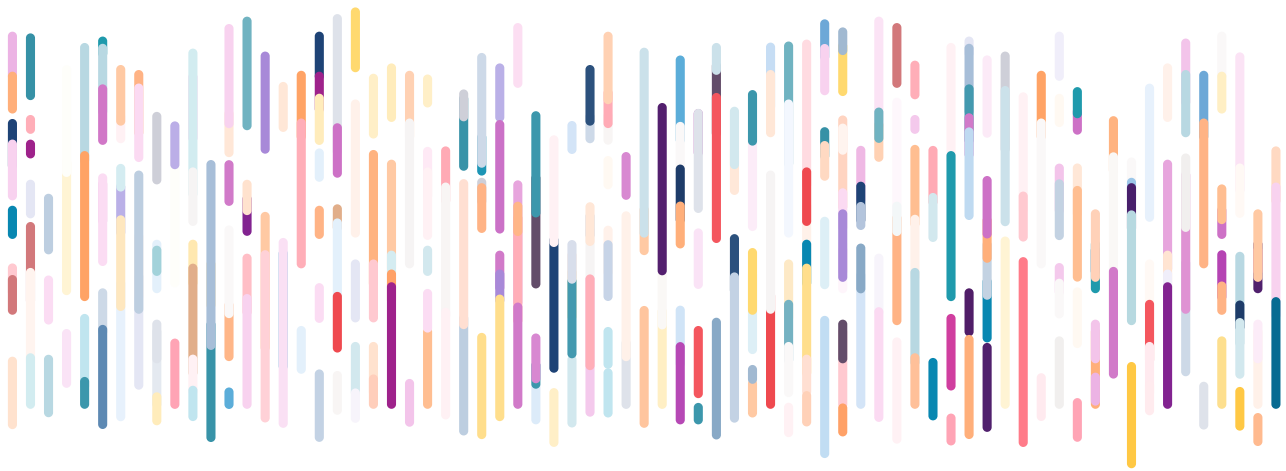
2026 Gene Forum

The 25th Annual International

25–26 August 2026 | in Tartu, Estonia



UNIVERSITY OF TARTU
Institute of Genomics

Eesti Geenikeskus
Estonian Genome Foundation

Funded by
the European Union

POSTER OVERVIEW

	CLINICAL GENETICS	Presenter
1	Insights from a Japanese hereditary tumor cohort: development of a machine learning–based risk prediction model for NF1 (iNForesight)	Mashu Futagawa
2	Genetic Characterization of Inherited Eye Diseases in a Lithuanian Cohort	Evelina Siavrienė
3	In vitro variant screening improves variant interpretation in ACTN2 and other proteins with homologous actin-binding domains	Johanna Ranta-Aho
4	From biobank to medical device: hereditary tumour cohort biobanks as validation infrastructure for genetic diagnosis of hereditary tumour syndromes by multigene panel testing	Akira Hirasawa
5	**Potential Founder BRCA2 Germline Genetic Variant in Estonia**	**Mikk Tooming Poster-talk**
6	Beyond established melanoma genes: interpretation of germline findings in suspected hereditary melanom	Valters Vegelis
7	Expression Patterns of Ataxia-Associated Genes During Human Cerebellum Development	Mariya Roy
8	CYP3A4 Pharmacogenetics and Cardiovascular Polypharmacy: Evidence from Our Future Health	Deniz Turkmen
	COMPLEX TRAIT GENETICS	
9	**Predictive utility of childhood polygenic scores for adult disease risk and inform early-life prevention**	**Francesca Rosamilia Poster-talk**
10	Genome-Wide Meta-Analysis of Psoriatic Arthritis Large Biobank and Population Cohorts	Aleksandra Judin
11	Improving the predictability and interpretation of polygenic scores for Big Five personality traits	Lisette Tagel
12	External exposome and metabolite associations in 152,000 participants of the Estonian Biobank	Karmel Teder
13	PRS-wide association study analysis provides insight into the genetic background of hormone sensitivity-related diagnoses in women	Anastasiia Bratchenko
14	A multi-ancestry evaluation of polygenic scores across global populations	Elia Tiso
15	Bidirectional causal relationships between thyroid function and depression are predominantly driven by early-onset MDD: a Mendelian randomization stu	Triinu Peters
16	Triangulating causal links between personality and health: Mendelian randomisation and within-family regression estimates	Kerli Ilves
17	Distinct genetic and phenotypic associations between inattention and hyperactivityimpulsivity symptom domains and educational attainment in the general adult population	Triinu Varvas
18	Comorbidities and genetic predisposition for pelvic organ prolapse	Valentina Rukins
19	Genetic haemochromatosis-associated multimorbidity: parallel analysis of Our Future Health and UK Biobank cohorts	Luke N Sharp
20	Molecular Lipid Signatures Define Metabolic Subtypes with Distinct Clinical and Genetic Risk Profiles	Soubantik Sengupta
21	Comparison of traditional cardiovascular risk factors and polygenic risk score between younger and older patients with first-ever myocardial infarction	Jana Smirnova
22	Novel loci and multi-omics risk models for rheumatoid arthritis through a million-participant genome-wide association meta-analysis	Galadriel Lucía Velázquez Silva
23	Does NAD play a causal role in aging and cognition? A Mendelian randomisation study	Yehor Chudovskyi
24	Covariate-aware genomic prediction of blood metabolite profiles using multi-task neural networks	Merve Nur Güler

	MICROBIOME & VIROME	Presenter
25	Reduced viral diversity and Enterococcaceae-phage enrichment distinguish the preterm infant gut virome	Radko Avi
26	Investigating the evolutionary dynamics underlying DNA virus immune response	Rutvi Rajpara
27	Reorganised, not expanded: genome-resolved metabolic potential and virome across ages 12–109	Taavi Päll
28	Polygenic Risk Links to the Gut Microbiome: The MAGI Catalog	Oliver Aasmets
29	Participant Recall to Explore the Gut Microbiome-Endotoxemia Axis as a Driver of Chronic Inflammation in Arthritis	Kreete Lüll
30	Microbial signals preceding depression onset: preliminary findings from the Estonian Microbiome Cohor	Kadi Vaher
31	Niche occupation by uncultivated RFN20 bacteria in the human gut	Kateryna Pantiukh
32	Multi-site microbiome mapping along the colorectal neoplasia continuum	Kertu Liis Krigul
33	Drug-Naive de novo Parkinson's Disease Microbiome Profiling by 16S rRNA Sequencing	Gizi Golt
	MOLECULAR & SPATIAL BIOLOGY	
34	Anticancer and Antimicrobial Properties of Silymarin from Milk Thistle (Silybum marianum): Phytochemical Analysis and Bioactivity Evaluation	Mohmmad Asif Gawhari
35	Spatial Organization of Cancer-Associated Fibroblast Subtypes in Gastrointestinal Tumors	Violeta Lisovska
36	High-Resolution Spatial Transcriptomics in Human Atherosclerotic Plaques: Dissection of Atherosclerotic Plaque Microanatomy	Tiit Örd
	BIOBANKS & SOCIETY	
37	Public Understanding and Readiness for Personalised Genetic Prevention: Genome of Europe (GoE) Baseline Survey in 5 European Countries	Merit Kreitsberg
38	Trust in the future genomic data - biobank participation and governance preferences in Estonia	Piret Hirv
39	Andmering: Building a national learning health system from routine health and genomic data using the OMOP Common Data Model	Riina Raudne
40	Bridging Biobanks: Introducing Japan's National Infrastructure and User-Centric Initiatives for Advancing Genomic Researche	Mizuki Morita
	POPULATION GENETICS	
41	**Maternal contacts across the Baltic Sea – thousands of mitogenomes reveal maternal genetic structure and recent expansions in the CircumBaltic region**	**Anne-Mai Ilumäe** **Poster-talk**
42	A genetic bridge over the gulf of finland: tracing the Origin of genetic connectedness between finns and Estonians	Roberto Didonna
	GENOMIC TECHNOLOGIES	
43	Scaling and democratising structure-based protein function prediction with metagenomic-deepFRI	Filip Schymik
44	F64: rapid customizable sequencing data filter	Alar Aints
45	A novel method of Single Cell Whole Transcriptome Sequencing from FFPE Samples	Ritika Kulshreshtha
46	RNA-Seq library preparation bias affects long transcript detection	Swethaa Natraj Gayathri
47	Click-Chemistry-Based High-Degree Sample Multiplexing for Large-Scale Single Cell Sequencing Dataset Generation	Nan Fang
48	EASIGEN-DS: Designing a distributed Research Infrastructure for Advanced Genomics Technologies	Kristiina Tambets

1

INSIGHTS FROM A JAPANESE HEREDITARY TUMOR COHORT: DEVELOPMENT OF A MACHINE LEARNING-BASED RISK PREDICTION MODEL FOR NF1 (INFORESIGHT)

Mashu Futagawa[1], Eiji Nakada[2], Tetsuya Okazaki[1], Sayoko Takeuchi[1], Kana Yamatogi[3], Hideki Yamamoto[1], Toshifumi Ozaki[2], Akira Hirasawa[1]

[1]Department of Clinical Genomic Medicine, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences
[2]Department of Orthopaedic Surgery, Okayama University Hospital
[3]Department of Clinical Genetics and Genomic Medicine, Okayama University Hospital

E-mail address of presenting author: mfutagawa@okayama-u.ac.jp

Background: The Japan Hereditary Tumor Cohort (JHTC) is a prospective multi-institutional cohort established to collect genomic, epidemiological, and biospecimen data from Japanese individuals with hereditary tumors (HT) and their families. Within this broader platform, disease-specific analyses can be conducted to generate evidence on HT such as Neurofibromatosis type 1 (NF1).

