# DYNALIFE

## WG1-WG2 Interaction Meeting

## Data driven evidence: theoretical models and complex biological data

5-7 June 2024

Teloglion Foundation of Arts

Thessaloniki, Greece

COST Action CA21169

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

ARISTOTLE UNIVERSITY OF THESSALONIKI

# Book of Abstracts

DYNALIFE, Thessaloniki 2024

# Application of the results of Symbolic Dynamics in the study of nullomers

Grácio[1], Clara

[1]*University of Évora and CIMA, Rua Romão Ramalho, 49, Évora, Portugal*

*e-mail: mgracio@uevora.pt*

**Key words:** Symbolic Dynamics, nullomers, topological invariants

The terms minimal absent words, nullomers and primes all describe sequences that do not occur in the entire genome or proteome of an organism. And one of the classical ways of describing a symbolic dynamical system is by using forbidden words.
Using an interval map we study quantitative measures of the complexity of sequences, as metric entropy and conductance. Taking into account this deep connection, it is intended to use the many results of Symbolic Dynamics in the study of nullomers

# Synonymous codon utilization in proteins: preserving function over sequence and investigating cross-species nullomer and antimicrobial peptide (AMP) correlations

Stefano Piotto

*Dept. Pharmacy, University of Salerno*

**Background**: Evolutionary pressures preserve protein function rather than specific amino acid sequences, with synonymous codons playing a crucial role in this adaptive process. Proteins, especially membrane proteins, depend heavily on their three-dimensional structure and environmental interactions (e.g., with membrane lipids [1]) for functionality. The flexibility and dynamics of protein structures are vital, often allowing for synonymous substitutions that maintain function despite sequence variability.

**Objective**: This study aims to redefine genomic and protein sequences by consolidating synonymous codons to a single representative codon for each amino acid. This approach focuses on preserving the structural and functional integrity of proteins while simplifying the genetic code used within the organism's genome.

**Methods**: We propose to transform protein and genomic sequences into simplified versions, where only one 'synonymous' codon per amino acid is retained. This transformation will standardize the genetic code across different datasets and facilitate direct comparisons. Further, we will investigate the presence of nullomers—sequences absent in one organism but potentially significant in another, such as antimicrobial peptides (AMPs) in contrasting species. The analysis will involve comparing these transformed sequences to identify functional overlaps or unique adaptations manifested as nullomers in one species that correspond to AMPs in another. We hypothesize that this method will reveal a fundamental conservation of functional sequences across species that transcends individual genetic variations. By comparing the standardized sequences, we aim to uncover evolutionary strategies that preserve protein functionality, even in the face of extensive sequence divergence. Additionally, we predict that some nullomers in one species may serve critical functions, like AMPs, in another, reflecting unique evolutionary solutions to common biological challenges.

**References:**

Transmembrane peptides as sensors of the membrane physical state, S Piotto, L Di Biasi, L Sessa, S Concilio Front. Phys., 24 May 2018 Sec. Membrane Physiology and Membrane Biophysics Volume 6 - 2018

# Mechanism of the proton transport mediated by ATP/ADP carrier

Jürgen Kreiter[1],&, Sanja Vojvodic[1], Mario Vazdar[2], Elena E. Pohl[1]

*[1] Physiology and Biophysics, Department of Biological Sciences and Pathobiology, University of Veterinary Medicine, 1210 Vienna, Austria*
*[2] Department of Mathematics, Informatics, and Cybernetics, University of Chemistry and Technology, 166 28 Prague, Czech Republic*
*§ Present address: Institute of Molecular and Cellular Physiology, Stanford Medical School, Stanford, CA, USA*

*e-mail: elena.pohl@vetmeduni.ac.at*

**Key words:** membrane proteins, mitochondrial transporter, ATP/ADP carrier, bilayer lipid membranes, reconstituted protein, MD simulations

The main function of the mitochondrial ATP/ADP carrier (adenine nucleotide translocase, ANT) is to exchange of ATP for ADP across the inner mitochondrial membrane. Recently, our group and others have shown in different experimental systems that ANT can also transport protons in the presence of different uncouplers (fatty acids, dinitrophenol, FCCP). The exact mechanism and significance of this process are not known. One of the hypotheses, the fatty acid cycle hypothesis, suggests that protons are transported by uncouplers, while the protein facilitates the transport of the anionic form of the protonophore. Based on conductance measurements of lipid bilayers reconstituted with the recombinant protein and MD simulations, we proposed a four-step mechanism for the "sliding" of the FA anion from the matrix into the mitochondrial intermembrane space. Understanding proton transport by ANT could lead to this protein being used as a drug target in the treatment of obesity, cardiovascular and neurodegenerative diseases.

# Reservoirs of ancestral Deltaviruses replicate in water molds

María José López-Galiano[1], Marco Forgia[2], Leticia Botella[3], Zofia Pasterny[1], Massimo Turina[4], Artem Babaian[5,6], Marcos de la Peña[1]

*[1] Instituto de Biología Molecular y Celular de Plantas, Universidad Politécnica de Valencia-CSIC, Valencia, Spain*
*[2] Institute for Sustainable Plant Protection, National Research Council of Italy, Torino, Italy*
*3 Department of Forest Protection and Wildlife Management, Faculty of Forestry and Wood Technology, Mendel University in Brno, Brno, Czech Republic*
*4 Institute for Sustainable Plant Protection, National Research Council of Italy, Brescia, Italy*
*5 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada*
*6 Terrence Donnelly Centre for Cellular & Biomolecular Research, University of Toronto, Toronto, ON, Canada*

*e-mail: rivero@ibmcp.upv.es*

**Key words:** circular RNA, ribozyme, deltavirus, viroid-like, oomycetes

The unique human Hepatitis Delta Virus (HDV), also known as Hepatitis D agent, is the smallest known virus composed by a minimal (~1,700 nt) circular RNA with a rod-like secondary structure. Historically regarded as the sole example of its class, recent data have unveiled novel deltaviruses among diverse metazoans, as well as widespread in environmental metatranscriptomes from unknown hosts. Following high throughput computational approaches, here we unveil novel and intriguing examples of deltavirus genomes in vertebrate (mammals, lizards and fish), invertebrate (sandfly, deep-sea mussel and termites), and diverse environmental metatranscriptomes. Moreover, minimal deltavirus-like genomes of around 1,200 nt of circular RNA were detected in transcriptomic data from two species of water molds (oomycetes). Molecular validation in the causal agent of the grapevine downy mildew, Plasmopara viticola, confirmed the presence of replicative circular forms for both polarities of the agent. Altogether, our data points to the water molds, and likely other fungal-like organisms, as potential reservoirs for Deltavirus-like agents in metazoans, highlighting the ecological and evolutionary complexity of these minimal replicators.

# The use of nanobodies to modulate the interaction between membrane receptors for cancer treatment

Nour Islem Yanis SAIDANI

*Aix Marseille Université*

*e-mail: islemsdn@gmail.com*

**Key words:** nanobodies, cancer, algorithms

The modulation of the interactions between different proteins has been a central strategy for the investigation of the function of specific receptors on the plasma membrane and for the treatment of cancer. We currently work to develop nanobodies, which are the product of an immune reaction developed by camelids and target very precisely a given protein. Lamas are used to immunize against a specific antigen, and subsequently nanobodies are identified, purified and used. Nevertheless, challenges in modelling and predicting the potential success of these nanobodies in inducing any pharmacological effect remains challenging. Therefore, there have been recently various algorithms trying to improve the effectiveness of nanobodies or other biomolecules that could serve to induce a conformational change and therefore modulate the activity of the receptor. These works are crucial in a sense they can improve shorten greatly the time of developing nanobodies, reduce the costs and target very precisely a given region of the receptor.

# Molecular dynamics and machine learning in the study of signal propagation in protein-kinase A

Claudia Arbeitman[1,2,3], Pablo Rojas[1], Oreste Piro4, Martin Garcia1

*1Theoretical Physics and Center of Interdisciplinary Nanostructure Science and Technology, Universität Kassel, Heinrich-Plett-Str. 40, 34132 Kassel, Germany*
*2 CONICET , Godoy Cruz 2290, Buenos Aires C1425FQB, Argentina*
*3 GIBIO-Universidad Tecnológica Nacional-Facultad Regional Buenos Aires, Medrano 951, Buenos Aires C1179AAQ, Argentina*
*4 Departament de Física, Universitat de les Illes Balears, Ctra. de Valldemossa, km 7.5, Palma de Mallorca E-07122, Spain*

*e-mail: claudia.arbeitman@physik.uni-kassel.de*

**Key words:** Protein Kinase A, Signalopathies, Mutations, Molecular Dynamics, Machine Learning.

Molecular Dynamics (MD) simulations and related analysis methods allow for the study of proteins states and transitions that are hidden in the static structures obtain by current experimental techniques. Specifically, they have the potential to enable the study of propagation of information within biomolecules upon a perturbation, such as the binding of a ligand. The propagation of signals from the binding site of the effector, to a distal active site (the region of the protein where its function is fulfilled, e.g. catalytic site), is known as allosterism, while the propagation pathway is known as allosteric pathway.
Important families of proteins, such as the Protein Kinase A (PKA), rely on allosteric regulation to control their level of activity. PKA enzyme is one of largest kinase families in eukaryotes, and has a central regulatory role in coordinating the cellular networks. The transition between active and inactive states of PKA, which is triggered by an allosteric mechanism, is disrupted by mutations or dysfunctional conformational dynamics, which alter allosteric networks and signaling pathways. The resulting dysregulation of PKA activity and are associated with the pathogenesis of various diseases, including cancer, neurodegenerative disorders, heart diseases, and endocrine pathologies.
I will address in this contribution key questions about the unknown precise allosteric mechanisms regulation PKA activity, as well as how small chances in genomic sequence, namely mutations, cause aberrant allosteric effects. I will show how can machine learning and the analysis of simulations-derived time series can shed light on the allosteric pathways and specific signal propagation.

# Genomic Sequences, Fractals and the ambiguity of Nullomers.

Iván Marqués Campillo1, Simone Giannerini2, Oreste Piro1, Stella Logotheti3, Alexandros Georgakilas3, Diego L. González4

*1Physics Department, University of Balearic Island, E-07071, Palma de Mallorca, Spain*
*2Department of Statistical Sciences, University of Bologna, Bologna, Italy*
*3 DNA Damage Laboratory, Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou, 15780 Athens, Greece*
*4CNR-IMM, Sezione di Bologna, Via Gobetti 101, Bologna, Italy*

*e-mail: iwmarquescampillo197@gmail.com*
*simone.giannerini@unibo.it*
*piro@imedea.uib-csic.es*
*stella_logotheti@mail.ntua.gr*
*Alexg@mail.ntua.gr*
*gonzalez@bo.imm.cnr.it*

**Key words:** nullomers, rarity, fractal, genomic, sequences, structure

The role of nullomers in the genome has long intrigued scientists, prompting the investigation of patterns in genomic sequences. In this work, we challenge the conventional understanding of nullomers, revealing their surprising ambiguity. We introduce a systematic methodology to characterize genomic sequences based on their presence within specific sets, such as the whole genome, the exome, or coding sequences. Our findings illuminate crucial aspects of genomic sequences:

1. Ambiguity in Rarity: Rare sequences, though present, exhibit properties akin to en-tirely absent sequences, emphasizing the complexity of genomic patterns.
2. Fractal Structure: Sequences arranged by frequency demonstrate a fractal structure, opening avenues for exploration using dynamical systems tools.
3. Generalized Chargaff's Rules: A generalized version of Chargaff's rules emerges when considering equivalence classes and chemical transformations.
4. Phylogenetic history: This methodology opens the door to future studies aiming to re-construct the phylogenetic history of extinctions across the eukaryotic tree of life.

While these aspects are under active investigation through collaborative efforts, we believe that our study, initiated within the framework of COST Action CA21169, lays the foundation for a profound understanding of life's origins from the perspective of theoretical biology.

