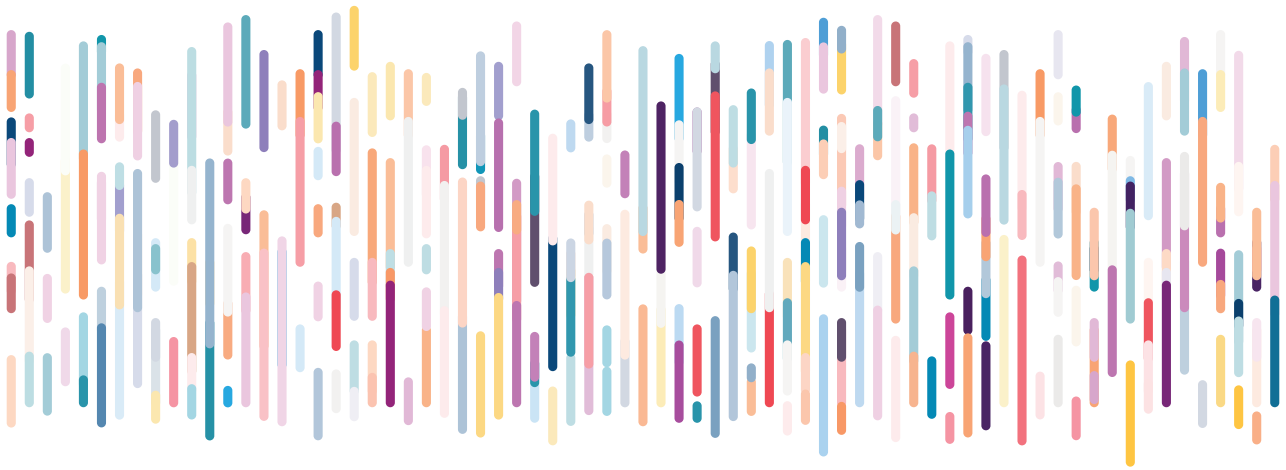
2024 Gene Forum

The 23rd Annual International

24–25 September 2024 Tartu, Estonia



UNIVERSITY OF TARTU
Institute of Genomics

Eesti Geenikeskus
Estonian Genome Foundation

Funded by
the European Union

POSTER OVERVIEW

No.	Title	Presenter
1	Comprehensive Benchmarking of Star Allele Calling Workflows for Pharmacogenomic Analyses	Sven van der Maas
2	The causes of economic outcomes: an examination using twins and molecular data	Alexandros Giannelis
3	Mendelian randomisation of personality and health behaviours	Kerli Ilves
4	Mapping brain-personality associations with genetic correlations reveals polylocalised brain maps	Uku Vainik
5	Assessing biases of using aDNA imputation for natural selection scans	Jose Rodrigo Flores Espinosa
6	Associations between genetic predisposition to psychiatric disorders and the severity of acute COVID-19 illness and other respiratory infections: A meta-analysis across five COVIDMENT cohorts	Kadri Kõiv
7	“Genetic investigation of self-reported cognitive symptoms in Mental Health online Survey cohort in the Estonian Biobank”	Triinu Varvas
8	Prevalence and impact of a protein-truncating POMC variant on obesity in the Estonian Biobank	Kanwal Batool
9	New insight into ovarian cancer (epi)genetic biomarkers	Ieva Vaicekauskaitė
10	Genetic and molecular mechanisms linking RSV infection to chronic lung diseases	Erik Abner
11	Shared Genetic Architecture Across Five Common Cardiovascular Diseases: Implications for Biology-Based Partitioned Polygenic Risk Scores	Tigist Demssew Adane
12	Genetic predisposition and antipsychotic treatment effect on metabolic syndrome in schizophrenia: a ten-year follow-up study using the Estonian Biobank	Maris Alver
13	Interpreting artificial neural networks to detect genome-wide association signals for complex diseases	Burak Yelmen
14	Autosomal dominant tibial muscular dystrophy in Estonia	Siiri Sarv
15	Accurate determination of clinically relevant structural variation in the CYP2D6 locus: from SNP microarrays to long-read sequencing	Maarja Jõeloo
16	R/Bioconductor data science framework for multi-omic data integration	Leo Lahti
17	Evaluating the Predictive Ability of Polygenic Risk Scores for Intrahepatic Cholestasis of Pregnancy in the Estonian Biobank	Fanny-Dhelia Pajuste
18	Effect of drug-metabolizing enzyme activity on survival in amyotrophic lateral sclerosis	Johanna Jaansoo
19	In vitro testing of genetic mutations of G protein coupled receptors	Tõnis Laasfeld
20	A history of repeated antibiotic usage leads to microbiota-dependent mucus defects.	Kertu Liis Krigul
21	Detecting embryo developmental potential by single blastomere RNA-seq	Monika Nõmm
22	Exploring drug-drug-gene interactions and the occurrence of adverse drug events based on electronic health records of Estonian Biobank participants	Laura Birgit Luitva
23	Genetic factors affecting the reporting of antidepressant side effects	Hanna Maria Kariis
24	The Predictors of Estonian Biobank Mental Health online Survey participation and the prevalence of mental health problems in the Estonian Biobank	Triinu Ojalo
25	Deciphering Suicide Risk in Major Depressive Disorder: A Polygenic Approach to Early Onset Cases	Siim Kurvits

Comprehensive Benchmarking of Star Allele Calling Workflows for Pharmacogenomic Analyses

Presenter: Sven van der Maas**,** University of Hasselt, Belgium
sven.vandermaas@uhasselt.be

Authors of the Research: Sven van der Maas, Pieter-Jan Volders, Gökhan Ertaylan

Short Summary of Work:

Adverse drug reactions (ADRs) are a burden to the healthcare system with a high societal cost and potentially detrimental effects for the individual patient. PGx-guided prescription aims to personalize drug therapy by tailoring the drug (dosage) to the individual's genetic makeup. This ameliorates the conventional "trial and error" approach of drug prescribing, thereby promising safer, more effective, and increasingly cost-effective drug treatment. Currently, there is no optimal workflow described for the identification of pharmacogenomic variants ("star alleles") from whole genome sequencing. Our research focuses on the benchmarking of different pharmacogenomics (PGx) workflows for star allele calling from whole genome sequencing data. This work is crucial for further enabling pharmacogenomic testing in clinical practice.

THE CAUSES OF ECONOMIC OUTCOMES: AN EXAMINATION USING TWINS AND MOLECULAR DATA

Presenter: Alexandros Giannelis, University of Minnesota, USA
giann071@umn.edu

Authors of the Research: Alexandros Giannelis, Robin Corley, Christian Hopfer, John K. Hewitt, Aldo Rustichini, Matt McGue, Scott Vrieze, James J. Lee

Short Summary of Work:

Cognitive and non-cognitive skills are known to correlate with socioeconomic outcomes. However, given the complexity of social interactions, it is uncertain whether these associations represent causal effects. We examine the effects of cognitive ability, educational attainment, and saving disposition, on economic outcomes. We model the effect of each predictor on income and financial distress, in a genetically informed sample of American twins. Exploiting the family structure of our sample, we control for familial (genetic or environmental) confounding. By employing polygenic scores as instrumental variables, combined with the twin design, we are able to provide more evidence for causal inference. We find that cognitive ability and education increase labor income, but do not decrease the risk of financial distress. Conversely, saving disposition, which is largely affected by the rearing environment, reduces financial distress.