Study question: Can a JHTC support both genotype-phenotype characterization of NF1 and the development of an interpretable disease-specific prediction tool?

Methods: JHTC collects clinical information and DNA samples through 51 collaborating institutions, with annual follow-up and de-identified biospecimen storage. Within this cohort, we first analyzed 44 individuals with suspected NF1 for genotype-phenotype correlations. We then used NF1 germline pathogenic variant (GPV) carriers from the cohort to develop iNForesight, an explainable machine-learning system for risk stratification of plexiform neurofibromas with respect to malignant transformation. The model used LightGBM, leave-one-out cross-validation, and Explainable AI.

Main results: In the cohort-based NF1 study, GPVs were identified in 36/44 cases (81.8%), including novel variants, while 40.0% of adults had malignancies. iNForesight achieved an accuracy of approximately 60–65% and enabled case-level visualization of how age, variant-related variables, smoking, and alcohol-related variables influenced model output.

Limitations: The NF1 subset was small, particularly for prediction modeling, and external validation of iNForesight has not yet been performed.

Wider implications: This study demonstrates how our cohort can serve as a shared research infrastructure that extends beyond a single disease entity. The JHTC platform enables the development of disease-specific studies, such as NF1 genotype-phenotype analysis and interpretable tool development, by integrating genomic, clinical, epidemiological, and biospecimen data.

2

GENETIC CHARACTERIZATION OF INHERITED EYE DISEASES IN A LITHUANIAN COHORT

Viktorija Gurskytė[1,2], Violeta Mikštienė[1,3,4], **Evelina Siavrienė**[1], Rasa Strupaitė-Šileikienė[2,3,5], Kristina Kazlauskaitė[6], Audronė Jakaitienė[1,6], Laima Ambrozaitytė[1,4], Evelina Marija Vaitėnienė[1,4], Ramūnas Dzindzalieta[3], Algirdas Utkus[1,3]

[1]Vilnius University, Faculty of Medicine, Department of Human and Medical Genetics, Institute of Biomedical Sciences, Santariškių str. 2, LT-08661, Vilnius, Lithuania
[2]Vilnius University Hospital Santaros Klinikos, Center of Eye Diseases, Santariškių str. 2, LT-08406, Vilnius, Lithuania
[3]Vilnius University, Life Sciences Center, Saulėtekio al. 7, LT-10257, Vilnius, Lithuania
[4]Vilnius University Hospital Santaros Klinikos, Center for Medical Genetics, Santariškių str. 2, LT-08406, Vilnius, Lithuania
[5]Vilnius University, Faculty of Medicine, Institute of Health Sciences, M. K. Čiurlionio str. 21, LT-03101, Vilnius, Lithuania
[6]Vilnius University, Faculty of Mathematics and Informatics, Naugarduko str. 24, LT-03225, Vilnius, Lithuania

E-mail address of presenting author: evelina.siavriene@mf.vu.l

Background: Inherited eye diseases (IEDs) comprise a diverse group of conditions causing significant visual impairment. Data on their genetic epidemiology in Baltic populations is scarce.

Study question: What are the most prevalent genetic causes of IEDs in a Lithuanian cohort?

Methods: Individuals with syndromic or non-syndromic IEDs were evaluated at a tertiary referral center between 2013 and 2025. Comprehensive ophthalmic examination and genetic testing were performed, using whole exome sequencing (WES) with analysis of Virtual Ophthalmic Gene Panel genes and/or Sanger sequencing. Variant pathogenicity was evaluated according to the American College of Human Genetics and Genomics guidelines. Data were deposited in the iGENLIT database.

Main results: 182 unrelated probands with confirmed molecular diagnoses were classified into five groups: rod-cone dystrophies (47.3%), macular dystrophies (23.6%), stationary retinal dystrophies (5.5%), cone-rod dystrophies (3.8%), and other IEDs (19.8%). Marked genetic heterogeneity was observed, with 168 disease-causing variants identified across 66 genes, including 29 autosomal recessive, 28 autosomal

dominant, 8 X-linked, and 1 mitochondrial gene. The most frequently altered genes were *ABCA4* (n=29), *USH2A* (n=19), *EYS* (n=13), and *RP1* (n=13). Notably, 43 variants (25.6%) were previously unreported.

Limitations: Segregation analysis was limited by incomplete family data in some cases. WES may have missed structural or deep intronic variants.

Wider implications: This is the first comprehensive analysis of IEDs in a previously underrepresented Lithuanian population. The high degree of genetic heterogeneity, combined with the identification of a substantial proportion of novel variants, reveals a distinct regional genetic spectrum and enriches variant databases.

IN VITRO VARIANT SCREENING IMPROVES VARIANT INTERPRETATION IN *ACTN2* AND OTHER PROTEINS WITH HOMOLOGOUS ACTIN-BINDING DOMAINS

Johanna Ranta-aho[1,2], Per Harald Jonson[1,2], Jaakko Sarparanta[1,2], Bjarne Udd[1,3], Marco Savarese[1,2]

[1]Folkhälsan Research Center, Haartmaninkatu 8, 00290, Helsinki, Finland
[2]Department of Medical Genetics, Medicum, University of Helsinki, Haartmaninkatu 8, 00290, Helsinki, Finland
[3]Tampere Neuromuscular Center, Tampere University and Tampere University Hospital, Biokatu 8, 33520, Tampere, Finland

E-mail address of presenting author: johanna.ranta-aho@helsinki.fi

Background: In the next-generation sequencing era, sequencing can be performed with unprecedented ease and lower costs, while detection of elusive variants has also improved. Hence, the diagnostic bottleneck in rare disease has shifted to data interpretation. As the functional impact of most missense variants remains unknown, the number of variants classified as variant of uncertain significance (VUS) has sharply risen in the last few years.

Study question: To improve variant interpretation in rare diseases, there is an urgent need to assess the functional impact of large numbers of variants.

Methods: We perform an *in vitro* functional screening of over 200 missense variants in ACTN2, a gene associated with myopathy and cardiomyopathy, to assess their effect on the expressed protein. Most of the variants included in the screening are in the actin-binding domain (ABD). Interestingly, homologous ABDs are found in other proteins. Thus, to test whether our results may apply to these other ABDs, we also test a few variants in FLNC.

Main results: We found over 20 variants in the ACTN2 ABD that cause protein aggregates in vitro. An FLNC variant equivalent variant to an aggregation-causing variants in ACTN2 also resulted in aggregates.

Limitations: This is a proof-of-concept study, specifically designed to test for protein aggregation. For more refined variant impact, further functional studies should be performed.

Wider implications: Our results improve variant interpretation in ACTN2 by providing functional data for over 200 variants. Preliminary data suggests that the findings may apply to other proteins with homologous ABDs, improving variant interpretation in many other genes as well.

FROM BIOBANK TO MEDICAL DEVICE: HEREDITARY TUMOUR COHORT BIOBANKS AS VALIDATION INFRASTRUCTURE FOR GENETIC DIAGNOSIS OF HEREDITARY TUMOUR SYNDROMES BY MULTIGENE PANEL TESTING

Akira Hirasawa[1], Sachio Nohara[2], Yasuyuki Hosono[3], Mashu Futagawa[1], Tetsuya Okazaki[1], Sayoko Takeuchi[1], Kana Yamatogi[1], Hideki Yamamoto[1], Hiroshi Nishihara[4]

[1]Department of Clinical Genomic Medicine, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan
[2]Bioinformatics Department, Mitsubishi Electric Software Corporation, Japan
[3]Department of Pharmacology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan
[4]Cancer Genomic Medicine Unit, Keio Cancer Center, Keio University School of Medicine, Tokyo, Japan

E-mail address of presenting author: hir-aki45@okayama-u.ac.jp

Background: Japan's multigene panel testing (MGPT) Guidance 2025 defines hereditary cancer syndrome as caused by carrying a germline pathogenic variant (GPV) in cancer susceptibility genes, regardless of whether cancer has developed. GPVs occur in ~10% of cancer patients, predispose to cancer in multiple organs and are shared with relatives, so diagnosis serves whole families. Yet every MGPT in Japan is foreign-built, analysed overseas and unapproved; Software as a Medical Device (SaMD) approval demands well-characterised germline references.