# Mitigating Class Imbalance in CAD Risk Prediction: A Model-Agnostic Ensemble Approach

Dimitrios Trygoniaris1, Vizirianakis Ioannis1, Malousi Andigoni1 & Mittas Nikolaos2

*1Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece*
*2Democretus University of Thrace, Kavala, 65404, Greece*

*e-mail: dtrygoni@auth.gr*

**Key words:** machine learning, class imbalance, CAD risk

Class imbalance, where positive cases are significantly outnumbered by negative cases, is a major obstacle in healthcare data analysis, particularly for tasks like predicting coronary artery disease (CAD) risk. Traditional machine learning models often perform poorly on imbalanced datasets, leading to biased predictions that may overlook high-risk individuals. This study investigates the application of a model-agnostic ensemble framework to address class imbalance in CAD risk prediction. Ensembles, by combining the strengths of multiple diverse models, offer improved robustness and generalizability compared to single models. The model-agnostic approach allows for the incorporation of various machine learning algorithms without prior assumptions about their underlying mechanisms. This research explores the impact of class imbalance on the accuracy of CAD risk prediction and the potential of model-agnostic ensembles to overcome this limitation. We propose a framework that utilizes a diverse set of machine learning algorithms to construct the ensemble. This diversity can be achieved by employing algorithms with different learning paradigms (e.g., Logistic Regression, K-Nearest Neighbors, Support Vector Machines, and Random Forest) or by manipulating hyperparameter settings within a single algorithm class. To address class imbalance specifically, appropriate classification metrics for class imbalance (e.g., balanced accuracy and F2-score) were employed as weighting coefficients for the baseline learners' predictions to contribute to the ensemble prediction. The evaluation of the ensemble framework will involve comparing its performance with single machine learning models on an imbalanced clinical dataset relevant to CAD risk prediction. Metrics like balanced accuracy, F2-score, and area under the ROC curve (AUC) were employed to assess the effectiveness of the ensemble in handling class imbalance and achieving reliable risk stratification. This research aims to demonstrate the potential of model-agnostic ensembles for improved CAD risk prediction in the presence of class imbalance. The findings can contribute to the development of more robust and accurate clinical decision support systems for cardiovascular disease prevention and management.

# Cancer radiotherapy-related cardiovascular diseases: providing multidisciplinary and omics-based solutions to a complex clinical condition

Stella Logotheti1, Hamid Khoshfekr Rudsari2, Athanasia Pavlopoulou3, Jerome Zoidakis4, Ilangko Balasingham2, Robert David5 and Alexandros G. Georgakilas1*

*1DNA Damage Laboratory, Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou, 15780 Athens, Greece*
*2Intervention Centre, Oslo University Hospital, Oslo, Norway*
*3Izmir Biomedicine and Genome Center (IBG), Balcova, Izmir 35340, Turkey*
*4Department of Biotechnology, Biomedical Research Foundation, Academy of Athens, Athens, Greece*
*5Department of Cardiac Surgery, Rostock University Medical Center, 18057 Rostock, Germany*

*e-mail: stella_logotheti@mail.ntua.gr, *alexg@mail.ntua.gr*

**Keywords:** radiation; tumor secretome; cardiooncology; CVD; intercellular communication; multimodal omics, machine learning

Advances in cancer therapeutics have improved patient survival rates. However, cancer survivors may suffer from adverse events either at the time of therapy or later in life. Cardiovascular diseases (CVD) represent a clinically important, but mechanistically understudied complication, which interfere with the continuation of best-possible care, induce life-threatening risks, and/or lead to long-term morbidity. These concerns are exacerbated by the fact that targeted therapies and immunotherapies are frequently combined with radiotherapy, which induces durable inflammatory and immunogenic responses, thereby providing a fertile ground for the development of cardiovascular diseases (CVDs). Stressed and dying irradiated cells produce 'danger' signals including, but not limited to, major histocompatibility complexes, cell-adhesion molecules, proinflammatory cytokines, and damage-associated molecular patterns. These factors activate intercellular signaling pathways which have potential detrimental effects on the heart tissue homeostasis. Herein, we present the clinical crosstalk between cancer and heart diseases, describe how it is potentiated by cancer therapies, and highlight the multifactorial nature of the underlying mechanisms. We particularly focus on radiotherapy, as a case known to often induce cardiovascular complications even decades after treatment. We provide evidence that the secretome of irradiated tumors entails factors that exert systemic, remote effects on the cardiac tissue, potentially predisposing it to CVDs. We suggest how multimodal omics, mathematical modeling and natural language processing can be combined with in vitro and in vivo experimentation to generate feasible experimental workflows that address the molecular mechanisms of radiotherapy-related cardiotoxicity at the organismal level. Such highly collaborative efforts at the wet/dry-lab interface hold promise to untangle the desirable immunogenic properties of cancer therapies from their detrimental effects on heart tissue. These results could be translated to next-generation regimens that maximize tumor control, minimize cardiovascular complications, and support quality of life in cancer survivors.

# Exploration in Reinforcement Learning

Adam Jedlička

*Department of Adaptive Systems, Institute of Information Theory and Automation, Pod Vodárenskou věží 1143, 182 00 Prague 8, Czech Republic*

*e-mail: jedlicka@utia.cas.cz*

**Key words:** Reinforcement Learning, Exploration methods, Markov Decision Processes

A wide variety of tasks including modeling biological problems can be modeled by Markov Decision Process (MDP). MDP consists of an agent interacting with the environment. The agent observes the environment state and influences it by purposefully selected actions.

Reinforcement learning (RL) is an approach to solving MDP. In RL the agent learns to make the optimal actions by using feedback (reinforcement) signal. Due to its ability to handle dynamic environments with high uncertainty, RL has been successfully applied to various biological problems: generating novel molecular structures in drug discovery [Liu, K., et al. (2018). "Deep Reinforcement Learning for de Novo Drug Design."], predicting protein folding [John Jumper et al. (202) "Highly accurate protein structure prediction with AlphaFold"], etc.

A very basic application of MDP terms in the example of the task of discovering new drugs mentioned earlier is as follows. A generative model (agent) learns a series of actions to create new molecules (states) for maximizing a score given by a predefined score function. RL is applied similarly in an example of genome assembly and other tasks in biology.

The exploration is an important task in RL that i) enables the agent to efficiently discover new "state-action" pairs; ii) improves the agent's ability to adapt to new, unseen, situations; iii) helps the agent to generalize knowledge acquired.

The so-called exploration-exploitation dilemma refers to optimizing the trade-off between discovering new states (exploration) and using already gathered knowledge for immediate reward The importance of the proper choice of this exploration algorithm lies in the potentially large improvement in the speed of convergence of the RL algorithm. The choice of a well-performing exploration algorithm is task and domain-specific thus there is no universal algorithm that would perform the best for every given task.

The proposed poster will i) briefly introduce a mechanism of how RL works along with the comprehensive implementation of a biology-related task into an MDP that is suitable to be solved by RL. ii) describe several exploration algorithms (from rather simple ε-greedy exploration to more complex methods such as the Intrinsic Curiosity Module (ICM)) along with their benefits and show how exactly they fit into the overall RL mechanism.

# Diffusion processes on periodic and disordered substrates

Marco Patriarca

*National Institute of Chemical Physics and Biophysics - Tallinn*

*e-mail: marco.patriarca@gmail.com*

**Key words:** (normal&anomalous) diffusion, disordered substrate potential, electrostatic effects.

In the present contribution I provide a short overview on some diffusion processes encountered in biological systems, where many processes depend on the motion of ions or molecules along periodic or disordered substrates. For example, molecular motors such as kinesins, dyneins, and myosins move on the periodic substrates associated to the microtubules [1,2]. Other systems interact with a disordered substrate, as in the case of enzymes such as RNA polymerases, exonuclease and DNA polymerases, and helicases; or ribosomes moving along mRNA; and RNA or DNA translocating through a pore.
The wide range of diffusion processes encountered in biological systems is associated to different underlying mechanisms of motion and result in statistically different laws of diffusion, e.g. normal or anomalous diffusion [3].
I will also discuss how at a microscopic level, such mechanisms underlying diffusion are related to and complicated by the fact that biological environment are electrolyte solutions in which electrostatic forces produce various nonlinear counter-intuitive phenomena such as overcharging of the macromolecules, like-charge attraction, and auto-ionization [4].

[1] R. D. Astumian and P. Hänggi, Brownian Motors, Phys. Today 55(11), 33 (2002).
[2] F. Jülicher, A. Ajdari, and J. Prost, Modeling molecular motors, Rev. Mod. Phys. 69, 1269 (1997).
[3] E. Heinsalu, M. Patriarca, I. Goychuk, G. Schmid, and P. Hänggi. Fractional Fokker-Planck dynamics: Numerical algorithm and simulations, Phys. Rev. E 73, 046133 (2006)
[4] A.Y. Grosberg, T.T. Nguyen, and B.I. Shklovskii, The physics of charge inversion in chemical and biological systems, Rev. Mod. Phys. 74, 329 (2002).

# Enhancing Clinical Predictions through Interpretable Machine Learning: An Analysis with XGBoost, LightGBM, and CATBoost

Andrej Novak[1,2], Marin Pavlov2, Ivan Zeljkovic2, Nikola Pavlović2, Šime Manola2

*1Department of Physics, Faculty of Science, University of Zagreb, Bijenička cesta 32, Zagreb, 10000, Croatia*
*2 Dubrava University Hospital, Avenija Gojka Šuška 6, Zagreb, 10000, Croatia*

*e-mail: andrej.novak@phy.hr or anovak@kbd.hr*

**Key words:** Interpretable Machine Learning, Clinical Predictors, Ensemble Methods

In this research, we benchmark interpretable machine learning (ML) models based on decision trees to analyze complex clinical datasets and identify key predictors of patient outcomes. Utilizing real patient data, our focus was on advanced ensemble methods including XGBoost, LightGBM, and CATBoost. These techniques are particularly adept at managing challenges such as non-linear distributions, multicollinearity, and confounding factors. Among them, XGBoost, augmented by SHAP (SHapley Additive exPlanations), was highlighted for its exceptional interpretability and robustness, offering deeper insights into data relationships that traditional statistical analyses often overlook. Our results confirm the efficacy of these ML approaches in providing meaningful clinical predictions and enhancing understanding of patient data, showcasing their potential in improving healthcare outcomes.

# Biological void is highly structured: nullomers and genomic rare sequences

Diego L. Gonzalez[1,2], Simone Giannerini[2], Greta Goracci[3], Iván Marqués Campillo[4], Oreste Piro[4,5]

*1Institute for Microelectronics and Microsystems (IMM-CNR) UOS Bologna, 40129, Italy*
*2Department of Statistical Sciences - UNIBO, 40126 Bologna, Italy*
*3Faculty of Economics and Management, Free University of Bozen-Bolzano - Italy*
*4Balearic Islands University, Spain*
*5IMEDEA Institute – CSIC - Spain*

*e-mail: gonzalez@bo.imm.cnr.it*

**Key words:** Nullomers, Dynamical Systems, Rare Sequence, Rare Diseases, Cancer, Omics, Fractals

The shortest sequences that do not occur in a genome or a proteome of an organism are called minimal absent words (MAWs) or nullomers (Pinho et al., 2009). The definition can be restricted to parts of the genome or proteome of a given organism, and also enlarged to all the known ones; this last possibility leads to the definition of primes. The actual role of nullomers is intensely debated but still remains unclear. From a biological point of view, the predominant opinion is that the absence of certain k-mers is related to evolutionary pressure for avoiding deleterious effects or features incompatible with biological functions. Nevertheless, some experimental findings have revealed relevant biological features associated to the existence nullomers. For example, the presence of two 5-amino-acid peptides codded in absent sequences would cause fatal damage to cancer cells (2), and the same MAWs have a lethal effect on cancer cell lines derived from nine different organs (1). Also, three minimal 12-mers that are absent in the human genome, do appear in two genes of Ebola virus genomes (9). On the other hand, significant human nullomers are frequently also absent in human viruses which suggest some sort of viral strategy to mimick or deceive the host. However, it still remains unclear whether the absence of a nullomer is an evolutionary consequence of its deleterious effects or just the product of randomness.

The investigation of nullomers is a very interesting subject both from the theoretical point of view and for the possible applications. Theoretically, nullomers open one door to connect with a very important tool of the theory of dynamical systems, namely symbolic dynamics, that allows to describe a sequence of symbols as the output of a dynamical system. Specifically, the idea of this strrategy is to map genomic sequences consisting of spatially ordered symbols from an alphabet of four (A;T;C;G) into a dynamically generated sequence of symbols temporally ordered. In this approach the absent sequences play the rol of a fingerprint to identify the generating system in much more informative way than are allowed sequences. On the other hand, from the point of view of the applications, absent genomic sequences might eventually be used as a tool for diagnosis and aetiology of some diseases, as well as to develop aptamers –short amino-acids sequences– with antibacterial capacity. Several collaborations along these lines, involving research on cancer, rare diseases, and aptamers have been initiated among members of the DYNALIFE COST action. The main problem regarding nullomers regards their definition and characterization in a rigorous and robust manner. This implies the development of advanced statistical approaches (it is a difficult problem to define mathematically the absence of a sequence), and also the implementation of semi-empirical analysis based on general properties of dynamical systems and appropriate representations (as the chaos game representation). The main objectives of the research are: i) to define rigorously nullomers with a quantitative estimation of the robustness of their determination; ii) to explore the organization of nullomers inside the genome and in relation to rare sequences (sequences that appear only a few times in a given genome); iii) to use nullomers to characterize some biological features of normal and pathological genomes; iv) to use nullomers for other related tasks (as for example in determining phylogenetic trees; v) to develop a data

Throughout our investigation, we employ advanced statistical methods and draw upon insights from the general theory of dynamical systems. While preliminary results offer promising proof-of-concept demonstrations, significant challenges persist, particularly concerning data quality and integration with existing annotation databases.

