GENOMICS@WUT

SYMPOSIUM 2022

PUBLISHED BY GENOMICS PLATFORM@WUT

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

June 2021

Contents

Introduction

To celebrate the second anniversary of its foundation, Genomics Platform@WUT organized its second symposium on 24-06-2022 as a hybrid event. Our venue was the Faculty of Mathematics and Information Sciences of the Warsaw University of Technology (Koszykowa 75, 00-662 Warszawa, Poland), Lecture Room 101.

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.

Keynote Speakers

Professor Sanghamitra Bandyopadhyay

About Speaker

- Professor, Machine Intelligence Unit
- Director, Indian Statistical Institute
- Fellow: FNA, FIEEE, FTWAS, FNAE, FNASc, FIAPR
- URL: http://www.isical.ac.in/~sanghami

Machine Learning in Computational Biology

Abstract: In this talk, we will focus on some Machine Learning Solutions in Computational Biology. In particular, we will first focus on a method of microRNA target prediction. Next, we will show its utilization for building a microRNA induced gene regulatory network, whose analysis reveals some novel markers for breast and colon cancer. We will then provide a brief overview of a single cell RNA-seq data clustering algorithm. Finally, we will conclude with an application of optimization technique for the drug design problem.

Professor Ujjwal Maulik

About Speaker

- Professor, Jadavpur University, Kolkata, India
- Former Chair Department of Computer Science and Engineering
- Fellow: IEEE, INAE, IAPR, Humboldt, ICTP
- URL: https://sites.google.com/site/drujjwalmaulik/

Dr. Ujjwal Maulik is a Professor in the Department of Computer Science and Engineering, Jadavpur University since 2004. He was also the former Head of the same Department. He held the position of the Principal in charge and the Head of the Department of Computer Science and Engineering, Kalayni Government Engineering College. Dr. Maulik completed B. Sc. in Physics and B. Tech in Computer Science from Calcutta University, Kolkata, and West Bengal, India in 1986 and 1989 respectively. He also received his M. Tech in Computer Science and Ph. D. in Engineering in 1992 and 1997 both at Jadavpur University, Kolkata, West Bengal, India.

Dr. Maulik worked in many universities and research laboratories around the world as visiting Professor/ Scientist including Los Alamos National Lab., USA in 1997, University of New South Wales, Australia in 1999, Universities of Texas at Arlington, USA in 2001, University of Maryland at Baltimore County, USA in 2004, Fraunhofer Institute for Autonome Intelligent Systems, St. Augustin, Germany in 2005, Tsinghua University, China in 2007, Sapienza University, Rome, Italy in 2008, University of Heidelberg, Germany in 2009, German Cancer Research Center (DKFZ), Germany in 2010, 2011 and 2012, Grenoble INP, France in 2010, 2013 and 2016, University of Warsaw in 2013 and 2019, University of Padova, Italy in 2014 and 2016, Corvinus University, Budapest, Hungary in 2015 and 2016, University of Ljubljana, Slovenia in 2015 and 2017, International Center for Theoretical Physics (ICTP), Trieste, Italy in 2014, 2017 and 2018. He is the recipient of Alexander von Humboldt Fellowship during 2010, 2011 and 2012 and Senior Associate of ICTP, Italy during 2012-2018. He is the Fellow of Indian National Academy of Engineers (INAE), India, International Association for Pattern Recognition (IAPR), USA and The Institute of Electrical and Electronics Engineers (IEEE), USA. His research interest include Machine Learning, Pattern Analysis, Data Science, Bioinformatics, Multi-objective Optimization, Social Networking, IoT and Autonomous Car. In

these areas he has published ten books, more than three hundred fifty papers, filed several patents and already guided twenty doctoral students.

Artificial Intelligence for Healthcare

Abstract: In this lecture first we will discuss some current trends in Artificial Intelligence (AI). We will briefly introduce advantages of Deep Learning Techniques and In this regard we will demonstrate applications of different machine learning algorithms including Deep Learning in real life situation. We will demonstrate how Deep Learning Techniques and Graph Neural Network. Subsequently we will demonstrate how these techniques can be used for healthcare applications.

