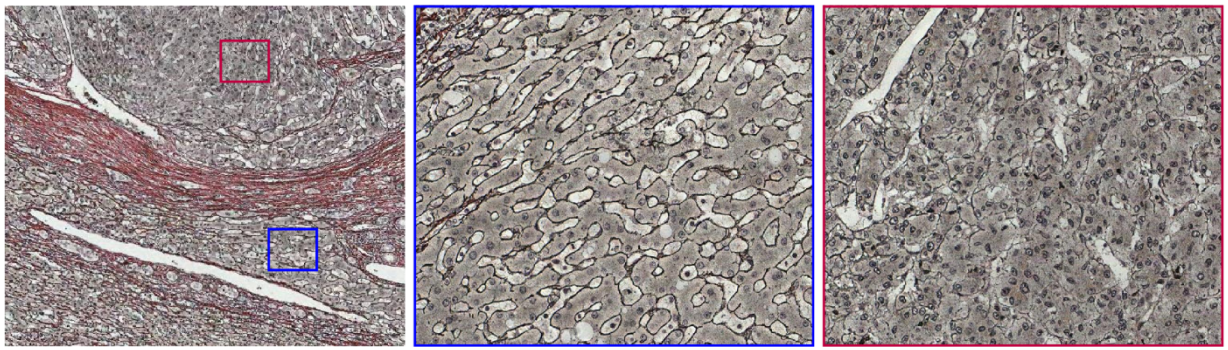
PROPOSED TOPICS FOR MASTER THESIS
Program: Systems biology
Admission year: 2022
Published: October, 2022

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: 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,	Julius Žilinskas	julius.zilinskas@mif.vu.lt

optimization problems for molecular shape similarity are investigated aiming at fast solution and acceptable accuracy.		
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: 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: Predictor of protein complex stability changes due to multiple mutations Abstract: Ability to accurately predict impact of multiple mutations upon a stability of a protein complex is nowadays very important as it can help to reveal which evolving COVID-19 lineages can have higher transmissibility or immune escape potential. There exist quite a lot of methods predicting impact due to single point mutations; however, there are few methods to predict effects of multiple mutation. The aim of master studies would be: i) create and compare performance of predictive models based on several approaches (linear regression, random forest, single layer perceptron) on ability to predict absolute free biding energies using features calculated based on static structures, ii) evaluate ability to exploit confrontational sampling using reduced presentation molecular dynamics calculations.	Gediminas Alzbutas	gediminas.alzbutas@mif.vu.lt
TITLE: Genetic relationship of Lithuanian and Indian populations using genomic data. Abstract: Cultural ties between India and Lithuania are believed to be rooted in ancient times. There is linguistic similarity between Lithuanian and Sanskrit. Have steps been taken to further deepen cultural links between India and Lithuania? To answer this question, the genetic data of Lithuanian and Indian populations will be used. Methods: kinship and inbreeding coefficient, genetic distances, principal component analysis, admixture.	Alina Urnikytė	alina.urnikyte@mf.vu.lt
TITLE: Evidence-based hexagonal grid parameter optimization procedure for various tissue pathology samples and analytics	Prof. A. Laurinavičius	arvydas.laurinavicius@vpc.lt
TITLE: Stand-alone tissue classifier based on enhanced image data sets (mission - exceed commercial HALO AI tool that we have)	Prof. A. Laurinavičius	arvydas.laurinavicius@vpc.lt

TITLE: Assessment of malignant and non-malignant micro-architecture in liver The human liver is an amazing organ as it is the only organ able to regrow itself, but unfortunately, the regeneration also leads to some inherent problems. Long-standing inflammatory liver damage, often caused by poor lifestyle choices, puts the liver in a state of constant regeneration. Over time the inflammatory state causes the liver tissue to separate into self-generating nodules surrounded by fibrous scar tissue, see panel a. Additionally, the self-generation process inside each nodule will eventually lead to uncontrolled growth, i.e. cancer, see panels b and c. Today we still haven't discovered any clear, visual markers of when/where this transition to a cancerous stage happens. This makes it is very difficult and time-consuming to determine if, and to what degree, liver patients have cancer in addition to any inflammatory diseases. The aim of this thesis is to use digital image analysis to detect fibrous tissue of different sizes/scales (between the nodules and the micro-structural fibers within the nodules) for detection of cancerous vs. non-cancerous nodules and/or for extracting quantities of fibrous tissue for investigating clinical impact. The thesis open-ended with regards to both overall goal and choice of method. This will eventually depend on the student's literature review while ensuring that the work fits within thesis time-frame and student effort and abilities. Some suggestions: • focus on accuracy of multi-scale fibers extraction and fiber classification and/or the subsequent analysis of possible clinical use. • choose deep learning and/or more classical image analysis methods and, of course, choose optimal/novel methods for subsequent clinical analysis.  A B C ***Figure 1**: Liver tissue: **A**: Fibrous scar tissue between nodules colored in red. The rectangles indicate by color regions with non-malignant and malignant fiber distributions shown in B and C. **B**: Fibers stained in black show homogeneous distribution of non-malignant tissue. **C**: Black fibers almost randomly distributed in malignant regions.*	Prof. A. Rasmusson	allan.rasmusson@vpc.lt