MENDELIAN RANDOMISATION OF PERSONALITY AND HEALTH BEHAVIOURS

Presenter: Kerli Ilves, Institute of Genomics, University of Tartu, Estonia

Authors of the Research: Kerli Ilves, Uku Vainik

Short Summary of Work:

Health behaviour plays a critical role in the development and persistence of many chronic diseases (CVDs, metabolic diseases, mental health disorders etc). Personality affects our behaviour and lifestyle choices and many observational studies have demonstrated an association between the personality traits and different types of health behaviours. Knowing the causal mechanisms of personality and health behaviours would potentially enable to improve the health behaviour interventions as well as study the mechanisms of personality development. For example, mendelian randomisation (MR) studies have already confirmed the observational hypothesis that trait Neuroticism is causally linked to smoking frequency as well as trait Extraversion is causally linked to smoking initiation. Our objective is to study the causal effects between Big Five personality traits and a range of health behaviours utilising genetic instruments as instrumental variables to confirm and refine the observational findings. We will apply a triangulation of methods such as phenotypic regression, allele score-based and two-sample MR, and within-family MR All the hypotheses being tested are based on associations demonstrated between health behaviours and personality traits in observational studies.

Mapping brain-personality associations with genetic correlations reveals polylocalised brain maps

Presenter: Uku Vainik, Institute of Psychology & Institute of Genomics, University of Tartu; Montreal Neurological Institute, McGill University
uku.vainik@ut.ee

Authors of the Research: Uku Vainik, Vincent Bazinet, René Mõttus

Short Summary of Work: Mapping normal personality trait variation to brain systems has been theoretically attractive for personality neuroscience, but empirically largely unsuccessful. Here, we used genetic correlations to relate genome-wide association studies of 13 brain structure modalities (N~36K) to the GWASes of Neuroticism (N~380K, UKB) and Big Five (N~75K, Estonian Biobank, EstBB).

We found several univariate associations, mostly with cortical surface area. Namely, UKB Neuroticism related negatively to occipital and orbitofrontal cortex. Conscientiousness related negatively to insular and inferior frontal cortex, and Openness positively to insular cortex. Further, the genetic correlation brain maps replicated - the UKB Neuroticism brain map was reproduced for EstBB Neuroticism across surface area and 12 other brain modalities (mean brain map r~.56).

To interpret the personality brain maps, we used the Neuromaps toolbox that integrates dozens of brain maps from the literature. Namely, we related the association maps that personality had with brain surface area to 6 common brain system components. Unfortunately, no brain system had consistent associations with personality brain maps. These analyses will be repeated as more powerful personality GWASes will become available.

In summary, personality is highly polylocalised in the brain, just as complex traits are highly polylocalised in the genome. The brain patterns may be relatable to common brain systems, suggesting a potential way forward in understanding the brain roots of personality.

ASSESSING BIASES OF USING ADNA IMPUTATION FOR NATURAL SELECTION SCANS

Presenter: Rodrigo Flores, Institute of Genomics, University of Tartu
rodrigo.flores@ut.ee

Authors of the Research: Rodrigo Flores[1], Liisa Loog[2], Anders Eriksson[1]
[1]Center for Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu
[2]Estonian Genome Center, Institute of Genomics, University of Tartu

Short Summary of Work:

Recent evolutionary studies have showed us that human evolution is largely characterized by a series of long-range expansions, followed by contacts between groups of different ancestries end exposure to various environmental pressures[1]. This has provided the very ingredients for natural selection to act on the genome.

Current methodologies allow us to use these ancestry patterns as well as ancient DNA (aDNA) to shed light on adaptive evolution in populations around the world. However, the fragmented nature of aDNA data is limiting the power and resolution of adaptive evolution inference. Using imputation based on present-day genomes offers a promising way forward, but cases have been reported where imputation would lead to erroneous conclusions of natural selection[2] , and there is currently a lack of formal understanding of what biases imputation may introduce in evolutionary inference.

This project aims to fill this gap using systematic and realistic simulations of natural selection in human populations from the last 50,000 years to investigate such biases and their causes. Here we present the results of various degrees of selection over a period ranging from 10,000 to 5,000 years. We model patterns of missing variation common for aDNA and perform imputation using a reference panel assembled from the last (present-day) sampling point. Using scenarios of selection acting on both de
novo mutations and standing variation, we assess how imputation accuracy decays as a function of sampling age, strength of selection, minor allele frequency (MAF), degree of missingness and distance from the selection sites.

In the absence of selection, our initial results indicate a slow decrease in accuracy with increasing age of the sample that can largely be attributed to greater genetic distance between older samples and the reference panel. In contrast, imputation accuracy is

markedly decreased in populations sampled during and before selective sweeps, and this is further increased by stronger levels of selection.

From these initial results we can already pinpoint cases where accuracy after imputation is not suitable for correct evolutionary inference. We plan next to show this by applying selection scans to these examples and confirm whether the initial selection signals can be detected or not.

References

1. Korunes KL., Goldberg A. Human genetic admixture. 2021. PLOS Genetics
2. Mathieson, S., Mathieson, I. FADS1 and the timing of human adaptation to agriculture. 2018. MBE

Associations between genetic predisposition to psychiatric disorders and the severity of acute COVID-19 illness and other respiratory infections: A meta-analysis across five COVIDMENT cohorts

Presenter: Kadri Kõiv, Estonian Genome Centre, Institute of Genomics, University of Tartu

Authors of the Research: Kadri Kõiv, Ragna Bugge Askeland, Ingibjörg Magnúsdóttir, Daniel McCartney, Lea Arregui Nordahl Christoffersen, Elis Haan, Helga Ask, Ole Birger Vesterager Pedersen, Unnur Anna Valdimarsdóttir, Kelli Lehto on behalf of the COVIDMENT consortium

Short Summary of Work:

Background: Psychiatric disorders have been associated with severity of COVID-19 illness, often measured by hospitalizations and mortality. To what extent shared genetic factors contribute to this association is poorly understood, especially in non-hospitalized general populations with milder COVID-19 illness course. We aimed to investigate the associations between polygenic liability to psychiatric disorders and acute COVID-19 illness severity using data from five population-based cohorts of the COVIDMENT consortium across Denmark, Estonia, Iceland, Norway, and the UK.

Methods: The total sample comprised 77 086 subjects (64% female, 41% SARS-CoV-2-positive) responding to questionnaires from May 2020 to March 2023. Acute COVID-19 severity was defined by 1) self-reported number of days bedridden (never, 1-6 days, >7 days), and 2) symptom count. Polygenic risk scores (PRS) for nine psychiatric phenotypes were constructed using PRS-CS. Estimates from cohort-based stepwise adjusted regression analyses were meta-analysed using random effects model.

Results: In the fully adjusted models, extended time bedridden was associated with the PRS of five psychiatric traits: anxiety (relative risk ratio 1.23 [1.10-1.38]), neuroticism (1.20 [1.07-1.35]), depression (1.19 [1.05-1.34]), schizophrenia (1.15 [1.03-1.30]), and bipolar disorder (1.07 [1.02-1.12]). Depression and neuroticism replicated using symptom count as the outcome. We additionally ran the models in COVID-19 negative individuals reporting flu-like symptoms, and seven out of nine psychiatric PRSs were associated with symptom count [incident rate ratio range: 1.01-1.04].

Conclusion: These findings imply that individuals with a genetic predisposition to psychiatric conditions may be at risk of a more severe acute course of COVID-19 and other respiratory viral infections.

