GENOMICS@WUT

SYMPOSIUM 2021

PUBLISHED BY GENOMICS PLATFORM@WUT

HTTPS://GENOMICS.MINI.PW.EDU.PL/CONFERENCE

June 2021

Contents

Introduction

Genomics Platform@WUT organized its first symposium on 23-06-2021 as a virtual conference to celebrate the first anniversary of its foundation. The event gathered almost 24 speakers and more than 40 participants.

About the Platform

The Genomics Platform@WUT was established on 11-05-2020 to integrate and promote the genomics-oriented community at Warsaw University of Technology. Coordinated by a Scientific Advisory Board headed by Dariusz Plewczyński (Laboratory of Bioinformatics and Computational Genomics), Genomics Platform@WUT consists of six research groups with diverse research areas:

1. Dariusz Plewczyński (Laboratory of Bioinformatics and Computational Genomics).
2. Robert Nowak (Artificial Intelligence Division).
3. Przemysław Biecek (MI2 Data Lab).
4. Małgorzata Adamczyk (Laboratory of Systems and Synthetic Biology).
5. Tomasz Gambin (biodatageeks).
6. Krzysztof Kaczmarski (GPU Programming Team).

Secretary of the Scientific Advisory Board: Michał Burdukiewicz.

The Platform is open to both experimental and computational scientists interested in genomics. It aims to integrate the genomics-oriented community of Warsaw University of Technology and promote its achievements.

Abstracts

3D chromatin methods

Presenting author: Michał Kadlof (Warsaw University of Technology)

Co-author(s): none

Presently, we observe enormous development of methods o studying chromatin structure. However, there is still lack of computational methods for transforming experimental data like Hi-C or ChIA-PET into meaningful 3D dimensional models. In this presentation, I'm going to present the algorithm for building generic polymer models that satisfy a certain set of constraints. Constraints may be obtained from several data sources like 1D genomics (ChIP-Seq), 2D genomics (Hi-C, ChIA-PET) and 3D Genomics (microscope imaging). All of this data types can be easily combined using python based modelling framework named Spring-Model. This method employs Molecular Mechanics and/or Dynamics, to perform data-driven modelling. It allows the user to flexibly include or exclude different data types and define interactions types. This algorithm is implemented as web-service, and for more advanced user as stand-alone version as an Open Source. It supports GPUs to accelerate the computations. It also allows performing flexible fitting into 3D microscope images (STED, EMISH Electron microscopy).

Algorithm for DNA sequence assembly by quantum annealing

Presenting author: Katarzyna Nałęcz-Charkiewicz (Warsaw University of Technology)

Co-author(s): Robert Nowak (Warsaw University of Technology)

The application of the quantum annealing algorithm to the problem of de novo assembly of DNA sequences will be presented. The task was performed on D-Wave quantum annealer and their SDK named Ocean. A Genomic Signal Processing approach was used. Overlaps between reads were detected by computing the correlation coefficient. The problem of sequence assembly in such a defined task comes down to the problem of finding in the similarity matrix a path meeting the minimum distance criterion.

Analysis of Oxford Nanopore output signal data for basecalling

Presenting author: Adam Napieralski

Co-author(s): Robert Nowak (Warsaw University of Technology)

Third-generation DNA sequencers using Oxford Nanopore technologies produce a raw current signal data that can be used for detecting nucleotides. This problem is called basecalling. Analysis of such data with machine learning models was performed, including Random Forest, SVM and deep neural networks - particularly seq2seq models. The models discover signal patterns of different k-mers. Quality of several models for different meta parameters was compared on data generated by the simulator.

Chromatin phase separation simulations by percolation and loop extrusion model at single loop resolution

Presenting author: Kaustav Sengupta (University of Warsaw)

Co-author(s): Michał Denkiewicz (Warsaw University of Technology), Teresa Szczepińska(Warsaw University of Technology), Ayatullah Faruk Mollah(Aliah University), Raissa D'Souza(University of California, Davis), Dariusz Plewczynski(Warsaw University of Technology)

We propose computational models to infer trajectories of chromosome folding into a hierarchical structure using graph theory, phase separation, and loop extrusion model. The folding trajectory is described by gradually introducing loops to the graph following various topological and data-driven methods. Moreover, the spatial and temporal reorganization of a loop by cohesin and CTCF for each chromatin contact is modeled using synthetic Loop Extrusion data.

