusty Tsigai, Pelagonia, Frizarta, Boutsiko, Mytlini, Ojalada) and South America (Creole) were analyzed to explore ROH and HRR patterns from different geographic clusters. Medium-density (Ovine SNP50) genotypes from SMARTER database (Cozzi et al., 2024) were analyzed using the detectRUNS R package with a “sliding-window” approach for ROH and “consecutive runs” for HRR. Total ROH/HRR count and length were estimated for each breed and SNPs with >1% occurrence in runs were selected to define ROH and HRR islands. Creole breed from Latin America showed the highest ROH number and length compared to the European breeds, likely due to intense historical selective pressures, while also sharing genomic components with European breeds as a descendant of those introduced by Spanish and Portuguese explorers during the colonial period (Revelo et al., 2022). On the other hand, Romanian Turcana and Rusty Tsigai sheep had the lowest ROH number and length, reflecting high genetic diversity.

poster #05

continued

Similar HRR patterns were observed across the sheep breeds. Common ROH islands were detected on chromosomes 2, 3, 8, 13 and 22, while common HRR islands were found on chromosomes 2, 3, 7 and 9, indicating potential functional importance of chromosomes 2 and 3. Greek Mytilini sheep demonstrated the highest number of ROH islands possibly linked to its geographical isolation. Hungarian Hortobagy Racka and Romanian Rusty Tsigai sheep had one ROH island, suggesting higher gene flow and crossbreeding events, while they share HRR hotspots with the Greek Pelagonia and Frizarta sheep. The Spanish breed Ojalada shared the highest number of ROH islands with other breeds, in consistence with domestic sheep migration from eastern Europe to Spain during the Roman period (Ciani et al., 2020). Ongoing identification and functional annotation of the genes on ROH and HRR islands aim to provide further insights into the relationship among the studied sheep breeds.

REFERENCES

Bosse, M., Megens, H. J., Frantz, L. A. F., Madsen, O., Larson, G., Paudel, Y., et al. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nature Communications 2014 5:1, 5(1), 1–8. https://doi.org/10.1038/ncomms5392

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. Nature Reviews Genetics, 19(4), 220–234. https://doi.org/10.1038/nrg.2017.109

Ciani, E., Mastrangelo, S., Da Silva, A., Marroni, F., Ferenčaković, M., Ajmone- Marsan, P., et al. (2020). On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genetics Selection Evolution, 52(1), 1–14. https://doi.org/10.1186/S12711-020- 00545-7/FIGURES/5

Cozzi, P., Manunza, A., Ramirez-Diaz, J., Tsartsianidou, V., Gkagkavouzis, K., Peraza, P., et al. (2024). SMARTER-database: a tool to integrate SNP array datasets for sheep and goat breeds. Gigabyte, 2024, 0–0. https://doi.org/10.46471/GIGABYTE.139

Revelo, H. A., Landi, V., López-Alvarez, D., Palacios, Y. A., Paiva, S. R., McManus, C., et al. (2022). New Insight into the Genome-Wide Diversity and Admixture of Six Colombian Sheep Populations. Genes, 13(8), 1415. https://doi.org/10.3390/GENES13081415/S1

Szpiech, Z. A., Mak, A. C. Y., White, M. J., Hu, D., Eng, C., Burchard, E. G., & Hernandez, R. D. (2019). Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity. The American Journal of Human Genetics, 105(4), 747–762. https://doi.org/10.1016/J.AJHG.2019.08.011

poster #06

ROH and HRR patterns across the genome of European goat and sheep breeds

Alexandros Pavlos Pavlidis1, Arianna Manunza 2, Johanna Ramirez-Diaz2, Paolo Cozzi2, Alessandra Stella2, Spiros Papakostas 3, Alexandros Triantafyllidis 1,4, Valentina Tsartsianidou1,4

1Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

2Institute of Agricultural Biology and Biotechnology, National Research Council 20133 Milano, Italy

3International Hellenic University, 57001 Thermi, Thessaloniki, Greece

4Center for Interdisciplinary Research and Innovation, Balkan Center 57001, Thessaloniki, Greece

email: apavlc@bio.auth.gr

Runs of Homozygosity (ROH) constitute continuous homologous genomic segments widely studied in farmed animals. Increased levels of homozygosity are usually formed by founder effects and geographic isolation, management practices and demographic history. Frequent ROH may indicate reduced reproduction success and fitness. Heterozygosity Rich Regions (HRR), which represent consecutive single nucleotide differences between paternal and maternal chromosomes, are also studied. HRR provide insights into the evolutionary history and genetic diversity, while facilitating the identification of genomic regions related to survival rate, immune response and fertility. The present study focuses on the exploration of these genomic patterns in 5 goat breeds (Fosses, Provencale, Eghoria, Skopelou, Swedish Landrace) and 6 sheep breeds (Bizet, Castellana, Churra, Chios, Solognote, Tsigai) from North, Central and South Europe, including Greek native populations.

poster #06

continued

LD-pruned 50K species-specific genotypes were obtained from SMARTER global database. ROH and HRR were detected with 'detectRUNS' R package, implementing "sliding window" and "consecutive runs" approaches. Hotspots were defined by selecting the SNPs with the highest frequency of occurrence within runs (top 1%). Most goat breeds exhibited predominantly short ROH, indicating ancient inbreeding events. The Swedish Landrace was characterised by the highest ROH number and length in contrast to the French Provencale and Fosses, which displayed this pattern in HRR. Greek breeds showed different patterns due to the geographic isolation of Skopelos compared to the mainland Eghoria. A common HRR island was found across all goat populations in chromosome 1, associated with embryonic development, whereas other shared islands between breeds have been related to immune response (chromosomes 3, 8, 12), survival in arid environments (chromosome 12) and meat and wool-related traits. The Swedish Landrace shared an island (ROH and HRR) with nearly every other breed, possibly due to historical connections to Viking trade. In sheep, Solognote displayed the highest ROH count and Churra the highest HRR count and length. All sheep breeds studied here share an island with the Spanish Castellana sheep, which has been attributed to migration from the Balkans to Spain along the Mediterranean route during the Roman period and the post-domestication dispersal of Merino sheep in Europe. Gene and functional annotation of the identified genomic regions are currently in progress to provide further information into the relationships between breeds and species.

REFERENCES

Cardoso, T. F., Amills, M., Bertolini, F., Rothschild, M., Marras, G., Geert Boink, Jordi Jordana, Capote, J., Carolan, S., Hallsson, J. H., Juha Kantanen, Pons, A., & Lenstra, J. A. (2018). Patterns of homozygosity in insular and continental goat breeds. Genetics Selection Evolution, 50(1). https://doi.org/10.1186/s12711-018-0425-7

Giorgio Chessari, Criscione, A., Marletta, D., Crepaldi, P., Baldassare Portolano, Manunza, A., Cesarani, A., Filippo Biscarini, & Mastrangelo, S. (2024). Characterization of heterozygosity-rich regions in Italian and worldwide goat breeds. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-023-49125-x

Manunza, A., Ramirez-Diaz, J., Cozzi, P., Lazzari, B., Gwenola Tosser-Klopp, Servin, B., Johansson, A. M., Grøva, L., Berg, P., Dag Inge Våge, & Stella, A. (2023). Genetic diversity and historical demography of underutilised goat breeds in North-Western Europe. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-48005-8

Ciani, E., Mastrangelo, S., Silva, A. D., Marroni, F., Maja Ferenčaković, Paolo Ajmone-Marsan, Baird, H., Barbato, M., Colli, L., Chiara Delvento, Dovenski, T., Gregor Gorjanc, Stephen, Hoda, A., Li, M.-H., Božidarka Marković, McEwan, J., Moradi, M. H., Otsanda Ruiz-Larrañaga, & Dragana Ružić-Muslić. (2020). On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genetics Selection Evolution, 52(1). https://doi.org/10.1186/s12711-020-00545-7

poster #07

Assessment of pharmacological correlation and clinical utility of physiologically based pharmacokinetic modeling of doxorubicin

George A. Mystridis1, George Batzias2, Ioannis S. Vizirianakis1,3

1Laboratory of Pharmacology, School of Pharmacy, Aristotle University of Thessaloniki, Greece

2Laboratory of Pharmacology, School of Veterinary Medicine, Aristotle University of Thessaloniki, Greece

3Department of Health Sciences, School of Life and Health Sciences, University of Nicosia, Cyprus