As we move forward, several critical considerations emerge:
i) Selection of suitable organisms for comprehensive testing, accounting for evolutionary dynamics.
ii) Addressing the computational demands inherent in robust nullomer determination, requiring both algorithmic optimization and access to substantial computational resources.
iii) For medical applications, access to genomes associated with specific diseases assumes paramount importance.

**Literature Review**

1) Alileche A, Hampikian G. The effect of Nullomer-derived peptides 9R, 9S1R and 124R on the NCI-60 panel and normal cell lines. BMC Cancer. 2017;17(1):533.
2) Alileche A, Goswami J, Bourland W, Davis M, Hampikian G. Nullomer derived anticancer peptides (NulloPs): differential lethal effects on normal and cancer cells in vitro. Peptides. 2012;38(2):302-311.
3) Pinho AJ, Ferreira PJ, Garcia SP, Rodrigues JM. On finding minimal absent words. BMC Bioinformatics. 2009;10:137.
4) Silva RM, Pratas D, Castro L, Pinho AJ, Ferreira PJ. Three minimal sequences found in Ebola virus genomes and absent from human DNA. Bioinformatics. 2015;31(15):2421-2425.

**Brief Data Description:**

Types of biological datasets: The main types of data that we are using are genomes; a natural extension is to proteomes. Estimated dataset sizes and availability of metadata: We are working mainly with genomes of very different size, from human to escherichia coli (the idea is to analyse also very short genomes (for example, M. Genitalium, mitochondria, etc.)
Bioinformatics platforms used for data production: GenBank, Nullomers data base

# Optimal penalized sparse PCA

Rosember Guerra-Urzola1, Soogeun Park.

*1Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands*

*e-mail: R.I.GuerraUrzola@tilburguniversity.edu*

**Key words:** sparse PCA, penalized PCA.

The advent of advanced technologies has vastly expanded our capacity to gather data, especially in the realm of biology. Yet analyzing such vast datasets poses significant challenges, such as computational instability and variable redundancy, particularly in high-dimensional settings. Sparse Penalized Principal Component Analysis (PCA) methods are frequently employed to mitigate these issues, valued for their computational efficiency and scalability. However, these methods often lack theoretical support for attaining optimal solutions, raising concerns about the reliability of their outcomes.

This study introduces an innovative approach to penalized PCA, utilizing cardinality as a sparsity-inducing penalty within a minorization-maximization framework. We theoretically demonstrate that our method consistently achieves a local optimum, ensuring convergence and stability under the condition that the covariance matrix is positive-definite. This condition is highly restrictive in practice, particularly when dealing with high-dimensional data, due to the covariance matrix being singular. We propose a straightforward correction procedure that guarantees robustness and stability across datasets of any size, including those in high-dimensional settings.

We employ the '16S' dataset from the MixOmics R-package to illustrate the relevance of this condition for the proposed algorithm to converge to an optimal solution. We then demonstrate how this correction efficiently maintains solution stability, representing a significant advancement in the analysis of high-dimensional data. This work not only fortifies the theoretical underpinnings of penalized PCA methods but also enhances their practical application in biological data analysis, facilitating the integration of complex data types and the advancement of theoretical models in biology.

# Exploring Federated Learning Approaches and Platforms in Medical AI models

Maria Katsioula1, Alexandra Kosvira2, Ioanna Chouvarda1

*1MSc student at Aristotle University of Thessaloniki, Greece*
*1Associate professor at Aristotle University of Thessaloniki, Greece*
*2PhD student at Aristotle University of Thessaloniki, Greece*

*e-mail: mkatsioula@auth.gr*

**Key words:** federated learning, deep learning, medicine, genomics

In recent years, the use of Artificial Intelligence models has become an important tool in data processing and analysis in the field of medicine and biomedical research. The need to protect the privacy of sensitive medical data has led to the use of new approaches that combine the efficiency of data analysis while preserving privacy. Federated learning is an innovative approach that allows to train models locally of distributed data without the need to share them. This is very important for real data coming from health as it opens new horizons in the collaboration of different data sources and contributes to personalized medicine. Genomics research holds immense promise for personalized medicine and understanding complex diseases but often faces significant privacy and security challenges. In this work we try to retrain a deep learning paradigm with a federated learning approach using as case study open gene expression data for breast cancer. We discuss about the framework of federated learning approach and the aggregation strategies. We compare the centralized approach of the deep learning model with the one that was trained to federated learning mode and discuss the results for prediction modeling and the challenges.

**Acknowledgement:** This work has received funding from the European Union under grant agreement No 101100633 and No 101137278.

# Integrating Graph Neural Networks and Pathway-level Biomarker Discovery in classifying High-Grade Serous Ovarian Carcinomas

A.Rousomanis1*, S. Plakias1, I. Boutalis1, A. Malousi2

*1Electrical & Computer Engineering Dept, Democritus University of Thrace, GREECE*
*2Medical School, Aristotle University of Thessaloniki, GREECE*

*atharous3@ee.duth.gr*

**Keywords:** Graph neural networks, Weighted Gene Co-expression Network Analysis (WGCNA), Gene expression analysis, Pathway-level biomarkers

Gene expression data stands at the forefront of understanding disease mechanisms, and machine learning offers unprecedented opportunities for predictive modeling. This study aims to construct a robust model for classifying High-Grade Serous Ovarian Carcinomas (HGSOC) with a particular emphasis on leveraging deep learning techniques. Given a set of 3,814 samples and 513 gene targets we employed a) a novel Graph Neural Network (GNN) architecture tailored specifically for processing graph-structured datasets, and b) a Weighted Gene Co-expression Network Analysis (WGCNA) to capture the nuanced gene interactions. Each sample is represented as a graph with the nodes corresponding to genes. To ensure precise information propagation within the graph structure, we compute the adjacency matrix $A$ using WGCNA. The matrix captures the connections between nodes by assigning weights to the correlation between expression levels. Moreover, we integrate the scale-free topology property into the adjacency matrix. Subsequently, the matrix $A$ undergoes a bisection process based on an optimized threshold, resulting in the creation of matrix $A$ that encodes the neighbours of each node. Diverse graph convolutional layers with attention mechanisms are then used to encode local gene-level features into higher-order co-functional gene pathway-level features facilitating the integration of gene-level and pathway-level features into fully connected layers. Compared to alternative models, our integrative approach achieves superior performance in predicting stages of HGSOC. Furthermore, to augment interpretability and glean insights into underlying biological mechanisms, we explore an innovative full-gradient graph saliency mechanism. This approach involves computing saliency scores of the inputs within the fully connected network with respect to the output, utilizing gradients. This enables discernment of input gene influence in predicting the actual patient class, thereby enhancing understanding of the model's decision-making process. This mechanism allows interpretation of model decisions at the pathway level, uncovering crucial pathway-level biomarkers, pivotal for fathoming disease progression in HGSOC.

# Addressing the uncertainty of machine learning models in genomic medicine using conformal predictions

Christina Papangelou1*, Konstantinos Kyriakidis 2, Pantelis Natsiavas 3, Ioanna Chouvarda 1, Andigoni Malousi 1

*1Medical School, Aristotle University of Thessaloniki, 54124, GREECE*
*2UC Santa Cruz Genomics Institute, Santa Cruz, CA 95060, USA*
*3Institute of Applied Biosciences, Center for Research and Technology Hellas*

**e-mail: papangec@auth.gr*

**Keywords:** machine learning, conformal prediction, genomic medicine, uncertainty quantification

The reliability of machine learning models in biomedicine is crucial, particularly in clinical applications in which inaccurate predictions can have life-threatening consequences. Several algorithms have been developed to mitigate the risk of prediction errors. Among these, conformal Prediction (CP) has gained attention due to its high validity and reduced computational cost. Unlike traditional methods, CP allows the generation of prediction sets, instead of point estimates, with guarantees on the error rate as a post-processing step to the already trained models. CP operates under the assumption of "independent and identically distributed random variables (i.i.d.)", emphasizing the exchangeability assumption.

In this study, we demonstrate the application of CP to detect unreliable predictions and improve the trustworthiness of the predictions in two genomic medicine applications. First, a support vector machine (SVM) model was trained to predict the pharmacological response to infliximab of 88 patients with Rheumatoid Arthritis and Crohn's disease using the expression levels of ~33k microarray gene targets. The error rate of the transductive CP model was 2.25%, compared to 12.5% prediction errors of the SVM model alone. Additionally, CP identified two out of 16 test cases as uncertain, prompting further evaluation by physicians.

Second, a multi-class XGboost model was developed to classify 1,311 patients with Diffuse Large B-cell Lymphoma (DLBCL) in three subtypes with distinct molecular signatures (ABC, MHG, GCB). 29,377 gene targets built the feature set and an inductive CP model was employed to test the effect on the performance of the XGboost model. The conformalized model resulted in 8.65% error rate, compared to 16.35% without CP, at 95% confidence level. Moreover, CP flagged 35.6% of the DLBCL test samples as uncertain for further manual classification.

In this work, we showed that the integration of conformal predictors in machine learning enables the quantification of model uncertainty, generating confident prediction outcomes applicable to clinical trials. CP coupled with learning models, can enhance the reliability, interpretability, and ethical use of predictive models in genomic medicine while identifying cases that require further examination by physicians.

# Enhancing Information-Based Modeling with Linear Methods for Prediction

Sandra Ferreira1, Dário Ferreira2

*1 Department of Mathematics and Center of Mathematics, University of Beira Interior, Covilhã, Portugal*
*2 Department of Mathematics and Center of Mathematics, University of Beira Interior, Covilhã, Portugal*

*e-mail: sandraf@ubi.pt*

**Key words:** Biological Information, Distributions, Gumbel, Linear models

This study presents an introduction into the challenges of translating the information metaphor into quantifiable predictive models, focusing on the application of linear models and the integration of biological data. By emphasizing the use of linear models for prediction, researchers aim to advance a comprehensive and empirically grounded approach to information-based analysis.

# Exploring Biological Networks with Deep Learning Methods and Omics data of Cancer Patients

Panagiotis Sarantidis1, Alexandra Kosvyra2, Ioanna Chouvarda2

*1MSc in Biomedical Engineering, Aristotle University of Thessaloniki, GR*
*2School of Medicine, Aristotle University of Thessaloniki, GR*

*e-mail: psarantid@ece.auth.gr*

**Key words:** Graph Convolutional Networks, Omics Data, Breast Cancer Subtypes, Deep Learning, Bioinformatics, Multi-Omics Integration

Cancer remains a major public health challenge worldwide, with numerous types exhibiting distinct molecular characteristics that affect treatment responses and outcomes. Accurate characterization of cancer grade and subtype is of utmost importance for the prognosis and treatment of cancer patients. There is a growing number of studies focused on the use of machine learning for accurate cancer classification. However, while most of them rely on traditional machine learning algorithms and single omics data, the rapidly expanding deep learning field presents a promising avenue for more accurate and robust solutions. The present study explores the use of Graph Convolutional Networks (GCNs) for the histological classification of breast cancer subtypes, specifically distinguishing between lobular and ductal carcinomas.

The methodology includes two approaches, utilizing single and multi-omics data, derived from TCGA, an open-access repository. The initial baseline model used only gene expression data and was later enhanced to integrate three types of omics datasets: Gene expression, methylation and copy number alterations (CNA). In the multi-omics context, two different model architectures were compared. One model employed a single-graph structure, while the other utilized three separate graph structures, each corresponding to one data type. Furthermore, the models were enhanced with the utilization of a knowledge enhancement layer, designed to apply adjustments to the predictions based on predefined knowledge.

In terms of results, models integrating multi-omics data exhibited superior accuracy, particularly the three-graph structure, which achieved a marginally higher validation accuracy of 86.5%, compared to the single graph's 85.8%, while the gene expression model achieved a validation accuracy of 82.9%. Incorporating the attention mechanism led to further enhancements, with the validation results of the multi-omics models reaching 87%. The interpretability of the model was also investigated, through feature perturbation methods, identifying gene signatures indicative of each cancer subtype. The top features of the datasets were used for enrichment analysis contributing to a more profound understanding.

Comparison with state-of-art methods shows that our approach for the histological classification of breast cancer is a novel contribution to the field. To our knowledge, no other studies have leveraged multi-omics data and deep learning architectures for this task. Our results not only underscore the potential of deep learning in improving predictive models for cancer classification but also establish the groundwork for future applications of our model in other cancer classification tasks.

**Acknowledgement:** This work has received funding from the European Union under grant agreement No 101100633 and No 101137278.

