**PROPOSED TOPICS FOR MASTER THESIS**

Program: Systems biology

Admission year: 2020

Published: October, 2020

|  |  |  |
| --- | --- | --- |
| **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: Estimation of effective population size from distributions of coalescent times**  The effective population size (*Ne*) is one of the most important natural population parameters providing an insight into the demographic history and dynamics of modern human populations through time .  Estimated effective population size as a function of time has many applications, for instance, better understanding of major ecological or historical events’ impact on humans such as glacial periods (Lahr and Foley 2001; Palkopoulou et al. 2013), agricultural shifts or technological advances (Boserup 1981). Knowledge about the demographic history is also important for studies of natural selection in order to avoid spurious finds (Li et al. 2012; Nielsen 2005).  Appropriate mathematical models have been developed for different genetic models to estimate *Ne* from genetic marker data. More recently, methods based on the sequentially Markovian coalescent (McVean and Cardin 2005; Marjoram and Wall 2006), such as PSMC (Li and Durbin 2011), MSMC (Schiffels and Durbin 2013), DiCal (Sheehan et al. 2013) and Bayesian approaches (Palacios et al. 2015) have advanced our ability to infer past population sizes.  **The main objective of this research proposal** is to infer the coalescent effective population size from genome-wide SNP genotyping data. | 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. The aim of this project is to establish statistical and computational methods for identification of behavioural biomarkers associated with different drinking styles that could be used to predict response to drug treatment in rats and may have a translational value. Data of detailed drinking-related behaviour (drinking frequency and speed, size and amplitude of drinking-bout, interbout interval, daily drinking fluctuations) and specific behavioural differences between rats (depression-like features, locomotor and circadian activity) will be used. Melatonin and lamotrigine are approved for human use to improve sleep quality and mood, respectively, and have been shown to reduce relapse-like alcohol drinking in rats. Neither drug is approved to treat AUD. However, sleep disturbances and low mood are among known predictors of increased alcohol consumption in humans. Our objective is to identify, if certain behavioural characteristics of the rat (depression-like behaviour, impaired circadian activity) would predict effectiveness of melatonin and lamotrigine in reducing alcohol consumption. We will also explore correlates between drinking style and efficacy of these drugs in reducing alcohol intake in rats. | 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: Application of Deep Neural Networks for Eye Fundus Image Analysis**  ABSTRACT: Aim of the study is to develop workflow for the eye fundus images analysis with the view to detect various anatomical changes. Early detection of the changes can indicate and even prevent such diseases as diabetes, glaucoma, cataract and etc. The images of known structure changes will be available as well as the consultations with the skilled ophthalmologist. Proposed solution has to be built on deep neural network application and tested. Students selecting the theme will face the Python as a primary programming language that will utilize the Keras and Tensorflow frameworks as those mostly used tools for deep neural network application with the strong community support. The student should have strong motivation to learn new skills all other knowledge will be acquired during the master study. | Jolita Bernatavičienė | jolita.bernataviciene@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 |