Genetic investigation of self-reported cognitive symptoms in Mental Health online Survey cohort in the Estonian Biobank

Presenter: Triinu Varvas, Institute of Genomics, University of Tartu, Estonia
triinu.varvas@ut.ee

Authors of the Research: Triinu Varvas, Kadri Kõiv, Kelli Lehto

Short Summary of Work:

Background: Cognitive deficits are one of the frequently present but often overlapping and therefore non-specific symptoms across many different neuropsychiatric disorders. Further research is needed to uncover the genetic architecture underlying specific cognitive problems. The aim of this study is to identify the genetic associations of self-reported cognitive symptoms, explore the genetic overlap between cognitive symptoms from different screening instruments and shared genetics with other neuropsychiatric traits and disorders.

Methods: The study uses Estonian Biobank (EstBB) Mental Health online Survey (MHoS) cohort (N = 86,244) questionnaire data on cognitive symptoms from the screening instruments of multiple different mental health problems (e.g. attention deficit disorder, post-traumatic stress disorder and depression). Main analyses used are factor analysis, genome-wide association studies (GWASs), post-GWAS analyses and polygenic risk score (PRS) association models.

Preliminary results: The ten cognitive items in the MHoS loaded into two larger factors called Mental sharpness problems (F1) and Concentration&Executive problems (F2). GWASs on the cognitive factors resulted in two risk loci for F1 (SNP-h2 = 0.06) and five risk loci for F2 (SNP-h2 = 0.10). Genetic correlations between subjective cognitive symptoms from different screening instruments were all significant (rg-s ranged from 0.43 to 1.03). Genetic correlations were the largest between cognitive factors and mood disorders and related traits (e.g., F1&miserableness-rg = 0.62; F2&anxiety for a month-rg = 0.44), while the rg-s with cognitive traits and disorders (e.g., education attainment and Alzheimer's disease) were smaller or non-significant. Mutually adjusted PRS association models with ten cognitive items and two factors as outcomes showed the largest and most consistent associations with neuroticism, major depression and subjective well-being (beta coefficients ranged from 0.02 to 0.06, 0.02 to 0.04, and -0.01 to -0.03, respectively).

Interim conclusions: These findings indicate that self-reported cognitive symptoms characteristic to different mental health problems have considerable genetic overlap and imply larger shared genetics with affective traits and disorders than with cognitive phenotypes.

Prevalence and impact of a protein-truncating POMC variant on obesity in the Estonian Biobank

Presenter: Kanwal Batool, Institute of Genomics, University of Tartu, Estonia

Authors of the Research: Erik Abner, Kanwal Batool, Nele Taba, Tiit Nikopensius, Kristi Läll, Anastasiia Alekseienko, Anders Erikssong, Joel Rämö, Hanna Maria Kariis, Liis Haljasmägi, Urmo Võsa, Taavi Tillmanna, Uku Vainika, Kelli Lehto, Kai Kisand, Estonian Biobank Research Team, Tõnu Esko

Short Summary of Work:

Population-specific genome-wide association studies can reveal high-impact genomic variants that influence traits like body-mass index (BMI). Using the Estonian Biobank BMI dataset (n=204,747 participants) we identified 214 genome-wide significant loci. Among those hits, we identified a common non-coding variant within the newly associated ADGRL3 gene (-0.18 kg/m^2; P = 3.06 × 10^{-9}). Moreover, the missense rare variant PTPRT:p.Arg1384His associated with lower BMI (-0.44 kg/m^2; P = 2.5 × 10^{-10}), while the protein-truncating variant POMC:p.Glu206* was associated with considerably higher BMI (+0.81 kg/m^2; P = 1.5 × 10-12), both likely affecting the functioning of the leptin-melanocortin pathway. POMC:p.Glu206* was observed in different North-European populations, suggesting a broader, yet elusive, distribution of this damaging variant. These observations indicate the novel roles of the ADGRL3 and PTPRT genes in body weight regulation and suggest an increased prevalence of the POMC:p.Glu206* variant in European populations, offering avenues for developing interventions in obesity management.

New insight into ovarian cancer (epi)genetic biomarkers

Presenter: Ieva Vaicekauskaitė, National Cancer Institute / Institute of Biosciences, Life Sciences Centre, Vilnius University, Vilnius, Lithuania, ieva.vaicekauskaite@gmc.stud.vu.lt

Authors of the Research: Ieva Vaicekauskaitė[1,2], Paulina Kazlauskaitė[1,2], Rugilė Gineikaitė[2], Rūta Čiurlienė[1], Juozas Rimantas Lazutka[2] and Rasa Sabaliauskaitė[1,2]

[1] National Cancer Institute, Vilnius, Lithuania;

[2]Institute of Biosciences, Life Sciences Centre, Vilnius University, Vilnius, Lithuania

Short Summary of Work:

Background/Objectives:

Ovarian cancer (OC) is the second deadliest gynecologic malignancy worldwide. The majority of fatalities are linked to high-grade serous ovarian cancer (HGSOC), an OC type most often diagnosed at an advanced stage (FIGO III-IV). At present, there are no dependable genetic or epigenetic biomarkers for OC.

Methods:

Tumor tissue from 66 patients with gynecologic cancers were analyzed (42 HGSOC, 15 other malignant gynecologic tumors, and 9 benign gynecologic tumors). 10 genes from Notch (*NOTCH1-4, HES1*, *DLL1*, and *JAG2*) and Wnt (*CTNNB1*, *FBXW7*) pathways, as well as chromatin remodeling complex SWI/SNF gene *ARID1A* expression, was analyzed by RT-qPCR, normalized to *GAPDH* expression. Promoter methylation of *ARID1A* and homeobox (HOX) related genes (*HOPX*, *ALX4*, *CDX2*) were evaluated via methylation specific PCR.

Results:

Significant changes in HGSOC gene expression compared to benign cases was detected in 8/10 genes analyzed ($p < 0.05$), while 7 Notch and Wnt pathway gene expression was also altered in HGSOC samples compared to other gynecologic tumors ($p < 0.05$). Promoter methylation was detected in 81,8% (54/66) of samples. *CDX2* promoter methylation was significantly more frequent in other gynecologic tumor group compared to benign cases ($p = 0.01$), and *HOPX* promoter methylation was more frequent in HGSOC cases compared to benign tumors ($p = 0.04$). *CTNNB1* expression proved the best single separator of HGSOC cases form the benign tumors (AUC = 0.97) out of 14 biomarkers analyzed.

Conclusion:

Gene expression and promoter methylation status combination could be useful as a diagnostic biomarker test for HGSOC, however, a more extensive analysis of these biomarkers in a larger cohort and non-invasive samples is needed to further validate the (epi)genetic HGSOC test model.

GENETIC AND MOLECULAR MECHANISMS LINKING RSV INFECTION TO CHRONIC LUNG DISEASES

Presenter: Erik Abner, Institute of Genomics, University of Tartu
erik.abner86@gmail.com

Authors of the Research: Erik Abner, Tarmo Annilo, CLARITY consortium

Short Summary of Work:

The CLARITY project investigates the causative role of Respiratory Syncytial Virus (RSV) in chronic lung diseases, particularly asthma. Asthma, a heterogeneous and chronic inflammatory airway disease, is influenced by both genetic and environmental factors. Early-life viral infections, particularly with RSV, are thought to be among significant non-genetic risk factors.

In the Horizon Europe project CLARITY, we will focus on identifying the genetic predispositions that increase susceptibility to RSV infection and the subsequent development of asthma. Leveraging data from the Estonian Biobank and other international cohorts, we aim to explore common genetic variants, such as those in the *GSDMB* gene, which is implicated in the regulation of inflammatory cell death and respiratory infections. By integrating genome-wide association studies (GWAS), polygenic risk scores, and functional validation, we seek to uncover how genetic variations and RSV strains modulate immune responses and trigger asthma development. The ultimate goal is to understand the underlying molecular mechanisms and to identify drug-like compounds that could reverse RSV-induced perturbations. The insights gained will support personalized prevention strategies and therapeutic development for virus-triggered asthma and potentially other chronic respiratory diseases.

