PROPOSED TOPICS FOR MASTER THESIS
Program: Systems biology
Admission year: 2023
Published: October, 2023

Topic with description	Contact person	Contact email
TITLE: Rare variant analysis of essential tremor-associated genes in general Lithuanian population Essential tremor (ET) is a common neurological disorder characterized by involuntary upper limb rhythmic movements. ET has a strong genetic basis that may develop via the additive contribution of risk variants of varying frequencies. About 20% of ET liability can be explained by common variants which are incorporated in polygenic risk scores (PRS) that quantify individual risk level (Liao et al. 2022). Several factors are associated with an altered risk of incident ET, such as environmental risk factors (e.g., smoking, exposure to pesticides). However, there is a lack of empirical data on whether these factors can identify a large group of persons at high risk for the disease from the general population. The hypothesis is that a genetic risk score based on currently identified risk loci would be a risk factor for incident ET in the general population, and that the genetic risk score would improve prediction of ET. Samples: whole genome sequencing data of **50** samples from individuals who reported a minimum of three generations of Lithuanian nationality. Average age of participants was 30 years. Genotyping data (700K SNPs) of **425** samples from individuals who reported a minimum of three generations of Lithuanian nationality. The average age of the participants was 53 years. Data on smoking, alcohol consumptions and other factors were collected.. Analysis: construct a genetic risk score for each individual, by adding up their number of risk alleles weighted by the log-transformed, reported risk-increasing or risk-decreasing effect size for the association with ET. Asociation between the genetic risk score and age, sex and smoking.	Alina Urnikytė	alina.urnikyte@mf.vu.lt
TITLE: Custom-built operant conditioning chamber for testing rat behaviour The term operant conditioning was coined by B.F. Skinner in the 1930s, and it means that behavior of all animals is controlled by its consequences (e.g., rewards or punishments). The most common apparatus used in conducting an operant conditioning study in rodents is the so-called ‘operant box’ (or ‘skinner box’). These chambers consist of a box, equipped with one or more manipulandi, transmitting the operant response and one or several devices	Valentina Vengelienė	valentina.vengeliene@gf.vu.lt

delivering the reinforcers (e.g., sweet rewards). These boxes are commercially available and their configurations can be adapted to specific needs including additional modules, which allow the programmed occurrence of other events (lights, tones, etc.) as discriminative stimuli and/or secondary reinforcers. Considering the manipulandi, these are usually levers, although systems based on more 'natural' responses (i.e., nose-poking for rats) have also been developed. Usually there is a manipulandum defined as 'active', meaning that it is linked to a reinforcer delivery, and another one referred to as 'inactive' (control), which lacks any programmed consequences. Our aim is to build (and validate) an operant conditioning chamber including several operant schedules - single versus complex, ratio versus interval, fixed (or progressive) versus variable. The most common program of reinforcement is the fixed ratio (FR, the reinforcer is delivered every time that a pre-selected number of responses are completed), followed by the progressive ratio (PR, the required ratio increases following a predefined progression, which usually is an arithmetic one) schedule. These schedules test rat's cognitive abilities and motivation to "work" for a reward. Functional operant chamber requires writing a program for different operant schedules and for collection and generation of a cumulative record after the data has been collected.		
TITLE: Elucidation of cell cycle transition during cardiac differentiation of human iPS cells using the genetically encoded FUCCI reporters and Incucyte imaging system Our lab has created human iPS cell lines constitutively expressing the genetically encoded cell cycle reporter FUCCI. Furthermore, our lab has established methods to differentiate human iPS cells into cardiomyocytes. In addition, our lab has an Incucyte system for live cell imaging and continuous monitoring of cell systems for up to several weeks. The project will implement image analyses algorithms through gaussian subtraction analyses or modern ML image analyses. The project will aim to elucidate the cell cycle distribution using FUCCI sensor during cardiac developmental stages cross-referencing it with key surface markers labelled with antibodies. This project has the possibility to be very collaborative with team members who developed the cardiac differentiation and genetically encoded reporters.	Jonathan Lee Arias Fuenzalida	jonathan.arias@gmc.vu.lt
TITLE: Cross-referencing of calcium oscillation and voltage changes on human iPS cell derived neurons and cardiomyocytes expressing the genetically encoded reporters GCaMP6 and ArcLight through Incucyte imaging system Our lab has created human iPS cell lines constitutively expressing the genetically encoded calcium reporter GCaMP6 and voltage reporter Marina. In addition, our lab has an Incucyte system for live cell imaging and continuous monitoring of cell systems for up to several weeks. The project will implement image analyses algorithms through gaussian subtraction analyses. F/F0 analyses and or modern ML image analyses. The project will aim to elucidate	Jonathan Lee Arias Fuenzalida	jonathan.arias@gmc.vu.lt

