3rd Symposium of Genomics Platform

Abstract Book

https://genomics.mini.pw.edu.pl/meeting2023/

Key Note speakers

Krzysztof Poterlowicz

ELIXIR-UK, University of Bradford

ELIXIR-UK: The UK powerhouse for FAIR life sciences data.

Krzysztof is Associate professor in Bioinformatics and Biostatistics, Academic Lead for High Performance Computing cluster and Director of Research for Computational and Data Driven Science at the Faculty of Life Sciences at the University of Bradford. Krzysztof is the training Co-ordinator and a management committee member of ELIXIR-UK, the national Node of ELIXIR in the United Kingdom, the intergovernmental organisation that brings together life science resources from across Europe. Krzysztof is the Principal Investigator on ELIXIR-UK: FAIR Data Stewardship training project, a UK national fellowship for Life Science Data Managers delivered together with collaborators from Cardiff University, Earlham Institute, University of Manchester, and University of Oxford.

Robert Nowak

Warsaw University of Technology

Assembling de-novo.

Robert Nowak, DSc PhD MSc(CS), is an associate professor and head of the Artificial Intelligence Division in the Institute of Computer Science in the Faculty of Electronics and Information Technology at Warsaw University of Technology. Research areas include software development, bioinformatics, artificial intelligence and data fusion. He is involved in developing software systems with efficient implementation of complex algorithms for WUT, European Defense Agency, Samsung, Braster Co., mBank Co., Octagonet Co., Milton-Essex Co., Filbico Ltd., Plum Ltd., Horizen Ltd., Aspartus Ltd., Shanghai Science and Technology Talents Development Center. Co-author of 160 articles, 3 books, 2 patents. He supervises 60 theses. e-mail: robert.nowak@pw.edu.pl

Pawel Stankiewicz

Baylor College of Medicine

Genomic hypermutation phenomenon and low-level somatic mosaicism.

Genomic Disorders: The focus of our research is pathogenetics of lung development and, in particular, the role of non-coding regulatory elements. We demonstrated that haploinsufficiency of the transcriptional factor FOXF1 gene on 16q24.1 results in a lethal neonatal diffuse developmental lung disorder, alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV). Moreover, we unraveled the role of the TBX4 and FGF10 genes in other developmental lung diseases, including acinar dysplasia and congenital alveolar dysplasia. We found that somatic mosaicism for CNVs that also contribute to germline mosaicism is significantly more common than previously thought. We showed that a considerable number of apparently de novo mutations causing genomic disorders occur in the previous generation as low-level somatic mosaicism and can thus be recurrently transmitted to future offspring. We continue to study the scale and clinical importance of this phenomenon.

Anna Gambin

University of Warsaw

From 48 to 46 chromosomes: a novel targeted assembler of segmental duplications unravels the complexity of the HSA2 fusion.

Professor Anna Gambin is deputy dean for research and international cooperation at the Faculty of Mathematics, Computer Science, and Mechanics at the University of Warsaw (term 2016-2024). In her scientific work, she deals with the mathematical modeling of molecular processes and efficient algorithms for analyzing biomedical data. Recently, her research is focused on computational methods supporting medical diagnostics based on genomic and proteomic data. She is the author of over 100 scientific publications and, to date, has supervised ten PhDs in computational biology.

Aleksandra Gruca

Silesian University of Technology

Low-complexity protein sequences: what can we learn from their analysis.

Aleksandra Gruca holds the position of Associate Professor in the Department of Computer Networks and Systems at the Silesian University of Technology. In addition, she serves as the Vice President of the Polish Bioinformatics Society. Her research primarily focuses on advancing machine learning techniques for analyzing protein sequences, with a particular interest in methods that search for similar low-complexity protein sequences with the aim to predict their function. She is also interested in application of predictive models for the integrated analysis of large-scale multi-omics datasets.

Abstracts

Teodor Buchner

Faculty of Physics, Warsaw University of Technology

Why do we need genomics in cardiology?

Abstract: The homeostasis of the human body (if it actually exists) depends on a great number of genetic and epigenetic factors. Thus even a response to repeatable stimuli may be incomparable between individuals. Whether we describe our human subjects, from the point of view of basic sciences, diagnose them or decide on their treatment, we often take mean values as the basis of our reasoning. The ideas such as pesonalized medicine or theranostics underline the fact, that humans are not equal. In cardiology the symptoms of many genetic diseases, such as channelopathies, may be observed directly and noninvasively, which opens a potential area of future research. A particular diversity in subject behavior will be presented, on the basis of feature analysis based on ECG recordings. Also the examples of popular genetic diseases, such as long QT syndrome, will be shown, with the typical alterations of a normal ECG morphology. The presented work ws developed with Karolina Rams and Michał Wierzchowski at the occasion of POB Biotechnology and Biomedical Engineering grant 1820/16/Z01/POB4/2021 of Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme.