Study question: Can hereditary tumour syndrome biobanks serve as validation infrastructure for a domestic whole-exome-sequencing-based MGPT device?

Methods: The Mitsubishi and Keio University biobanks (200 tumour/normal pairs; 1,055 cases) supply sequence data for software development. MeJeTo (Mid-West Japan Hereditary Tumour Cohort; https://cgm.hsc.okayama-u.ac.jp/en/cohort/), holding DNA, tissue and clinical data from 51 institutions and supported by OKADAI BIOBANK (Okayama University Hospital Biobank; https://biobank.ccsv.okayama-u.ac.jp), provides ~400 germline DNA samples characterised by domestic laboratories as reference standard.

Main results: Building on software for tumour/normal comprehensive genomic profiling (PleSSision Exome), we redesigned algorithms adding copy-number variant and mosaicism detection, and implemented a pipeline annotating against ClinVar,

gnomAD and the Japanese reference 2KJPN. PMDA consultations have been conducted; performance evaluation is in preparation.

Limitations: Performance evaluation is not yet executed; sensitivity and specificity remain undetermined. Biobank reference variants were characterised across laboratories and platforms, introducing heterogeneity; variants of uncertain significance remain hard to adjudicate.

Wider implications: Hereditary tumour syndrome biobanks are not merely research resources but regulatory infrastructure, supplying reference material device approval requires. Population-specific references such as 2KJPN improve interpretation for underrepresented populations — transferable to other national biobanks.

POTENTIAL FOUNDER *BRCA2* GERMLINE GENETIC VARIANT IN ESTONIA

Mikk Tooming[1,2], Kanwal Batool[3], Kadri Rekker[2], Kadri Toome[2], Piret Laidre[2], Estonian Biobank Research Team[3], Erik Abner[3], Katrin Õunap[1,2], Tiina Kahre[1,2]

[1]Department of Genetics and Personalized Medicine, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
[2]Genetics and Personalized Medicine Clinic, Tartu University Hospital, Tartu, Estonia
[3]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

E-mail address of presenting author: mikk.tooming@kliinikum.ee

Background: Pathogenic variants (PV) in *BRCA1* and *BRCA2* are major contributors to hereditary breast and ovarian cancer (HBOC). While founder variants have been described in several populations, population-specific *BRCA* variants in Estonia remain largely unexplored.

Study question: Does the NM_000059.4(BRCA2):c.8572C>T p.(Gln2858*) pathogenic variant represent a founder variant in the Estonian population?

Methods: We analyzed a clinical cohort of 7,087 individuals referred for molecular testing at Tartu University Hospital between 2007 and 2023, including 2,856 breast cancer cases, 759 ovarian cancer cases, and 3,472 healthy family members. Genetic testing utilized APEX microarrays, Sanger sequencing, and next-generation sequencing panels. Population prevalence and geographic distribution were evaluated in the Estonian Biobank (EstBB; n=207,954) using genotyping array and imputed genetic data.

Main results: The *BRCA2* c.8572C>T variant was the most frequent *BRCA2* PV in the clinical cohorts, identified in 20 breast cancer (0.7%), 7 ovarian cancer patients (0.9%), and 48 healthy relatives (1.4%). In EstBB, 146 carriers were detected, corresponding to an estimated prevalence of approximately 1 in 1,424 individuals. The variant is rare in other populations and exhibited marked geographic clustering in southeastern Estonia. Population structure analyses demonstrated that carriers clustered within the local Estonian genetic background, supporting a shared ancestral origin.

Limitations: Haplotype analysis has not yet been performed to directly confirm a common ancestral chromosome carrying the variant.

Wider implications: These findings strongly support *BRCA2* c.8572C>T as a likely Estonian founder PV originating from southeastern Estonia. Recognition of this variant could improve targeted genetic testing strategies, cancer risk assessment, genetic counseling, and cascade screening in Estonia.

Funding: PRG2040, PRG1291, TT17, PLTGIARENG24

BEYOND ESTABLISHED MELANOMA GENES: INTERPRETATION OF GERMLINE FINDINGS IN SUSPECTED HEREDITARY MELANOMA

Valters Vegelis[1], Linda Gailīte[2], Dace Pjanova[1]

[1]Institute of Microbiology and Virology, Riga Stradins University, Latvia
[2]Scientific Laboratory of Molecular Genetics, Riga Stradins University, Latvia

E-mail address of presenting author: n/a

Background: Genetic testing for hereditary melanoma primarily targets established melanoma predisposition genes. Exome-based testing may identify variants related to other cancer syndromes, but their relevance remains uncertain.

Study question: Does interpretation beyond established melanoma genes add relevant information, and how could these findings be clinically categorised?

Methods: We performed exome sequencing* in 80 individuals (69 predisposed melanoma patients and 11 relatives), followed by analysis of a 36-gene virtual panel: ACD, AKT1, APC, ATM, BAP1, BRCA1, BRCA2, CDK4, CDKN2A, CHEK2, MC1R, MITF, OCA2, MLH1, MSH2, MSH6, NF1, PALB2, PMS2, POLE, POT1, PTEN, RAD50, RB1, SMAD4, STK11, TERF2IP, TERT, TERF1, TSC1, TSC2, TP53, TYR, TYRP1, XPA, and XPC. Pathogenic/likely pathogenic (P/LP) variants were grouped as: (I) melanoma predisposition, (II) broader hereditary cancer predisposition, or (III) pigmentation-related variants with possible modifying effects. Interpretation considered disease association, inheritance, phenotype, and family context. Pathogenicity was not assumed to indicate melanoma causality.

Main results: Among melanoma patients, 11 of 69 (15.9%) carried nine distinct P/LP variants of CDK4, POT1, TP53, BRCA2, CHEK2, NF1, TYR and TYRP1. Group I occurred in three individuals (4.3%; CDK4, n=2; POT1, n=1), Group II in four (5.0%; TP53, BRCA2, CHEK2, NF1, n=1 each), and Group III findings in four (5.8%; TYR, n=5, TYRP1, n=1). Eight of 11 variant-positive melanoma patients (73%) fell outside Group I. Among relatives, three carried CDK4 (n=1) or TYR (n=2) variants. One CDK4 carrier had multiple primary melanomas.

Limitations: This analysis was limited lack of segregation analyses.

Wider implications: Findings beyond established melanoma genes may prompt syndrome-specific assessment, while pigmentation gene findings require cautious, phenotype and family-based interpretation.

*****This work benefited from access to the University of Łódź, part of the BBMRI-ERIC research infrastructure, through canSERV (EU Horizon Europe Grant No. 101058620). We thank the Latvian Biomedical Research and Study Centre and the Genome Database of the Latvian Population for providing biological samples, and data.

7

EXPRESSION PATTERNS OF ATAXIA-ASSOCIATED GENES DURING HUMAN CEREBELLUM DEVELOPMENT

Mariya Roy, Mari Sepp

Institute of Genomics, University of Tartu, Estonia

E-mail address of presenting author: roymarrria@gmail.com

Background: Ataxias are a heterogeneous group of neurological disorders characterized by impaired balance and coordination, leading to progressive disability. They most commonly arise from dysfunction or degeneration of the cerebellum, a brain region essential for motor control. Many forms of ataxia have genetic causes with more than a hundred genes associated with hereditary ataxias.

Study question: Cerebellar Purkinje cells represent a particularly vulnerable neuronal population across many genetic forms of ataxia, but the mechanisms underlying this selective vulnerability remain incompletely understood.

Methods: This study examines the expression of ataxia-associated genes across developmental stages and cell types of the human cerebellum. A curated list of ataxia-associated genes was compiled from the OMIM database and analyzed using bulk and single-nucleus RNA-sequencing datasets. Gene expression was compared across developing organs, developmental trajectories within the cerebellum, and cerebellar cell types.

Main results: Ataxia-associated genes showed higher expression during cerebellar development than in other organs. Their expression increased during cerebellar development and cellular differentiation and was particularly high in Purkinje cells. Genes associated with autosomal dominant ataxias showed greater cerebellar and Purkinje cell specificity than genes associated with autosomal recessive ataxias.