Abstracts

3D genome structures reconstruction

Presenting author: Krzysztof Banecki (Warsaw University of Technology)

Co-author(s): Stanisław Wokulski (University of Warsaw)

Reconstruction of the 3D chromatin folding from contact map and/or microscopy data is highly demanded. But the task is plagued with problems such as substantial missing data or limited resolution. Here we present some of the algorithms used to deal with 3D structure reconstruction from contact map data as well as some of our own results in this this area.

A Stochastic Monte-Carlo model for Loop Extrusion of Chromatin

Presenting author: Sevastianos Korsak (Warsaw University of Technology)

Co-author(s): pf. dr. hab. Plewczynski

In this project, we created a model which simulates the dynamics of loop extrusion for large regions of chromatin. Firstly, we gathered the biophysical laws that governs the formation of loops, which are mediated by two main proteins: cohesin and CTCF. Having these biophysical laws, we can constructed a stochastic model where proteins move along the polymer network of our simulation, and we can define a Monte Carlo stochastic simulation, that can simulate the motion of proteins on the network. Using Chip-Seq data we can explore the microstates of specific regions and reconstruct the heatmaps.

Comparative Analysis of Long Versus Short-Read Sequencing Technologies in Terms of SV Inference of TRIOS

Presenting author: Sachin Gadakh (Warsaw University)

Co-author(s): Sachin Gadakh(University of Warsaw) , Dr. Karolina Jodkowska (Center of New Technologies), Mateusz Chiliński (Warsaw University of Technology), Prof.Dariusz Plewczyński (Center of New Technologies, Warsaw University of Technology)

We present a comprehensive analysis of Whole-genome sequencing of TRIOS samples in-house sequenced, using long-read such as Oxford Nanopore sequencing technology 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 defined by the 1000 Genomes project. We designed an ensemble-based method that takes structural variants output from multiple SV callers and trains them using neural networks and benchmarked on the high-quality SVs from the 1000 Genomes Project.

Comparative analysis of the 3D structure of the genome of selected human lymphoblastoid cells based on ChIA-PET and HiChIP data

Presenting author: Anastasiya Gurova (Warsaw University of Technology)

Co-author(s): Michał Łaźniewski (CEZAMAT), Dariusz Plewczyński (MiNI PW)

Among the indirect biochemical techniques used to capture the 3D structure of the genome, we can distinguish the most popular Hi-C method, methods using immunoprecipitation (ChIA-PET and HiChIP) and methods

independent of ligation (GAM). In this talk, we will analyze the results of two complementary ChIA-PET and HiChIP experiments to identify CTCF-mediated interactions. The project will identify DNA loops using a single tool - ChIA-PIPE. These will be used to compare the 3D structure of the genome of related people.

Compartmentalisation detected by Hi-C and GAM - orthogonal methods in 3D structural genomics

Presenting author: Teresa Szczepińska (CEZAMAT, Warsaw University of Technology)

Co-author(s): Christoph Thieme (BIMSBMax-Delbrück Centre for Molecular Medicine, Berlin), Sachin Gadakh (CeNT, University of Warsaw), Alexander Kukalev (BIMSBMax-Delbrück Centre for Molecular Medicine, Berlin), Warren Winick-Ng (BIMSBMax-Delbrück Centre for Molecular Medicine, Berlin), Rieke Kempfer (BIMSBMax-Delbrück Centre for Molecular Medicine, Berlin), Thomas M. Sparks (BIMSBMax-Delbrück Centre for Molecular Medicine, Berlin), Miao Yu (Ludwig Institute for Cancer Research, La Jolla CA), Bing Ren (Ludwig Institute for Cancer Research, La Jolla CA), Dariusz Plewczynski(Warsaw University of Technology), Ana Pombo (BIMSBMax-Delbrück Centre for Molecular Medicine, Berlin)

Understanding the limitations of chromatin mapping techniques is important for discovering the function of chromatin structure and efficiently assessing long-range gene regulation. Whereas Hi-C contact maps are based on the frequency of proximity ligation in millions of cells, GAM extracts spatial information by sequencing the genomic content of hundreds to thousands ultra-thin, randomly oriented nuclear slices. We have performed GAM on H1 human embryonic stem cells and compared GAM data with Hi-C data from the same cell line. The two methods show differences in the assigned compartments.