SHARED GENETIC ARCHITECTURE ACROSS FIVE COMMON CARDIOVASCULAR DISEASES: IMPLICATIONS FOR BIOLOGY-BASED PARTITIONED POLYGENIC RISK SCORES

Presenter: T.D (Tigist) Adane, Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands

Authors of the Research: T.D (Tigist) Adane [1], J.B.J. van Meurs [1], A.G. Uitterlinden [1], J.G.J. van Rooij [1]

[1]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands

Short Summary of Work:

Background/Objectives: Cardiovascular disease (CVD), the leading cause of death in Europe, is an umbrella term for several complex CVDs involving gene-environment interactions with coronary artery disease (CAD), stroke, heart failure (HF), atrial fibrillation (AF), and peripheral arterial disease (PAD) as the most prevalent. They share common clinical risk factors, and overlapping etiologies have been recognized. Genome-wide association studies (GWAS) have identified hundreds of genetic risk factors for each complex disorder, which can be used to identify underlying biological mechanisms. Additionally, these genetic factors allow for disease risk prediction in the form of polygenic risk scores (PRS), where all identified genetic variants, with each small effect on disease risk, are taken together to compute the total genetic risk for CVDs. Our study aims to investigate the interconnection between these five common CVDs using their genetic underpinning and PRSes.

Methods: We used the summary statistics of the largest GWAS for CAD, Stroke, HF, PAD, and AF, primarily involving European or predominantly European populations and five common cardiovascular diseases: CAD, stroke, HF, PAD, and AF, defined based on the International Classification of Diseases criteria. Restricted PRS were calculated for each CVD within the genetic data of 484,472 participants of the UK biobank using genome-wide significant variants. All PRSs were standardized and investigated for association with their respective outcome (lifetime disease prevalence, age at onset). Additionally, we explored the association of each standardized PRS with other CVDs. We used logistic regression for both analyses.

Results: This study explored the association between five PRSes (CAD, stroke, HF, PAD, and AF) and each of their corresponding CVDs.We observed that all PRSs were significantly associated with the five CVDs, except for PAD PRS, which did not significantly associate with AF. Particularly, CAD PRS demonstrated the highest odds with CAD (OR = 1.50, 95% CI: 1.49-1.52) and PAD (OR = 1.32, 95% CI: 1.26-1.38). Stroke PRS, in comparison, had a stronger association with AF (OR = 1.13, 95% CI: 1.10-1.15)) than with stroke itself (OR = 1.05, 95% CI: 1.03-1.07). Meanwhile, HF PRS (OR = 1.15, 95% CI: 1.12-1.19), CAD PRS (OR

= 1.15, 95% CI: 1.12-1.19), and AF PRS(OR = 1.18, 95% CI: 1.14-1.21) demonstrated the same odds of HF.PAD PRS strongly associated With PAD (OR = 1.31, 95% CI: 1.25-1.37) followed by CAD (OR = 1.12, 95% CI: 1.11-1.13). Lastly, AF PRS is most significantly associated with AF, showing the highest OR (OR = 1.57, 95% CI: 1.53-1.61) followed by HF (OR = 1.18, 95% CI: 1.14-1.21).

Conclusion: Our results highlighted robust associations of specific PRS with their respective CVDs and were informative for other CVDs. Individuals with a high PRS for either one of five common CVDs may have an increased risk for multiple CVDs due to the underlying shared etiology. The observed variability in odds ratios, particularly with HF and stroke, may reflect heterogeneity in disease phenotypes and classification, emphasizing the complex genetic architecture. The findings suggest that PRS can be used for risk assessment, and the overlap between common CVDs should be considered for screening and preventive measures. Finally, partitioning the variants used in CVD PRS might allow the separation of disease-specific or disease-overlapping biological pathways, potentially allowing for more fine-grained prediction of disease, potential symptoms, comorbidities, and prognosis. However, more work is needed in this area.

Keywords: PRS, Shared genetics, Cardiovascular diseases, Risk assessment

Genetic predisposition and antipsychotic treatment effect on metabolic syndrome in schizophrenia: a ten-year follow-up study using the Estonian Biobank

Presenter: Maris Alver, Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia
maris.alver@ut.ee

Authors of the Research: Maris Alver, Silva Kasela, Liina Haring, Laura Birgit Luitva, Estonian Biobank Research Team, Health Informatics Research Team, Krista Fischer, Märt Möls, Lili Milani

Short Summary of Work:

Background: Schizophrenia (SCZ) patients exhibit 30% higher prevalence of metabolic syndrome (MetS) compared to the general population with its suboptimal management contributing to increased mortality. Large-scale studies providing real-world evidence of the underlying causes remain limited.

Methods: To address this gap, we used real-world health data from the Estonian Biobank, spanning a median follow-up of ten years, to investigate the impact of genetic predisposition and antipsychotic treatment on the development of MetS in SCZ patients. Specifically, we set out to characterize antipsychotic treatment patterns, genetic predisposition of MetS traits, MetS prognosis, and body mass index (BMI) trajectories, comparing SCZ cases (n=677) to age- and sex-matched controls (n=2,708). Findings: SCZ cases exhibited higher genetic predisposition to SCZ (OR=1·75, 95%CI 1·58–1·94), but lower polygenic burden for increased BMI (OR=0·88, 95%CI 0·88–0·96) and C-reactive protein (OR=0·88, 95%CI 0·81–0·97) compared to controls. While SCZ cases showed worse prognosis of MetS (HR 1·95, 95%CI 1·54–2·46), higher antipsychotic adherence within the first treatment year was associated with reduced long-term MetS incidence. Linear mixed modelling, incorporating multiple BMI timepoints, underscored the significant contribution of both, antipsychotic medication, and genetic predisposition to higher BMI, driving the substantially upward trajectory of BMI in SCZ cases.

Interpretation: These findings contribute to refining clinical risk prediction and prevention strategies for MetS among SCZ patients and emphasize the significance of incorporating genetic information, long-term patient tracking, and employing diverse perspectives when analyzing real-world health data.

INTERPRETING ARTIFICIAL NEURAL NETWORKS TO DETECT GENOME-WIDE ASSOCIATION SIGNALS FOR COMPLEX DISEASES

Presenter: Burak Yelmen, Institute of Genomics, University of Tartu, Estonia
burak.yelmen@ut.ee

Authors of the Research: Burak Yelmen, Maris Alver, Silva Kasela, Lili Milani

Short Summary of Work:

Investigating the genetic architecture of complex diseases is challenging due to the highly polygenic and interactive landscape of genetic and environmental factors. Although genome-wide association studies (GWAS) have identified thousands of variants for multiple complex phenotypes, conventional statistical approaches can be limited by simplified assumptions such as linearity and lack of epistasis models. In this work, we trained artificial neural networks for predicting complex diseases using both simulated and real genotypes from the Estonian Biobank. We assessed the weights and features learned by these models and extracted feature importance scores to identify potential genomic loci associated with the target phenotype. We used simulated data under various scenarios to test the capacity of our approach to detect causal loci. Simulation results demonstrated that causal loci can be detected with good precision using strict selection criteria but downstream analyses are required for fine-mapping the exact variants due to linkage disequilibrium. By applying our approach to the schizophrenia cohort in the Estonian Biobank (1814 cases), we were able to detect multiple novel loci related to this highly polygenic and heritable disorder. With further improvements in model optimization and confidence measures, artificial neural networks can enhance the identification of genomic loci associated with complex diseases, providing a more comprehensive approach for GWAS and serving as initial screening tools for subsequent functional studies.