Keywords: PBPK modeling; drugs; doxorubicin; population pharmacokinetics

Recently, quantitative systems pharmacology (QSP) is in the spotlight of modern pharmacology by providing, through physiology based pharmacokinetic (PBPK) modelling, mechanistic insights into the pharmacokinetics (PK) processes towards achieving better drug efficacy and safety profiles in the clinical setting. PBPK models are based on in vitro-in vivo correlation (IVIVC) procedures. Striving to be as mechanistic as possible in nature, they are based on the underlying anatomical, physiological, and biochemical characteristics of an organism. In this context, precision medicine could be achieved in the clinical setting by connecting PBPK models with pharmacodynamic (PD) prediction models capable of exploiting simulation that includes the molecular biomarker information for individuals and for specific populations. To this end, it is important to note that despite being an “old” drug, doxorubicin (DOX) remains an important and valuable therapeutic agent in cancer therapy. Its clinical use, however, is limited due to safety issues correlated to the cumulative dose used and manifested mainly as serious cardiotoxicity that could potentially lead to congestive heart failure causing up to 50% mortality. To address these therapeutic issues of DOX and its metabolites in the body, the aim of this study has been to develop a PBPK model to predict the plasma concentration profile of DOX as a first step towards the development of a model for more complicated DOX formulations, such as liposomes. To do so, the Simcyp Simulator was used for modeling. The PK data incorporated into the simulator were extracted from published DOX-related clinical studies and the needed physicochemical data was extracted from public databases. One clinical study was used as a training dataset, and another was used as a validation dataset for the model. To construct the model, a “middle-out” logic was used, to optimize the model parameters and enhance the predictive capability. To predict individual parameters, a reverse translation approach was implemented by physiologically relevant mathematical equations and simulator’s predictive modules for every construction step. The model used a simulated population to predict individual patient variability values. Following the construction, the calculated parameters were tested using the training dataset. The last step was the overall performance evaluation of the model using the independent values of the validation dataset. The candidate models exhibited satisfying predictive capacities, with one model showing the best overall performance. The results obtained support the importance of such a methodology to implement DOX delivery within the context of personalized medicine decisions in the clinical setting.

poster #07

continued

REFERENCES

Mystridis, G.A., et al., Physiologically Based Pharmacokinetic Modelling and Simulation to Predict the Plasma Concentration Profile of Doxorubicin. Pharmaceutics, 2022. 14(3):541.

Vizirianakis, I.S., et al. Tackling Pharmacological Response Heterogeneity by PBPK Modeling to Advance Precision Medicine Productivity of Nanotechnology and Genomics Therapeutics. Expert Review in Precision Medicine & Drug Development, 2019. 4:139–151M

Vizirianakis, I.S., et al., Enabling Personalized Cancer Medicine Decisions: The Challenging Pharmacological Approach of PBPK Models for Nanomedicine and Pharmacogenomics (Review). Oncology Reports, 2016. 35(4):1891-904.

poster #08

Genome informatics approaches for infectious disease and cancer therapy

Antonios T Tsanakas1, Harmen JG van de Werken2, Christos A Ouzounis1, Peter D Katsikis2

1BCCB Group, Artificial Intelligence & Information Analysis Laboratory, School of Informatics, Aristotle University of Thessaloniki, Thessalonica, Greece

2Department of Immunology, Erasmus University Medical Center, Rotterdam, Netherlands

We investigate a range of computational methodologies and artificial intelligence (AI) algorithms for the analysis of large-scale biomolecular datasets from viral diseases and cancer. The primary focus is on genome informatics and specifically cancer transcriptomics.

We are currently examining gene fusion events and neoantigen candidates in melanoma, lung mesothelioma, and pancreatic adenocarcinoma using PacBio long-read RNAseq, Illumina short-read RNAseq, and Whole Exome Sequencing (WES). A novel AI-based approach is developed to estimate the immunogenicity of identified neoantigens, supporting next-generation cancer vaccine development.

Additionally, we have developed predictive diagnostics for viral diseases, including a machine learning model for COVID-19 severity prognosis based on immunological biomarkers. This model, designed for explainability, demonstrates high accuracy in risk assessment for hospitalized patients.

Our work,a collaboration between the Compuster Science Department AUTH and Erasmus University Medical Center, contributes to personalized medicine by elucidating molecular disease mechanisms and supporting innovative therapeutic strategies with bioinformatics approaches.

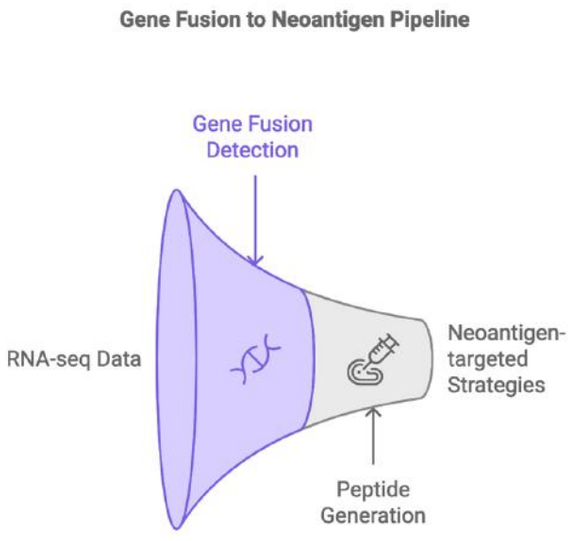


Figure: Fusion-to-Neoantigen Pipeline Using RNA-seq Data

poster #08

continued

REFERENCES

Chen Y, Wang Y, Chen W, Tan Z, Song Y, HGSV Consortium, Chen H & Chong Z (2023) Gene Fusion Detection and Characterization in Long-Read Cancer Transcriptome Sequencing Data with FusionSeeker. Cancer Res. 83(1), 28–33. https://doi.org/10.1158/0008-5472.CAN-22-1628

Qin Q, Popic V, Yu H, White E, Khorgade A, Shin A, Wienand K, Dondi A, Beerenwinkel N, Vazquez F, Al'Khafaji AM & Haas B J (2024) CTAT-LR-fusion: accurate fusion transcript identification from long and short read isoform sequencing at bulk or single cell resolution. bioRxiv 2024.02.24.581862. https://doi.org/10.1101/2024.02.24.581862

Sakamoto Y, Sereewattanawoot S & Suzuki A (2020) A new era of long-read sequencing for cancer genomics. J Hum Genet 65(1), 3–10. https://doi.org/10.1038/s10038-019-0658-5

poster #09

Cyanobacteria identification with deep convolutional neural networks

Sofoklis Keisaris, Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thermi, Thessaloniki, 60361 - 57001, Greece
Fotis Psomopoulos, Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thermi, Thessaloniki, 60361 - 57001, Greece
Spyros Gkelis, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece
email: sofoskeisaris@gmail.com

Cyanobacteria are ubiquitous microorganisms and while some serve beneficial purposes as primary producers, bio-indicators on monitored ecosystems or biotechnological assets, others can form harmful algal blooms and produce a variety of toxins, posing general risks. Therefore, accurate cyanobacteria species identification holds great significance in an abundance of distinct cases. Traditional methods [1] of cyanobacteria recognition and classification are usually labor-intensive and costly [2]. The main focus of this study is designing and implementing an image recognition process that will facilitate identification and categorization of distinct cyanobacterial species, that belong to the genera Microcystis, Dolichospermum (Anabaena), and Arhtrospira (commercially known as “Spirulina”), a set never put before in the test of AI image recognition, within provided microscopy images [3,4].

The methodology employed encompasses several key stages, including image segmentation to isolate cyanobacterial cells, manual annotation for labeling these cells, and the implementation of an object detection model of the YOLO family [5,6], for the automated recognition and classification of cyanobacterial taxa. Results show that the employed deep learning model has the potential to effectively discern the presence of cyanobacteria within an image and accurately classify the specific cyanobacterial species involved.