the relationship between calcium transients and voltage changes on human relevant cell systems. This project has the possibility to be very collaborative with team members who developed the differentiation systems and genetically encoded reporters.		
TITLE: Quantum computing for modelling biological structures The typical search of the conformational space of for example proteins is a NP-hard problem. The quantum computing algorithms in some problems can reduce the computational complexity. In this topic we investigate modelling structures and having defined conformational spaces then using quantum algorithms for search/predictions in those spaces. The quantum search and quantum machine learning algorithms will be investigated.	Linas Petkevičius	linas.petkevicius@mif.vu.lt
At the National Center of Pathology (VPC) pathologists inspect microscopy images of tissue samples to quantify any information that will aid with diagnosis, therapy decisions and outcome prediction for patient with a wide range of diseases. A lot of research is done on these images to further our understanding of the diseases and to extract any information that may aid the pathologists, clinicians, and patients. This research covers everything from lab-based development of new tissue preparation and staining methods to applying the latest statistical, AI and image analysis/computer vision methods. As a master's student at VPC we will make you part of our daily research team and let you work with state-of-the art research that uses statistical analyses, AI/deep learning methodologies (both evaluation and implementation) in a medical image setting. As such, your thesis work will give you a solid foundation from which you can undertake more complex medical imaging research tasks. Instead of assigning a fixed task, we prefer to decide the specific topic as you develop your own understanding and ideas during the literature review. We are open to any ideas and know-how you may already have; otherwise, we offer the topics below as a starting point. **TITLE: Kidney** Several research groups around the world are currently aiming to develop tools that allow for constructing a computational model of the human kidney. The aim is to extract all the kidney microarchitectural compartments into a comprehensive model for inference of properties about kidney diseases and allograft pathology. There are several areas of research of interest here, the most natural ones are segmentation of kidney tissue regions, feature extraction and classification into sub-types. Also, as it's not possible to extract all information about the components of the kidney from a single image, information must be combined from multiple images. A particular "hot" research area regards transfer of tissue stains between different images using AI/deep learning methods. The following references may serve as a starting point:	Allan Rasmusson	allan.rasmusson@vpc.lt

• Holscher, D. L., et al. (2023). "Next-Generation Morphometry for pathomics-data mining in histopathology." Nat Commun 14 (1): 470. https://www.nature.com/articles/s41467-023-36173-0, https://www.nature.com/articles/s41467-023-36173-0 • Jayapandian, C. P., et al. (2021). "Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains." Kidney Int 99(1): 86-101. https://www.sciencedirect.com/science/article/pii/S0085253820309625 • Cazzaniga, G., et al. (2023). "Time for a full digital approach in nephropathology: a systematic review of current artificial intelligence applications and future directions.", J Nephrol. https://link.springer.com/article/10.1007/s40620-023-01775-w		
At the National Center of Pathology (VPC) pathologists inspect microscopy images of tissue samples to quantify any information that will aid with diagnosis, therapy decisions and outcome prediction for patient with a wide range of diseases. A lot of research is done on these images to further our understanding of the diseases and to extract any information that may aid the pathologists, clinicians, and patients. This research covers everything from lab-based development of new tissue preparation and staining methods to applying the latest statistical, AI and image analysis/computer vision methods. As a master's student at VPC we will make you part of our daily research team and let you work with state-of-the art research that uses statistical analyses, AI/deep learning methodologies (both evaluation and implementation) in a medical image setting. As such, your thesis work will give you a solid foundation from which you can undertake more complex medical imaging research tasks. Instead of assigning a fixed task, we prefer to decide the specific topic as you develop your own understanding and ideas during the literature review. We are open to any ideas and know-how you may already have; otherwise, we offer the topics below as a starting point. **TITLE: Tumor Heterogeneity in Breast Cancer (BC)** The heterogeneity of tumors has proven to play a significant role in the development of cancers and thus it is extremely relevant in the process of diagnosis, therapy decision and outcome prediction, not only in BC. A substantial dataset from breast cancer is available and ready for investigating the properties and relevance of the most common heterogeneity measures obtained by a grid-based sub-sampling approach. Depending on your findings in the literature review, you may also choose to look into novel ways of defining and quantifying BC tumor heterogeneity, for instance by developing AI/deep learning methods. • Tumor Heterogeneity in Breast Cancer - 2017 Review: https://www.frontiersin.org/articles/10.3389/fmed.2017.00227/full • Breast cancer intratumour heterogeneity: current status and clinical implications - 2018: https://onlinelibrary.wiley.com/doi/10.1111/his.13642	Allan Rasmusson	allan.rasmusson@vpc.lt