# Coupling functions for inference of interaction mechanisms: application to brain and cardiovascular oscillatory interactions

Tomislav Stankovski[1,2]

*1Faculty of Medicine, Ss Cyril and Methodius University, Skopje 1000, North Macedonia*
*2Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*

*e-mail: t.stankovski@ukim.edu.mk*

**Key words:** coupling functions, interactions, information flow, neuroscience, cardiovascular

Coupling functions describe the detailed information about the functional mechanisms underlying the interactions and prescribe the physical rule specifying how an interaction occurs [1]. The functional mechanisms can also describe in detailed the causality and the information flow between nonlinear dynamical systems. By focusing on dynamical oscillators and how they interact, we investigated different biological interactions. In particular, we used a method based on dynamical Bayesian inference [2] in order to model and reconstruct the coupling functions from data of interacting biological systems. The method accounts also for potential time-varying dynamics and noise interferences, similar to those encountered in biological systems. The effectiveness of the method is demonstrated on three cases: neural cross-frequency coupling in sleep [3], cardio-respiratory coupling during ramped time-varying breathing [2,4] and brain-heart-lungs interactions during general anaesthesia [5]. These applications demonstrate the advantages of using coupling functions for quantifying directional information flow in biological interactions.

[1] Stankovski, T., Pereira, T., McClintock, P. V., & Stefanovska, A. (2017). Coupling functions: Universal insights into dynamical interaction mechanisms. Rev. Mod. Phys., 89(4), 045001.
[2] Stankovski, T., Duggento, A., McClintock, P.V., & Stefanovska,A. (2012). Inference of time-evolving coupled dynamical systems in the presence of noise. Phys. Rev. Lett., 109(2), 024101.
[3] Manasova, D., & Stankovski, T. (2023). Neural Cross-Frequency Coupling Functions in Sleep. Neuroscience, 523, 20-30.
[4] Lukarski, D., Stavrov, D., & Stankovski, T. (2022). Variability of cardiorespiratory interactions under different breathing patterns. Biomed. Signal Process. Control., 71, 103152.
[5] Stankovski, T.,et al. (2016). Alterations in the coupling functions between cortical and cardio-respiratory oscillations due to anaesthesia with propofol and sevoflurane. Philos. Trans. Royal Soc. A ., 374(2067), 20150186.

# Polygenic Risk Score Analysis for Juvenile Idiopathic Arthritis

Hamid Khoshfekr Rudsari1, Piotr Pawel Jaholkowski2, Helga Sanner1

*1Oslo University Hospital, Sognsvannsveien 20, 0372 Oslo, Norway*
*2 University of Oslo, Problemveien 11, 0313 Oslo, Norway*

*e-mail: h.k.rudsari@studmed.uio.no & khoshfekr1994@gmail.com*

**Key words:** Polygenic risk score, genome-wide association studies, juvenile idiopathic arthritis.

Juvenile Idiopathic Arthritis (JIA) is a heterogeneous autoimmune disorder with a complex genetic etiology. Understanding the genetic factors contributing to JIA susceptibility is crucial for improving diagnosis and treatment outcomes. Polygenic Risk Score (PRS) offers a promising approach to dissect the genetic architecture of JIA by aggregating the effects of multiple genetic variants across the genome. We aim to investigate the utility of PRS in predicting susceptibility to JIA and identifying genetic markers associated with disease risk. Genetic data from genome-wide association studies (GWAS) were utilized to construct PRSs for JIA in the Norwegian Mother, Father, and Child Cohort Study (MoBa). MoBa included pregnant women (1999-2008) from the general population with data from regular follow-up contacts from children (n>108 000), mothers (n>95 000) and fathers (>70 000). Umbilical cord blood was taken at birth. Main aim of MoBa is to understand the etiology of complex diseases through estimation of specific exposure - outcome associations. We identified JIA cases using Norwegian Patient Registery (NPR). We defined JIA as at least two entries of relevant diagnostic codes (ICD10 codes: M08.0- M08.9 reflecting JIA subgroups) in NPR. The control cases are defined as MoBa children participants without a JIA diagnosis. The genotyping, quality control, and imputation of the MoBa sample genetic data are described and done elsewhere. We applied PRSice, version 2.3.38 to calculate the PRSs from GWAS of JIA. Logistic regression models were used to test the association between JIA diagnosis and each PRS. Covariates included 10 genome-wide principal components, age, sex, and genotyping batches. Our findings reveal a significant association between PRS and JIA risk, with higher PRS associated with increased odds of developing JIA. The odds of developing JIA for children with high PRS were 1.78 times higher (95% CI [1.57, 2.03]) compared to those with low PRS. This finding suggests a statistically significant positive association between PRS and JIA risk ($p < 0.05$). The results underscore the potential of PRS as a valuable tool for risk stratification and early detection of JIA. Incorporating PRS into clinical practice may facilitate personalized treatment approaches and improve patient outcomes. However, further studies are required to validate our findings in larger cohorts and elucidate the functional significance of identified genetic variants. In conclusion, our study demonstrated the utility of PRS in elucidating the genetic basis of JIA and underscores its potential clinical implications. By leveraging PRS, we can move towards a more personalized approach to managing JIA and ultimately improve patient care.

# Using therotical models in agricultural biotechnology: a pathway to sustainable innovation

*Deniz Ekinci*

*Ondokuz Mayıs University, Samsun, Türkiye*

*e-mail: deniz.ekinci@omu.edu.tr*

**Key words:** agricultural biotechnology, pest management, crop science, nutrition

Agricultural biotechnology stands as a cornerstone in addressing the global challenges of food security, sustainability, and environmental conservation. The integration of theoretical models into the field of agricultural biotechnology offers a promising avenue for advancing innovation and optimizing resource utilization. My presentation will deal with the possible applications of theoretical models in agricultural biotechnology, encompassing diverse domains such as crop improvement and pest management. Theoretical frameworks, ranging from mathematical models to systems biology approaches, provide invaluable insights into complex biological processes, enabling scientists to decipher intricate genetic mechanisms and predict phenotypic outcomes. Through the integration of computational simulations and empirical data, theoretical models may facilitate the design of tailored interventions for enhancing crop productivity, resilience, and nutritional quality. Moreover, theoretical modeling serves as a powerful tool for optimizing the deployment of biotechnological interventions, ensuring their efficacy while minimizing potential risks to ecosystems and human health. By fostering interdisciplinary collaborations and leveraging cutting-edge computational techniques, the integration of theoretical models promises to unlock new frontiers in agricultural biotechnology, driving sustainable agricultural practices and fostering resilience in the face of global environmental challenges.

# Statistical Integration of Multi-Omics for Outcome Variables

He Li1, Said el Bouhaddani2, Zhujie Gu3, Jeanine Houwing-Duistermaat[1,4]

*1Dept. of Mathematics, Radboud University, 6525 AJ, The Netherlands*
*2Dept. of Data Science and Biostatistics, Julius Centre, UMC Utrecht, 3584 CX, The Netherlands*
*3Medical Research Council Biostatistics Unit, University of Cambridge, CB2 1TN, UK*
*4Dept. of Statistics, University of Leeds, LS2 9JT, UK*

*e-mail: he.li@ru.nl*

**Key words:** Multi-Omics, Data Integration, Latent Variable Models, Dimension Reduction, Polygenic Risk Score

Several integration methods are available in multi-omics to obtain variables which represent the genetic components of an omic dataset. These variables can be included as covariates in models. However, in the integration stage, choosing the appropriate method remains a challenge. Here, the following three integration approaches are considered: the univariate polygenic risk scores (PRS), and the multivariate O2PLS and PO2PLS. The genetic joint components are the genetic scores. The multivariate approaches decompose the datasets into joint, data-specific and residual parts. O2PLS estimates the parameters by an algorithm, while PO2PLS uses maximum likelihood under a probability assumption. Our aim is to evaluate the performance of the three integration models by simulation.

We consider three simulation designs with various settings of sample sizes, noise levels, number of omic variables and heterogeneity levels between the omic and genetic datasets, as well as various models for the relationship between the genetic scores and the outcome. In Design 1, genetic and omic variables follow a joint multivariate normal distribution. In this setting, the genetic variables, for example, principal components derived from SNPs within a gene, are assumed to be normally distributed. For Designs 2 and 3, the genetic datasets are fixed and simulated only once. Design 2 employs a multivariate normal distribution for the genetic variables, while Design 3 uses Cosi2 software to generate more realistic categorical genetic data. We evaluate the performance of the three integration methods by the amount of explained variance by the genetic scores or joint components. Finally, we applied the three methods to genetic and metabolomic datasets from the ORCADES cohort. We use the obtained genetic score in a model for body mass index (BMI).

For Design 1, PO2PLS shows the best performance, especially for the noisy and high-dimensional settings. In Design 2, PRS shows relatively better performance compared to the multivariate methods. In Design 3, PRS and O2PLS perform similarly to Design 2, whereas PO2PLS performs worst. This might be attributed to the non-normality of the data. When applied to the ORCADES dataset, PRS outperforms the others, possibly due to the low dimension of metabolomic datasets.

# DiffSegR: an RNA-seq data driven method for differential expression analysis using changepoint detection

Arnaud Liehrmann[1,2,3], Etienne Delannoy[1,2], Alexandra Launay-Avon[1,2], Elodie Gilbault[4], Olivier Loudet[4], Benoît Castandet[1,2], Guillem Rigaill[1,2,3]

*1Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, Université Evry, Gif sur Yvette, 91190, France*
*2Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris Cité, CNRS, INRAE, Gif sur Yvette, 91190, France*
*3Laboratoire de Mathématiques et de Modélisation d'Evry (LaMME), Université d'Evry-Val-d'Essonne, UMR CNRS 8071, ENSIIE, USC INRAE, Evry, 91037, France*
*4Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000, Versailles, France*

*e-mail: guillem.rigaill@inrae.fr*

**Key words:** multiple changepoints detection, RNA-seq, differential expression analysis

To fully understand gene regulation, it is necessary to have a thorough understanding of both the transcriptome and the enzymatic and RNA-binding activities that shape it. While many RNA-Seq-based tools have been developed to analyze the transcriptome, most only consider the abundance of sequencing reads along annotated patterns (such as genes). These annotations are typically incomplete, leading to errors in the differential expression analysis. To address this issue, we present DiffSegR - an R package that enables the discovery of transcriptome-wide expression differences between two biological conditions using RNA-Seq data. DiffSegR does not require prior annotation and uses a multiple changepoints detection algorithm to identify the boundaries of differentially expressed regions in the per-base log2 fold change. In a few minutes of computation, DiffSegR could rightfully predict the role of chloroplast ribonuclease Mini-III in rRNA maturation and chloroplast ribonuclease PNPase in (3′/5′)-degradation of rRNA, mRNA and tRNA precursors as well as intron accumulation. We believe DiffSegR will benefit biologists working on transcriptomics as it allows access to information from a layer of the transcriptome overlooked by the classical differential expression analysis pipelines widely used today. DiffSegR is available at https://aliehrmann.github.io/DiffSegR/index.html and published (https://doi.org/10.1093/nargab/lqad098).

# Agent-based cooperation in shared environment

Aleksej Gaj1, Miroslav Kárný1

*1Department of Adaptive Systems, Institute of Information Theory and Automation, Pod Vodarenskou vezi 4, Prague 8*

*e-mail: gaj@utia.cas.cz*

**Key words:** implicit cooperation, multi–agent systems, Bayesian learning, fully probabilistic design, decision–making theory

Understanding the cooperation mechanism in biological systems is one of the important aspects of their modelling. Transport processes in biomolecules as well as molecules interactions involve cooperation at the molecular level [1], [2]. For instance modelling protein-protein interactions requires understanding how binding and other factors influence these interactions. So far computational modeling techniques like molecular dynamics simulations, Monte Carlo simulations, or kinetic modeling approaches were used.
The contribution addresses the problem of the implicit cooperation in multi-agent environment formulated and solved as a decision making task. Methodologically agents are modelled as dynamic self-interested independent Bayesian agents that use fully probabilistic design (FPD) [3] to find the optimal DM strategy. The agents cooperate via sharing their DM rules (patterns). The neighbour's DM rule, gathered by the co-operating agent, is used as an external predictor to enrich the agent's knowledge about the neighbour. This scenario describes the mechanisms that occur in biological systems.