Limitations: The limitations of this study include the incomplete catalog of ataxia-associated genes and gaps in the current transcriptomic data, which do not fully capture cerebellar cellular diversity across all developmental stages.

Wider implications: Together, the findings of this study reveal potential links between gene expression specificity and selectivity of neuronal vulnerability, and suggest directions for future research into the mechanisms underlying ataxia pathogenesis.

CYP3A4 PHARMACOGENETICS AND CARDIOVASCULAR POLYPHARMACY: EVIDENCE FROM OUR FUTURE HEALTH

Deniz Turkmen, Luke Pilling on behalf of GEMINI

AGE Group, Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, UK

E-mail address of presenting author: d.turkmen2@exeter.ac.uk

Background: Polypharmacy is common in people with multiple long-term conditions and increases the risk of drug-drug interactions and adverse drug reactions. CYP3A4 metabolises many commonly prescribed cardiovascular medicines, and genetic variation may alter drug exposure and treatment response. However, the contribution of CYP3A4 genetic variation to cardiovascular medication burden and treatment patterns is poorly understood.

Study question: We investigated whether genetic variants affecting CYP3A4 activity are associated with cardiovascular medication burden, treatment duration and discontinuation, and whether effects differ among individuals receiving multiple CYP3A4-metabolised medicines.

Methods: We used genetic and linked dispensed prescribing data from ~730,000 Our Future Health participants. We identified commonly prescribed CYP3A4-metabolised cardiovascular medicines and derived measures of medication use, drug burden, treatment duration and discontinuation. Regression and time-to-event models were used to assess associations between CYP3A4 genetic variants and medication outcomes, adjusting for demographic and genetic covariates.

Main results: *CYP3A4* pharmacogenetic variation was not associated with overall cardiovascular polypharmacy or CYP3A4 medication burden. However, the CYP3A4*22 decreased-function variant was significantly associated with shorter amlodipine treatment duration and increased discontinuation.

Limitations: Dispensed prescribing data indicate medication supply rather than actual medication use or adherence, and treatment duration may not reflect continuous exposure. CYP3A4 metabolism varies between drugs and may be influenced by other metabolic pathways and drug-drug interactions.

Wider implications: Initial findings suggest that CYP3A4 genetic variation may influence treatment patterns for specific cardiovascular medicines rather than overall polypharmacy. Further work is needed to evaluate CYP3A4 genetic effects across a broader range of medicines and to determine their clinical relevance for personalised prescribing in multimorbidity.

PREDICTIVE UTILITY OF CHILDHOOD POLYGENIC SCORES FOR ADULT DISEASE RISK AND INFORM EARLY-LIFE PREVENTION

Rosamilia Francesca[1,2], Fiorito Giovanni[1], Vartiainen Pekka[2], Tantari Giacomo[3], Rubinacci Simone[2], Uva Paolo[1], Ceccherini Isabella[4]

[1]Clinical Bioinformatic Unit, IRCCS Giannina Gaslini Institute, Genoa, Italy
[2]Institute for Molecular Medicine Finland, Helsinki, Finland
[3]Pediatric Clinic, IRCCS Giannina Gaslini Institute, Genoa, Italy
[4]UOSD Research Laboratories Aggregation Area, IRCCS Giannina Gaslini Institute, Genoa, Italy

E-mail address of presenting author: francesca.rosamilia@helsinki.fi

Background: Many adult diseases trace their origins to childhood, yet whether pediatric genetic risk profiles can effectively predict adult-onset conditions remains poorly characterized.

Study question: We investigated if pediatric polygenic scores (PGS) can identify individuals at increased risk for adult-onset diseases, revealing shared genetic architecture throughout life.

Methods: We constructed 73 PGS for 30 pediatric traits using GWAS summary statistics and the PGS Catalog. We tested associations between pediatric PGS and 25 adult phenotypes in FinnGen and UK Biobank (replication phase). Linkage disequilibrium score (LDSC) regression and multi-trait conditional joint analysis (mtCOJO) were used to distinguish genetic pleiotropy from effects mediated by genetically related traits.

Main results: We identified 27 significant pediatric–adult associations in FinnGen, primarily linking childhood autoimmune genetic risk to adult cardiometabolic and autoimmune phenotypes. These associations robustly replicated in the UK Biobank. Notably, the pediatric type 1 diabetes PGS strongly predicted adult rheumatoid arthritis, with individuals in the highest PGS decile showing a 1.8-fold increased risk compared with those in the lowest decile (OR=1.88, 95% CI=1.76–2.01; $P=1.09\times10^{-75}$). LDSC and mtCOJO analyses confirmed that these overlapping risks represent direct genetic links rather than mediation by confounding traits.

Limitations: Analyses were limited to individuals of European ancestry and to currently available pediatric GWAS studies, limiting the transferability of results.

Wider implications: Systematic evaluation of the ability of the polygenic genetic score (PGS) derived from childhood traits aims to improve the prediction of complex diseases in adulthood by adding information beyond the adult-derived PGS and clinical risk factors.

10

GENOME-WIDE META-ANALYSIS OF PSORIATIC ARTHRITIS LARGE BIOBANK AND POPULATION COHORTS

Aleksandra Judin[1], Anastasiia Alekseineko[5], Erik Abner[5], Galadriel Lucia Velazquez Silva[1], Tanel Traks[2], Triin Laisk[1], Estonian Biobank research team[5,*], Külli Kingo[3,4], Reedik Mägi[1], Ene Reimann[1]

[1]Institute of Genomics, University of Tartu, Estonia
[2]University of Tartu, Institute of Clinical Medicine, Clinical Research Centre
[3]University of Tartu, Institute of Clinical Medicine, Dermatology Clinic
[4]Tartu University Hospital, Clinic of Dermatology
[5]Estonian Genome Centre, Institute of Genomics, University of Tartu

[*]Andres Metspalu, Lili Milani, Tõnu Esko, Reedik Mägi, Mait Metspalu, Mari Nelis and Georgi Hudjashov

E-mail address of presenting author: aleksandra.judin@ut.ee

Background: Psoriatic arthritis (PsA) is a chronic immune-mediated inflammatory disease affecting the skin, joints, tendons, and entheses. Although genetic factors contribute to PsA susceptibility, its genetic architecture remains incompletely understood.

Study question: We aimed to identify PsA-associated genetic risk loci and prioritise candidate genes and biological pathways contributing to disease pathogenesis.

Methods: We performed a GWAS meta-analysis across six predominantly European cohorts: the Estonian Biobank, Million Veteran Program, FinnGen release 12, UK Biobank, the Aterido *et al.* 2019 PsA GWAS cohort, and the Soomro *et al.* 2022 multi-cohort PsA GWAS dataset. The analysis included 1,409,655 individuals: 15,794 PsA cases and 1,393,861 controls. Summary statistics were harmonised and meta-analysed using an inverse-variance weighted fixed-effect model in GWAMA. Downstream no-MHC analyses included variants present in at least three cohorts. Loci were defined using FUMA, gene-level analyses were performed with MAGMA, candidate genes were prioritised using FLAMES, and *HLA* alleles were analysed separately in the Estonian Biobank.

Main results: The meta-analysis identified 25,045 genome-wide significant variants and 68 genomic risk loci, including 31 putative novel PsA loci. Prioritised genes included *GADD45A, PTPRC, NEK7, REL, TP63, FYN, DDX58, FOXO1, AKAP13, SOCS1*, and *PTPN2*, implicating immune regulation, inflammatory signaling, skin biology, and

musculoskeletal pathways. The Estonian Biobank *HLA* analysis identified 19 Bonferroni-significant alleles, with the strongest association for *HLA-C*06:02*.

Limitations: Most participants were of European ancestry, and prioritised genes require functional validation. *HLA* analyses were not conditional or haplotype-based.

Wider implications: These findings expand the genetic landscape of PsA and support a shared immune-skin-joint axis in disease pathogenesis.

IMPROVING THE PREDICTABILITY AND INTERPRETATION OF POLYGENIC SCORES FOR BIG FIVE PERSONALITY TRAITS

Lisette Tagel[1], Uku Vainik[1,2], Estonian Biobank Research Team[2], René Mõttus[1,3]

[1]Institute of Psychology, University of Tartu, Tartu, Estonia
[2]Institute of Genomics, University of Tartu, Tartu, Estonia
[3]Department of Psychology, Edinburgh University, Edinburgh, United Kingdom

E-mail address of presenting author: lisette.tagel@ut.ee

Background: Typically, behavioural outcomes have poor prediction with polygenic scores (PGS). A recent genome-wide association study of >1M participants enabled creating PGS for the Big Five personality traits. Still, these PGSs predicted the traits' phenotypic scores with lower accuracy (*r*~.10-.19) than what is observed for other behavioural phenotypes such as educational attainment (*r*~.35). Further, the nature of the Big Five PGSes is unclear — what are the specific behaviours that they encompass?