AUTOSOMAL DOMINANT TIBIAL MUSCULAR DYSTROPHY IN ESTONIA

Presenter: Siiri Sarv, Department of Genetics and Personalized Medicine, Institute of Clinical Medicine, University of Tartu, Estonia
siiri.sarv@ut.ee

Authors of the Research: Siiri Sarv[1], Tiia Reimand[1,2], Eve Õiglane-Shlik[1,3], Sanna Puusepp[4], Lisanna Põlluaas[1], Laura Mihkla[2], Sander Pajusalu[1,2], Sandra Ütt[5], Katrin Gross-Paju[5], Marco Savarese[5,6], Peter Hackman[5,6], Bjarne Udd[5,6,7], Katrin Õunap[1,2]

[1]Department of Clinical Genetics, Institute of Clinical Medicine, Faculty of Medicine, University of Tartu, Tartu, Estonia
[2] Genetics and Personalized Medicine Clinic, Tartu University Hospital, Tartu, Estonia
[3] Neurology Clinic, Tartu University Hospital, Tartu, Estonia
[4] Department of Pathology, Tartu University Hospital, Tartu, Estonia
[5] West Tallinn Central Hospital, Tallinn, Estonia
[5] Folkhälsan Research Center, Helsinki, Finland
[6] Medicum, University of Helsinki, Helsinki, Finland
[7] Tampere Neuromuscular Center, Tampere University Hospital, Tampere, Finland

Short Summary of Work:

Introduction: Tibial muscular dystrophy (TMD; MIM#600334, ORPHA:609) is an adult onset distal myopathy that was first described by Udd *et al* in Finnish patients. An 11-bp insertion/deletion in the last exon (exon 364/Mex6) of the *TTN* (2q31.2) is the Finnish founder variant (FINmaj). In heterozygous state, FINmaj causes autosomal dominant (AD) TMD with slow progression affecting preferentially the tibialis anterior muscle. Facial muscles, upper extremities, and proximal muscles are usually spared. To date, FINmaj variant is not described outside the Finnish population. After identification of a first family, this study aimed to retrospectively find more TMD patients and to estimate the FINmaj frequency in the Estonian population.

Material and methods: We retrospectively analyzed all large gene panel (TruSight One (n=4800 genes) and Trusight One Expanded (n=6700 genes), Illumina) and whole exome sequencing (WES) data from 2014-2024 to find FINmaj carriers in the cohort of Genetics and Personalized Medicine Clinic (GPMC) in Tartu University Hospital (TUH) (n=11,530) and in the cohort of unsolved myopathy/neuropathy patients in West Tallinn Central Hospital cohort (n=52). We also analyzed Estonian Genome Center of University of Tartu (EGCUT) cohort (n=4,845), to find FINmaj carriers from Estonian general population. Magnetic resonance imaging (MRI) was performed on seven FINmaj patients.

Results: We found five families with 17 patients who carry the heterozygous FINmaj variant: two patients with autosomal recessive limb-girdle muscular dystrophy type LGMD R10 and 15 patients with AD TMD. All the TMD patients in the cohort carried the FINmaj variant and had clinical symptoms by the age of 50 years. MRI analysis showed TMD-specific findings in >50yo patients. FINmaj variant incidence in the Estonian GPMC TUH cohort is 1/2306. We did not find any individuals carrying FINmaj in the EGCUT cohort. Haplotype analysis is at work.

Conclusion: AD TMD is the one of the most common hereditary muscle diseases in Estonia and is still underdiagnosed. To gain a clearer understanding, additional investigations, including haplotype analysis, are required.

Funding: EJP-RD IDOLS-G; PRG471, PRG2040.

ACCURATE DETERMINATION OF CLINICALLY RELEVANT STRUCTURAL VARIATION IN THE CYP2D6 LOCUS: FROM SNP MICROARRAYS TO LONG-READ SEQUENCING

Presenter: Maarja Jõeloo, Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia, maarja.joeloo@ut.ee

Authors of the Research: Maarja Jõeloo, Kadri Maal, Kristi Krebs, Fanny-Dhelia Pajuste, Benjamin Stone, Burak Yelmen, Viktorija Kukuškina, Georgi Hudjašov, Lili Milani

Short Summary of Work:

Cytochrome P450 2D6 is a liver enzyme that contributes to the metabolism of over 20% of commonly used drugs. The corresponding gene, CYP2D6, is known for being highly polymorphic. The repetitive elements flanking the gene give rise to structural variants (SVs), and hybrid genes can arise with a neighbouring highly homologous pseudogene CYP2D7. Here, we perform comparisons of complex CYP2D6 haplotypes in the Estonian Biobank (EstBB) using microarray and whole-genome sequencing (WGS).
EstBB comprises over 210,000 individuals genotyped with Illumina microarray. Additionally, ~2,800 and 150 samples are sequenced with Illumina short reads (srWGS) and PacBio long reads (lrWGS), respectively. First we estimated that, based on srWGS, 10.3% of CYP2D6 alleles in EstBB contained an SVs/hybrids and 75 samples contained rare or complex genotypes that we were unable to resolve without lrWGS. Secondly, we used Stargazer software to assign CYP2D6 star alleles to the full genotyped EstBB cohort. We observed high concordance (99.4%) with srWGS for diplotypes without SVs or hybrid genes. Popular star allele callers cannot leverage microarray data to detect structural rearrangements, and other tools utilised for this purpose suffer from high false positive and negative rates in this region. Nevertheless, our preliminary observations suggest unique array intensity signatures for star alleles encompassing SVs or hybrid genes. This indicates the possibility of estimating true CYP2D6 effects for biobank participants without WGS data.

In conclusion, while long reads are essential for resolving complex CYP2D6 haplotypes, microarray and short reads provide excellent data that contribute to the advancement of personalized medicine. We anticipate that thousands of samples with lrWGS will allow us to assemble a unique multi-platform dataset for developing advanced methodologies to distinguish microarray-based signals corresponding to specific SVs and hybrid structures in clinically relevant genes.

R/BIOCONDUCTOR DATA SCIENCE FRAMEWORK FOR MULTI-OMIC DATA INTEGRATION

Presenter: Leo Lahti, Department of Computing, University of Turku, Finland
leo.lahti@utu.fi

Authors of the Research: Leo Lahti

Short Summary of Work:

Heterogeneous and hierarchically structured multi-assay data arises in several fields of life sciences. Recently, next-generation data containers have emerged to enhance the incorporation of such data in custom data science workflows. These containers store relations between interlinked data tables in a highly optimized way, and provide functions to access, manipulate and integrate data at varying resolutions for seamless downstream analysis within statistical and probabilistic programming workflows. This work will highlight recent advances in multi-omic data integration methodology in R/Bioconductor, with examples from contemporary microbiome research.

Evaluating the Predictive Ability of Polygenic Risk Scores for Intrahepatic Cholestasis of Pregnancy in the Estonian Biobank

Presenter: Fanny-Dhelia Pajuste, University of Tartu, Estonia
fanny-dhelia.pajuste@ut.ee

Authors of the Research: Fanny-Dhelia Pajuste, Kristi Läll, Triin Laisk, Prof Reedik Mägi

Short Summary of Work:

Background: Intrahepatic cholestasis of pregnancy (ICP) is a liver disorder occurring in the late second and early third trimester of pregnancy, characterized by elevated liver enzymes and severe itching. The condition poses significant risks to both maternal and fetal health, including stillbirth and preterm birth. Recent studies have identified genetic factors contributing to ICP susceptibility, creating an opportunity to assess the predictive value of polygenic risk scores (PRS) for ICP.