[1] Stefan MI, Le Novère N (2013) Cooperative Binding. PLOS Computational Biology 9(6)
[2] Zhou, R., & Berne, B. J. (2002). What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Protein Science, 11(11), 2869-2881.
[3] Kárný Miroslav, Guy TatianaV.: Fully probabilistic control design, Systems and Control Letters vol.55, 4 (2006), p. 259-265 [2006]
[4] Kárný Miroslav: Towards fully probabilistic control design. Automatica, 32(12):1719–1722, 1996

# Analysis of spatial omics data using functional data analysis

Mark D. Robinson

*SIB Swiss Institute of Bioinformatics and Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

*e-mail: mark.robinson@mls.uzh.ch*

**Key words:** spatial (transcript)omics, functional data analysis, functional PCA

Many spatial omics technologies, which measure transcriptome, proteome or epigenome layers from tissue slices, are now emerging. Interestingly, the variant of technology used dictates to some extent the data analysis strategy. For example, array-based approaches (e.g., Visium or Visium HD) can be represented as regular lattice data whereas most imaging-based approaches (e.g., MERSCOPE) result in molecular-level measurements that are often segmented into cell-level summaries; the latter could be represented as irregular lattice data. After cell typing, some spatial datasets can also be represented as spatial point patterns. In the analysis of these emerging spatial omics datasets, spatial statistics methods that have been used for decades in ecology and geography offer many possibilities. In this talk, I will give an overview of the technologies generating spatial omics datasets and describe our early efforts to combine analyses of high dimensional omics datasets with classical spatial statistics methods. This includes representing spatial summaries as functional data and applying functional data analysis strategies

# Selection of methods derived from dynamical systems theory

Anna Krakovská

*Institute of Measurement Science of Slovak Academy of Sciences,*
*Dúbravská cesta 9, 841 04 Bratislava 4, Slovakia*

*e-mail: krakovska@savba.sk*

**Key words:** dynamical systems, time series, causality, complexity, entropy

Complex dynamics can emerge through various mechanisms. Chaos theory demonstrates that we only need a few simple nonlinear differential equations to produce unpredictable behaviour. Conversely, agent-based models with a multitude of individual entities, each following simple rules, can also give rise to highly intricate structures. Another intriguing aspect is that complexity can manifest in spatial or temporal dimensions - either as complicated patterns that remain constant over time or as surprising time varying behaviour.
This presentation lists some methods inspired by dynamical systems and asks whether they are suitable for addressing the specific goals of the COST Action CA21169 Information, coding, and biological function: The dynamics of life (DYNALIFE).
Methodology from dynamical systems often relates to one-dimensional measurements, but represented in a multi-dimensional state space through a process called phase space embedding. The reconstructed state portrait is topologically equivalent to the attractor of the underlying system. This allows important information about the original system, such as the minimum number of variables needed to model or the attractor's complexity (entropies, fractal dimensions), to be obtained from the reconstruction. The multi-dimensional reconstruction is also useful for noise reduction and predicting the system's evolution. Estimated measures such as complexity and predictability can be effectively utilized for classification or in the causal analysis of complex dynamical systems and networks. Prediction based causal inference was introduced by Clive Granger in 1969 already. Namely, if information about the past of the time series x helps improve the prediction of y, then we say that the x causes y. Granger causality is a powerful tool in the case of stochastic autoregressive processes. However, if deterministic dynamics play a dominant role in the data, alternative causal approaches such as conditional mutual information (transfer entropy), comparison of fractal dimensions of attractors, and cross-prediction methods become more prominent. The methods can be useful in bivariate causal detection and also for studying large dynamical networks with nodes represented by time series data or complexity estimates. Pairwise causal analysis can help to identify strongly connected components of the graph (sets of mutually reachable vertices), representing distinct subsystems interconnected via one-way driving links.

Acknowledgement: Supported by the project APVV-21-0216, and VEGA 2/0023/22.

# Understanding biological data through in silico studies: the hIFNγ glycosylation puzzle

Nevena Ilieva1, Elena Lilkova1, Peicho Petkov2, and Leandar Litov2

*1Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences, 25A, Acad. G. Bonchev Str., Sofia 1113, Bulgaria*
*2Faculty of Physics, Sofia University "St. Kl. Ohridski", 5, J. Bourchier Blvd., Sofia 1164, Bulgaria*

*e-mail: nevena.ilieva@iict.bas.bg*

**Keywords:** hIFNγ, glycosylation, activity modulation, molecular dynamics

Human interferon-gamma (hIFNγ) serves as a pivotal immunomodulating secretory glycoprotein. This natural cytokine is glycosylated, featuring two independently and differently glycosylated N-glycosylation sites in each monomer chain. While glycosylation is not essential for its activity, experimental evidence suggests that it facilitates the folding and dimerization of the recombinant protein and shields it from proteolytic degradation, thereby prolonging its circulatory half-life.
Recently, it was found that when labelled with a His6FLAG tag added to their N-termini, both glycosylated (expressed in insect cells) and non-glycosylated (expressed in E. Coli) forms of the fusion protein exhibit 100 times lower than expected biological activity. Upon removal of the tag, the non-glycosylated forms recover their activity, while in the glycoforms the tag becomes enterokinase-resistant and cannot be removed. Utilizing computational studies, we elucidated the mechanism by which glycosylation maintains the integrity of the cytokine molecule. Our findings unveil a dual-sided mechanism of glycosylation-mediated activity modulation rooted in the glycans' restrictive impact on the C-termini conformation space: firstly, by reducing the entropy of the initial state, which stabilizes the cytokine and extends its circulatory half-life at typical activity levels; secondly, by increasing the enthalpy of the final state, leading to the emergence of differently truncated forms, some of which exhibit enhanced biological activity. Ultimately, both mechanisms prove advantageous for the biological function of the cytokine, affirming glycosylation as pivotal for the stability and preservation of high activity in hIFNγ. These insights hold particular significance in the design of intracellular protein inhibitors, whether based on the entire cytokine or restricted to its C-terminal domain.

Acknowledgments This research was partly supported by the Bulgarian National Science Fund under Grants DN-11/20/2017 and KP-06-N71/3/2023. Computational resources were provided at the HPC Cluster BioSim at the Faculty of Physics of Sofia University "St. Kl. Ohridski".

# Study of dynamical systems by Modified Optimal Homotopy Asymptotic Method

Nicolina Pop1, Remus Daniel Ene2

*1 Department of Physical Foundations of Engineering, Politehnica University of Timisoara, 2 Vasile Parvan Blvd, 300223 Timisoara, Romania*
*2 Department of Mathematics, Politehnica University of Timisoara, 2 Victoria Square, 300006 Timisoara, Romania*

e-mail: nicolina.pop@upt.ro

**Key words:** Optimal homotopy asymptotic method; boundary layer flow; symmetries; Hamilton–Poisson realization

The synchronization or optimization of nonlinear system performance with applications in medicine or electrical engineering are based on the study of dynamical systems.
Based on the mathematical model development in [1], the Modified Optimal Homotopy Asymptotic Method (mOHAM) technique [2] is used to obtain effective and accurate dual analytic approximate solutions taking into account of the thermal effects and to investigate the chemically reactive solute transfer problem in a viscous fluid over an exponentially stretching sheet [3].
The heat and mass transfer problem is analytically explored by using the modified mOHAM. By similarity, the motion equations are reduced to a set of nonlinear ordinary differential equations. Based on the numerical results, there are dual analytic approximate solutions within mass and heat transfer problem.
The aim of this work is to investigate the effective and accurate dual analytic approximate solutions taking into account of the thermal effects. The influence of the physical parameters (the Prandtl number and the temperature distribution parameter) over the temperature profile is analytically explored for both solutions: the first approximate solution and the corresponding dual solution [4].
The advantage of this procedure consists in independence of small or large parameters, and provides us with a simple way to optimally control the convergence of the approximate solutions.
Obtained results are in a good agreement with the numerical results and show that our procedure is effective, accurate and easy to use in applications.

[1] Bararnia H., Gorji M., Domairry G. , Acta Appl Math. 2009;106:125–133
[2] Ene R.D., Marinca V. Appl Math Comput. 2015;269:389–401.
[3] Ene R.D., Pop N. Wave Random Complex. 2021; 1–23; DOI: 10.1080/17455030.2021.1971328
[4] Ene R.D., Pop N. Mathematics,2023,DOI: 10.3390/math11051124.

# Comparing Nonlinear Growth Curve Models for Chickens

Nurinisa Esenbuga1,

*1Department of Animal Science, Faculty of Agriculture, Atatürk University, Erzurum-TURKEY*

*e-mail: nesenbuga@gmail.com*

**Key words:** Growth curve, nonlinear functions, broiler

This study was conducted to find the best-fitted non-linear model for broilers. Eight non-linear models including Von Bertalanffy, Brody, Gompertz, Logistic, Morgan-Mercer-Flodin, Negative exponential, Richards, and Weibull were fitted on body weight-age data and compared adjusted coefficient of determination (R2), mean square error (MSE), iteration number (IN), Akaike's information criterion (AIC) and Bayesian information criterion (BIC) for determining the most appropriate model describing the growth curve for broilers. Based on all the criteria used for comparing these models in broilers, it can be established that the Richards and Gompertz function gave the best fit for the age-body relationship although Bertalanffy, Logistic, MEF and Weibull functions were equally good in predicting the growth curves of the chickens. Richards and Gompertz models have the highest R2, and the lowest HKO, AIC, and BIC values. However, it was observed that the Richards model had higher iteration values (64.9) than Gompertz model (9.6). Brody and Negative exponential functions were not good in fitting chicken growth data in this study to parameter estimates, R2, MSE, IN, AIC, and BIC values.

# Evaluation of systematic reviews of prognostic models for COVID-19: an overview of systematic reviews

Persefoni Talimtzi, Sofia Tsokani, Anna-Bettina Haidich

*Department of Hygiene, Social-Preventive Medicine and Medical Statistics, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, University Campus, 54124, Thessaloniki, Greece.*

*e-mail: haidich@auth.gr*

**Key words:** Prognostic model; overview of systematic review; COVID-19; reporting completeness; transparency

Background: During the COVID-19 pandemic, there was an abundance of prognostic models for the diagnosis and prognosis of this new coronavirus. Many studies have tried to review and critically appraise these models.
Objectives: The purpose of this study is to evaluate the reporting completeness and transparency of systematic reviews (SRs) of prognostic models for COVID-19.
Methods: MEDLINE, Scopus, Cochrane Database of Systematic Reviews, and Epistemonikos (epistemonikos.org) were searched for systematic reviews of prognostic models for COVID-19 until December 31, 2022. The risk of bias in systematic reviews (ROBIS) tool was used to assess the methodological quality of the selected SRs. This overview was registered in OSF as https://osf.io/7y94c.
Results: Ten SRs were retrieved containing from 4 to 310 primary studies and from 6 to 606 prognostic models; none of the SRs synthesized the results in a meta-analysis. Only three of the SRs had a pre-registered and publicly available protocol in a repository for protocols and one of them had its data publicly available on the website. The majority of SRs (70%) had an overall high risk of bias resulting more often from concerns in the synthesis and reporting of findings. The overall corrected covered area (CCA) was 6.3% which shows a small amount of overlapping.
Conclusions: The reporting completeness and transparency of SRs of prognostic prediction models for COVID-19 was poor and should follow certain reporting guidelines to enhance transparency so that clinicians can rely on them to select the appropriate prognostic model for use in each individual patient. Pre-established protocol, detailed information on both methodology and process followed, as well as clear reporting of findings are essential aspects to which attention should be paid.

# Evaluation of Growth Characteristics in Awasi and Morkaraman Sheep with Non-Linear Functions

Omer Cevdet Bilgin1, Nurinisa Esenbuga2,

*1Department of Statistics, Ataturk University,Erzurum, Turkey*
*2Department of Animal Science, Faculty of Agriculture, Atatürk University, Erzurum-TURKEY*

*e-mail: ocbilgin@atauni.edu.tr*

**Key words:** Growth curve, body weight, Sheep

Body weight–age data from fifty-four Morkaraman and seventy-nine Awassi sheep, from birth to thirty-six months of age, were used to obtain a growth pattern for the breeds, with the use of Brody, logistic, Gompertz, Bertalanffy, and Richards non-linear functions. In both breeds, the Brody and Richards models, with equal R2 as 0.99, were superior to the others. However, Brody was precise on its parameters and easy to interpret than Richards function. As a result of heterogeneity, on the other hand, no curve computed using mean body weights in this study had the ability to uniformly represent every individual ($p<0.001$). Biological interpretations of the model parameters were discussed.

# A comparıson nonlınear models to descrıbe the growth of laying hens

Mahmut Sinan Taspinar, Nurinisa Esenbuga

*Department of Animal Science, Faculty of Agriculture, Atatürk University, Erzurum-TURKEY*
*taspinar@atauni.edu.tr*

**Keywords:**Lohmann layers, nonlinear models, growth curves

In this study, five nonlinear models, namely Lojistik, Brody, von Bertalanffy, Richards, and Gompertz were used to determine the variation of growth in terms of body weight from day 40 after hatch until they were days 210 days of age in 89 Lohmann LSL layers. Five nonlinear models were used: Logistics, Brody, von Bertalanffy, Richards, and Gompertz. The accuracy of the models was determined the mean squares errors (MSE), Coefficient of Determination (R2), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). While Richards and Gompertz models with the highest R2, lowest MSE, AIC and BIC values showed the best fit in Lohmann layers, the models showing the worst fit were Brody and Logistic models. In this study, the Richards and Gompertz nonlinear mixed model showed the best goodness of fit for the data set, and is considered the model of choice to describe and predict the growth curve of Lohmann LSL commercial layers.