Study question: We hypothesized that the poor prediction problem is partly due to a large proportion of random and systematic measurement error in typically-used (often brief) self-report questionnaire scores.

Methods: To overcome phenotypic measurement error, we combined self-reported (*N*=73,983) and informant-rated (*N*=21,986) personality data with Big Five PGS-s. To further increase predictability, we deconstructed the Big Five traits by allowing 198 single personality items to predict the Big Five PGS-s. Such personality-wide association study (aka pheWAS) allowed us to describe the specific traits most strongly captured by the Big Five PGSs.

Main results: All five PGS-s predicted their corresponding latent trait scores representing the shared and hence consensually valid variance of self- and informant-ratings (*r*~.10-.21; +12.5% over previous findings). Moreover, predicting PGS-s from individual items' consensually valid variance further increased accuracy (*r*~.12-.29; +50% over previous findings). Items most strongly correlating with each PGS unpacked their psychological interpretation, providing important context to the original genome-wide association study.

Wider implications: In summary, combining comprehensive multi-rater data suggested that the Big Five PGSs, likely to be used widely over the next few years, capture considerably more phenotypic personality trait variance than canonical

findings have suggested. A personality-wide pheWAS approach also aided in clarifying the psychological content of those PGSs.

External exposome and metabolite associations in 152,000 participants of the Estonian Biobank

Hanna-Maria Kukk[1,2,3], **Karmel Teder**[1], Estonian Biobank research team[1], Kees de Hoogh[4], Roel Vermeulen[4], Krista Fischer[1,2], Mara Delesa-Velina[2], Jaanika Kronberg[1,2]

[1]Institute of Genomics, University of Tartu
[2]Institute of Mathematics and statistics, University of Tartu
[3]Institute of Computer Science, University of Tartu
[4]Institute of Risk assessment Sciences, Utrecht University

E-mail address of presenting author: karmel.teder@ut.ee

Background and study question: Our aim was to study associations of external exposome, consisting of air pollution and built environment domains from the EXPANSE project[1], with metabolic markers quantified by Nightingale NMR-based platform[2].

Methods: We analysed 249 metabolites of 152,000 Estonian Biobank participants, and 19 exposome variables linked with participants' address at recruitment. External exposome and metabolite associations were calculated for each single exposure with all single metabolites using linear regression. Additionally, we considered principal components of air pollution and built environment domains with different levels of adjustment. Final model was adjusted for age, sex, smoking, education and time of day.

Main results: We identified 2,202 significant associations at Bonferroni-corrected p-value of 0.01. Most (241) metabolic features were associated with at least one exposome variable. Distance from sea, NO_2 and urban index were associated with largest number of metabolites, while metabolites associated with largest number of exposome variables were polyunsaturated fatty acids, linoleic acid percentage, ratio between polyunsaturated and monounsaturated fatty acids, total branched chain fatty acids, isoleucine and valine. Exposure-metabolite signatures were interpreted with enrichment for disease signatures.

Limitations: Main limitations are that the external exposures only include a small subset of environmental factors. Available exposure variables also only take into account the person's address and not time spent outside home. Personal exposures, such as indoor air quality, are not considered.

Wider implications: Metabolites as intermadiate phenotype are associated with environmental exposures and these signatures are associated with diseases, which is a valuable starting point for future causal studies of environmental exposures, with metabolites as potential mediators.

References

1. https://doi.org/10.1097/EE9.0000000000000162
2. https://doi.org/10.1038/s41586-026-10532-5

PRS-WIDE ASSOCIATION STUDY ANALYSIS PROVIDES INSIGHT INTO THE GENETIC BACKGROUND OF HORMONE SENSITIVITY-RELATED DIAGNOSES IN WOMEN

Anastasiia Bratchenko[*], Merli Koitmäe[*], Jelisaveta Džigurski, Reedik Mägi, Estonian Genome Centre, Kristi Läll[#], Triin Laisk[#]

Institute of Genomics, University of Tartu, Estonia

[*]Shared first authors
[#]Shared last authors

E-mail address of presenting author: anastasiia.bratchenko@ut.ee

Background: Fluctuations in sex hormone levels are a normal part of women's lives, but they are frequently associated with significant physical and psychological symptoms. Common manifestations include premenstrual symptoms (PMS), postpartum depression (PPD), vasomotor symptoms (VMS), and postmenopausal atrophic vaginitis (PAV), with most women experiencing one or more symptoms of these conditions during their lifetime.

Study question: The goal of the study was to investigate the association between hormone-sensitivity phenotypes and genetic predisposition to other traits using a PRS-wide association study approach.

Methods and main results: We applied data from the Estonian Biobank (N=136,791 women) and 4,793 polygenic risk scores from the PGS catalogue, and regression models were adjusted for age. After FDR correction (Benjamini-Hochberg method, $p<0.05$), VMS showed the strongest polygenic overlap (525 associations across 126 traits), followed by PAV (494 associations across 120 traits), PMS (315 associations across 83 traits), and PPD (18 associations across 8 traits). All hormone-sensitivity phenotypes were positively significantly associated with genetic predispositions to depression and neuroticism. Interestingly, opposite patterns were observed across hormone-sensitivity phenotypes. For example, genetic predisposition to obesity and coffee consumption were negatively associated with VMS and PAV but positively associated with PMS. Beyond depression and neuroticism, hormone-sensitive phenotypes showed substantial overlap with genetic predispositions related to body measurements, abdominal and back pain, and anxiety-related traits.

Limitations and wider implications: Although the observed associations do not establish causality, these findings provide insight into the genetic background of hormone-sensitive phenotypes and may inform future precision medicine.

A MULTI-ANCESTRY EVALUATION OF POLYGENIC SCORES ACROSS GLOBAL POPULATIONS

Elia Tiso[1], Linda Repetto[1], Brooke Wolford[2], Kristi Läll[1], Veronica Tozzo[3], Kristina Zguro[4], Masahiro Kanai[5], Ying Wang[6], Pedro M. De Macedo Gomes[7], Muhammad Khir A. Kohilan[8], Radja Badji[8], Kuang Lin[9], Shuifeng Huang[10], Benjamin Wingfield[11], Max Tamlander[12], Aoife McMahon[11], Yue Ji[11], Sarah Finer[13], Zhengming Chen[14], Kristian Hveem[15], Yukinori Okada[16], Bogdan Pasaniuc[3], Robin G. Walters[9], Derrick Bennett[9], David A. Van Heel[13], Ren-Hua Chung[17], Henrike O. Heyne[18], Alicia Martin[19], Nina Mars[4], Andrea Ganna[4], Samuli Ripatti[4], Stavroula Kanoni[13]

[1]University of Tartu, Estonia
[2]Norwegian University of Science and Technology, Norway
[3]David Ge en School of Medicine at UCLA, USA
[4]University of Helsinki, Finland
[5]Massachusetts General Hospital & Broad Institute of MIT and Harvard, USA
[6]Massachusetts General Hospital, USA
[7]University of Potsdam, Germany
[8]Qatar Precision Health Institute, Qatar
[9]University of Oxford, UK
[10]Institute of Population Health Sciences, National Health Research Institutes, Taiwan
[11]European Bioinformatics Institute, United Kingdom
[12]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Finland
[13]Queen Mary University of London, UK
[14]University of Oxford, UK
[15]NTNU, Trondheim, Norway
[16]Osaka University, Japan
[17]National Health Research Institutes, Taiwan
[18]Hasso Plattner Institute, Germany
[19]Massachusetts General Hospital, USA

E-mail address of presenting author: tiso@ut.ee

Background: Polygenic scores (PGS) are increasingly used for genetic risk stratification, but most have been developed and evaluated in European populations. The extent to which PGS effects and their age- and sex-specific heterogeneity generalise across ancestries remains unclear.

Study question: Using data from the INTERVENE consortium, we investigated the transferability and heterogeneity of PGS across diverse ancestry groups.

Methods: We analysed harmonised genomic and electronic health record data from over 1.3 million participants across 11 biobanks, including European, East and South Asian, African, Middle Eastern, and admixed American ancestries. Associations between PGS and incident disease outcomes were assessed within ancestry, age, and sex strata using cohort-specific Cox models, followed by meta-analysis to quantify heterogeneity within and between ancestries.

Main results: PGS were consistently associated with increased disease risk across populations, but effect sizes varied substantially by ancestry. Coronary heart disease (CHD) and type 2 diabetes (T2D) were used as representative common diseases, with similar patterns observed across other outcomes. For CHD, heterogeneity was greatest in European ancestry and lower in East Asian and African groups. T2D demonstrated stronger ancestry-specific variability, particularly among Europeans. Ancestry-normalised scores produced only modest changes and did not meaningfully reduce cross-ancestry heterogeneity.