Computational modeling of human-nCoV interaction network

Presenting author: Anup Kumar Halder (Warsaw University of Technology)

Co-author(s): Sovan Saha (IEM,Kolkata); Soumyendu Sekhar Bandyopadhyay (Jadavpur University); Piyali Chatterjee(NSEC); Mita Nasipuri (Jadavpur University); Subhadip Basu (Jadavpur University)

The identification of virus and host protein–protein interactions could be beneficial in understanding the disease transmission behavior of the virus. The present work focuses on developing a computational model for the nCoV-Human protein interaction network, using the experimentally validated SARS-CoV-Human protein interactions. Human spreader are identified using Susceptible-Infected-Susceptible (SIS) model as potential human targets for nCoV baits. A GO-based fuzzy affinity function has been used to construct the nCoV-Human protein interaction network at a high specificity threshold.

ConsensuSV - HPC-ready, ML-enhanced automated pipeline for Illumina-based variant detection

Presenting author: Mateusz Chiliński (Warsaw University of Technology)

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

We have developed two bioinformatics pipelines: ConsensuSV-core, and ConsensuSV-pipeline for the identification of several types of mutations in human DNA sequences of the whole genomes. The first algorithm servers the purpose of merging the calls from multiple, independent Structural Variant (SV) callers using consensus approach. The second algorithm servers as a wrapper for the first one - it is high performance computing (HPC) ready pipeline for variant detection developed using luigi framework. Availability: https://github.com/SFGLab/ConsensuSV-pipeline

Drug repurposing for identification of potential spike inhibitors for SARS-CoV-2 using molecular docking and molecular dynamics simulations

Presenting author: Michał Łaźniewski (Warsaw University of Technology)

Co-author(s): x

This work aims to identify potential drugs using an in silico approach. Molecular docking was carried out on both approved drugs and substances previously tested in vivo. This step was followed by a more detailed analysis of selected ligands by molecular dynamics simulations. Because the SARS-CoV-2 virus evolves rapidly due to a plethora of immunocompromised hosts, the compounds were tested against five different known lineages. As a result, we could identify substances that work well on individual lineages and these showing broader efficacy.

Generative Adversarial Networks for Dataset Augmentation

Presenting author: Jakub Józefowicz (CeNT)

Co-author(s): Bartosz Garguliński (UW), Jakub Skrajny (UW), Maciej Wojtala (UW)

Generative Adversarial Networks have shown success with enabling many unsupervised tasks to be trained in a supervised setting. In this talk I would like to present a case study of placing the adversary in a more central role to allow for expanding training datasets. We will see how this gives the potential to train other models where the data is lacking the multiplicity for reliable convergence or where we would like to expand the training space to encompass plausible, but yet undiscovered points, increasing the models resilience when applied to temporally evolving systems.

Intrinsic linking of chromatin in human cells

Presenting author: Michał Denkiewicz (Warsaw University of Technology)

Co-author(s): Maciej Borodzik (University of Warsaw); Krzysztof Spalinski (Warsaw University of Technology); Kamila Winnicka (University of Warsaw); Kaustav Sengupta (Warsaw University of Technology); Marcin Pilipczuk (University of Warsaw); Michał Pilipczuk (University of Warsaw); Yijun Ruan (The Jackson Laboratory for Genomic Medicine); Dariusz Plewczynski (Warsaw University of Technology)

The principles of organization of the chromatin (DNA) strand in the nucleus plays a significant role in its function. One unexplored possibility is that CTCF-mediated chromatin loops, an important building block of this organization, might be linked together, impacting the resulting 3D structure. We propose a practical approximate algorithm based on graph theory that finds such links, based on searching for K6 minors in graphs constructed from pairwise chromatin interaction data obtained from ChIA-PET experiments. We find multiple candidate linked regions and study their characteristics.

Multi-scale phase separation by explosive percolation with single chromatin loop resolution

Presenting author: Kaustav Sengupta (Warsaw University of Technology)

Co-author(s): Kaustav Sengupta(Warsaw University of Technology), Michał Denkiewicz(Warsaw University of Technology), Mateusz Chiliński(Warsaw University of Technology), Teresa Szczepińska(Warsaw University of Technology), Ayatullah Faruk Mollah(Warsaw University of Technology), Sevastianos Korsak(Warsaw University of Technology), Raissa D'Souza(University of Davis), Yijun Ruan(Jackson Laboratory), Dariusz Plewczynski(Warsaw University of Technology)

We propose models of dynamical genome folding into hierarchical components. Our models are based on explosive percolation theory. The chromosomes are modeled as graphs where CTCF chromatin loops are represented as edges. The folding trajectory is simulated by gradually introducing loops to the graph following various edge addition strategies. Finally, we propose the genome folding model - a biophysical pseudo-time process guided by a single scalar order parameter.