This success paves the way for future work, which will focus on extending the model's capabilities to identify these taxa within mixed images containing both the taxa of interest along with non relative cells, further advancing its practical utility.

poster #09

continued

REFERENCES

Komárek, J., Kaštovský, J., Mareš, J., & Johansen, J. R. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia, 86(4), 295-335. https://doi.org/10.1002/9781118402238.ch2

Li, C., Wang, K., & Xu, N. (2019). A survey for the applications of content-based microscopic image analysis in microorganism classification domains. Artificial Intelligence Review, 51(4), 577–646. https://doi.org/10.1007/s10462-017-9572-4

Rani, P., Kotwal, S., Manhas, J., Sharma, V., & Sharma, S. (2022). Machine Learning and Deep Learning Based Computational Approaches in Automatic Microorganisms Image Recognition: Methodologies, Challenges, and Developments. Archives of Computational Methods in Engineering, 29(3), 1801–1837. https://doi.org/10.1007/s11831-021-09639-x

Zhang, J., Li, C., Yin, Y., Zhang, J., & Grzegorzek, M. (2023). Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. Artificial Intelligence Review, 56(2), 1013–1070. https://doi.org/10.1007/s10462-022-10192-7

Ultralytics. (2023). Ultralytics YOLOv8 (v8.0) [Computer software]. GitHub. https://github.com/ultralytics/ultralytics

Ultralytics. (2024). Ultralytics YOLOv11 (v11.0) [Computer software]. GitHub. https://github.com/ultralytics/ultralytics

poster #10

Classification of NSCLC subtypes using lung microbiome from resected tissue based on machine learning methods

Pragya Kashyap, Department of Bioscience & Bioengineering, Indian Institute of Technology, Jodhpur, Rajasthan, India

email: neetu25iitj@gmail.com

Classification of adenocarcinoma (AC) and squamous cell carcinoma (SCC) poses significant challenges for cytopathologists, often necessitating clinical tests and biopsies that delay treatment initiation. To address this, we developed a machine learning-based approach utilizing resected lung-tissue microbiome of AC and SCC patients for subtype classification. Differentially enriched taxa were identified using LEfSe, revealing ten potential microbial markers. Linear discriminant analysis (LDA) was subsequently applied to enhance inter-class separability. Next, benchmarking was performed across six different supervised-classification algorithms viz. logistic-regression, naïve-bayes, random-forest, extreme-gradient-boost (XGBoost), k-nearest neighbor, and deep neural network. Noteworthy, XGBoost, with an accuracy of 76.25%, and AUROC (area-under-receiver-operating-characteristic) of 0.81 with 69% specificity and 76% sensitivity, outperform the other five classification algorithms using LDA-transformed features. Validation on an independent dataset confirmed its robustness with an AUROC of 0.71, with minimal false positives and negatives. This study is the first to classify AC and SCC subtypes using lung-tissue microbiome.

REFERENCES

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E. & Adjei, A. A. Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. Mayo Clin. Proc. 83, 584–594 (2008).

Zheng, M. Classification and Pathology of Lung Cancer. Surg. Oncol. Clin. N. Am. 25, 447–468 (2016).

Seyfried, T. N. & Huysentruyt, L. C. On the origin of cancer metastasis. Crit. Rev. Oncog. 18, 43–73 (2013)

Uramoto, H. & Tanaka, F. Recurrence after surgery in patients with NSCLC. Transl. Lung Cancer Res. 3, 242–249 (2014).

Stock-Martineau, S., Magner, K., Jao, K. & Wheatley-Price, P. Challenges of Immunotherapy in Stage IV Non–Small-Cell Lung Cancer. JCO Oncol. Pr. 17, 465–471 (2021).

poster #11

Predicting protein interactions between p53 isoforms, MDM2 and p21 using Alphafold Server

Nikos Kouvarakis1,2, Konstantina Psatha1,2,4, Michalis Aivaliotis1,2,3

1Functional Proteomics and Systems Biology (FunPATh), Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Thessaloniki, Greece

2Laboratory of Biological Chemistry, School of Health Sciences, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

3Basic and Translational Research Unit, Special Unit for Biomedical Research and Education, School of Medicine, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece

4Laboratory of Medical Biology - Genetics, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece

Email: aivaliotis@auth.gr

Tumor suppressor protein p53 constitutes an evolutionary orchestrator of vital pathways for cancer protection. It responds to various stress conditions, leading to many response options, such as cell cycle arrest, activation of DNA repair mechanisms and apoptotic cell death. One of those responses is the upregulation of cyclin-kinase inhibitor p21. Highly influential to the tumor suppressor's p53 response is its isoform stoichiometry. p53 isoforms are produced by alternative splicing, alternative promoter usage, alternative translation initiation, and post-translational modification. This study is an attempt to predict the putative interacting and structural differences of wt-p53 and its isoforms with its physiological negative regulator E3 ubiquitin-protein ligase MDM2 and cyclin-kinase inhibitor p21 using Alphafold Server, aiming to acquire a better view of their dynamic biological mechanisms. The sequences were uploaded from the UniProt database. Alphafold analysis produced a predicted quaternary structure for every match between p53 isoforms and proteins MDM2 and p21. Structural differences were recorded, indicating differential biological function. A parallel analysis will be conducted in RStudio to identify protein overlaps, enabling cross-comparison of results for more robust conclusions. These predictions warrant experimental validation, therefore, future validation studies will include IP-MS to isolate complexes, complemented by hydrogen/deuterium exchange (HDX-MS) for interface analysis and chemical cross-linking (XL)-MS for distance constraints, thereby bridging computational and empirical findings.

poster #11

continued

REFERENCES

Leitner, A., Faini, M., Stengel, F., Aebersold, R., 2016. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. Trends in Biochemical Sciences 41, 20–32. https://doi.org/10.1016/j.tibs.2015.10.008

Psatha, K., Kollipara, L., Drakos, E., Deligianni, E., Brintakis, K., Patsouris, E., Sickmann, A., Rassidakis, G.Z., Aivaliotis, M., 2023. Interruption of p53-MDM2 Interaction by Nutlin-3a in Human Lymphoma Cell Models Initiates a Cell-Dependent Global Effect on Transcriptome and Proteome Level. Cancers 15, 3903. https://doi.org/10.3390/cancers15153903

Steffens Reinhardt, L., Groen, K., Newton, C., Avery-Kiejda, K.A., 2023. The role of truncated p53 isoforms in the DNA damage response. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1878, 188882. https://doi.org/10.1016/j.bbcan.2023.188882

Zhang, W., Zhong, T., Chen, Y., 2017. LC-MS/MS-based targeted proteomics quantitatively detects the interaction between p53 and MDM2 in breast cancer. Journal of Proteomics 152, 172–180. https://doi.org/10.1016/j.jprot.2016.11.002

poster #12

Examining selective pressure and motif occurrence in de novo genes of the *Saccharomycotina* subphylum

Eirini M. Tzermpou 1,2, Emilios Tassios1 & Nikolaos Vakirlis1

1: Hellenic Pasteur Institute, Athens, Greece
2: Deree-The American College of Greece, Athens, Greece

eirinitzermpou@gmail.com

How new genes and protein functions evolve is a fundamental biological question. For a long time, we thought new genes resulted exclusively from tinkering of pre-existing ones, via processes such as gene duplication or exon shuffling. But we now know that novel genes can also evolve from entirely non-coding sequences, through a process known as de novo gene emergence. Yet much about how this process unfolds as well as its overall evolutionary impact remain unclear. Here, we conducted the largest scale computational investigation of de novo emerged novel genes to date, exploiting a rich dataset comprised from 332 budding yeast genomes, spanning the entire biodiversity of the Saccharomycotina subphylum.

The depth of the dataset, along with our rigorous pipeline, allowed us to accurately determine and adjust the evolutionary origin timings of all gene families, identify genes exclusive to the Saccharomycotina subphylum, and define a solid set of 5,662 de novo genes. Our analysis examines how selective pressure acts on these genes compared to a set of conserved genes using nonsynonymous to synonymous substitution rate. Additionally, to investigate how motifs in intergenic regions affect the evolution of novel genes de novo, we constructed a motif detection and analysis pipeline to examine the presence of different types of motifs within these genes of different ages for all species in the subphylum.

REFERENCES

Tassios, E., Nikolaou, C. & Vakirlis, N. Intergenic Regions of Saccharomycotina Yeasts are Enriched in Potential to Encode Transmembrane Domains. Molecular Biology and Evolution 40, msad059 (2023).

Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, Boudouris JT, Schneider RM, Langdon QK, Ohkuma M, Endoh R, Takashima M, Manabe RI, Čadež N, Libkind D, Rosa CA, DeVirgilio J, Hulfachor AB, Groenewald M, Kurtzman CP, Hittinger CT, Rokas A. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell. 2018 Nov 29;175(6):1533-1545.e20.

poster #13

Pangenomic analysis of thermophiles

Iouliana Parisi, Alexios Loukas & Ilias Kappas*

Department of Genetics, Development & Molecular Biology, School of Biology, Aristotle University of Thessaloniki, 541 24, Greece

iouliana@bio.auth.gr, aloukas@bio.auth.gr, * ikappas@bio.auth.gr

Abstract

In recent years, there has been growing interest in pangenomic studies, driven by the advancement of high-throughput next-generation sequencing and computational technologies. The analysis of pangenomes offer valuable insights into the genomic diversity and evolutionary dynamics within taxa. In this project, we analyze the pangenome of the largest, up to date, collection of thermophiles from Archaea and Bacteria. Thermophiles thrive at temperatures between 50 and 122°C and are typically found in thermal vents and hot springs, where other life forms cannot survive. This pangenomic study brings focus to this group of microorganisms by selecting 200 complete genomic sequences of thermophilic strains from the NCBI Genomes Database. A highly diverse and open pangenome with a remarkably small core genome was uncovered, potentially reflecting the evolutionary history and selective pressures required for life in extreme temperatures.

Keywords: thermophile, pangenome, extremophile, comparative genomics, micropan.

Introduction

Pangenomics has emerged as a novel discipline at the intersection of biology, computer science, and applied mathematics (Tettelin & Medini, 2020) and a powerful approach for uncovering genomic diversity across organismal lineages. The pangenome is defined as the totality of genes present in all genomes across a group of organisms. This collection can be divided into genetic elements universally shared by all members, the core genes, those present in only some, the accessory genes (Brockhurst et al., 2019), and those found in a few or a single genome, the unique genes. While many pangenome studies focus on taxonomic clades, fewer investigate specific ecological groupings such as the extremophiles.

poster #13

continued

Extreme environments account for much of Earth's habitable space and host organisms from all three domains of life. However, prokaryotes are most abundant in these niches (thermal vents and hot springs) due to their exceptional diversity and adaptability. Thermophiles are differentiated into hyperthermophiles which can grow optimally up to 105°C and moderate thermophiles, typically growing between 50°C and 70°C (Soppa, 2013). They can be found within most major prokaryotic lineages covering immense taxonomic, functional, biochemical, physiological, and ecological diversity (Zhou et al., 2022). Currently, most of this variation remains poorly understood, but high-performance DNA sequencing and bioinformatics applications enable the recovery of high-quality genomes.

Among the tools developed for such pangenomic analysis, micropan, an R package, offers a user-friendly platform for pangenome construction, visualization, and exploration across large prokaryotic datasets.

Data & Methods

To conduct the pangenomic analysis, first, a genome table was compiled, consisting of 200 genomes of thermophiles. Genomic sequences were retrieved from the NCBI Genomes Database with the criteria for their selection being: all organisms were prokaryotes (Bacteria or Archaea), had an optimum growth temperature of ≥ 55°C and their level of assembly was strictly complete. The dataset was not limited to species level but included genomes of strains and isolates from multiple thermophilic species found in Thermobase (http://togodb.org/db/thermobase), to capture as much of the thermophiles genetic diversity as possible.

The computational part of the analysis was performed using the R package micropan. Protein sequences were extracted from all the genomes using the program prodigal with a cut-off default score of 40 and subsequently clustered into orthologous groups. Protein clustering was performed using a distance-based threshold of 0.75 (default parameters). The Heaps' law was used to estimate pangenome openness and the rarefaction curve diagram was constructed using the rarefaction.r application. Cluster data were grouped based on occurrence patterns using a binomial mixture model estimate and pie charts were used to visualize the distribution of core, accessory, unique, and other cluster types across the dataset. Additionally, Manhattan distances between the genomes were computed and a weighted Manhattan distance dendrogram was produced to visualize their relations. Accuracy was ensured by appointing weights on shell clusters found amongst genomes. The web tool iTOL (Letunic & Bork, 2021) was used to annotate the weighted dendrogram for the taxonomic status of the organisms with the NCBI Taxonomy Database as reference.

poster #13

continued

Results

A total of 200 thermophilic genomes were analyzed, uncovering 40.200 gene clusters. The estimate for the total number of clusters existing in this group was 57.462 gene clusters, a rather accurate assessment using such a large dataset. The application of Heaps' law on the data gave an alpha value of 0.474 indicating an open pangenome. This is also demonstrated from the rarefaction curve in Fig. 1, which illustrates the number of new clusters observed in every genome. With the use of a binomial mixture model estimate, cluster data best fit in 8 categories based on their occurrence frequencies in the genomes. A visualization of the thermophilic pangenome can be seen in Fig. 2, where a significantly small core is observed. The majority - nearly 75% - of the pangenome consists of unique clusters. Based on genome distances computed from micropan, a dendrogram was created. To enhance its accuracy, weights were appointed to shell (core) genes. As seen in Fig. 3, the dendrogram is taxonomically reliable as a clear differentiation between Bacteria and Archaea is observed.

Discussion

The observable widely open pangenome potentially reflects high genomic diversity among thermophiles, likely reflecting adaptation to varied and extreme environments. A significantly small core suggests that only a few genes have been conserved during prokaryotic diversification. Housekeeping and thermo-adaptive functions are expected within the core and extended core, though these may vary across genomes. On the other hand, the large number of unique clusters depicts the vast genomic diversity of thermophiles, which is expected given the taxonomic breadth of the dataset spanning both Bacteria and Archaea. Conversely, prokaryotes demonstrate tremendous variation in gene content even within individual clones (Tettelin & Medini, 2020). The study of microbial extremophiles at the upper thermal boundary of life is of enormous interest considering the evolutionary implications of gene transfer and genetic exchange, their biotechnological applications, and their implication in theories on the origin of life on Earth or elsewhere (de la Haba et al., 2022).

poster #13

continued

REFERENCES

Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. The ecology and evolution of pangenomes. Curr Biol. 2019;29(20):R1094–R1103.

de la Haba RR, Antunes A, Hedlund BP. Extremophiles: Microbial genomics and taxogenomics. Front Microbiol. 2022;13:984632.

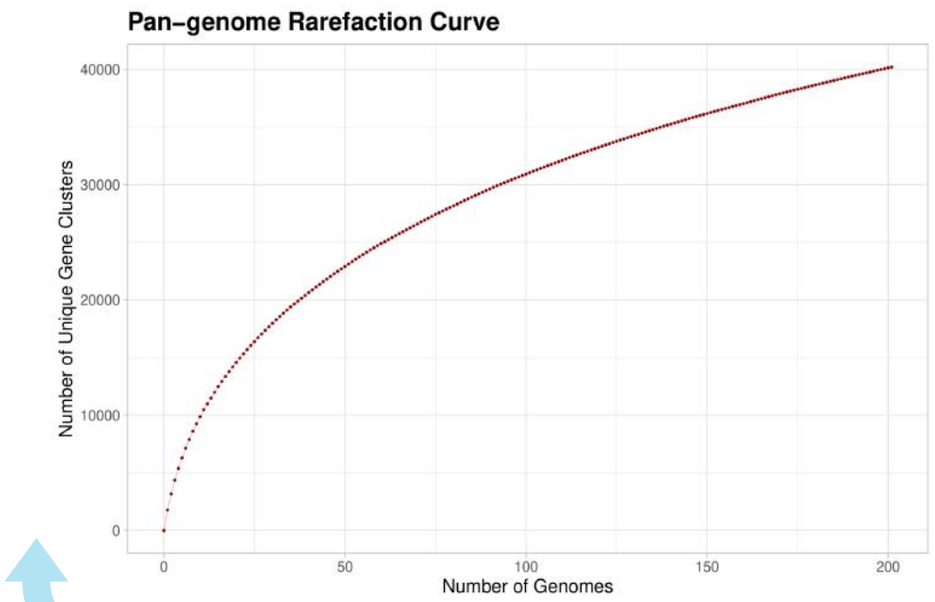
Letunic I, Bork P. Interactive Tree Of Life (iTOL) v6: recent updates and new developments. Nucleic Acids Res. 2021;49(W1):W293–W296.