Methods: Utilizing the genome-wide association study (GWAS) summary statistics including 1,138 cases and 153,642 controls with European ancestry (Dixon et al., 2022), we calculated PRS for ICP for the participants of the Estonian Biobank. Our analysis included 1,152 women diagnosed with ICP (ICD-10 code O26.6) and 47,775 female controls who had late pregnancy or delivery related ICD-10 codes, but no ICP. The association between ICP risk and 10 PRS quantiles was assessed using logistic regression adjusted for age at recruitment and the first 10 genetic principal components (PCs) to control for population stratification. A time-to-event analysis was conducted on 138 pregnancies with ICP and 10,073 control pregnancies using Cox proportional hazards model, accounting for age at pregnancy, birth year, and the first 10 PCs, to determine the hazard of ICP onset across five PRS quantiles. Associations between ICP PRS and maximum liver enzyme levels were evaluated using linear regression, while logistic regression was employed to determine the risk of bile acid levels exceeding 40 µmol/L for 5 PRS quantiles, with both models adjusted for the first 10 PCs.

Results: ICP prevalence increased with PRS quantiles, with a prevalence rate of 6.05% in the highest decile compared to 0.92% in the lowest, indicating a strong association between PRS and ICP prevalence. The risk for ICP was increased for the 30% of individuals with the highest PRS values compared to the middle 20%. The odds ratio for the highest PRS decile (top 10%) compared to the middle was 3.25 (95% CI: 2.70-3.91, $p=1.8\times10^{-35}$).

The odds ratio corresponding to the highest PRS decile compared to the lowest decile was 6.92 (95% CI: 5.19-9.61, $p=4.4\times10^{-33}$). The hazard ratio for ICP onset was 1.94 (95% CI 1.67 -2.26, $p=1.4\times10^{-17}$) across the 5 PRS quantiles, indicating that with each increase to a higher PRS quantile, the risk of developing ICP nearly doubles. There was a positive association between ICP PRS and levels of alanine aminotransferase, aspartate aminotransferase, and bile acids, with the strongest effect observed for bile acids. In addition, the top 20% of individuals by PRS showed a higher risk of elevated bile acid levels (>40 µmol/L).

Conclusions: Our findings indicate that PRS can be an effective tool for predicting the risk of developing ICP as well as the likelihood of elevated bile acid levels, associated with a higher risk of pregnancy complications and stillbirth. These scores could help identify high-risk women, facilitating enhanced monitoring strategies and potentially improving maternal and fetal outcomes.

References:

Dixon, P.H., Levine, A.P., Cebola, I. et al. GWAS meta-analysis of intrahepatic cholestasis of pregnancy implicates multiple hepatic genes and regulatory elements. Nat Commun 13, 4840 (2022). https://doi.org/10.1038/s41467-022-29931-z

Effect of drug-metabolizing enzyme activity on survival in amyotrophic lateral sclerosis

Presenter: Johanna K Jaansoo, Department of Basic and Clinical Neuroscience, King's College London, United Kingdom (UK)
johanna.k.jaansoo@kcl.ac.uk

Authors of the Research: Johanna K Jaansoo[1], Ammar Al-Chalabi[1], Ahmad Al Khleifat[1]
[1] Department of Basic and Clinical Neuroscience, King's College London, United Kingdom (UK)

Short Summary of Work:

Background: One of the most important pharmacogenes is CYP2D6, an enzyme from the cytochrome P450 enzyme superfamily that is involved in the metabolism of ~70-80% of common drugs, with CYP2D6 alone being involved in ~25% of all common drug metabolism. The CYP2D6 locus is highly polymorphic, with its genetic variation contributing to the enzyme's heterogeneous functional activity. Variation in the CYP2D6 locus is associated with adverse drug reactions and reduced drug efficacy with over 40 drugs already having CYP2D6 activity-specific dosing instructions supported by the FDA. In addition to drug metabolism, CYP2D6 is involved in the metabolism of neurochemicals in the CNS with the enzyme's dysfunction being implicated in neurodegenerative diseases, such as Parkinson's disease (1) and Alzheimer's disease (2).

Objectives: In this study, we set out to investigate the effect variation in CYP2D6 and other pharmacogenes has on the progression and survival of people with ALS.

Methods: We obtained data from the Project MinE sequencing consortium. We used the Cyrius tool for genotyping the CYP2D6 gene as it has been designed to overcome the gene's highly polymorphic nature and was most accurate at this locus compared to other genotyping methods. In our preliminary analyses, we used the Cox proportional hazards regression to analyse the effect of CYP2D6 genotype-derived metaboliser phenotype on survival in ALS.

Results: When controlling for sex, age of onset, site of onset, ALS phenotype and C9orf72 expansion status, our preliminary investigations in the UK dataset have found a small increase in the risk of earlier mortality (Hazard Ratio (HR): 1.78 (95% CI: 1.11-2.85),

p=0.017) within the group of individuals with increased CYP2D6 metabolic activity (Ultrarapid Metabolisers, UM).

Discussion: The findings indicate that the ultrarapid CYP2D6 enzyme activity seems to increase the risk of shorter survival in ALS. However, it is unknown whether the effect is independent or mediated through Riluzole metabolism. Currently, there is work undergoing expanding this analysis to a larger European dataset to investigate the extent of this effect internationally and a smaller repetition of this analysis in a dataset where Riluzole use is known. In addition, there is work underway investigating the effect of variation in other pharmacogenes in people with ALS.

References:

Mann A, Miksys SL, Gaedigk A, Kish SJ, Mash DC, Tyndale RF. The neuroprotective enzyme CYP2D6 increases in the brain with age and is lower in Parkinson's disease patients. Neurobiol Aging. 2012;33(9):2160-2171.

Cacabelos R, Martínez R, Fernández-Novoa L, Carril JC, Lombardi V, Carrera I, Corzo L, Tellado I, Leszek J, McKay A, Takeda M. Genomics of dementia: APOE-and CYP2D6-related pharmacogenetics. International Journal of Alzheimer's Disease. 2012;2012(1):518901

IN VITRO TESTING OF GENETIC MUTATIONS OF G PROTEIN COUPLED RECEPTORS

Presenter: Tõnis Laasfeld, University of Tartu, Institute of Chemistry, Estonia
tonis.laasfeld@ut.ee

Authors of the Research: Tõnis Laasfeld, Maris-Johanna Tahk, Maris Alver, Lili Milani, Ago Rinken, Anni Allikalt

Short Summary of Work:

G protein-coupled receptors (GPCRs) are seven transmembrane receptors that play a crucial role in transmitting signals across the cell membrane. GPCRs are one of the most important drug targets, with more than 30% of all approved drugs targeting these proteins. Our workgroup has been focusing on developing assays to identify novel drug candidates that can modulate GPCR signaling, leading to potential new therapies. As a result, we have developed several fluorescence-based assays that enable us to describe various aspects of GPCR signaling. On the one hand, fluorescence anisotropy assay allows monitoring ligand binding in real time in a simplified model system by using receptors expressed on viral particles. On the other hand, we use fluorescence microscopy with live cells to follow ligand binding in more native-like expression systems, enabling quantification of kinetic parameters as well as monitoring ligand-receptor complex trafficking. For this assay, we have developed automatic machine learning based workflows that allows us to quantify whole cell and cell membrane fluorescence intensity. The receptors' activation can be measured using FRET-based biosensors through monitoring the level of second messenger cAMP. All the assays have been validated with several different wild-type GPCRs and corresponding fluorescent ligands.