Comparative Analysis of Long (Oxford Nanopore Technologies 1D) Versus Short-Read Sequencing Technologies in Terms of SV Inference

Presenting author: Sachin Gadakh (Centre of New Technologies, University of Warsaw ; Doctoral School of Exact and Natural Sciences, University of Warsaw)

Co-author(s): Mateusz Chiliński (Warsaw University of Technology) ; Dariusz Plewczyński (Centre of New Technologies, University of Warsaw ; Warsaw University of Technology)

We present a comprehensive analysis of Oxford Nanopore sequencing technology as compared with short-read techniques, such as Illumina. In our study, we focus on the structural variants, at least 50 bp segments of DNA in length that are unique for personal genomes, as identified by the 1000 Genomes project. Moreover, we designed an ensemble-based method that can process both long-read and short-read sequencing data for the variant discovery process to accurately analyze the human genome.

Comparison of 3D Chromatin Contacts in Human ESC-H1 Mapped Using GAM or Hi-C

Presenting author: Teresa Szczepińska (Warsaw University of Technology, and Max-Delbrück Centre for Molecular Medicine)

Co-author(s): Christoph Thieme (Max-Delbrück Centre for Molecular Medicine), Alexander Kukalev (Max-Delbrück Centre for Molecular Medicine), Warren Winick-Ng (Max-Delbrück Centre for Molecular Medicine), Rieke Kempfer (Max-Delbrück Centre for Molecular Medicine), Thomas M. Sparks (Max-Delbrück Centre for Molecular Medicine), Miao Yu (Ludwig Institute for Cancer Research, La Jolla CA), Luca Fiorillo (Università di Napoli Federico II, and INFN Napoli), Bing Ren (Ludwig Institute for Cancer Research, La Jolla CA), Mario Nicodemi (Università di Napoli Federico II, and INFN Napoli), Dariusz Plewczynski (Warsaw University of Technology and Warsaw University), Ana Pombo (Max-Delbrück Centre for Molecular Medicine)

The biological aspects of chromatin structure that can be discovered using different techniques are a major interest of the NIH 4D Nucleome consortium. Whereas Hi-C contact maps are based on the frequency of proximity ligation, GAM extracts spatial information about 3D genome topology by sequencing the genomic content of randomly orientated nuclear slices. We have carefully compared topologically associated domains, compartments, and pairwise contacts at 50 kb resolution between GAM and HiC.

Estimated nucleotide reconstruction quality symbols of basecalling tools for Oxford Nanopore sequencing

Presenting author: Wiktor Kuśmirek (Warsaw University of Technology, Institute of Computer Science)

Co-author(s): none

In our research we compared the estimated nucleotide reconstruction quality symbols (signs from every fourth line of the FASTQ file) reported by another basecallers. The conducted experiments consisted in basecalling the same raw data sets from the nanopore device by another basecallers and comparing the provided quality symbols denoting the estimated quality of the nucleotide reconstruction.

Genomic map assembly algorithm utilizing binary sequences

Presenting author: Przemysław Stawczyk (Warsaw University of Technology)

Co-author(s): none

Optical Mapping (OM) was found very useful in validating genomic assemblies, correction and scafffolding, single raw OM read describe long fragment of DNA molecule. Raw OM data are could by asembled to create consensus maps. The result of such assembly are maps that can cover entire chromosome. I propose new algorith for creation of consensus maps based on binary seqences. In my algorithm I explore possibility for using binary representation for genome maps.

How can we use mathematical modelling to complete understanding of cellular metabolism in living organisms

Presenting author: Małgorzata Adamczyk (Warsaw University of Technology)

Co-author(s): none

High-throughput data are hoped to complete understanding of cellular systems. However, it is not a trivial task, to predict the behaviour of dynamic metabolic reaction networks based on transcriptomics, proteomics or metabolomics data. A variety of modelling approaches have been developed, including constraint-based modelling in large scale, genome-wide models to comprehensively describe metabolism at the system level.