Limitations: PGS excluded rare variants, harmonised phenotypes may overlook country-specific coding practices, available GWAS remain disproportionately European, and demographic or recruitment biases may affect estimates. Age-specific risks were also modelled using broad age categories.

Wider implications: These findings highlight the importance of evaluating PGS in diverse populations and developing improved methods to support equitable and accurate genetic risk prediction across ancestries.

BIDIRECTIONAL CAUSAL RELATIONSHIPS BETWEEN THYROID FUNCTION AND DEPRESSION ARE PREDOMINANTLY DRIVEN BY EARLY-ONSET MDD: A MENDELIAN RANDOMIZATION STUDY

Triinu Peters[1,2,3], Raphael Hirtz[4], Lars Dinkelbach[3,5], Anke Hinney[1,2,3], John Shorter[6]

[1]Section of Molecular Genetics in Mental Disorders, University Hospital Essen, Essen, Germany
[2]Center for Translational Neuro- and Behavioural Sciences, University Hospital Essen, Essen, Germany
[3]Institute of Sex and Gender-Sensitive Medicine, University Hospital Essen, Essen, Germany
[4]Helios University Medical Centre Wuppertal-Children's Hospital, Witten/Herdecke University, Wuppertal, Germany
[5]Department of Pediatrics II, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
[6]Department of Science and Environment, Roskilde University, Roskilde, Denmark

E-mail address of presenting author: triinu.peters@uni-due.de

Background: Previous MR studies reported inconsistent evidence on causal relationships between thyroid function and depression, largely using broadly defined MDD phenotypes.

Study question: We investigated whether causal relationships differ according to age at MDD onset.

Methods: We performed a bidirectional two-sample Mendelian randomization (MR) analysis. Genetic instruments were obtained from the largest available GWAS for thyroid traits and hypothyroidism and from the age-at-onset MDD GWAS (Shorter et al., 2025). Robust MR methods (Contamination Mixture, MR-Lasso, MR-APSS) were applied when heterogeneity was detected. Multivariable MR (MVMR) adjusted for BMI, C-reactive protein, interleukin-6, smoking initiation, and alcohol consumption was performed to assess direct effects.

Main results: Higher FT4 levels showed a causal effect on the risk of early-onset MDD (IVW: OR=1.09, p=0.010), remaining significant in MVMR analyses. A weaker causal association was observed for all-age MDD. Liability to hypothyroidism showed a causal effect on the risk of all-age MDD (IVW: OR=1.03, $p=4.9\times10^{-4}$), which remained robust in MVMR analyses. Conversely, liability to early-onset and all-age MDD showed a

causal effect on the risk of hypothyroidism: IVW: OR=1.09, $p=1.1\times10^{-8}$; OR=1.13, $p=3.2\times10^{-14}$, respectively. No evidence for causal effects was observed for normal-range TSH, FT3, or late-onset MDD.

Limitations: Two-sample MR assumes a linear association between exposure and outcome. Most GWAS participants were of European ancestry. FT3 analyses were underpowered.

Conclusion and wider implications: Our findings suggest that bidirectional causal relationships between thyroid function and depression appear to be largely driven by early-onset MDD. Thyroid monitoring in early-onset depression may be clinically relevant.

TRIANGULATING CAUSAL LINKS BETWEEN PERSONALITY AND HEALTH: MENDELIAN RANDOMISATION AND WITHIN-FAMILY REGRESSION ESTIMATES

Kerli Ilves[1,2]

[1]Institute of Genomics, University of Tartu
[2]National Institute for Health Development

E-mail address of presenting author: kerli.ilves@ut.ee

Background: Personality traits have replicable associations with a range of health behaviours and outcomes, but predominantly correlational evidence limits causal interpretation. Recent well-powered GWAS of the Big Five enables to apply genetically informed methods to further investigate these associations.

Study question: We investigate potential causal relationships between personality traits and key health behaviours related to chronic disease using genetic variants as instrumental variables.

Methods: We conducted two-sample bidirectional Mendelian randomisation (MR) analyses of Big Five personality traits and physical activity, BMI, smoking initiation and alcohol consumption. Genetic instruments were genome-wide significant, LD-pruned, and Steiger-filtered. We triangulated evidence across three complementary estimators: weighted mode, weighted median and MR-CAUSE. Associations were considered robust when estimates were directionally consistent and at least two 95% confidence intervals excluded zero. We further examined associations using within-family design with proband and parental polygenic indices (PGI) in the Estonian Biobank.

Main results: We found evidence for potential causal effects of personality on several important health behaviours, although effect sizes were generally small. For example, higher genetic liability to Extraversion increased alcoholic drinks per week, while Neuroticism showed a positive effect on BMI. Adjustments for parental PGIs attenuated some associations, suggesting that Neuroticism-BMI relationship may partly reflect indirect genetic or dynastic effects.

Limitations: MR lies on strong assumptions that are particularly challening to test for complex behavioural traits. Despite triangulation across methods, findings should be considered preliminary.

Wider implications: Genetically informed triangulation can clarify pathways linking personality and health. Distinguishing direct from familial effects helps to improve understanding of personality-health mechanisms and inform more targeted prevention strategies.

DISTINCT GENETIC AND PHENOTYPIC ASSOCIATIONS BETWEEN INATTENTION AND HYPERACTIVITY-IMPULSIVITY SYMPTOM DOMAINS AND EDUCATIONAL ATTAINMENT IN THE GENERAL ADULT POPULATION

Triinu Varvas[1], Elis Haan[1], Laura Hegemann[2,3], Natàlia Pujol-Gualdo[1,4], Sofiya Babok[1], Katri Pärna[1], Siim Kurvits[1], Kadri Kõiv[1], Melissa Vos[5], Martje Bos[5], Catharina Hartman[5], Helga Ask[2,6], Kelli Lehto[1]

[1]Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia
[2]PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway
[3]Psychiatric Genetic Epidemiology Group, Research Department, Lovisenberg Diaconal Hospital
[4]Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Catalonia
[5]University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), Groningen, the Netherlands
[6]PROMENTA Research Center, Department of Psychology, University of Oslo, Norway

E-mail address of presenting author: triinu.varvas@ut.ee

Background: Attention-deficit/hyperactivity disorder (ADHD) has been strongly associated with lower educational outcomes. Limited knowledge exists about relationship between ADHD symptomatology in the general population.

Study question: What are the genetic and phenotypic links between inattention (INA) and hyperactivity-impulsivity (HYP/IMP) symptom domains and educational attainment (EA) in adults?

Methods: We utilized the GWAS meta-analysis summary results of the Adult ADHD Self-Report Scale, conducted in the Estonian Biobank (EstBB) and the Norwegian Mother, Father and Child Cohort Study (N_total=141,505) for genetic correlations (r_g), MiXeR, Mendelian randomization (MR), and for polygenic score (PGS) predictions in an independent EstBB subset (N=72,589). Phenotypic analyses with educational levels were conducted in the EstBB sample (n=77,276).

Main results: There was negative r_g between EA and HYP/IMP (r_g=-0.28, 95% CI:-0.33,-0.22) and positive r_g between the INA domain (r_g=0.24, 95% CI:0.20,0.29). MR showed the same pattern in both directions. MiXeR revealed different proportions of

concordant effects within the shared component between EA (HYP/IMP=38%, INA=61%). HYP/IMP and INA PGSs showed dose-response associations with educational level in an independent sample, but again in the opposite directions ($PGS_{HYP/IMP}$: OR=0.94, 95% CI:0.92,0.95; PGS_{INA}: OR=1.05, 95% CI:1.04,1.07). On the phenotypic level, higher HYP/IMP was associated with lower education ($p<0.0001$), while INA displayed a U-shaped relationship, being higher in both basic and advanced level ($p=0.13$).

Limitations: HYP/IMP items primarily capture hyperactivity rather than impulsivity.

Wider implications: Emphasizing a link between higher self-reported inattention problems and advanced education in the general adult population. Importance of finer granularity on the symptom dimensions and the population when screening for ADHD features.

COMORBIDITIES AND GENETIC PREDISPOSITION FOR PELVIC ORGAN PROLAPSE

Valentina Rukins[1,*], Iveta Mikeltadze[2,3,*], Estonian Biobank Research Team[1], Kristiina Rull[2,4,#], and Triin Laisk[1,#]

[1]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
[2]Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
[3]Department of Surgical and Gynecological Oncology, Tartu University Hospital, Tartu, Estonia
[4]Women's Clinic, Tartu University Hospital, Tartu, Estonia

[*,#]Equal contribution

E-mail address of presenting author: valentina.rukins@ut.ee

Background: Pelvic organ prolapse (POP) is a prevalent gynecological disorder that significantly impairs quality of life. Despite its substantial clinical and societal impact, POP and its related disease burden remain largely underrecognized.