The focus of our current study is the genetic variation of GPCRs and specifically dopaminergic receptors. For these receptors we have several different fluorescent ligands available. Therefore, the tools we have developed can potentially be used to test the impact of missense mutations on ligand binding and receptor activation of specific GPCRs in vitro. Moreover, by testing already approved drugs on various genetic mutants and comparing the results with the wild-type receptors we can obtain information about differences of receptor signaling at the molecular level. In vitro testing allows for the prediction of drug responses, potentially helping to identify drugs which would be optimal for a particular patient in terms of better response or reduced side-effects.

A HISTORY OF REPEATED ANTIBIOTIC USAGE LEADS TO MICROBIOTA-DEPENDENT MUCUS DEFECTS

Presenter: Kertu Liis Krigul, Institute of Genomics, University of Tartu, Estonia
kertu.krigul@ut.ee

Authors of the Research:
Kertu Liis Krigul & Rachel H. Feeney, Supapit Wongkuna, Oliver Aasmets, Sandra M. Holmberg, Reidar Andreson, Fabiola Puértolas-Balint, Kateryna Pantiukh, Linda Sootak, Tõnis Org, Tanel Tenson, Elin Org & Bjoern O. Schroeder

Short Summary of Work:

Recent evidence indicates that repeated antibiotic usage lowers microbial diversity and ultimately changes the gut microbiota community. However, the physiological effects of repeated – but not recent – antibiotic usage on microbiota-mediated mucosal barrier function are largely unknown. By selecting human individuals from the deeply phenotyped Estonian Microbiome Cohort (EstMB), we here utilized human-to-mouse fecal microbiota transplantation to explore long-term impacts of repeated antibiotic use on intestinal mucus function. While a healthy mucus layer protects the intestinal epithelium against infection and inflammation, using *ex vivo* mucus function analyses of viable colonic tissue explants, we show that microbiota from humans with a history of repeated antibiotic use causes reduced mucus growth rate and increased mucus penetrability compared to healthy controls in the transplanted mice. Moreover, shotgun metagenomic sequencing identified a significantly altered microbiota composition in the antibiotic-shaped microbial community, with known mucus-utilizing bacteria, including *Akkermansia muciniphila* and *Bacteroides fragilis*, dominating in the gut. The altered microbiota composition was further characterized by a distinct metabolite profile, which may be caused by differential mucus degradation capacity. Consequently, our proof-of-concept study suggests that long-term antibiotic use in humans can result in an altered microbial community that has reduced capacity to maintain proper mucus function in the gut.

DETECTING EMBRYO DEVELOPMENTAL POTENTIAL BY SINGLE BLASTOMERE RNA-SEQ

Presenter: Monika Nõmm, Chair of Animal Breeding and Biotechnology, Institute of Veterinary Medicine and Animal Sciences Estonian University of Life Sciences, Tartu, Estonia

Authors of the Research: Monika Nõmm[1], Marilin Ivask[1, 4], Pille Pärn[1, 3], Ene Reimann[2], Ülle Jaakma[1, 3], Sulev Kõks[4,5,6]

[1]Chair of Animal Breeding and Biotechnology, Institute of Veterinary Medicine and Animal Sciences Estonian University of Life Sciences, Tartu, Estonia

[2]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

[3]Competence Centre on Health Technologies, Tartu, Estonia

[4]Department of Pathophysiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

[5]Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA, Australia

[6]Perron Institute for Neurological and Translational Science, Perth, WA, Australia

Short Summary of Work:

Selecting high quality in vitro produced (IVP) bovine embryos for transfer is a difficult task when using only visual embryo observation. We aimed to develop a single blastomere biopsying technique at morula stage allowing embryo selection at early development without compromising embryo production and detecting differences in gene expression profiles of these biopsied morulae to determine if there are distinctions in embryo development between embryos developing into blastocysts and embryos with arrested development in morula stage after biopsying.

In vitro produced bovine embryos were cultured in groups till day 5 when 65 morulae were biopsied with a microneedle and one blastomere was aspirated. The biopsied morulae were further individually cultured in culture media droplets under mineral oil until day 8 when blastocyst formation was recorded. For whole transcriptome sequencing six biopsy samples from embryos arrested in morula stage and six biopsy samples from embryos developing to the blastocyst stage were chosen. The samples were sequenced using the SOLiD 5500 Wildfire platform together with paired-end sequencing chemistry (50 bp forward and 50 bp reverse). Raw data was mapped to the bovine reference genome bosTau7 using Lifescope software (Thermo Fisher Scientific). For statistical and functional analysis Bioconductor packages DeSeq2 and ReactomePA implemented in R were used.

Out of 65 biopsied morulae 32 developed to blastocysts (49.2%). Out of 108 760 successfully mapped genes, 1204 showed a difference in mRNA expression level. Out of these, 155 genes were up-regulated in embryos developing to blastocysts. The most expressed genes were for example HSBP1 and ATP5G3. The pathway enrichment analysis of embryos developing to blastocysts revealed significant enrichment in “organelle biogenesis and maintenance”, “mRNA splicing” and “mitochondrial translation” pathways. These findings suggest principal differences in gene expression patterns and functional networks of embryos able to reach the blastocyst stage compared to embryos arrested in development. Single blastomere biopsy at morula stage allows embryo selection early without compromising embryo production.

This study was supported by Enterprise Estonia grant EU30020, Institutional research funding IUT 8-1 and Horizon 2020 Project SEARMET 692299.

Exploring drug-drug-gene interactions and the occurrence of adverse drug events based on electronic health records of Estonian Biobank participants

Presenter: Laura Birgit Luitva, Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia
laura.birgit.luitva@ut.ee

Authors of the Research: Laura Birgit Luitva, Kristi Krebs, Dage Särg, Maarja Jõeloo, Kadri Maal, Saskia Kuusk, Krista Fischer, Merve Nur Güler, Burak Yelmen, Maris Alver, Lili Milani

Short Summary of Work:

Genetic variation, multimorbidity, and drug-drug interactions significantly impact drug efficacy and can elevate the risk of adverse drug events (ADEs). Research on these factors has been fragmented and primarily limited to small clinical studies. The potential of real-world data has not been explored yet for the systematic analysis of the effects of drug-drug-gene interactions on drug response. As part of the SafePolyMed project we are working on the development of a novel, evidence-based risk scoring system using machine learning on large real-world datasets to identify patients at risk. We have extracted patient reported outcome measures from questionnaires and electronic health records using natural language processing. Information on prescriptions and purchases of drugs with warnings for drug-drug interactions have been retrieved from the database of the Estonian Health Insurance Fund. These datasets are currently being combined with genotype data and polygenic risk scores for diseases and adverse events to build novel models for personalized pharmacotherapy.

The SafePolyMed project receives funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101057639.

GENETIC FACTORS AFFECTING THE REPORTING OF ANTIDEPRESSANT SIDE EFFECTS

Presenter: Hanna Maria Kariis, Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
laura.birgit.luitva@ut.ee

Authors of the Research: Hanna Maria Kariis, Dage Särg, Kristi Krebs, Kadri Kõiv, The Estonian Biobank Research Team, Maris Alver, Kelli Lehto, Lili Milani

Short Summary of Work:

Introduction: Antidepressant side effects are prevalent, leading 37% of patients to discontinue treatment. The genetic basis of these side effects is not well understood. A deeper understanding of the underlying mechanisms could help identify individuals at risk of side effects and improve treatment outcomes.