# Exploring Molecular Symmetry, Language Modelling, and Data Integrity in Biological and Computational Sciences

Luis Mandel

*NTT DATA*
*Hans-Döllgast-Straße 26*
*80807 Munich*
*Germany*

*e-mail: luis.mandel@nttdata.com*

**Key words:** Context free grammars, molecular symmetry, data integrity

This work explores the intricate relationships among molecular symmetry, language modelling, and data integrity across biological and computational domains. We examine the roles of palindromes in DNA/mRNA, Stochastic Context-Free Grammars, and Cyclic Redundancy Check algorithms.
Palindromes in DNA/mRNA sequences not only showcase molecular symmetry but also contribute significantly to gene regulation and RNA folding processes. Through their symmetrical structures, they serve as recognition sites for various biomolecules, influencing gene expression and RNA stability. In computational linguistics, SCFGs and CFGs offer powerful tools for modelling language structures. Context Free Grammars provide insights into the grammatical structure of mRNA sequences, shedding light on their regulatory mechanisms.
The discussion extends to data integrity verification, where CRC algorithms play a pivotal role. These algorithms, inspired by principles of error detection in digital data, ensure the accuracy and reliability of transmitted information. Interestingly, parallels can be drawn between CRC algorithms and mRNA translation fidelity mechanisms. Both employ redundancy and error-checking mechanisms to maintain data integrity, whether in digital systems or biological processes.
By understanding these interconnections, we gain deeper insights into the fundamental principles governing molecular biology, computational linguistics, and data science. This knowledge not only enriches our understanding of biological processes and linguistic structures but also informs the development of robust algorithms for error detection and correction in various technological applications.

# Predicting structural disorder of proteins using the wavelet transforms

Nataša Ž. Mišić1, Nevena Ćirić2, Jovana Kovačević2, Miloš Milovanović3

*1 R&D Institute Lola, Kneza Višeslava 70a, Belgrade, Serbia*
*2 Faculty of Mathematics, University of Belgrade, Studentski trg 16, Belgrade, Serbia*
*3 Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, Belgrade, Serbia*

*e-mail: nmisic@rcub.bg.ac.rs*

**Key words:** protein disorder, primary sequence, wavelet transform, classification

Prediction of protein structures, functions and interactions based on its primary sequence has been a significant challenge in bioinformatics and structural biology for decades. Recent progress has been significantly accelerated by the widespread incorporation of artificial intelligence and machine learning-based approaches with various methodologies, assessments, and databases. This study proposes a computational method for detecting protein structural order and disorder, i.e., whether a protein 3D structure is stable or not, according to its primary sequence, using simple sequence-based assessment measures for protein structure and interaction prediction with continuous and discrete wavelet transform feature extraction. For protein sequence representation based on continuous wavelet (CW) transform, the singular value decomposition method is used to extract descriptors from the CW images (scalograms). In an additional approach based on the discrete wavelet (DW) transform of the same protein sequences, the joint statistics of wavelet coefficients across scales is effectively characterized by the wavelet-domain hidden Markov model (WHMM). The choice of the optimal wavelet is defined by maximization of global complexity that is information contained in the hidden Markov tree (HMT), which is a measure of self-organization representing the increase of local complexity across scales. The Viterbi algorithm is used to detect hot spot residues in the protein sequence, which relate prominent values of detail coefficients corresponding to singularities. In the final step, various machine learning models will be explored to tackle the classification of proteins as ordered or disordered based on their feature representation.

Acknowledgment
This research was supported by Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant No. 451-03-66/2024-03/ 200066 and 451-03-47/2023-01/200104, as well as partly by the COST Actions DYNALIFE CA21169 and ML4NGP CA21160.

# Data segmentation: Univariate mean change and beyond

Haeran Cho1

*1University of Bristol, UK*

*e-mail: haeran.cho@bristol.ac.uk*

**Key words:** data segmentation, change point, high dimensionality, time series analysis

Data segmentation a.k.a. multiple change point analysis has received considerable attention due to its importance in time series analysis and signal processing, with applications in a variety of fields including natural and social sciences, medicine, engineering and finance. I will first review the canonical data segmentation problem which aims at detecting and localising multiple change points in the mean of univariate time series. Arguably, the canonical segmentation problem has been the most popular framework to propose new data segmentation algorithms and study their efficiency in the last decades. The second part motivates the importance of attaining an in-depth understanding of strengths and weaknesses of methodologies for the change point problem in a simpler, univariate setting, as a stepping stone for the development of methodologies for more complex problems. Extensions towards high-dimensional change point problems are also discussed. I will showcase some of the change point detection methodologies with examples including an application to DNA copy number aberration detection.

# Directed self-assembly, genomic assembly complexity and the formation of biological structure, or, what are the genes for nacre?

Julyan Cartwright[1,2]

*1Instituto Andaluz de Ciencias de la Tierra, CSIC–Universidad de Granada, Armilla, Granada 18100, Spain*
*2Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, Granada 18071, Spain*

*e-mail: julyan.cartwright@csic.es*

**Key words:** complexity, DNA, information, self-assembly, self-organization

Biology uses dynamical mechanisms of self-organization and self-assembly of materials, but it also choreographs and directs these processes. The difference between abiotic self-assembly and a biological process is rather like the difference between setting up and running an experiment to make a material remotely compared with doing it in one's own laboratory: with a remote experiment—say on the International Space Station—everything must be set up beforehand to let the experiment run 'hands off', but in the laboratory one can intervene at any point in a 'hands-on' approach. It is clear that the latter process, of directed self-assembly, can allow much more complicated experiments and produce far more complex structures than self-assembly alone. This control over self-assembly in biology is exercised at certain key waypoints along a trajectory and the process may be quantified in terms of the genomic assembly complexity of a biomaterial.

# Genome editing and gene conversion

Atle M. Bones

*Cell, Molecular and Genomics group, Department of Biology, Norwegian University of Science and Technology, Norway*

*e-mail: atle.m.bones@ntnu.no*

**Key words:** Diploid, allelic variation, DNA repair, polymorphism

Genome editing technology, introduced over the last decade for editing of DNA in a range of species, has become a game changer for functional analyses of genes and functional elements in nucleic acids. Together with utilization in medicine and biotechnology this has opened possibilities unprecedented with the previous technologies. CRISPR/Cas9 genome editing is very precise when it comes to the target site. The guide sequence will lead the Cas9 nuclease to the correct site of the DNA and the nuclease will make a double strand break at the predicted cut site. This corresponds to look up text in a very large book at a specific page, line and make a cut next to the specified letter. In a standard CRISPR/Cas9 protocol this will introduce a break that most often will be repaired by non-homologous end joining repair (NHEJ). NHEJ repair can result in wt characteristics if the reading frame is restored. But it will also often result in the productions of mutants with indels of variable size that are out of frame and providing non-functional mutants. The further processing of these observed mutants with indels and the outcome will be discussed with focus on gene conversion and loss of heterozygosity.

# Characterization of short single-stranded DNAs

Lev Sukovatyi1, Mahipal Ganji2, Rafayel Petrosyan[1, 3], *

*1 L. A. Orbeli Institute of Physiology NAS RA, 0028 Yerevan, Armenia*
*2 Department of Biochemistry, Indian Institute of Science, 560 012 Bangalore, India.*
*3 Zaven & Sonia Akian College of Science and Engineering, American University of Armenia, 0019 Yerevan, Armenia*

*e-mail: rafayel.petrosyan@aua.am*

**Key words:** DNA base stacking energies, single-stranded DNA, molecular dynamics simulations

Characterization of short ssDNAs will help us to design and synthesize DNA nanostructures more efficiently. In order to address this problem, we conducted 1 μs long, all-atom molecular dynamics simulations of 7 base-long ssDNAs (with different sequences) in explicit solvent. We calculated the hydrodynamic radii and related it to the relative mobility of these ssDNAs. Recently, two single-molecule experimental studies have been published where authors reported the base staking energies of different combinations of two DNA bases. In these two experimental studies, there is disagreement in some base stacking energies. From the molecular dynamics simulations, we calculate the base stacking energies of bases that compose ssDNAs. We show that simulations of the same system with different force fields can provide significantly different results.

# Migration of single excitation and loss of information contained in DNA

Dalibor Chevizovich

*Laboratory for Theoretical and Condensed Matter Physics, "VINČA" Institute of Nuclear Sciences, National Institute of the Republic of Serbia, P. O. Box 522, University of Belgrade, 11001 Belgrade, Serbia*

*e-mail: cevizd@vinca.rs*

**Key words:** Biomolecules, excitation self-trapping, excitation migration

DNA is the basic carrier of information that manage the processes taking place inside a living cell. Ionizing radiation and some other processes can damage biomolecule significantly affecting the information contained in it. In the case of the DNA and RNA, such damages can reduce the fidelity of DNA molecule replication and, therefore, can become a source of genomic mutations. It is interesting that mutations do not have to occur only in the place where the DNA molecule is directly damaged. They can also occur in places that are significantly distant from the site of the molecule's injury because single electron (or some other types of excitations) generated by the damage can migrate through the molecular structure. Consequently, taking into into account the migration processes of the induced excitation in DNA, RNA, and proteins, is important for consideration of the effects of ionizing radiation on living organisms.

Here, we will present the results of the quantum mechanical study of the kinematics of single excitation through simplified biomolecular chain. We will also consider the resulting quantum correlations that appear between the structural elements of the macromolecular chain by means of the formed polaron. We assume that such correlations can describe the migration of the injected excitation through the structure. In other words, the stability of excitation transfer through the bio-molecular chain should be a purely quantum mechanical effect. We take into account main structure parameters of biomolecules, and the influence of ambient temperature. Theoretical framework for our investigation is Holstein molecular crystal model and the concept of non-adiabatic polaron. Such quasiparticle can propagate over long distances with minimal energy loss. The influence of ambient temperature on transport process was taken into account by the mean field approximation.

The analysis of the proposed model indicates that the maximum distance reached by excitation is related to the magnitudes of the basic parameters of the structure through which the excitation migrates. On the other hand, the exact values of many of these parameters are not precisely known, but only the ranges in which they are found are known. As it involves a large number of different parameters, the estimation of the distance range of excitation is not easy to do using standard numerical methods. Because, we suggest the employment the deep learning methods in further studying of such problems.

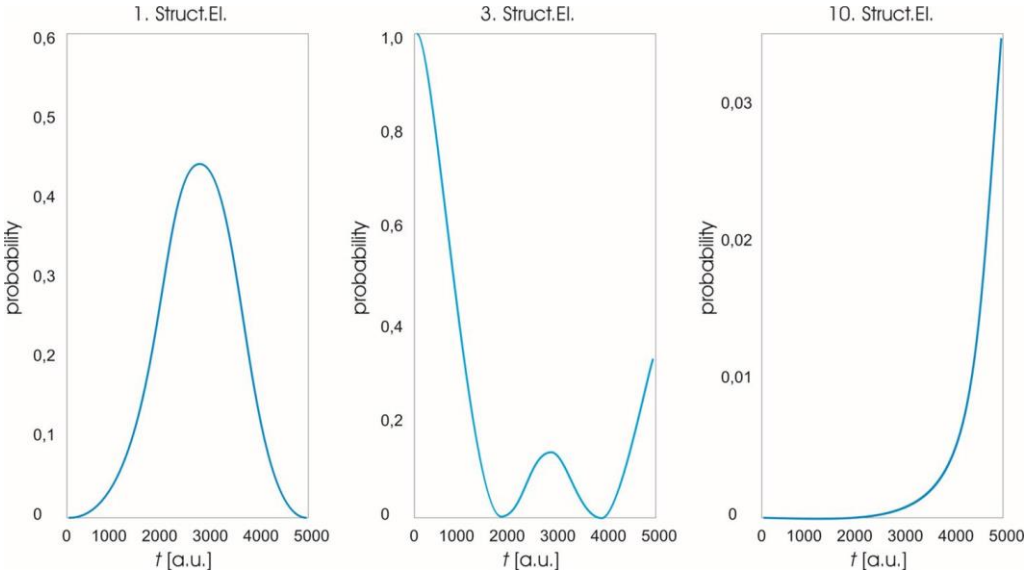


Fig.1. The distribution of the probability of the exctitation detection along the biomolecule. The excitation is injected on the 3.rd structure element.