Bio: Teodor Buchner is an academic lecturer at Faculty of Physics, Warsaw University of Technology, R&D Expert at EXATEL S.A. and a member of Polish Cardiac Society. Since 14 years he chairs an international workshop Cardiology Meets Physics and Mathematics, typically in March/April in Zakopane, accessible at kasprowisko.pl, where all the translational and interdisciplinary subjects are warmly welcome. His recent scientific interests are located in the theory of biopotentials, where he tries to uncover physics, which was buried under mathematics, which was patient enough to eagerly agree to describe many human ideas on life, the universe and everything.

Marek Wiewiórka

Institute of Computer Science, Warsaw University of Technology

Genomic Data Lakehouse

Abstract: Recent advances in high-throughput sequencing (HTS) have contributed to an unprece-dented growth in the amount of generated multiomics datasets. It opened the door to many population-scale, national genome initiatives such as 100,000 Genomes Project in the United Kingdom or 1000 Polish Genomes.

On the other hand, many popular bioinformatics methods for secondary and tertiary data analyses exhibit high computational complexity. To compound the situation, great number of the existing genomic analysis tools and algorithms are intrinsically sequential and incapable of exploiting the power of distributed computing. In particular, there have not been yet proposed truly scalable methods for common genomic operations such as calculating depth of coverage, summarizing short reads in a form of pileup and joining datasets using interval intersections.
Furthermore, very little studies have addressed the challenges of architecting genomic cloud platforms for distributed processing and analysis of HTS data. Equally, the idea of unified declarative programming approaches for expressing genomic operations using Structured Query Language (SQL) and DataFrame API have not gained enough traction yet.
This talk aims at presenting Genomic Data Lakehouse concept together with SeQuiLa project that implements novel, scalable methods for the aforementioned computationally intensive routines.

Bio: Marek is a seasoned Big Data and Cloud Architect with 15+ years of experience in designing and implementing modern data and MLOps platforms.

Currently Chief Data Architect @GetInData | Part of Xebia and Research Assistant at Warsaw University of Technology putting the finishing touches to his Phd dissertation.

Privately - a keen long distance runner, gravel bikes enthusiast and absolutely in love with the Italian Lakes!

Krzysztof Banecki

Warsaw University of Technology

Simulation of 3D chromatin conformation using genomic and microscopic data

Abstract: 3D modelling of the chromatin is a rapidly developing field. 3-dimensional models of chromatin in the nucleus are crucial in providing quantitative framework for testing hypotheses concerning the chromatin conformation’s role and behavior. Two most popular data types used in the field of 3D chromatin modelling are microscopic and genomic data. Although a multitude of methods have been proposed to incorporate those datasets into chromatin modelling only very few of them are capable of using both data sources simultaneously. Here we present our approach for combining the microscopic 3D-EMISH images with Hi-C genomic data in one single simulation. Our approach relies on the computations on the graph induced by a 3D-EMISH data followed by a Simulated Annealing algorithm that runs towards optimizing the similarity between the transformed distance matrix of a current structure and the target Hi-C matrix.

Bio: Krzysztof Banecki graduated wtih a degree of Mathematics studies with a specialization in Mathematical Statistics and Data Analysis. Currently he is a PhD student in the Faculty of

Mathematics and Information Science at the Warsaw University of Technology and a memeber of the Laboratory of Bioinformatics and Computational Genomics. His main interests comprise of the 3D genome structure, chromatin interactions and its implications in phylogenetics and evolutionary processes.