• Tumour Heterogeneity of Breast Cancer: From Morphology to Personalised Medicine - 2018: https://karger.com/pat/article/85/1-2/23/266538/Tumour-Heterogeneity-of-Breast-Cancer-From • Breast cancer global tumor biomarkers: a quality assurance study of intratumoral heterogeneity – 2019 - https://www.sciencedirect.com/science/article/pii/S0893395222010882?via%3Dihub		
TITLE: Elucidating the expression profile of circulating micro-RNAs in cardiovascular disease patients Cardiovascular diseases (CVDs), including coronary artery disease (CAD) and myocardial infarction (MI), represent the leading causes of both global mortality and morbidity. In this study, we embark on elucidating the expression profiles of circulating microRNAs in the blood plasma of CVD patients, employing high-throughput sequencing as a non-invasive method for disease monitoring. After adhering to standard data preprocessing and quality control procedures, we will proceed with read alignment to the reference genome, followed by microRNA identification and quantification. When applicable, we will conduct differential expression detection, target gene prediction, and functional enrichment analysis to comprehend the molecular pathways and processes associated with differentially expressed microRNAs and their predicted target genes. Furthermore, our investigation will encompass the analysis of correlations between microRNA expression and various CVD risk factors, encompassing age, sex, weight, body mass index (BMI), smoking status, the presence of atrial fibrillation, arterial hypertension, hypertension grade, congestive heart failure, previous percutaneous coronary intervention (PCI), prior myocardial infarction, and laboratory test values, such as total cholesterol, LDL, HDL, triglycerides, creatinine, ALT, AST, glucose, HbA1c, and echo-EF. By doing so, we aim to uncover the relationships between microRNA expression and the onset and progression of CVD non-invasively, providing new insights into its diagnosis, prognosis, and potentially also treatment. (*Researchers involved: MD Knoka E., Dr. rer. nat. Vilne B., Dr. med. Gailīte L., assoc. prof. Trušinskis K.*)	Baiba Vilne	baiba.vilne@rsu.lv
TITLE: Enhancing the CIBERSORT analysis: creating a comprehensive mononuclear phagocyte signature matrix for tumour microenvironment assessment Tumor-infiltrating leukocytes (TILs) constitute a crucial component of the tumour microenvironment, and their presence has been linked to prognosis and treatment response. The CIBERSORT analytical tool, developed by Newman et al., offers a means to estimate the abundance of various leukocyte cell types within mixed cell populations through gene expression data. Initially, this computational method was validated using a signature matrix	Baiba Vilne	baiba.vilne@rsu.lv