# Integration of single cell transcriptome sequencing data of head and neck cancer cells

Dragana Dudic1, Diana Domanska2, Nicolina Sciarffa3, Francisca Hofman-Vega4, Serafina Reif5 and Nico Trummer5

*1Faculty of Computer Science and Informatics, University Union Nikola Tesla, Belgrade, Serbia*
*2Department of Pathology, University Hospital, Oslo, Norway*
*3Advanced Data Analysis Group, Ri.MED Foundation, Palermo, Italy*
*4Department of Otorhinolaryngology, University Hospital Essen, Essen, Germany*
*5TUM School of Life Sciences, Technical University of Munich, Freising, Germany*

*e-mail: ddudic@unionnikolatesla.edu.rs*

**Key words:** scRNA-seq, head and neck cancer, data integration

Head and neck cancer is the seventh most common cancer in the world with squamous cell carcinoma as the most common histology. This heterogeneous group of tumors with aggressive malignancy is characterized by a specific tumor microenvironment which can be explored using a single-cell transcriptome sequencing. Single-cell transcriptome sequencing is a relatively new technology that provides high-resolution insight into transcriptomes using high-throughput sequencing. Over time, multiple single-cell sequencing platforms were developed: different versions of CEL-seq and Smart-seq, Drop-seq, inDROP, 10x Genomics, Seq-well and BD Rhapsody, where 10x Genomics platform is the most used for single-cell transcriptome sequencing.

In order to create an atlas of head and neck cancer cells, we collected publicly available single-cell transcriptome datasets of head and neck cancer cells produced by different single-cell transcriptome sequencing platforms. For every dataset we performed dataset specific preprocessing based on dataset characteristics, but in order to keep sensitive low abundant cells. Also, we collected dataset metadata from scientific papers, datasets description and from communication with authors. So far, we created a heterogeneous collection of 713,764 head and neck cancer cells obtained from 214 patients and 17 different studies.

In order to integrate and visualize such a heterogeneous single-cell datasets collection, we created a SIMBA pipeline which incorporates several integration methods, cell filtering and automated annotation tools. Input for SIMBA pipeline is a dataset collection composed of rds and/or h5ad files with predefined structure, where for every file the quality control parameters are given in csv formatted file. Output of the SIMBA pipeline is a cellxgene instance, which is publicly available for our head and neck cancer cells datasets collection.

# On the stable difference schemes solution of nonlinear system of sine-Gordon equations corresponding to DNA dynamics

Ozgur Yildirim1

*1Deprartment of Mathematics, Yildiz Technical University,*
*Davutpasa Campus, Istanbul, Turkey*
*e-mail: ozgury@yildiz.edu.tr*

**Key words:** DNA, wave equations, numerical analysis, stability

The system of sine-Gordon equations which have some powerful applications in life sciences, and in particular to the DNA dynamics is considered. A special case of this system, which describes the open states in DNA double helices is studied. The numerical solution of this system is obtained by finite difference method with fixed point iteration. Analysis of the data and results are presented for the numerical experiments

# Emergent computational control of molecular processes via genetic coding

Peter R Wills1

*1Integrative Transcriptomics*
*Interfaculty Institute for Bioinformatics and Medical Informatics (IBMI)*
*Faculty of Science, University of Tübingen*
*Sand 14*
*D-72076 Tübingen, Germany*

*e-mail: peter.wills@guest.uni-tuebingen.de*

**Key words:** aaRS enzymes, protein structure, code expansion, paleo-phylogeny, quasi-species

At first sight the complexity of biological data is overwhelming but the genetic code gives us an example of how biology itself developed low-dimensional rules to interpret the data in genetic sequences to produce some of the "artefacts", proteins, that are central to the processes of cellular self-construction. Emergent genetic coding provided a form of the autonomous purposeful control using very high-dimensional structures (fine details of dynamic quantum mechanical environments inside proteins) whereby the internal chemical events inside cells are marshalled and integrated to maintain life. We have developed a new approach to the paleo-phylogeny of the aminoacyl-tRNA synthetase (aaRS) coding enzymes, whose function is to match amino acids to their cognate tRNAs. We construct the joint family trees of the Class I and II aaRS on the assumption that the root cause of each branching event is not a random genetic mutation, these happen continually and largely without any effect. Rather, the essential cause of aaRS-tree branching (code expansion) is the simultaneous occurrence of (i) particular "mutations" within the genetic quasi-species, and (ii) the local chance appearance within the protein quasi-species of individual aaRS types able to interpret the newly available genetic sequence as instructions for their synthesis using the still-nascent expanded alphabets of codons and amino acids. The large array of data we use, now available at www.aars.online, consists of the sequences and structures of aaRS enzymes spanning the entire Tree of Life. Likewise, our heuristic assumption has required us to develop tree-building methods that employ amino acid substitution matrices of progressively reduced dimensions to represent events in epochs closer to the evolutionary root of the Class I and II aaRS trees when the genetic code accommodated fewer than the canonical proteogenic 20 amino acids of the modern alphabet. The fundamental theoretical perspective of these studies unites physics, bioinformatics and biochemistry: we model progressive genetic code expansion as a series of self-organising saltations occurring in a thermodynamically driven, spatially distributed, reaction-diffusion "GRT" system that couples the two fundamental processes of molecular biology, replication and translation.

# Augmenting multi-omics integration with prior-informed imputation and drug information

Said el Bouhaddani1, Jeanine Houwing-Duistermaat2

*1Dept. Data science & Biostatistics, UMC Utrecht, Heidelberglaan 100, 3584CX, Netherlands*
*2Dept. Mathematics, Radboud University Nijmegen, Heyendaalseweg 135, 6525AJ, Netherlands*

*e-mail: s.elbouhaddani@umcutrecht.nl*

**Key words:** Omics data integration, machine learning, drug screen, gene-gene interactions, dimension reduction, POPLS-DA, genomics

The integration of multiple omics datasets is becoming increasingly popular in modern biological research. A new promising direction in multi-omics research is augmenting data integration with prior and external information, such as gene-gene interactions and drug targets. However, typically, parts of the multi-omics data are incomplete or missing.
Our motivational example is the study of multiple system atrophy (MSA) and Parkinson's disease (PD) utilizing cell lines. The aim is to better understand the molecular basis of MSA and PD and their interplay with drugs. To this end, transcriptomics, proteomics, and drug screening data were measured in human brain cell lines under a disease-inducing and control environment. Previous integrative analyses1 were limited to a smaller subset of genes and proteins that were available in both datasets, due to incomplete proteomics data. This results in a potential loss of many relevant genes and proteins and a suboptimal integration of the datasets.
We propose an integrative workflow to jointly analyze multi-omics and drug data. First, we develop a novel prior-informed chained multiple imputation algorithm to impute incomplete protein data, using StringDB protein-protein interaction networks. Next, we apply Probabilistic OPLS discriminant analysis, POPLS-DA, to model the (imputed) multi-omics datasets in terms of joint, omics-specific, and residual components. These components consist of weighted combinations of genes and proteins that best separate cases from controls across all omics data. Finally, we apply a 'direct drug neighbor' approach to augment POPLS-DA top genes and proteins with drug screening data. Here, DrugBank is used as prior information on potential drug targets. An augmented gene-gene interaction network is used to visualize which top genes and proteins are relevant drug targets.
We apply our workflow to the multi-modal data from cell lines. We compare our chained multiple imputation algorithm with alternatives.
In conclusion, our integrative workflow yields a list of top genes and (imputed) proteins that best separate cases from controls and are targeted by toxicity-reducing drugs. Compared to the restricted analysis, we identify additional genes and proteins relevant to PD and MSA.

1. el Bouhaddani S, et al. (2024) Statistical integration of multi-omics and drug screening data from cell lines. PLoS Comput Biol 20(1): e1011809.

# Investigation of the existences of “non-existent sequences” in cis-regulatory elements

Isil Takan 1, Athanasia Pavlopoulou[2,3], Alexandros G Georgakilas4, Carsten O. Daub[1,5]*

*1 Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden.*
*2Izmir Biomedicine and Genome Center, Izmir, Turkey.*
*3Izmir International Biomedicine and Genome Institute, Dokuz Eylül University, Izmir, Turkey.*
*4DNA Damage Laboratory, Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou Campus, 15780 Athens, Greece*
*5Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden.*

*e-mail: isil.takan@ki.se*

**Key words:** nullomer, 12-mer, cis-regulatory elements, enhancers, FANTOM5

Enhancers are dynamic cis-regulatory elements which cooperate with promoters to regulate gene expression. The existence of enhancer RNA (eRNA) makes it possible to examine the sequence features of these elements thoroughly. Of importance, in recent years, research on nullomers (i.e. short DNA sequences absent from the genome of a given organism) has gained a lot of interest. In this study, we aimed to investigate whether enhancers harbor nullomer sequences and how do they correlate with expression levels. Collectively, 180825 human intronic nullomer sequences were obtained from the UCSF nullomer database (https://pharm.ucsf.edu/nullomers); approximately notably, ~79.5% of them are not present in the protein coding regions in the human genome. The nullomers not found in coding regions were tested in order to determine whether they are located in enhancer regions in the human genome. A total of 63284 enhancers (61.4% located in introns and 29.2% in intergenic regions) were obtained from the publicly available resource FANTOM5 (https://fantom.gsc.riken.jp/5/). Out of these enhancers, 925 (16.76% located in introns and 69.95% in intergenic regions) explicitly include nullomers of size 12 bp; 766 of these 12-mers are found exclusively in the enhancer sequences. Our preliminary results indicate that a comprehensive analysis of the enhancers would shed light on the relationship between regulatory elements and nullomers in humans and provide detailed information on the effect of nullomers on functional elements and expression levels of genes that are regulated by these elements.

# Context Matters: A Clinical Study Utilizing LLM-Powered Image Analysis for the Detection of Cardiac Pathologies

Matea Novak[1,2], Ivan Zeljkovic[2,3], Ana Jordan[3], Šime Manola[3]

*1RIT Croatia (Rochester Institute of Technology), Zagreb, 10000 Croatia*
*2School of Medicine, Catholic University of Croatia, Zagreb, 10000 Croatia*
*3Dubrava University Hospital, Avenija Gojka Šuška 6, Zagreb, 10000 Croatia*

*e-mail: mnovak@unicath.hr*

**Key words:** Large Language Models (LLMs), Image Analysis, Clinical Diagnostics

Recent advancements in artificial intelligence have led to the exploration of Large Language Models (LLMs) like GPT-4.0 in various domains, including image data analysis in medicine, biology and evolution. This study focuses on the application of LLMs to interpret complex medical imaging data, specifically electrocardiograms (ECGs), which are crucial for diagnosing coronary artery disease (CAD) and other cardiac conditions. CAD remains a leading cause of global mortality, despite declining rates in several EU countries. In this investigation, we utilized the GPT- 4.0 model to analyze a human dataset comprising 120 ECGs that represented a diverse range of cardiac pathologies, including CAD, various arrhythmias, conduction system diseases, and cardiomyopathies. The study evaluated the LLM's efficacy in several key areas: contextual understanding of medical imaging, automated diagnostic reporting from ECG data, and zero-shot detection of cardiovascular diseases. Our results indicate that GPT-4.0 can accurately identify different cardiac conditions, engage in advanced clinical discussions, but also is highly dependent on the supporting textual input. The findings suggest that integrating LLMs like GPT-4.0 into clinical practice could potentially enhance the diagnostic processes by providing robust, AI-driven insights and analysis.

# Modelling a biocompatible hybrid material for oxygen generation against hypoxia-reperfusion injury

Pierro C. I. 1, Albanese P. 2, Lavecchia A.M. 3, Xinaris C.3, Altamura E. 1, Mavelli F.1*

*1Chemistry Department University of Bari A. Moro, Italy*
*2Department of Earth, Environmental and Physical Sciences, University of Siena,Siena,Italy*
*3Department of Molecular Medicine, Institute Mario Negri, Bergamo,Italy*

**e-mail: fabio.mavelli@uniba.it*

**Key words:** Photosynthesis, Thylakoids, Photoproduction of O2, Poloxamer P407, hydrogel

This work aims to develop and characterize a smart soft biomaterial, capable to photo-produce O2 under continuous irradiation to counteracting the effects of the absence of oxy-gen on cell tissues. It consists of thylakoids (1), photosynthetic membranes extracted from chloroplast of Spinacia oleracea leaves, embedded in a thermogel matrix composed of a triblock copolymer: the poloxamer Kolliphor© P407 (2). The thylakoid dispersion in the poloxamer solutions (Fig. 1) does not influence the polymer physico-chemical properties al-ways exhibiting a sol-gel transition as confirmed by rheological analysis. On the other hand, the incorporated thylakoids are not perturbed by the sol-gel transition and appear homoge-neously dispersed in the gel phase, as evidenced by the fluorescent confocal microscopic analysis. The photoactivity of the photosynthetic apparatus, specifically photosystem II (PSII), embedded in the gel phase has been also tested by the Hill assay (5). PSII absorbs photon energy by catalysing the water molecule oxidation with the production of molecular O2 (the water-splitting reaction) and, therefore, donates electrons to an external electron ac-ceptor, the Hill's reagent (2,6-dichlorophenol indophenol, DPIP). The O2 photoproduction of thylakoid was quantified through oxygraphy resulting in about 700 nmol/mL under continu-ous lighting for about 1 hour. The O2 photoproduction was first tested in a thylakoid bulk dispersion then, then in vitro on cardiomyocytes of the left ventricle (AC16) in hypoxia con-dition. The first results of these experiments will be shown and discussed along with the first tentative of modelling poloxamer gels embedding thylakoids.