Filip Korzeniewski

Prediction of amino acids mutations in SARS-CoV-2 spike protein using recurrent neural networks

Abstract: The global outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) urgently requires an effective vaccine for prevention. The primary immunological target of COVID-19 vaccines is the SARS-CoV-2 spike (S) protein. The goal of this work was to develop a method that predicts which mutations are likely to occur in the next periods using historical spike protein sequence data. The main challenges were to process the data SARS-CoV-2 strains in a way it can be later processed by Recurrent Neural Network and to create appropriate models based on RNN able to predict mutations. For this purpose, there was implemented a pipeline using Python programming language that includes preprocessing of raw data, clustering of virus sequences, training a RNN model, and evaluation. The preprocessing consists of data filtering, conversion and creating final datasets structures. Clustering part concerns finding correlations between sequences in consecutive periods by grouping them into resemblant clusters using k-means algorithm. Training part contains the implementation of the ML models used to conduct prediction and other auxiliary training functions. The results achieved by prepared models were compared and discussed. The final outcome indicates that applied approach seems to be prominent in epitopes mutation prediction. The areas of potential improvements were distinguished in the final section.

Bio: Filip Korzeniewski is a former student and a graduate of Warsaw University of Technology in 2 majors: Computer Science; Mechanics and Machine Design. He is also a future graduate of Warsaw School of Economics in Project Management. As his diploma projects he worked on bio-technical subjects. His most recent work, supervised by prof. Tomasz Gabin, considered combination of bioinformatics and AI – prediction of epitopes mutations in SARS-CoV-2 spike protein using models based on RNNs.

Alexander Myronov

Faculty of Mathematics and Information Sciences, Warsaw University of Technology

BERTrand - peptide-TCR binding prediction using BERT and random pairing

Abstract: The developments in T cell receptor (TCR) sequencing techniques allowed for a significant increase in the amount of peptide:TCR binding data available. High-quality prediction models for a fixed epitope sequence are possible, provided enough known binding TCR sequences are available. However, the performance drops significantly for previously unseen peptides. BERTrand is a language model with custom pre-training for peptide:TCR tasks and random pairing for negative decoy examples. It can reliably predict peptide:TCR binding for previously unseen peptides and may be useful for in-silico TCR engineering for novel cancer therapies.

Bio: Born in Ukraine, came to Poland 12 years ago. Studied Computer Science at the Jagiellonian University in Cracow. 10 years in software development, 7 years of experience working with Machine Learning projects. Enrolled in applied PhD studies at MiNI PW 5 years ago, have been researching ML applied to immunology.

Krysztof Kucharczyk

Krzysztof Kucharczyk Techniki Elektroforetyczne Sp. z o.o.

TBD

Abstract: TBD

Bio: TBD

Sevastianos Korsak

Faculty of Mathematics and Information Sciences, Warsaw University of Technology

Multiscale Modelling of Chromatin

Abstract: It is known that chromatin has a hierarchical fractal multiscale nature. The quantum of chromatin is nucleosomes and they are composed by histone atoms which are subject to some chemical reactions called histone modifications. The next scale of interest is the scale

of loops a TADs, where the main players of the game are Smc complexes and CTCF proteins. Finally, chromatin can be separated in compartments A and B, which correspond to strong and weak interactions, which are usually described by strong and weak Leonard Jones potential. In this presentation, we will explore the multiscale structure of chromatin and simulation methods which correspond to each one of the scales.

Bio: Sebastian Korsak started his scientific path with his bachelor in physics in the specialization of astrophysics and mechanics, in the National and Kapodistrian University of Athens. His bachelor thesis was related to astrophysical and experimental plasmas, with focusing on magnetohydrodynamic instabilities. Moreover, he received the title of the master of science by the Technological University of Athens, in the specialization of data science, statistics, and he wrote a thesis related to statistical physics of Ising model. Together with his studies, he participated in an internship about the search of Erdos motifs in DNA sequence. Finally, he ended up working with Plewczynski lab in Warsaw.

Krzysztof Fornalski

Faculty of Physics, Warsaw University of Technology

The Avrami-Dobrzyński phase transition model of cellular cancer transformation.

Abstract: The nucleation and growth theory, described by the Avrami equation, and usually used to describe crystallization and nucleation processes in condensed matter physics, was applied to cancer physics. This can enhance the popular multi-hit model of carcinogenesis to volumetric processes of single cell's DNA neoplastic transformation. The presented approach assumes the transforming system as a DNA chain including many oncogenic mutations. Finally, the probability function of the cell's cancer transformation is directly related to the number of oncogenic mutations. This creates a universal sigmoidal probability function of cancer transformation of single cells, as observed in the kinetics of nucleation and growth, a special case of a phase transition process, as proposed by prof. Ludwik Dobrzyński (1941-2022). The proposed model, which represents a different view on the multi-hit carcinogenesis approach, is tested on clinical data concerning gastric cancer and breast cancer. The results also show that cancer transformation follows DNA fractal geometry, but the exact physical way of that process is still unknown.