comprising 22 functionally defined human immune subsets (LM22). However, there is room for improvement and expansion of this matrix by generating additional, more refined leukocyte signatures. These enhancements can provide deeper insights into the functional significance of these cells within the tumour microenvironment. Therefore, the primary objective of this project is to construct a novel mononuclear phagocyte signature matrix. This matrix will be employed to explore correlations between defined cell gene signatures and critical clinical outcomes such as overall survival, treatment response, and effector lymphocyte populations. Utilising freely available data, such as the TCGA database, this research aims to enhance our understanding of the intricate interplay between immune cell populations and their impact on cancer progression and treatment outcomes. (*Researchers involved: prof. Dace Pjanova, Dr. rer. nat. Vilne B., Vaivode K. PhD*)		
TITLE: Optimizing the survival of motor neuron 1 (SMN1) deletion detection from exomes Homozygous deletion of the survival of motor neuron 1 (SMN1) gene is responsible for a group of rare and severe monogenic progressive neuromuscular disorders known as autosomal-recessive proximal spinal muscular atrophies (SMA). Identifying SMN1 deletions from short-read next-generation sequencing (NGS) data has been challenging so far due to the high level of homology with the survival of motor neuron 2 (SMN2) gene. As a result, this testing method is not routinely employed in diagnostics. In recent years, novel bioinformatic tools and concepts have emerged, enabling the detection of SMN1 deletions from whole-genome sequencing or targeted short-read NGS data. However, there is limited information available regarding the applicability of these approaches to exome data, as well as their sensitivity and specificity. This study aims to evaluate several existing tools and potentially develop novel ones for the detection of SMN1 deletions from exome data, followed by the validation of their clinical utility. (*Researchers involved: MD Rots D., Dr. rer. nat. Vilne B., Dr. med. Gailīte L.*)	Baiba Vilne	baiba.vilne@rsu.lv
TITLE: Integrated Systems Analysis for DNA Microarray Quality Assurance and Troubleshooting Ensuring the robustness, accuracy, and reproducibility of complex biochemical testing platforms - such as DNA microarrays - is of paramount importance. This thesis topic aims to employ a holistic approach that merges systems biology with quality assurance and troubleshooting methodologies to develop a system that elucidates DNA microarray testing component interactions, potential failure points, and their ramifications. By integrating the profiling of genomic sample responses with assay component mapping and troubleshooting, a comprehensive framework can be established to detect, understand and address potential inconsistencies in DNA microarray quality tests. Rapid troubleshooting, guided by an	Julija Baginskaitė	julija.baginskaite@thermofisher.com

integrated model, will reduce downtime and enhance the reliability of quality testing outcomes.		
TITLE: Single-potential ECG signal analysis to identify bursts and classes The standard arrhythmia classes are used in the study according to the Association for the Advancement of Medical Instruments (AAMI) recommendations for arrhythmia classification systems. The literature mainly distinguishes them based on the characteristic points of QRS parameters, RR intervals, morphological and statistical properties, and dimension reduction methods. The research will examine machine learning methods for recognizing classes of arrhythmias based on the MIT BIH database. 90% of the heart murmurs in the MIT-BIH database belong to class N, 3% to SVEB, 6% to VEB, and 1% to F. There is no data for class Q. The uneven distribution of bursts and noisy ECG data will introduce scientific uncertainties in the automatic recognition and classification of bursts. This paper will investigate the automatic classification of the three largest classes (N, SVEB and VEB). An interpatient division scheme will be applied, i.e., each patient's heartbeats can receive only one of the training or testing sets. The division scheme is considered an accurate way to assess the quality of classification methods for classifying heart arrhythmia objectively. The study will also evaluate the size of the analysis window of the ECG signal and its effect on the classification accuracy, and the obtained results will be validated using independently collected data.	Povilas Treigys	Povilas Treigys
TITLE: Classification of medical signals using near-infrared spectroscopy (NIRS) signals in small data samples Near-infrared spectroscopy (NIRS) is a noninvasive clinical tool that can provide continuous and real-time information regarding regional tissue oxygen saturation (rSpO2) and indirectly information on regional organ perfusion. It is based on the relative transparency of biological tissues (bone, skin, and soft tissue) and the ability to diferentiate oxygenated haemoglobin from deoxygenated haemoglobin, as they have distinct absorption spectra. The ratio of oxygenated and total haemoglobin represents the rSpO2. Numerous clinical studies have been conducted to determine normal NIRS values for preterm neonates and to assess the use of NIRS to evaluate the regional brain and kidney oxygenation in diferent pathologic conditions. Despite the growing clinical use of cerebral and peripheral NIRS monitoring in premature infants, its routine use and importance in determining patent ductus arteriosus signifcance is still controversial. The main aim of master thesis will be to develope machine learning models or other data analysis technique for classification of patent ductus arteriosus status in preterm neonates using NIRS spectroscopy signals. Literature: https://doi.org/10.1186/s12887-021-03036-w; https://doi.org/10.3390/medicina580404.	Tadas Žvirblis	tadas.zvirblis@mf.vu.lt