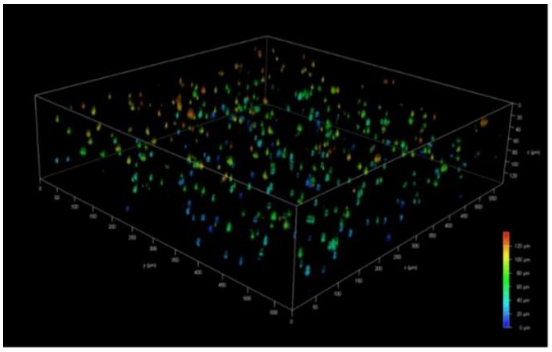

**Figure 1. 3D representation of thylakoids embedded in P407 gel, obtained by the convolution of 141 confocal images**.

1. Johnson M. P., Essays in Biochemistry, 60 (2016), 225-273.
2. Dumortier G., et al., Pharmaceutical Research, 23 (2006), 2709-2728.
3. Miller T.E., et al., Science, 368 (2020), 649-654.
4. Liu T. and Chu B., Journal of Applied Crystallography, 33 (2000), 727-730.
5. Hill R., Nature, 139 (1937), 881.

# Policy Learning via Fully Probabilistic Design

Siavash Fakhimi Derakhshan, Tatiana Valentine Guy

*Adaptive System Department, Institute of Information Theory and Automation, Czech Academy of Sciences, Prague 18200, Czech Republic*

*e-mail: Fakhimi@utia.cas.cz*

**Key words:** Fully probabilistic design, imitation learning, Kullback-Liebler divergence, learning from demonstration, optimal policy.

The discovery of new molecules and materials with desired properties, known as inverse design problems, helps expand the horizons of novel and innovative real-life applications. There are a variety of inverse problems in chemistry encompassing various subfields like drug discovery, retrosynthesis, structure identification, etc. The domain size of the state variables on the chemical space makes it infeasible to search through all possible molecules. Recent developments in modern machine learning methods have shown great promise in tackling problems of this kind. Inverse reinforcement learning is one of the effective methods that has been used frequently in this domain.

Choosing the proper reward function is the challenging part of the problem and has a great impact on the performance of the learning mechanism in the inverse reinforcement learning technique, especially when we are dealing with systems that have high uncertainty. On the other hand, although the decision design algorithm based on fully probabilistic design (FPD) performs well in dealing with systems that are stochastic representation, it requires a high computational and processing cost.

Applying formalism of fully probabilistic design, we propose a new general data driven approach for finding a stochastic policy from demonstrations. The approach infers a policy directly from data without interaction with the expert or using any reinforcement signal. The expert's actions generally need not to be optimal. The proposed approach learns an optimal policy by minimising Kullback-Liebler divergence between probabilistic description of the actual agent-environment behaviour and the distribution describing the targeted behaviour of the optimised closed loop. We demonstrate our approach on simulated examples and show that the learned policy: i) converges to the optimised policy obtained by FPD; ii) achieves better performance than the optimal FPD policy whenever a mismodelling is present.

# Learning from dynamical systems through observed dynamics

Pablo Rojas1, Claudia Arbeitman[1,2], Martin Garcia1, Oreste Piro[1,3,4]

*1Theoretical Physics, University of Kassel, Kassel, Germany*
*2CONICET, Argentina*
*3Department of Ecology and Marine Resources, Institut Mediterrani d'Estudis Avançats, IMEDEA (CSIC–UIB), Balearic Islands, Spain*
*4Departament de Física, Universitat de les Illes Balears, Ctra. de Valldemossa, km 7.5, Palma de Mallorca E-07122, Spain*

*e-mail: pablo.rojas@uni-kassel.de*

**Key words:** dynamical system reconstruction, network inference, recurrent neural network, reservoir computing, time series

The reconstruction of dynamical systems and their features from observed trajectories is a complex inverse problem, and a central research topic of a many scientific disciplines. It is often the case that the experimental manipulation of relevant variables is challenging or unachievable with state-of-the-art techniques, making causal studies difficult. In those scenarios, the ability to infer features of dynamical systems from collected data becomes relevant to their mechanistic understanding.
A particular case of the reconstruction from data consists in the analysis of sequential data (time series, symbolic sequences, etc.), in which the order encodes the generation process. Delayed observations can embedd underlying attractors, a result formalized in Takens Theorem. Profiting on this property, recurrent neural networks (RNNs) have the potential to generate embeddings of dynamical systems by storing past values of inputs. A specific class of RNNs, whose architecture is known as Reservoir Computing, combines the embedding properties provided by its fading memory, the expressivity of neural networks, and data-cheap training properties provided by its randomly generated recurrent part and the linear readout, to form a dynamical system with arbitrary small approximation errors to other dynamical systems, when driven by the observed time series. In the reservoir computing paradigm, a recurrent neural network with fixed weights and nodes with memory of past states (the reservoir) is driven by ordered sequences from its input layer, and outputs through a linear readout which is fitted to the next steps in the sequence, thereby synchorizing with the dynamical system. In this talk, I will sketch an application of Reservoir Computing to infer networks from multivariate dynamical systems, and I will provide an overview of is application to protein dynamics and genomic sequences, generalizing on the notion of constructed reservoir.

# Helicoidal-Peyrard-Bishop model for DNA dynamics and micromanipulation experiments

Slobodan Zdravković

*Institut za Nuklearne Nauke Vinča, Univerzitet u Beogradu, Serbia*
*e-mail: szdjidji@vin.bg.ac.rs*

**Key words:** DNA, Parameter selection, Micromanipulation experiments

A key problem in DNA modelling is a parameter selection. For this paper, the relevant parameters are a harmonic constant of the longitudinal interaction k , and the depth D and the inverse width a of the Morse potential. We demonstrate why k should be bigger than $a^2D$ . A segment of DNA having N base pairs represents two parallel series of N springs of the springconstants (SCs) k . An equivalent SC is $k_e = 2k/N$. From $F = k_e\Delta x$ we obtain $k = FN/2\Delta x$ .
Different values for k can be found in literature. They span from 0.032 N/m to 24 N/m . To the best of our knowledge, the only tedious estimations were done in Ref. [1], where the authors suggested k = 12 N/m . Fig. 1 emerges from a micromanipulation experiment on a single DNA [2]. Such experiments can help to determine k. From the second part of the figure, which is linear, we can estimate the ratio $F/\Delta x$. As the experiment was done on N =48.5x103 base pairs [2], we get $k \approx 1.2$ N/m .
The extension of DNA may result on the stretching of the nucleotides themselves. To model such effect, we can introduce an additional series of N springs of SCs $k_0$ . Following the procedure explained above, we obtain a new value for k which is $k_n = \frac{kk0}{k0 - k}$

For $k_0 >> k$, we get $k_n = 2k$, while $k_0 = {}^{10}/_{9}k$ brings about $k_n = 10k$

This issue deserves further research. The experiment mentioned above was not intended to estimate k. For example, it was pursued for twisted DNA. To test the value k , DNA should be untwisted, i.e., prepared as a ladder, which is possible. Also, it would be a big challenge to determine the ratio $k_0/k$ . It is very likely that biochemists or molecular biologists could help.

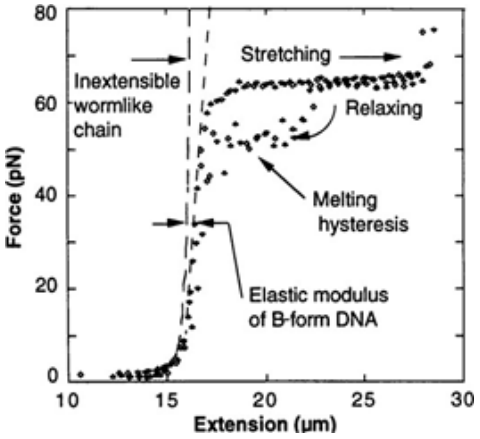


Fig. 1. A force applied on the DNA segment vs. its stretching (Ref. [2]).

[1] S. Zdravković and M.V. Satarić, Phys. Lett. A 373 (2009) 2739.
[2] S.B. Smith, Y. Cui and C. Bustamante, Science 271 (1996) 795.

# Topology and dynamics of transcriptome (dys)regulation

Michel Planat

*CNRS, Institut FEMTO-ST, Université de Franche-Comté, Besançon, France*

*michel.planat@femto-st.fr*

**Key words:** group theory, character variety, Painlevé equations, transcriptome, microRNAs, diseases, cancer research.

This talk follows from paper https://www.preprints.org/manuscript/202403.1741/v1 (in press in MDPI IJMS).
RNA transcripts play a crucial role as witnesses of gene expression health. Identifying dis-ruptive short sequences in RNA transcription and regulation is essential for potentially treat-ing diseases. Let's delve into the mathematical intricacies of these sequences.
We've previously devised a mathematical approach for defining a "healthy" sequence (10.14218/GE.2023.00079). This sequence is characterized by having at most four distinct nucleotides (denoted as nt < 5). It serves as the generator of a group denoted as fp.
The desired properties of this sequence are as follows:
fp should be close to a free group of rank nt-1, it must be aperiodic and should not have iso-lated singularities within its SL(2,C) character variety (specifically within the corresponding Groebner basis).
Now, let's explore the concept of singularities. There are cubic surfaces associated with the character variety of a four-punctured sphere denoted as S(2,4). When we encounter these singularities, we find ourselves dealing with some algebraic solutions of a dynamical second-order differential (and transcendental) equation known as the Painlevé VI equation (10.3390/dynamics4010001). In certain cases, S(2,4) degenerates, in the sense that two punc-tures collapse, resulting in a "wild" dynamics governed by the Painlevé equations of an in-dex lower than VI.
We provide examples of these fascinating mathematical structures within the context of mi-cro RNAs. Specifically, we find a clear relationship between decorated character varieties of Painlevé equations and the character variety calculated from the seed of oncomirs. These findings find applications in cancer research and the control of other diseases.
The work is being completed by an algebraic geometric approach of the miRNA-target regu-latory network and other networks using current data bases and statistical tools. It corre-sponds to WG1 and WG2 subgroups of Dynalife program.

DYNALIFE WG1-WG2 INTERACTION MEETING

DATA DRIVEN EVIDENCE: THEORETICAL MODELS AND COMPLEX BIOLOGICAL DATA

The dimension and diversity of biological data are ever-increasing. How to summarize this kind of data and how to use it to answer relevant biological questions in the framework of theoretical biology are open research questions. To integrate and test theoretical models on biological molecular mechanisms using this data is the ultimate challenge.

The primary objective of this interdisciplinary workshop is to bring together scientists from the fields of statistics, physics, mathematics, bioinformatics, biology, and computer sciences to address the following topics:

Linking WG1 (theoretical models) and WG2 (data analysis approaches), i.e. integrating and testing theoretical models with data. nullomers and free groups, transport processes in biomolecules, and modelling interactions between molecules using agents.

Addressing modern biological problems by analysing all kind of (high-dimensional) data. Specifically approaches for change point detection, omics data integration, machine learning approaches for genomics, and Markov chain theory for genomics.

Rather than being a series of informative talks, our goal is to create meaningful opportunities for collaborative work that address the specific objectives of DYNALIFE. Time will be allocated in the programme to discuss opportunities for ongoing and future collaborations among participants.

The conference will primarily address the specific objectives of WG2 and their interaction with WG1. However, it will be open to participants from all working groups who are aligned with the main challenge and specific objectives outlined for this grant period. The conference foresees plenary talks of 50 minutes (45 + discussion), long presentations of 25 minutes (20 + discussion), short presentations of 15 minutes (10 + discussion), poster sessions, and Working Tables. Parallel sessions will be avoided as much as possible. Slots will be allocated to the in-person Management Committee Meeting (MC), to WG meetings, and to a mixed (in-person-virtual) Core Group Meeting. Cross-field interaction and excellence will be also promoted through the plenary lectures by prominent keynote speakers.

Haeran Cho, School of Mathematics, University of Bristol

Marcos de la Peña, IBMCP, Polytechnic University of Valencia

Mark Robinson, Department of Molecular Life Sciences, University of Zurich

Nevena Ilieva, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

SCIENTIFIC COMMITTEE
Jeanine Houwing-Duistermaat
Greta Goracci
Simone Giannerini
Clara Gracio
Diego Luis Gonzalez
Andigoni Malousi
Natasa Misic

ORGANIZING COMMITTEE
Andigoni Malousi
Giorgos Tzimagiorgis
Bettina Haidich
Ioanna Chouvarda
Dimitrios Trigoniaris
Thanos Rousomanis
Anna Korda

COST Action CA21169

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

ARISTOTLE UNIVERSITY OF THESSALONIKI