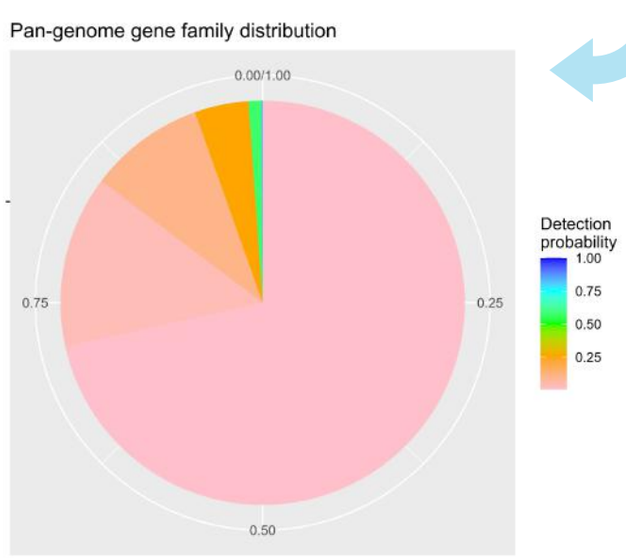
Soppa J. Protein adaptations in archaeal extremophiles. Res Microbiol. 2013;164(6):518–527.

Tettelin H, Medini D, editors. The Pangenome: Diversity, Dynamics and Evolution of Genomes. Cham: Springer; 2020.

Zhou Y, Zhang Y, Wang L, et al. Diversity of thermophilic prokaryotes. Front Microbiol. 2022;13:984632.

Figure 2. Distribution of gene clusters (gene families) in the pangenome based on their detection frequencies in all the genomes.





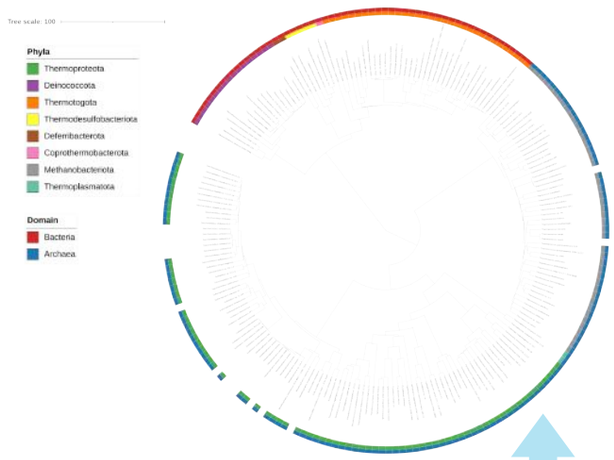


Figure 3. Weighted circular dendrogram of the thermophilic pangenome, annotated using the web tool iTOL for domain and phylum level.

Figure 1. Rarefaction curve of the thermophilic pangenome, demonstrating its openness, as the curve continues to rise with the inclusion of additional genomes.

poster #14

Advancing Synthetic Biology through Integrated Modeling and Software: A Retrospective on iGEM Thessaloniki's Contributions

Ioanna Karampela, Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Kyriakos Marios Papamanolis, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Michalis Aivaliotis, Laboratory of Biological Chemistry, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

email: karampela@cheng.auth.gr

Over the past seven years, iGEM Thessaloniki has made significant strides in synthetic biology by developing innovative modeling techniques and software tools. These efforts have included designing genetic circuits, creating predictive models of biological systems, and building software tools. By combining computational methods with experimental biology, the team has improved the design and analysis of synthetic constructs, contributing to a deeper understanding of complex biological processes. More specifically, in 2017 the team developed a modeling framework to simulate DNA strand displacement reactions, utilizing tools like Visual DSD and KinDA to predict system behavior. The 2019 project introduced POSEIDON, a DNA-based computational system designed to quantify DNA-protein interactions through toehold-mediated strand displacement circuits, employing modeling tools like KinDA and Visual DSD for simulation. In 2021, a Python-based tool was developed, integrating NUPACK and ViennaRNA, to design and analyze toehold switches for miRNA detection, facilitating early diagnosis of pancreatic cancer. The 2022 "Theriac" project featured a Python script for designing Y-shaped DNA hairpins to detect glioblastoma-specific microRNAs. In 2023, the team designed a modeling simulation to analyze the dynamics of a dual kill-switch system and quorum sensing mechanism within an engineered microbial consortium, ensuring biosafety and controlled activity in post-wildfire soil restoration applications. These developments highlight iGEM Thessaloniki's dedication to integrating computational and experimental approaches, advancing synthetic biology research and applications.

Key words: modeling, software, iGEM, bioinformatics, biology, Systems Biology, machine learning, computational biology, synthetic biology

poster #14

continued

REFERENCES
iGEM Thessaloniki 2017 Wiki, Project's Model Page,
https://2017.igem.org/Team:Greece/Model

iGEM Thessaloniki 2019 Wiki, Project's Model Page,
https://2019.igem.org/Team:Thessaloniki/Model

iGEM Thessaloniki 2021 Wiki, Software Page,
https://2021.igem.org/Team:Thessaloniki/Software

iGEM Thessaloniki 2022 Wiki, Project's Model Page,
https://2022.igem.wiki/thessaloniki/model

iGEM Thessaloniki 2023 Wiki, Project's Model Page,
https://2023.igem.wiki/thessaloniki/model

poster #15

metagRoot: A comprehensive database of protein families associated with plant root microbiomes

Maria N. Chasapi1, Iro N. Chasapi1, Fotis A. Baltoumas1, Eleni Aplakidou1,2, Evangelos Karatzas1,3, Ioannis Iliopoulos2, Dimitrios J. Stravopodis4, Ioannis Z. Emiris5,6, Aydin Buluç7,8, Ilias Georgakopoulos-Soares9, Nikos C. Kyrpides10, Georgios A. Pavlopoulos1,*

1 Institute for Fundamental Biomedical Research, BSRC "Alexander Fleming", Vari 16672, Greece.
2 Division of Basic Sciences, University of Crete Medical School, Heraklion 71110, Greece.
3 European Bioinformatics Institute, Hinxton, UK
4 Department of Biology, School of Sciences, National and Kapodistrian University of Athens, 15701 Athens, Greece.

5 Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 16122, Athens, Greece.
6 ATHENA Research Center, Aigialias & Chalepa, 15125, Marousi, Greece.
7 Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

8 Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA.
9 Institute for Personalized Medicine, Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, PA, USA.

10 DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. * Corresponding authors: pavlopoulos@fleming.gr

The plant root microbiome plays a vital role in plant health, nutrient uptake, and environmental resilience (1). To explore and harness this diversity, we present metagRoot, a specialized and enriched database focused on the protein families of the plant root microbiome. Designed to bridge genomic and ecological data, metagRoot integrates metagenomic, metatranscriptomic, and reference genome sequences from root-associated microbiomes, offering a robust platform for functional and structural exploration. All data were sourced from the Integrated Microbial Genomes & Microbiomes (IMG/M) database (2). After stringent quality control steps, including removal of truncated sequences, filtering of low-complexity regions, and exclusion of sequences under 35 amino acids (3), the dataset included 1,199 metagenomes, 327 metatranscriptomes, and 2,979 reference genomes. Protein families were generated using MMseqs2 Linclust (30% identity, 80% coverage), and only clusters with at least 100 members were retained. To improve alignment quality and reduce redundancy, hhfiltering was applied using a 95% identity threshold and 70% coverage, ensuring that each family retained its original diversity while minimizing sequence bias.

poster #15

continued

The resulting catalog comprises 71,091 protein families, richly annotated with Pfam domains, CRISPR arrays, Cas elements, multiple sequence alignments, taxonomic classifications, habitat metadata, and predicted 3D structures via AlphaFold2. Structures were validated through Foldseek comparisons to experimentally resolved entries in CATH, PDB, and AlphaFoldDB, and clustered into 9,463 superfamilies. This led to the identification of 28 novel structures, highlighting the potential of metagRoot to uncover new aspects of microbial functional diversity.

Environmental metadata were mapped using the GOLD classification, enabling protein family categorization across diverse plant-associated niches such as the rhizosphere, endosphere, rhizoplane, nodules, bulbs, and aerial roots. Taxonomic annotation leveraged a multi-tool approach (Kraken2, MMseqs taxonomy, Whokaryote, EukRep, GeNomad), yielding a dataset primarily bacterial (94.9%) with representation from archaea, eukaryotes, viruses, and plasmids.

metagRoot is accessible via a user-friendly web interface, built using an ASP.NET MVC architecture with a MySQL backend and an interactive front-end. The platform is designed to accommodate both expert and non-expert users and offers flexible search and filter options based on environmental origin, taxonomic classification, presence of CRISPR-Cas systems, and combinations of these parameters, greatly enhancing data exploration and discovery potential. Interactive visualization tools including MSAViewer, Mol*, SkyLign, and OpenStreetMap further enrich the user experience by providing integrated sequence alignments, 3D molecular structures, HMM logos, and geospatial context.