Study question: We aim to characterize POP-associated comorbidities across different life stages and assess the joint effect of POP diagnosis and genetic susceptibility on the overall comorbidty burden.

Methods: We leveraged electronic health records and genomic data of Estonian Biobank participants (137,440 women), including 10,483 cases with clinically confirmed POP and 126,957 controls. To characterize the comorbidity patterns across the life course, we conducted an age-stratified phenome-wide association study. We further assessed the interaction between the comorbidity burden, POP diagnosis and previously established polygenic score (PGS Catalog: PGS002288) using linear regression.

Main results: Across all age groups, 276 diagnoses were more frequent among cases ($P<0.05/13,594$). Age-stratified analysis revealed distinct age-dependent comorbidity patterns in women with POP. Among younger participants (<40 years), POP clustered with pregnancy- and delivery-related conditions, whereas in older age groups it was associated with disorders of genitourinary, digestive, and musculoskeletal systems. Comorbidity burden was higher in cases ($mean_{cases}=41.5$ vs $mean_{controls}=34.1$, $P=1.94\times10^{-279}$). However, we observed a negative interaction between POP status and PGS on comorbidity burden (interactive term -0.56±0.21, P=0.01).

Limitations: Digital records span approximately the past two decades, limiting pregnancy and delivery data for older participants. Moreover, they lack clinically relevant details, such as prolapse severity.

Wider implications: Our findings suggest that pelvic organ prolapse diagnosis may reflect broader systemic vulnerability to biologically related conditions.

GENETIC HAEMOCHROMATOSIS-ASSOCIATED MULTIMORBIDITY: PARALLEL ANALYSIS OF OUR FUTURE HEALTH AND UK BIOBANK COHORTS

Janice L Atkins[1], **Luke N Sharp**[1], Robin N Beaumont[1], João Delgado[1], Caroline F Wright[1], David J Hunter[2], Iain Turnbull[2], Jeremy Shearman[3], Luke C Pilling[1]

[1]Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, Devon, EX1 2LU, UK
[2]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[3]Department of Gastroenterology, South Warwickshire NHS Foundation Trust, UK

E-mail address of presenting author: l.sharp@exeter.ac.uk

Background: The iron-overload disease haemochromatosis is predominantly caused by *HFE* p.C282Y homozygosity (prevalence ~1 in 150 individuals of Northern European genetic ancestry) and is associated with excess morbidity, yet penetrance and expressivity are highly variable.

Study question: Analysing >1.2M individuals from the Our Future Health (OFH) and UK Biobank (UKB) cohorts can we determine the excess disease burden and multimorbidity attributable to *HFE* p.C282Y homozygosity?

Methods: Participants with genotype and health data from OFH (N=755,000) and UKB (N=480,000) included 3,536 and 2,890 p.C282Y homozygotes, respectively. We ascertained haemochromatosis diagnoses plus 86 long-term conditions from medical records. Sex stratified adjusted logistic regression estimated genotype-disease associations.

Main results: In OFH, cumulative incidence of haemochromatosis (within p.C282Y homozygotes) by age 80 was 56.5% in males and 48.1% in females (compared to 49.4% and 33.9% in UKB). Excess disease burden was identified, for example a quarter of osteoarthritis cases in p.C282Y homozygotes could be prevented if they had similar risk as non-homozygotes (OFH estimate=27.5%, 95% CI: 22.1-32.9%; UKB estimate=22.1%, 95% CI: 18.22-5.6%). p.C282Y homozygotes exhibited more multimorbidity with higher proportions with 2+, 3+, 4+, and 5+ common conditions by 60yrs compared to non-homozygotes. In all analyses the estimates were higher for males.

Limitations: Participation bias could lead to an underestimation of disease burden.

Wider implications: *HFE* p.C282Y homozygotes exhibited a significant excess burden of common diseases and multimorbidity. The attributable disease is preventable if risk in genotype carriers can be reduced. Additionally, the increased disease burden and non-uniform prevalence have implications for case-finding.

20

MOLECULAR LIPID SIGNATURES DEFINE METABOLIC SUBTYPES WITH DISTINCT CLINICAL AND GENETIC RISK PROFILES

Soubantik Sengupta[1], Mathias J. Gerl[2], Christian Klose[2], Matti Pirinen[1,3,4], Kai Simons[2], Elisabeth Widén[1], Samuli Ripatti[1,3,5], Rubina Tabassum[1]

[1]Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland
[2]Lipotype GmbH, Dresden, Germany
[3]Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[4]Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland
[5]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA

E-mail address of presenting author: soubantik.sengupta@helsinki.fi

Background: Reclassification of cardiometabolic risk by utilizing high-dimensional lipidomic profiles could offer better risk assessment for cardiometabolic diseases.

Study question: In 7,266 individuals, we asked whether — (1) plasma lipidomics identifies clinically distinct metabotypes, (2) polygenic scores (PGS) for cardiometabolic traits differ among metabotypes, and (3) genome-wide association study (GWAS) reveals genetic heterogeneity.

Methods: We utilized Self-Organizing Maps to cluster individuals using inverse-ranked transformed levels (adjusted for covariates) of 179 lipid species. Comparisons across metabotypes were performed using one-way ANOVA, t-tests, chi-square tests or regression models as appropriate. GWAS were performed using REGENIE with cluster membership as binary outcome.

Main results: We identified five metabotypes with distinct lipidomic profiles – C1 (high SMs, low TAGs); C2 (higher PCs, higher PCO-s); C3 (low PCs, low SMs); C4 (high PCs, low SMs); C5 (high PCO-s, low PCs). Significant differences in metabolic outcomes were observed across metabotypes (FDR<0.05). C2 had highest risk profile (mean BMI (27.88±5.21kg/m^2), type 2 diabetes (T2D) prevalence (9.07%) and hypertension prevalence (54.25%)) whereas C5 had the lowest risk with T2D and hypertension prevalence (2.49% and 35.29%, respectively). Consistently, C2 had higher PGSs for T2D and coronary artery disease while C5 had lower genetic risk for them (FDR<0.05).

GWAS analyses showed differences in association of known lipid loci including FADS1-3, APOE and others with metabotypes.

Limitations: Findings require external validation and longitudinal studies to establish generalizability and causality.

Wider implications: Integrating lipidomic and genomic profiles could unveil metabolic heterogeneity underlying cardiometabolic diseases.

COMPARISON OF TRADITIONAL CARDIOVASCULAR RISK FACTORS AND POLYGENIC RISK SCORE BETWEEN YOUNGER AND OLDER PATIENTS WITH FIRST-EVER MYOCARDIAL INFARCTION

J. Smirnova[1], T. Ainla[2], T. Marandi[2], M. Blöndal[3], M. Kals[4]

[1]University of Tartu
[2]North Estonia Medical Centre, Centre of Cardiology
[3]Tartu University Hospital, Department of Cardiology
[4]Estonian Genome Centre, Institute of Genomics, University of Tartu

E-mail address of presenting author: jana.smirnova@ut.ee

Background: Risk profiles differ between patients with early- and late-onset myocardial infarction (MI), highlighting the need to better define age-specific determinants, including genetic susceptibility. Such knowledge could improve the identification of high-risk young individuals and optimize primary prevention strategies.

Study question: How do traditional cardiovascular risk factors and polygenic risk for coronary artery disease differ between patients with early- and late-onset MI?

Methods: Data on incident MI cases (2012–2023) were retrospectively obtained by linking the Estonian Biobank (EstB, n≈210,000) and Estonian Myocardial Infarction Registry (EMIR, n≈30,000). Patients identified in both databases (n=3,031) were included. Clinical characteristics were obtained from EMIR and polygenic risk score (PRS) from EstB. Early- vs late-onset MI was defined using the European Society of Cardiology premature cardiovascular disease cut-offs: ≤55 years in men and ≤60 years in women. The multi-ancestry PRS_{CAD} developed by Patel et al. and validated in EstB by Puusepp et al. was used. Multivariable logistic regression estimated adjusted odds ratios (aORs) for individual risk factors and their differential association with MI across age groups. $p<0.05$ was considered statistically significant.

Main results: Among 3,031 participants, 780 had early- and 2,251 late-onset MI (Table 1). In multivariable analysis, current smoking showed the strongest association with earlyonset MI (aOR 3.84, 95% CI 3.14–4.70), followed by obesity (aOR 1.92, 1.49–2.46), high PRS (aOR 1.85, 1.51–2.27), and male sex (aOR 1.28, 1.05–1.56). Overweight, former smoking, and dyslipidemia did not differ significantly between groups. Diabetes (aOR

0.62, 0.47–0.79) and hypertension (aOR 0.40, 0.33–0.49) were associated with late-onset MI (Figure 1).