Identification of potential mechanisms of gene expression regulated by DNA methylations within transcription factor binding sites for breast cancer

Presenting author: Michal Wlasnowolski (Warsaw University of Technology)

Co-author(s): Marta Jardanowska (Institute of Computer Science Polish Academy of Sciences) Michal Dabrowski (Institute of Computer Science Polish Academy of Sciences) Indrajit Saha (The National Institute of Technical Teachers Training and Research (NITTTR), Kolkata, India) Dariusz Plewczynski (Warsaw University of Technology)

We investigated the potential mechanism of breast cancer, mediated by DNA methylations located within Transcription Factor Binding Sites. We selected significant mRNA, miRNA, and DNA methylations from TCGA data by using the MCFS-ID method. We identified TFBS affected by these methylations inside genes, and which TFs combined with methylations are good predictors of gene expression. Then we performed long-range chromatin interactions analysis that could also affect gene expression changes.

Molecular modeling and simulation of potential drug-repurposing for SARS-CoV-2 (COVID-19) coronavirus

Presenting author: Doni Dermawan (Faculty of Chemistry, Warsaw University of Technology)

Co-author(s): Muchtaridi Muchtaridi (Faculty of Pharmacy, Universitas Padjadjaran) Syahrul Hidayat (Faculty of Pharmacy, Universitas Padjadjaran) Michal Lazniewski (Faculty of Mathematics and Information Science, Warsaw University of Technology) Dariusz Plewczynski (Faculty of Mathematics and Information Science, Warsaw University of Technology)

This study aims to identify the most favorable potential drug against SARS-CoV-2 using bioinformatics methods. Molecular docking, pharmacophore modeling, and molecular dynamics simulations were employed for the selected approved, clinical trial, and pre-clinical trial drugs. The results show that the most promising agents for each group are Stanozolol CPI-0610, and RS-PPCC, respectively. These ligands also interact with the key residues at the S protein's binding site, G339, N343, V367,and S373.

Negative data set sampling as the source of bias in prediction of antimicrobial peptides

Presenting author: Michał Burdukiewicz (Medical University of Białystok)

Co-author(s): none

Antimicrobial peptides (AMPs) are molecules that participate in host defense and/or microbial competition by disrupting negatively charged bacterial membranes. To reduce the costs of experimental research, robust machine learning models for AMP prediction are essential. We implement several methods for sampling negative data sets and investigated their impact on the performance of machine learning models.

Network analysis of 3D genomic data from population-averaged and single cell experiments

Presenting author: Michał Denkiewicz (Faculty of Mathematics and Information Science, Warsaw University of Technology)

Co-author(s): Dariusz Plewczyński (Faculty of Mathematics and Information Science, Warsaw University of Technology)

We analyze structural genomics data (ChIA-PET, HiChIP, ChIA-Drop) using network science methods. Our aims include: identification important nodes, duplicated pathways, and creating a model integrating multiple datasets into a consensus network, which can be used for 3D modeling. Moreover, we use our approach to compare multiple cell lines. Our work will advance the understanding of multilevel organization of the human genome and its relation to function.

Nucleotide methylation classification using machine learning methods on the Oxford Nanopore based data

Presenting author: Łukasz Neumann (Warsaw University of Technology)

Co-author(s): Robert Nowak (Warsaw University of Technology)

During the project various machine learning algorithms were used for methylation classification in genomic data. Our research is based on the provided dataset as well as data generated via the DeepSimulator sequencer simulator. We report our findings with regards to the quality of the classifiers and the sizes of the dataset. Additionally, we discuss the correlation between different datasets and its influence on the results of the classifiers.

Optimisation of HiChIP method to generate high-resolution map of chromatin contacts mediated by CTCF and cohesin in the human genome

Presenting author: Karolina Jodkowska (Warsaw University of Technology, University of Warsaw)

Co-author(s): Zofia Parteka (University of Warsaw), Michał Łaźniewski (Warsaw University of Technology), Dariusz Plewczyński (Warsaw University of Technology)

Recent advances in the field of 3G genomics have revealed that human genome structure is tightly linked to the fundamental cellular processes such as transcription, replication or DNA repair. Chromatin Conformation Capture (3C) based techniques serve to study chromatin structure in the nucleus. In this project we aim at optimisation of HiChIP method to obtain a map of chromatin contacts mediated by cohesin and CTCF - major players in the establishment of human genome topology.