Use-case studies in the database showcase how combining sequence, structure, and ecological data can reveal novel Cas protein families, niche-specific symbiotic elements, and functional annotations for previously uncharacterized proteins. Collectively, metagRoot bridges gaps in microbial protein knowledge by focusing on the often-overlooked plant root microbiome, providing a valuable platform for research in plant-microbe interactions, structural biology, and microbiome-informed agriculture.

REFERENCES

Chen, I-Min A et al. “The IMG/M data management and analysis system v.7: content updates and new features.” Nucleic acids research vol. 51,D1 (2023): D723-D732. doi:10.1093/nar/gkac976

Vandenkoornhuyse, Philippe et al. “The importance of the microbiome of the plant holobiont.” The New phytologist vol. 206,4 (2015): 1196-206. doi:10.1111/nph.13312

Pavlopoulos, Georgios A et al. “Unraveling the functional dark matter through global metagenomics.” Nature vol. 622,7983 (2023): 594-602. doi:10.1038/s41586-023-06583-7

poster #16

Integrative Analysis of Multi-Omics Data to Identify Deregulated Molecular Pathways and Druggable Targets in Chronic Lymphocytic Leukemia

Dimitra Mavridou1,2, Konstantina Psatha1,2,3, Michalis Aivaliotis1,2,4*

1Laboratory of Biochemistry, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

2Functional Proteomics and Systems Biology (FunPATh)—Center for Interdisciplinary Research and Innovation (CIRI-AUTH), GR-57001 Thessaloniki, Greece

3Laboratory of Medical Biology—Genetics, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

4Basic and Translational Research Unit, Special Unit for Biomedical Research and Education, School of Medicine, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

*Correspondence: aivaliotis@auth.gr

Chronic Lymphocytic Leukemia (CLL), a heterogeneous B-cell malignancy with diverse genetic profiles and clinical outcomes, remains a significant clinical challenge, necessitating deeper molecular insights and potential therapeutic proposals. Multi-omics approaches, combining MS-based proteomics and bioinformatics analyses, hold promise to elucidate disease-related mechanisms and identify therapeutic targets. This study aimed to integrate proteomic, transcriptomic, and patient survival data to detect deregulated pathways in CLL and propose drug repurposing candidates for potential therapeutic use.

Publicly available proteomic (35 CLL patients, 12 controls) and transcriptomic (53 CLL patients, 16 controls) datasets were analyzed. Differentially expressed proteins (DEPs) were identified and correlated with transcriptomic data. Protein-protein interaction networks and functional enrichment analyses were performed using STRING and Cytoscape. Survival analysis (Kaplan-Meier) assessed prognostic protein markers and drug repurposing analysis (PANDRUGS) prioritized FDA-approved compounds targeting deregulated pathways.

poster #16

continued

Proteomic profiling unraveled 1,023 DEPs (608 upregulated, 415 downregulated) discriminating CLL from healthy B-cell controls. Top upregulated proteins (e.g., YEATS2, PIGR, SNRPA) correlated with poor survival, suggesting prognostic value. Multi-omics integration pinpointed limited correlation between proteomic and transcriptomic data ($r^2 = 0.306$), indicating post-transcriptional regulation. Pathway enrichment analysis demonstrated that upregulated proteins were involved in RNA processing, stress response, and metabolism, while downregulated proteins were implicated in immune evasion and vesicle transport. Drug repurposing analysis prioritized 168 FDA-approved drugs, including bosutinib (targeting CDC37, SMAD2), vorinostat (HDAC2/7), and bortezomib (proteasome subunits), suggesting a multi-target mechanism of action with potential for CLL therapy.

Our study demonstrated the power of integrative multi-omics bioinformatics in uncovering deregulated CLL-related proteins and pathways, supporting drug repurposing as a promising strategy for therapeutic development. YEATS2 and SNRPA were identified as novel prognostic markers, while targeted therapies such as bosutinib and bortezomib show potential for clinical translation. This pipeline not only facilitates biomarker discovery but also accelerates the identification of repurposed drugs with therapeutic relevance in CLL.

REFERENCES

1. Mavridou D, Psatha K, Aivaliotis M. Proteomics and Drug Repurposing in CLL towards Precision Medicine. Cancers (Basel). 2021;13(14):3391. Published 2021 Jul 6. doi:10.3390/cancers13143391

2. Mavridou D, Psatha K, Aivaliotis M. Integrative Analysis of Multi-Omics Data to Identify Deregulated Molecular Pathways and Druggable Targets in Chronic Lymphocytic Leukemia. J Pers Med. 2024;14(8):831. Published 2024 Aug 6. doi:10.3390/jpm14080831

poster #17

Unraveling the Molecular Landscape of Mantle Cell Lymphoma Through Integrative Omics and Drug Repurposing

Georgia Orfanoudaki1,2, Konstantina Psatha1,2,3, Michalis Aivaliotis1,2,4,5*

1Functional Proteomics and Systems Biology (FunPATh), Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Balkan Center, GR-54124 Thessaloniki, Greece

2Institute of Molecular Biology and Biotechnology Foundation for Research and Technology—Hellas, GR-70013 Heraklion, Greece

3Laboratory of Medical Biology—Genetics, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

4Laboratory of Biological Chemistry, School of Medicine, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece 4

5Basic and Translational Research Unit, Special Unit for Biomedical Research and Education, School of Medicine, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

Mantle cell lymphoma (MCL) is a rare and aggressive subtype of B-cell non-Hodgkin lymphoma, characterized by significant inter-/intra-tumoral heterogeneity and limited therapeutic options. Despite advances in MCL management, relapses and drug resistance remain major challenges, underscoring the critical need to elucidate its molecular drivers to enhance diagnosis, optimize treatment strategies, and overcome loss of treatment responsiveness. This study aimed to dissect MCL's molecular landscape across the disease stages, identify key proteins and pathways involved in MCL progression, prioritize high-value therapeutic targets, and propose novel drug-repurposing strategies using integrative network-based approaches. Towards this end, publicly available transcriptomic datasets from MCL patients and healthy controls were reanalyzed, employing a multi-step computational pipeline: (1) differential gene expression analysis, (2) pathway enrichment (KEGG, Reactome, GO), (3) co-expression network construction and spectral clustering, and (4) drug repurposing (DrugBank). Network proximity metrics were applied to predict synergistic drug combinations.

poster #17

continued

Our findings revealed the "butterfly effect", starting from a small set of differentially expressed genes, leading to escalating gene regulation alterations in progressive MCL stages and numerous deregulated processes. Aggressive stages recorded >2,600 differentially expressed genes (DEGs), including VEGFA, SPARC, and SOX11. Early-stage MCL involved Hedgehog and p53 signaling, while advanced stages activated MAPK, VEGF, and B-cell receptor pathways. Co-expression modules revealed stage-specific networks, notably involving ribosomal and cytokine signaling genes. Drug repurposing analysis prioritized 388 FDA-approved drugs, including kinase inhibitors (e.g., fostamatinib targeting SYK/BTK) and antipsychotics (e.g., serotonin receptor antagonists). Three clinically used MCL drugs (ibrutinib, acalabrutinib, zanubrutinib) were validated computationally. Finally, network-based screening predicted effective drug pairs (e.g., cannabinoid receptor modulators with tyrosine kinase inhibitors) for synergistic targeting of MCL pathways.

This integrative approach unraveled stage-specific MCL mechanisms and recommended repurposed drugs and combinations with therapeutic potential. The findings highlight VEGFA/SPARC as key, putative biomarkers and propose experimental validation of predicted therapies to improve clinical outcomes in MCL. This paradigm accelerates the implementation of precision medicine-oriented solutions, aiming to optimize therapeutic efficacy and patient prognosis.

