PROPOSED TOPICS FOR MASTER THESIS
Program: Systems biology
Admission year: 2021
Published: October, 2021

Topic with description	Contact person	Contact email
TITLE: DNA properties of computationally mapped nucleosome positions ABSTRACT. Nucleosome positioning DNA sequence patterns (NPS) - usually distributions of particular dinucleotides or other sequence elements in nucleosomal DNA - at least partially determine chromatin structure and arrangements of nucleosomes that in turn affect gene expression. Statistically, NPS are defined as oscillations of the dinucleotide periodicity with about 10 base pairs (bp) which reflects the double helix period. Recently few distinctive patterns in nucleosomal sequences were observed that can be termed as packing and regulatory referring to distinctive modes of chromatin function [1]. Our working hypothesis for the future studies is that packing patterns tend to be preferred by evolutionary lower organisms and regulatory by higher organisms. Given vast amount of publicly available nucleosome maps in various organisms [2,3] it is possible investigate computational maps of nucleosomes at the regulatory elements of the genomes and characterize sequence dependant rigidity and plasticity of DNA at these locations. REFERENCES 1.Pranckeviciene E, Hosid S, Liang N, Ioshikhes I (2020) Nucleosome positioning sequence patterns as packing or regulatory. PLoS Comput Biol 16(1): e1007365. https://doi.org/10.1371/journal.pcbi.1007365 2. Teif V.B. (2016). Nucleosome positioning: resources and tools online. Briefings in Bioinformatics 17, 745-757. [https://generegulation.org/nucleosome-positioning-database/] 3. Yongbing Zhao et al., NucMap: a database of genome-wide nucleosome positioning map across species, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, [https://bigd.big.ac.cn/nucmap/]	Erinija Pranckevičienė	erinija.pranckeviciene@mf.vu.lt
TITLE: Comprehensive Characterization of Human Genome Variation by High Coverage Whole-Genome Sequencing of Lithuanians.	Alina Urnikytė	alina.urnikyte@mf.vu.lt
TITLE: Predicting drug treatment response in alcohol use disorder Inter-individual differences between patients with alcohol use disorder (AUD) underlie their different response to drug treatment and are the likely cause for poor treatment efficacy in some of them. Low sleep quality is one predictor of increased alcohol consumption in humans. Melatonin is a hormone approved for human use to improve sleep quality and have been shown to reduce relapse-like alcohol drinking in rats. Our objective is to identify, if impaired circadian activity of the rat would predict effectiveness of melatonin in reducing voluntary alcohol consumption.	Valentina Vengelienė	valentina.vengeliene@gf.vu.lt

TITLE: Molecular shape similarity in virtual screening for drug discovery ABSTRACT. Virtual Screening can drastically accelerate drug discovery processes. Molecular shape similarity is essential in virtual screening for drug discovery. Shape similarity is used to compare in detail the shape of a query molecule against a large database of potential drug compounds. In order to evaluate shape similarity accurately the molecules should be optimally adjusted. In this work, optimization problems for molecular shape similarity are investigated aiming at fast solution and acceptable accuracy.	Julius Žilinskas	julius.zilinskas@mif.vu.lt
TITLE: Applying ANN and machine learning for crystal property prediction ABSTRACT: Modern methods of machine learning allow to detect latent features and regularities in large amounts of data which were formerly inaccessible for automated analysis. The Crystallography Open Database (COD, https://www.crystallography.net/), the world's largest collection of open access crystal structure data for small molecules. This data set open excellent opportunity to train various machine learning tools (ANNs, SVMs, Bayesian classifiers) to predict various properties of materials, starting from unit cell volume, melting point, dielectric properties from material structures. In this work, several different machine learning methods should be quantitatively investigated for their suitability for the crystal property prediction tasks. The trained networks should then be used to validate the COD data base and incoming structures to detect possible inaccuracies or errors in data.	Saulius Gražulis	saulius.grazulis@bti.vu.lt
TITLE: Computing chemical structures in P1 cell for the Crystallography Open Database ABSTRACT: In crystallographic structure descriptions, a smallest unique structure fragment – an asymmetric unit – is usually refined and presented; such data set fully describes a crystal structure together with unit cell parameters and crystal symmetry information. The crystallographic description however is not always convenient for chemiinformatic investigations, where researchers are concerned with molecules, their connectivity, proximity and geometry. Thus a full molecular description must be always restored from crystallographic description before further chemoinformatic computations can be carried out. Repeating such computations every time is time and energy consuming, can introduce additional errors and makes a barrier of entry into chemiinformatics research higher for chemists who were not trained in crystallographic computations. In this work we propose to compute a database of full molecule descriptions together with data provenance from the Crystallography Open Database (COD, https://www.crystallography.net/), and build a novel open database of chemical structures using models of P1 crystal cells. In this database, investigation of optical isomers (their frequency of occurrence, presence in co-crystal, conformational variability) should be then carried out.	Saulius Gražulis	saulius.grazulis@bti.vu.lt
TITLE: Non-invasive intracranial blood volume pulse wave monitoring and signal analysis Abstract: An ultrasonic methods often are used to non-invasively measure intracranial blood volume (IBV) pulse waveforms. This technology has shown a strong association between invasively recorded ICP pulse waves and non-invasively recorded IBV pulse waves. The objective of the master thesis was	Linas Petkevičius	linas.petkevicius@mif.vu.lt

to investigate the signal analysis methods and functional parametrizations. Later using deep learning techniques create the IBV estimation models and apply for patients risks estimation. This research will require the technical knowledge of Python and Pytorch frameworks, as well as wide range medical data analysis by creating new mathematical models.		
TITLE: Deep learning architecture comparison for the detection of de novo mutations in next generation sequencing data. Abstract: De novo mutations (DNMs) are an important factor for determining the cause of the disease. Unfortunately, identifying reliable DNMs remains a major challenge due to sequence errors, uneven coverage, and mapping artifacts. We developed a deep convolutional neural network (CNN) DNM caller (DeNovoCNN), that encodes alignment of sequence reads for a trio as 160×164 resolution images. The aim during the master thesis would be evaluating additional neural network architectures such as transformer models for the detection of DNMs and their comparison to DeNovoCNN. This project requires knowledge of Python and Pytorch or Keras frameworks. References: DeNovoCNN: A deep learning approach to de novo variant calling in next generation sequencing data. G. Khazeeva, K. Sablauskas, et al. https://www.biorxiv.org/content/10.1101/2021.09.20.461072v1	Karolis Šablauskas	karolis.sablauskas@santa.lt
TITLE: Investigation of rare disorders via analysis of target RNA and transcriptome sequencing. Abstract: Rare genetic disorders and congenital anomalies are significant causes of chronic illness. More than 7000 rare disorders are known to date, about 80 % of them have genetic etiology. The advent of genome wide molecular technologies and application of whole exome/genome sequencing in clinical practice significantly accelerated the identification of genetic causes of rare hereditary conditions. However, massively parallel sequencing-based tests generate large amounts of data, thereby presenting a challenge to determine the single variant responsible for the disease. The purpose of this work is to proceed the research of rare genetic disorders beyond DNA studies to get a deeper understanding of the functional impact of DNA variants of interest on the mRNA level and provide insights into biological processes underlying the disease.	Eglė Preikšaitienė	egle.preiksaitiene@mf.vu.lt