Structural Variants Detection by Consensus Algorithm

Presenting author: Mateusz Chiliński (Warsaw University of Technology)

Co-author(s): Dariusz Plewczyński (Warsaw University of Technology)

We present various methods and tools used for the detection of Structural Variation in human genomes. The attempt to improve the quality of the Structural Variants identification from the whole genome sequencing (WGS) experiments is done using the novel ConsensuSV algorithm. The method integrates the SV sets using machine learning, by combining decision trees and neural networks. The biological samples being analysed originate from the 1000 Genomes Project.

Super-resolution imaging of chromatin loop structures in human cells

Presenting author: Zofia Parteka (MiNI, Warsaw University of Technology, CeNT University of Warsaw)

Co-author(s): Jufen Zhu (Jackson Laboratory, Farmington, USA), Karolina Jodkowska (CeNT, University of Warsaw), Byoungkoo Lee (Jackson Laboratory, Farmington, USA), Ping Wang (Jackson Laboratory, Farmington, USA), Teng-Leong Chew (Janelia Research Campus, Ashburn, VA, USA), Jesse Aaron (Janelia Research Campus, Ashburn, VA, USA), Yijun Ruan (Jackson Laboratory, Farmington, USA), Dariusz Plewczyński (MiNI, Warsaw University of Technology, CeNT, Univesity of Warsaw)

The 3D structure of chromatin determines genome compaction in the nucleus. Chromatin fibers form a variety of dynamic conformations to achieve high packing density. However, the spatial organization of chromatin fiber in a single nucleus has been difficult to observe. We present single chromatin loop modeling based on iPALM microscopy. We present analysis of raw peaks positions and 3D images of single chromatin loop from different human nuclei. Our goal is to compute image driven models.

The 13C metabolic flux analysis as a tool in validating the interplay of gene expression networks and metabolism

Presenting author: Róża Szatkowska (Chair of Drug and Cosmetics Biotechnology, Faculty of Chemistry, Warsaw University of Technology, Warsaw, Poland)

Co-author(s): Małgorzata Adamczyk (Chair of Drug and Cosmetics Biotechnology, Faculty of Chemistry, Warsaw University of Technology, Warsaw, Poland)

Metabolism has a key influence on the global coordination of the transcriptional activity of cells and defines their biochemical activity. Fluxomics using 13C-labeled substrate is an established technique to track the complete metabolic pathways providing a global perspective on combined regulation at the level of transcription, translation and metabolic flux. The method is specially useful for studying cancer, metabolic diseases such as diabetes, and metabolic engineering of microorganisms.

The cohesin-mediated looping of the human genome across cell lines.

Presenting author: Abhishek Agarwal (Warsaw University of Technology)

Co-author(s): Dariusz Plewczyński (Professor, Warsaw University of Technology)

The multiscale hierarchical spatial structure of the mammalian genome is defined by the chromatin loops, TADs, compartments, & chromosomal territories. The chromatin looping is observed, where CTCF protein arrests the loop extrusion process driven by the ring-like cohesin molecular motor. These chromatin interactions vary among cell types and conditions, such spatial variability is correlated with the differences in gene expression, that contributes to transcription and DNA replication processes

List of speakers

1. Adamczyk, Małgorzata
2. Agarwal, Abhishek
3. Burdukiewicz, Michał
4. Chiliński, Mateusz
5. Denkiewicz, Michał
6. Dermawan, Doni
7. Gadakh, Sachin
8. Gambin, Tomasz
9. Jodkowska, Karolina
10. Kaczmarski, Krzysztof
11. Kadlof, Michał
12. Kuśmirek, Wiktor
13. Łaźniewski, Michał
14. Nałęcz-Charkiewicz, Katarzyna
15. Napieralski, Adam
16. Neumann, Łukasz
17. Nowak, Robert
18. Parteka, Zofia
19. Sengupta, Kaustav
20. Stawczyk, Przemysław
21. Szatkowska, Róża
22. Szczepińska, Teresa
23. Wiewiórka, Marek
24. Wlasnowolski, Michal


Genomics Platform@WUT