REFERENCES

Orfanoudaki G, Psatha K, Aivaliotis M. Insight into Mantle Cell Lymphoma Pathobiology, Diagnosis, and Treatment Using Network-Based and Drug-Repurposing Approaches. International Journal of Molecular Sciences. 2024; 25(13):7298. https://doi.org/10.3390/ijms25137298

poster #18

@@title

Ljubisa Miskovic

ljubisa.miskovic@epfl.ch

@@text

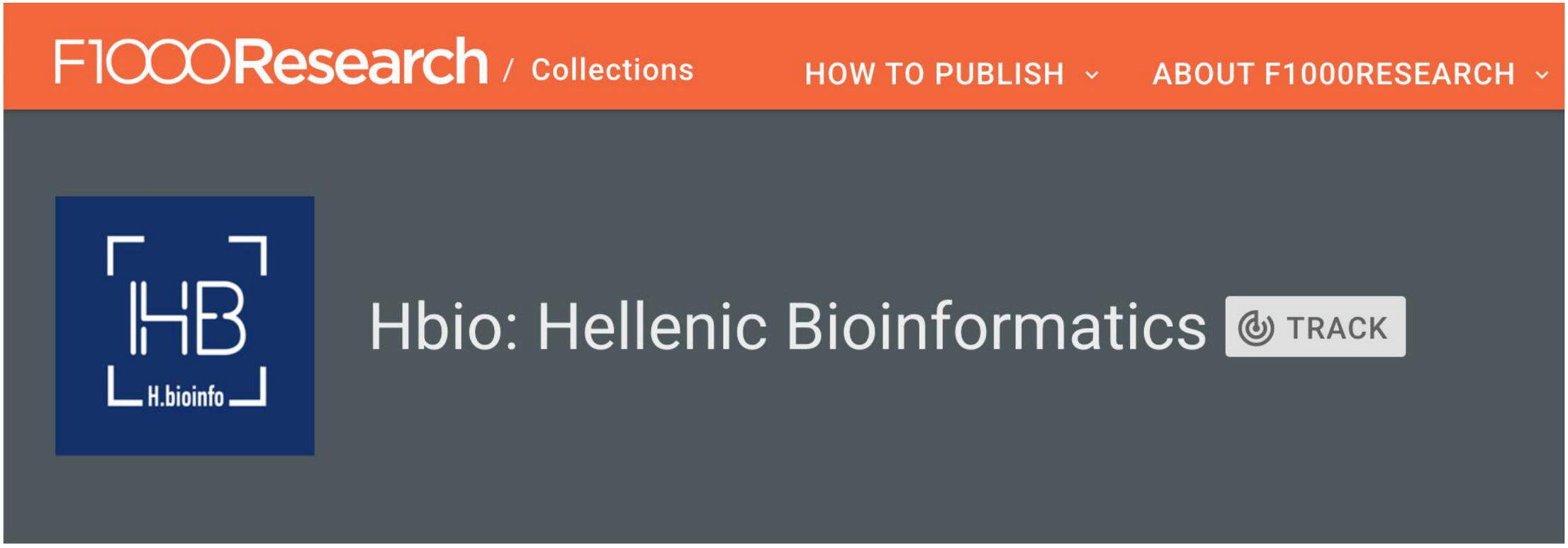
poster #18

continued

@@

Hbioinfo F1000 collection

https://f1000research.com/collections/Hbio/about-this-collection



this booklet will be available online after the conference

participant index

in alphabetical order

excludes speakers, most committee members, on-site and late registrations

Surname:	Name:	Affiliation	
Afxenti	Sotiroula	Bioinformatics Department, The Cyprus Institute of Neurology & Genetics	CY
Akrioti	Naya	AppBio, Department of Biology, AUTH	GR
Aktselis	Ioannis	School of Biology, AUTH	GR
Alektoridis	Georgios	School of Chemistry, AUTH	GR
Amoutzias	Grigoris	Department of Biochemistry & Biotechnology, U Thessaly	GR
Anagnostopoulou	Dafni	Institute for Fundamental Biomedical Research, BSRC Alexander Fleming	GR
Angelopoulou	Athina	School of Pharmacy, AUTH	GR
Antari	Ioanna	Aristotle University of Thessaloniki (AUTH)	GR
Aplakidou	Eleni	University of Crete & BSRC Alexander Fleming	GR
Apostolou	Evangelos	School of Informatics, AUTH	GR
Asvesta	Olga-Dimitra	School of Biology, AUTH	GR
Athanasiadis	Alkiviadis	Department of Botany, School of Biology, AUTH	GR
Billini	Maria	School of Pharmacy, AUTH	GR
Bouas	Christos	School of Informatics, AUTH	GR
Charitidis	Nireas-Iridanos	School of Informatics, AUTH	GR
Chasapi	Iro	BSRC Alexander Fleming	GR
Chasekidou	Eva	School of Biology, AUTH	GR
Chatzikallinikidis	Georgios	School of Biology, AUTH	GR
Chatzinikolaou	Niovi	School of Biology, AUTH	GR
Chondrogiannis	Anestis	School of Biology, AUTH	GR
Christou	Maria	School of Biology, AUTH	GR
Damerla	Annita-Maria	School of Pharmacy, AUTH	GR
Damkali	Marina	AppBio, Department of Biology, AUTH	GR
Daoudaki	Anastasia-Maria	Molecular Biology & Genetics, DUTh	GR
Daoudakis	Petros	Department of Electrical & Computer Engineering, AUTH	GR
Dimou	Maria	School of Biology, AUTH	GR
Eftychiou	Marios	VIB Center for AI & Computational Biology, KU Leuven	BE
Eleftheriadou	Aikaterini	Faculty of Agriculture, Forestry & Natural Environment, AUTH	GR
Eleftheriou	Dimitra-Anastasia	Aristotle University of Thessaloniki (AUTH)	GR
Evmorfidou	Dorothea	School of Biology, AUTH	GR
Foutsitzi	Paraskevi	School of Biology, AUTH	GR

Georgiadou	Maria	School of Pharmacy, AUTH	GR
Georgiou	Grigoris	Bioinformatics Department, The Cyprus Institute of Neurology & Genetics	CY
Ghiotto	Gabriele	Department of Biology, University of Padua	IT
Giannopoulos-Dimitriou	Alexandros	Department of Pharmacology, School of Pharmacy, AUTH	GR
Gkiatas	Vasileios	School of Biology, AUTH	GR
Gkini	Ioanna	School of Biology, AUTH	GR
Gkoutzamanis	Antigonos	AppBio, AUTH & International Hellenic University	GR
Hatzipanagiotou	Maria	Aristotle University of Thessaloniki (AUTH)	GR
Iakovidis	Dionysios	School of Biology, AUTH	GR
Iliopoulos	Ioannis	School of Medicine, University of Crete	GR
Ioannidou	Alexandra	School of Chemistry, AUTH	GR
Kalaitzi	Xeni	Biological Applications & Technologies, FORTH	GR
Kallipoliti	Maria	School of Medicine, University of Crete	GR
Kappas	Ilias	School of Biology, AUTH	GR
Karagianni	Polina	AppBio, Department of Biology, AUTH	GR
Karampela	Ioanna	Department of Chemical Engineering, AUTH	GR
Karaniadi	Androniki	Department of Biochemistry & Biotechnology, U Thessaly	GR
Karathanasis	Nestoras	Bioinformatics Department, The Cyprus Institute of Neurology & Genetics	CY
Kirmizi	Natalia	School of Biology, AUTH	GR
Koniari	Efstathia	School of Biology, AUTH	GR
Kotina	Vasiliki-Maria	Aristotle University of Thessaloniki (AUTH)	GR
Kotoula	Artemis	University of Crete	GR
Kotroni	Eleni	Aristotle University of Thessaloniki (AUTH)	GR
Kotsopoulos	Christos	Laboratory of Biological Chemistry, Department of Medicine, AUTH	GR
Kougioumtzis	Iordanis	School of Biology, AUTH	GR
Kovacheva	Victoria	The University of Salford	UK
Lagkouvardos	Ilias	Department of Clinical Microbiology, School of Medicine, University of Crete	GR
Lathira	Vasiliki	School of Biology, AUTH	GR
Lili	Chrystalla	IPPS Precision Medicine, AUTH	GR
Loukas	Alexios	Hellenic Centre for Marine Research (HCMR)	GR
Lyrou	Fani	Forest Genetics Lab, Faculty of Agriculture, Forestry & Natural Environment, AUTH	GR
Malousi	Andigoni	School of Medicine, AUTH	GR