Aim: Here, we aim to investigate the role of CYP2C19 variation, polygenic risk scores (PRS) for psychiatric disorders, antidepressant response, and traits respective to side effects on the reporting of antidepressant side effects. We meta-analyse the findings with the Australian Genetics of Depression Study.

Methods: Data were pooled from two questionnaires – the Estonian Biobank Mental Health online Survey (N=86,244), the Estonian Biobank Vaccines and Medications Adverse Effect study (N= 49,366) and from unstructured electronic health records using natural language processing (N=2,035). A sensitivity analysis was conducted on 8,566 individuals with a recorded depression diagnosis.

Results: Among 10,478 antidepressant users, 68.4% reported at least one side effect. In a subgroup of 7,244 individuals taking antidepressants metabolised by CYP2C19, poor metabolisers had 74% higher odds of reporting a side effect compared to normal metabolisers (OR= 1.738, 95%CI= 1.184-2.612). Increased genetic predisposition to major depressive disorder, and schizophrenia, was associated with higher reporting of side effect among all antidepressant users (n=10,478). Genetic predispositions to higher BMI, anxiety, and systolic blood pressure, were associated with respective side effects among all antidepressant users.

Conclusion: Our findings underscore the role of genetic factors in antidepressant side effects and have potential implications for personalised medicine approaches that could improve treatment outcomes.

THE PREDICTORS OF ESTONIAN BIOBANK MENTAL HEALTH ONLINE SURVEY PARTICIPATION AND THE PREVALENCE OF MENTAL HEALTH PROBLEMS IN THE ESTONIAN BIOBANK

Presenter: Triinu Ojalo, Estonian Genome Center, Institute of Genomics, University of Tartu

Authors of the Research: Triinu Ojalo[1], Elis Haan [1,2], Kadri Kõiv[1], Hanna Maria Kariis[1], Kristi Krebs[1], Helen Uusberg[3], Helene Alavere[1], EstBB research team[1], Lili Milani[1], Kelli Lehto[1]

[1]Estonian Genome Center, Institute of Genomics, University of Tartu

[2]Viljandi Hospital Psychiatry Clinic, Viljandi, Estonia

[3]Institute of Psychology, University of Tartu, Estonia

Short Summary of Work:

Introduction: Mental health problems and disorders are widespread worldwide, impairing the quality of life of the affected individuals. Only a few large biobanks have collected self-report data to collect symptom-level information from participants. Estonian Biobank Mental Health online Survey (EstBB MHoS) was conducted in the spring of 2021 to significantly expand phenotyping on a broad range of mental health symptoms and related phenotypes available in EHR to facilitate research in psychiatric genomics.

Aim: We aimed to explore the differences between EstBB MHoS respondents and non-respondents and estimate the prevalence of common psychiatric disorder risk based on questionnaires and electronic health record (EHR) data.

Methods: Participants and non-participants were compared based on socioeconomic and lifestyle factors based on the EstBB baseline questionnaire and the EHR-based psychiatric diagnoses and psychiatric polygenic risk scores (PRS). Prevalence of mental disorders risk was assessed based on cut-off thresholds of the self-report brief screening instruments of current and/or lifetime depression, anxiety, mania/hypomania, posttraumatic stress disorder, disordered eating, alcohol abuse, attention-deficit/hyperactivity disorder, adverse childhood experiences (ACE). This was compared to the EHR diagnose-based prevalence for respective psychiatric disorders.

Results: MHoS survey had 86,244 participants, and the response rate was 47%. The mean age of participants was 49 years, and 70.6% of the participants were of the female sex. Comparing participants with non-participants regarding sociodemographic variables,

there was a bias towards the female sex (non-participants 61.4%) and a university degree (participants 51.6%, non-participants 34.3%). There was no significant difference between participants and non-participants regarding EHR diagnoses of any mental health disorder. In MHoS participants, the highest rates of risk of mental disorders were depression (current 26.6%, lifetime 36.4%) and generalised anxiety (current 23.3%, lifetime 38.2%). Comparing the self-reported data with EHR diagnoses, it was comparable in Depressive disorder lifetime diagnosis (24.7%) but noticeably lower in lifetime Generalized Anxiety disorder diagnosis (3.3%).

Conclusions: The participants of the Estonian Biobank MHoS, when compared to non-participants, were similar in terms of age and lifestyle factors, but females and individuals with a university degree were somewhat overrepresented. The participants and non-participants did not differ significantly regarding the mental health diagnoses in the EHR. As self-reported mental health disorder risk was mainly higher in MHoS than the prevalence of actual diagnoses in EHR, the results indicate that some disorders may be underdiagnosed; however, due to the nature of screening instruments used, it may reflect subclinical cases as well.

Deciphering Suicide Risk in Major Depressive Disorder: A Polygenic Approach to Early and Late Onset Cases

Presenter: Siim Kurvits, Research group of neuropsychiatric genomics, Institute of Genomics, University of Tartu, Estonia

Authors of the Research: Siim Kurvits, Tryggve consortium, Triinu Ojalo, Toomas Haller, Lili Milani, Kelli Lehto

Short Summary of Work:

Major Depressive Disorder (MDD) is a complex and heterogeneous disease with diverse manifestations. Early onset MDD (eoMDD) has been associated with self-harm, suicide and comorbidity with other psychiatric disorders, while late onset MDD (loMDD) correlates with cardiovascular disease risk and declines in memory and language. This study aims to explore the association between the polygenic risk scores (PRS) for eoMDD, loMDD and suicide attempts, utilizing data from Estonian electronic health records and self-reported Mental Health online Survey conducted in the Estonian Biobank (EstBB).

Our research draws on the meta-analysis of genome-wide association studies (GWASs) conducted in cohorts from five Nordic countries (Estonia, Finland, Sweden Norway and Denmark) with detailed electronic health records and harmonized phenotype definitions, encompassing approximately 46,000 cases of eoMDD and 37,000 cases of loMDD. Distinct diagnostic criteria were applied to define eoMDD and loMDD, requiring an initial diagnosis before the age of 26 for eoMDD and a first diagnosis at age 50 or later (ICD-10 codes F32 or F33) for loMDD. Our focus is on understanding the specific impact of the eoMDD PRS on suicide attempts. We employed LDPred2 to calculate polygenic risk scores for Estonian Biobank participants based on transNordic eoMDD and loMDD GWAS meta-analysis results, excluding EstBB. LD Score regression was used to estimate genetic correlations. Suicide ideation and attempts were measured using the Paykel'ssuicide scale from the self-reported mental health online survey, along with ICD-10 codes for suicide attempts, providing a comprehensive evaluation of the risk for suicide attempts in high and low genetic predisposition groups.

The genetic correlations between eoMDD and loMDD show only moderate overlap (rg=0.58, SE=0.04). Notably, the genetic correlation with suicide attempts is significantly larger for eoMDD (rg=0.9, SE=0.05) than for loMDD (rg=0.4, SE=0.03). Preliminary results indicate a statistically significant correlation between the PRS for eoMDD and increased

susceptibility to suicide attempts for self-reported and electronic health record-based definitions. The risk model for suicide attempt stratified using PRS for eoMDD is demonstrated at the conference.

This study emphasizes the role of polygenic risk scores in predicting suicide risk among individuals with MDD. The integration of electronic health records and online questionnaires ensures the reliability of our findings for both self-reported and clinical data. Our findings contribute to a deeper understanding of the nuanced relationship between MDDs age of onset, genetic predisposition, and suicide attempts, providing a basis for informed clinical strategies and further research initiatives.

"Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them."