Bio: Dr Krzysztof W. Fornalski graduated Faculty of Physics, Warsaw University of Technology (2007) in the field of nuclear and particle physics. He graduated his PhD (with honors) in 2012 in the National Centre for Nuclear Research (NCBJ) under the supervision of prof. Ludwik Dobrzyński. His PhD thesis was related to the radiation biophysics. After the doctoral degree, he started to work parallelly in the industry (Polish nuclear power plant project). Since 2022 he is an adiunct (assistant professor) at the Faculty of Physics, Warsaw University of Technology. His main fields of interest are related to the radiation biophysics, low doses of ionising radiation and cancer physics.

Katarzyna Nałęcz-Charkiewicz

Warsaw University of Technology

De novo assembly as a solution to the Traveling Salesman Problem in bounded-degree graphs - in search of best QUBO formulation

Abstract: The traveling salesman problem (TSP) is one of the most frequently analyzed Computer Science problems over the last few decades. It belongs to the class of NP-complete problems. Hence, despite the availability of a number of classical solutions (exact and approximate), new approaches are still sought, including those using the quantum paradigm. We focus on the special case of STSP, namely the search for the optimal Hamiltonian cycle in regular graphs with bounded degree of a graph vertex. A task formulated in such a way has a number of practical applications in bioinformatics, among others the problem of assembling DNA sequences or multiple sequence alignment. The aim of this work was to try to answer the question, which of the available formulations of the TSP problem in bounded-degree graphs are the best to be solved using the quantum annealer - a special type of quantum computer, dedicated to optimization tasks, where the parallelism of calculations is somehow inscribed in its nature.

Bio: Katarzyna Nałęcz-Charkiewicz - MSc. (Computer Science), in 2014 graduated Warsaw University of Technology with distinction. In 2012-2020, worked as a programmer and systems architect, gaining experience in commercial projects of various sizes (from mobile applications to complex order management systems) in various technologies (Objective-C, C/C++, Python, Java). From 2020, a doctoral student at the Doctoral School of the Warsaw University of Technology. From 02.2023, an assistant at the Institute of Computer Science, in the Department of Artificial Intelligence (Faculty of Electronics and Information Technology). Scientific interests include bioinformatics and quantum computing.

Anup Kumar Halder

Faculty of Mathematics and Information Sciences, Warsaw University of Technology

Chromatin Loop Prediction using Deep Hybrid Learning

Abstract: With the development of rapid and cheap sequencing techniques, the cost of WGS has dropped significantly. However, the complexity of the human genome is not limited to the pure sequence – and additional experiments are required to learn the human genome's influence on complex traits. One of the most exciting aspects is the spatial organisation of the genome – which can be discovered using spatial experiments (e.g., Hi-C, ChIA-PET). The information about the spatial contacts helps in the analysis and brings new insights into

our understanding of the disease developments. Deep Neural Networks have revolutionized every aspect of life. In this work, we are applying them along with other classical machine learning algorithms to create a robust artificial intelligence algorithm that can predict interactions in the nucleus. We have used an ensemble of deep learning with classical machine learning algorithms. The deep learning network we used was DNABERT, which utilises the BERT language model (based on transformers) for the genomic function. The classical machine learning models included SVMs, RFs, and KNN. The whole approach was wrapped together as deep hybrid learning (DHL). We found that the DHL approach has increased the metrics on CTCF and RNAPII sets. DHL approach should be taken into consideration for the models utilising the power of deep learning.

Bio: Anup Kumar Halder is a Postdoctoral Fellow at Warsaw University of Technology in the Faculty of Mathematics and Information Science, Poland. He received his ME and PhD degrees in Computer Science and Engineering from Jadavpur University in 2014 and 2021, respectively. He was a recipient of the prestigious Visvesvaraya PhD Fellowship award from MeitY, GOI. His recent achievements cover qualitative and quantitative biological data analysis, proteomics, biomolecular interaction network analysis, pattern recognition, machine learning and deep learning in 3D genomics.

Sponsors

Gold Sponsor

Genomed®

Silver Sponsor

KUCHARCZYK

techniki elektroforetyczne

Patron

POLSKIE TOWARZYSTWO BIOCHEMICZNE · OD 1958 · POLISH BIOCHEMICAL SOCIETY · SINCE 1958 ·