Matsouka	Katerina	School of Biology, AUTH	GR
Michailidis	Alexandros	School of Informatics, AUTH	GR
Minoudi	Styliani	School of Biology, AUTH	GR
Misailidis	Michail	School of Biology, AUTH	GR
Mpikos	Ioannis	School of Biology, AUTH	GR
Mystridis	Georgios	School of Pharmacy, AUTH	GR
Nicolaidou	Eleni	Bioinformatics Department, The Cyprus Institute of Neurology & Genetics	CY
Olo Ndela	Eric	Lawrence Berkeley National Lab	USA
Oulas	Anastasis	The Cyprus Institute of Neurology & Genetics	CY
Ouzounis	Christos	School of Informatics, AUTH	GR
Panou	Eva	AppBio, Department of Biology, AUTH	GR
Panou	Eva	School of Biology, AUTH	GR
Papadimitriou-Tsantarliotou	Aliki	School of Pharmacy, AUTH	GR
Papadopoulos	Panagiotis	Institute of Molecular Biology & Biochemistry, FORTH	GR
Papadopoulos	Nikolaos-Tryfon	School of Biology, AUTH	GR
Papadopoulos	Ioannis	School of Biology, AUTH	GR
Papadopoulos	Stefanos	School of Biology, AUTH	GR
Papadopoulou	Theodora	Center for Diabetes Research, ULB	BE
Papamichail	Evangelia	Aristotle University of Thessaloniki (AUTH)	GR
Papasakellariou	Konstantinos	Technical University of Denmark (DTU)	DK
Papastergiou	Margarita	School of Biology, AUTH	GR
Papavasiliou	Danae	School of Biology, AUTH	GR
Papazoglou	Panagiota	Department of Pharmacology, School of Pharmacy, AUTH	GR
Parisi	Iouliana	School of Biology, AUTH	GR
Parisis	Efthymios	School of Biology, AUTH	GR
Pavlidis	Alexandros-Pavlos	School of Biology, AUTH	GR
Pavlopoulos	Georgios	BSRC Alexander Fleming	GR
Pazis	Angelos	School of Biology, AUTH	GR
Pechlivanis	Nikos	Centre for Research & Technology Hellas	GR
Petmezas	Athanasios	Faculty of Agriculture, Forestry & Natural Environment, AUTH	GR
Prifti	Antea	School of Biology, AUTH	GR
Promponas	Vasilis	University of Cyprus	CY

Psomopoulos	Fotis	Institute of Applied Biosciences, CERTH	GR
Psychos	Christos	School of Biology, AUTH	GR
Pyrgelis	Nikolaos	Hellenic Pasteur Institute	GR
Rajić	Martina	National Institute of Chemistry, Ljubljana	SI
Ramopoulou	Aikaterini	University of Ioannina	GR
Rebi	Aggeliki	School of Biology, AUTH	GR
Saiti	Katerina	Department of Pharmacology, School of Pharmacy, AUTH	GR
Savva	Kyriaki	Bioinformatics Department, The Cyprus Institute of Neurology & Genetics	CY
Sfakianaki	Eirini	Pharmacy/Medicine, AUTH	GR
Siosiou	Anna	School of Biology, AUTH	GR
Siskou	Andrianna	School of Pharmacy, AUTH	GR
Sivvas	Anastasios	AppBio, Department of Biology, AUTH	GR
Stagaki	Marianna	School of Medicine, University of Crete	GR
Stamatellos	Georgios	AppBio, AUTH & International Hellenic University	GR
Stampouli	Argyro	Faculty of Agriculture, Forestry & Natural Environment, AUTH	GR
Symeonidou	Maria	*Undergraduate Student*	GR
Tassios	Aimilios	Hellenic Pasteur Institute, University of Ioannina	GR
Theodosiou	Eleni	Institute of Applied Biosciences, CERTH	GR
Tourvas	Nikolaos	Laboratory of Forest Genetics & Tree Breeding, AUTH	GR
Tsialta	Kassiani	School of Biology, AUTH	GR
Tzaka-Fragkouli	Kyriakoula	School of Biology, AUTH	GR
Tzermpou	Eirini-Maria	Hellenic Pasteur Institute	GR
Tziastas	Christos	School of Biology, AUTH	GR
Vakirlis	Nikolaos	Hellenic Pasteur Institute	GR
Valtopoulou	Anastasia	Department of Electrical & Computer Engineering, AUTH	GR
Varveris	Alexandros	School of Biology, AUTH	GR
Vasileiadis	Spyridon	School of Biology, AUTH	GR
Vizirianakis	Ioannis	School of Pharmacy, AUTH	GR
Xenos	Alexandros	Barcelona Supercomputer Center	ES
Xiros	Georgios	School of Biology, AUTH	GR
Yiangou	Athina	Faculty of Agriculture, Forestry & Natural Environment, AUTH	GR
Zirdeli	Iordana	Institute of Molecular Biology & Biochemistry, FORTH	GR

author index

in alphabetical order

minor errors may be present

X

Y

Z

notes

notes

notes

notes

notes

notes

notes

notes

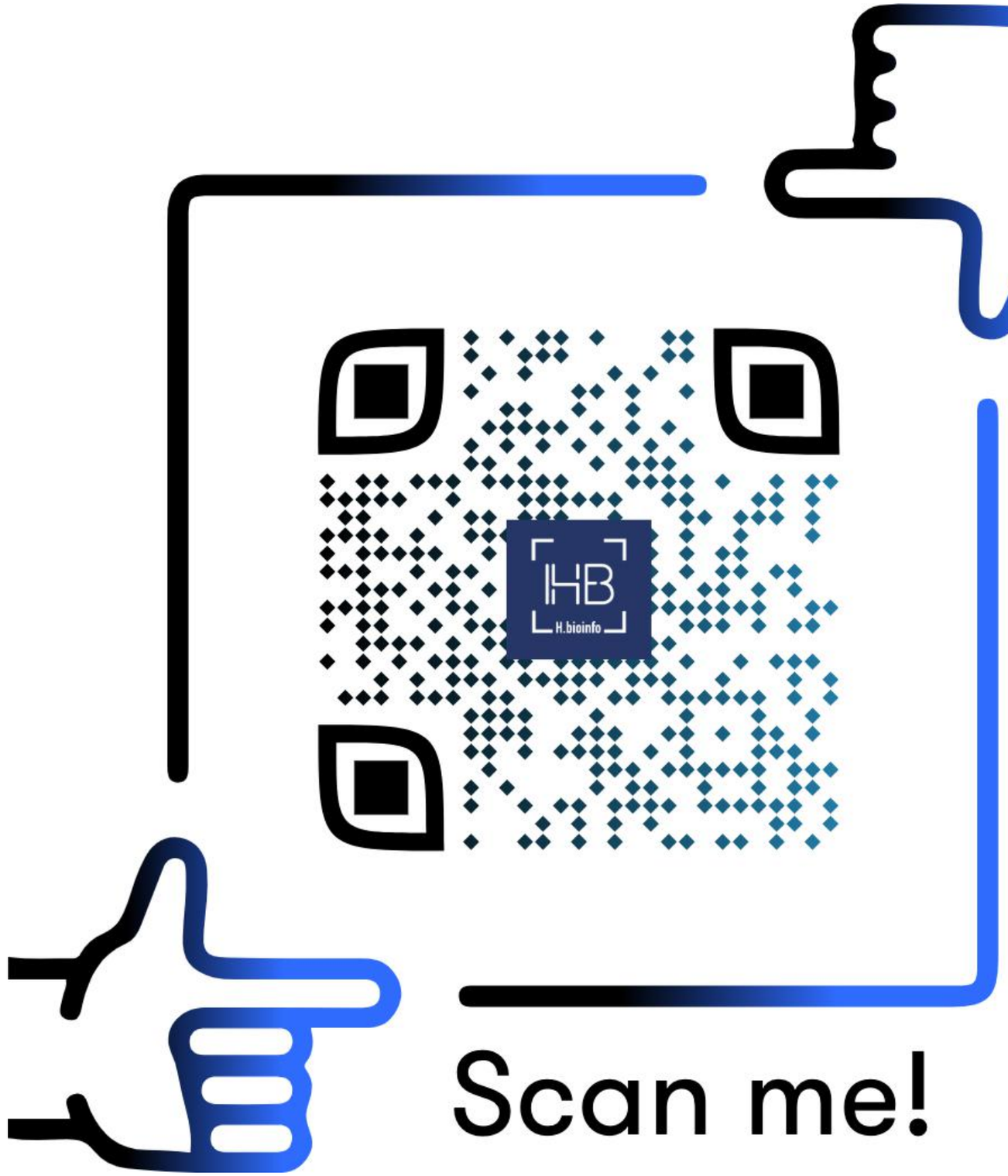
May 2025

Hellenic Bioinformatics

Greece & Cyprus

http://hbio.info

this may not work after May 1 ;)



hbio2025.hbio.info