PartSeg - tool for reproducible segmentation of chromatin

Presenting author: Grzegorz Bokota (Centre of New Technologies, university of Warsaw)

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

PartSeg is a novel tool for reproducible segmentation from high-resolution microscopy images. During its development, we focused mainly on algorithms for chromatin segmentation in the nucleus, but one could easily extend it for other multichannel data. This presentation will show a few examples of PartSeg usage on chromatin-related data.

Super-resolution visualization of chromatin loop folding in human lymphoblastoid cells using interferometric photoactivated localization microscopy

Presenting author: Zofia Tojek (Warsaw University of Technology, CeNT UW)

Co-author(s): Jacqueline Jufen Zhu (The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, 06030, USA); Byoungkoo Lee (The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, 06030, USA); Karolina Jodkowska (Centre of New Technologies, University of Warsaw, S. Banacha 2c, 02-097, Warsaw, Poland); Ping Wang (The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, 06030, USA); Jesse Aaron (Advanced Imaging Center, Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA, 20147, USA); Teng-Leong Chew (Advanced Imaging Center, Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA, 20147, USA); Krzysztof Banecki (Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland); Dariusz Plewczyński (Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland; Centre of New Technologies, University of Warsaw); Yijun Ruan (The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, 06030, USA)

Recent studies in 3D genomics inferred the very basic functional chromatin folding structures known as chromatin loops. To visualize the looping structure of chromatin, we applied interferometric photoactivated localization microscopy (iPALM) to image a specific chromatin loop.To reconstruct the chromatin structures from captured images, we modeled them using traveling salesman problem solver. We then compared the physical distances in image models with contact frequencies generated by ChIA-PET and Hi-C examine the concordance.

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

Presenting author: Abhishek Agarwal (Centrum Nowych Technologii Uniwersytetu Warszawskiego)

Co-author(s): Abhishek Agarwal (Center of New Technologies, University of Warsaw, Poland), Dariusz Plewczyński (Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland)

The multiscale hierarchical spatial structure of the mammalian genome is defined by the chromatin loops, Topological Associated Domains, compartments, & chromosomal territories. The chromatin looping is observed when CTCF protein participates in the loop extrusion process driven by the ring-like cohesin molecular motor. The cohesin-mediated chromatin interactions vary among cell types and conditions. Such spatial variability correlates with the differences in gene expression between those cellular states and contributes to the microscale transcription and DNA replication processes.

The dynamic role of Cohesin in the maintaining the genome architecture

Presenting author: Ashutosh Choudhury (University of Warsaw)

Co-author(s): Abhishek Agarwal (University of Warsaw); Sebastian Korsak (Warsaw University of Technology); Karolina Jodkowska (University of Warsaw); Dariusz Plewczynski (Warsaw University of Technology, University of Warsaw)

It is fascinating to know that a two-metre-long human DNA is packed into a micrometre-sized nucleus. The cohesin complex primarily contributes to the formation of loops anchored by CTCF and mediated by other genome organisers which fold the genome in a hierarchal order. Cohesin-driven higher-order chromatin structures are dynamic and explained by the loop extrusion model (LEM). Cohesin also plays a role in gene expression and regulation. This talk explains our current understanding of the dynamic nature of the Cohesin-DNA complex and its dependence of genome maintenance.

List of speakers

1. Agarwal, Abhishek
2. Bandyopadhyay, Sanghamitra
3. Banecki, Krzysztof
4. Bokota, Grzegorz
5. Chiliński, Mateusz
6. Choudhury, Ashutosh
7. Denkiewicz, Michał
8. Gadakh, Sachin
9. Gurova, Anastasiya
10. Halder, Anup Kumar
11. Józefowicz, Jakub
12. Korsak, Sevastianos
13. Łaźniewski, Michał
14. Maulik, Ujjwal
15. Sengupta, Kaustav
16. Szczepińska, Teresa
17. Tojek, Zofia


Genomics Platform@WUT