Table 1. Baseline characteristics with comparison of early- and late-onset myocardial infarction patients.

	Young n=780	**Old** n=2251	**p-value**
Age, mean (SD)	49.8 (6.3)	70.4 (8.7)	<0.001
Male sex	553 (71%)	1299 (58%)	<0.001
Dyslipidaemia	583 (75%)	1669 (74%)	0.78
Diabetes	100 (13%)	494 (22%)	<0.001
Hypertension	457 (59%)	1784 (79%)	<0.001
Smoking			<0.001
current	435 (56%)	524 (23%)	
former	115 (15%)	469 (21%)	
never	230 (29%)	1258 (56%)	
BMI			0.083
obese	311 (40%)	786 (35%)	
overweight	301 (39%)	932 (41%)	
normal	165 (21%)	517 (23%)	
underweight	3 (0.4%)	16 (0.7%)	
Previous PAD	18 (2.3%)	203 (9.0%)	<0.001
Previous stroke	16 (2.1%)	129 (5.7%)	<0.001
Previous CHF	51 (6.5%)	421 (19%)	<0.001
PRS_{CAD}			<0.001
0-90%	554 (71%)	1828 (81%)	
90-100%	227 (29%)	423 (19%)	

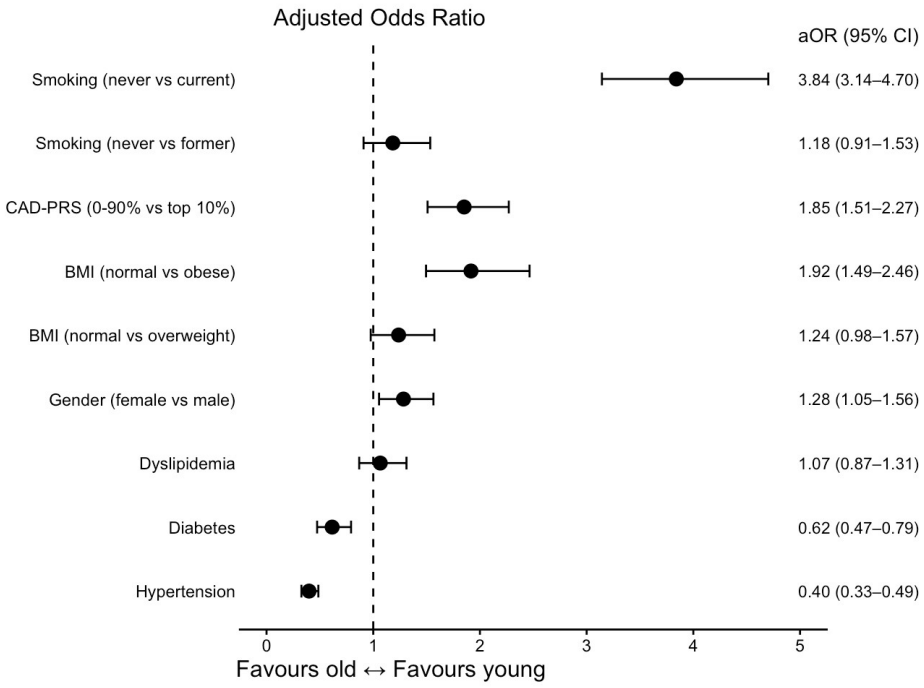


Figure 1. Differential risk factor profile for early- versus late-onset myocardial infarction. Data is presented as mean±standard deviation or count (percent). Early-onset – male ≤55y, female ≤60y; Late-onset – male >55y, female >60y; MI – myocardial infarction; BMI – body mass index: under 18.5kg/m2 considered underweight, 25 to 29.9kg/m2

considered overweight, 30.0kg/m2 and over considered obese; PAD – peripheral artery disease; CHF – chronic heart failure; PRSCAD – polygenic risk score for coronary artery disease, percentiles.

Limitations: Small group sizes, limiting statistical power and precision. Risk factors were assessed at the time of MI, prior exposure and exposure duration were unavailable. Some risk factors were self-reported (e.g., smoking), introducing potential reporting bias.

Wider implications: Incorporating polygenic risk into established clinical risk scores may improve identification of individuals at high risk for premature MI and support earlier, targeted prevention.

NOVEL LOCI AND MULTI-OMICS RISK MODELS FOR RHEUMATOID ARTHRITIS THROUGH A MILLION-PARTICIPANT GENOME-WIDE ASSOCIATION META-ANALYSIS

Galadriel Lucía Velázquez Silva[1], Jelisaveta Džigurski[1], Urmo Võsa[1], Nele Taba[1], Aare Märtson[2], Kaspar Tootsi[2], Katrin Ulst[2], Raili Müller[2], Estonian Biobank research team[1], Elin Org[1], Triin Laisk[1], Reedik Mägi[1], Kristi Läll[1,#], Ene Reimann[1,#]

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

[#]Equal contribution

E-mail address of presenting author: galadriel.lucia.velazquez.silva@ut.ee

Background: Rheumatoid arthritis (RA) has no cure and can cause irreversible joint damage, which a substantial number of patients already present with at the time of diagnosis, making early risk prediction valuable.

Study question: Which genetic factors contribute to RA and seropositive RA risk, and can genomics and metabolomic integration improve risk prediction beyond established clinical factors?

Methods: In this work, genome-wide association study meta-analyses were performed for RA and seropositive RA, comprising approximately one million participants of European ancestry. Novel disease-specific polygenic risk scores (PRSs) were constructed and evaluated over traditional models. In parallel, integrating metabolomic data into high-dimensional models was evaluated.

Main results: Eight and six novel genomic risk loci were defined for RA and seropositive RA, and candidate causal genes were identified, highlighting relevant biological pathways, including established immune pathways and estrogen metabolism. The novel disease-specific PRSs enhanced the predictive performance over clinical risk factors (incremental C-statistics of 2.7 and 5.1 for RA and seropositive RA, respectively). In addition, multi-omics risk scores integrating genomics, clinical risk factors and metabolomics enhanced risk stratification over models based on clinical risk factors and genomics, particularly for seropositive RA, where the hazard ratio of the highest decile increased from 4.9 to 5.7.

Limitations: This study is based on European ancestry populations, potentially limiting generalisability to other populations. Metabolite levels were measured only at baseline, which may limit the long-term predictive accuracy of metabolomic models.

Wider implications: These findings expand the understanding of genetic factors underlying RA and support the value of including PRSs in risk assessment, while suggesting metabolomic integration may further enhance risk stratification, particularly for seropositive RA.

DOES NAD PLAY A CAUSAL ROLE IN AGING AND COGNITION? A MENDELIAN RANDOMISATION STUDY

Yehor Chudovskyi[1], Uku Vainik[1,2,3]

[1]Institute of Psychology, University of Tartu
[2]Institute of Genomics, University of Tartu
[3]Department of Neurology and Neurosurgery, McGill University, Canada

E-mail address of presenting author: yehor.chudovksyi@ut.ee

Background: Nicotinamide adenine dinucleotide (NAD) levels are reported to decline with age, purportedly affecting cognitive function. There is interest in NAD-replenishing therapies for combating these effects. Yet, the current scientific backing of such treatments lacks quality.

Study question: In this study, we seek to evaluate the effects of NAD levels on cognition and its decline using Mendelian Randomization (MR).

Methods: In the absence of NAD levels GWAS we used GWAS summary statistics for levels of NAD metabolites, mostly in blood plasma, blood plasma pQTL and brain eQTL GWAS of proteins involved in NAD metabolism to indirectly estimate NAD's effects on cognition and cognitive decline. We utilized Alzheimer's (AD) and Parkinson's disease (PD), frailty (the poor cognition component) as cognitive decline outcomes, and Intelligence as age-independent cognition outcome. Metabolome MR was conducted both in genome-wide and cis-regime (using variants in NAD-related genes and their neighborhoods); in pQTL and eQTL MR only cis-loci were used. Where applicable, we tested a number of variations of the MR method.

Main results: We identified that only quinolinate, 2PY and 4PY metabolites show causal effect on the studied outcomes, almost exclusively on PD. The underlying variants are likely pleiotropic, sitting in the ACMSD region known to influence PD's risk not via NAD levels. Only CD38 protein shows causal effects on PD, an association previously known.

Limitations: Used exposures have limitations, like low power (metabolites); measurements from blood may not reflect brain level and age is not corrected for in any exposures.

Wider implications: We suggest conducting research with tissue-specific, well-powered GWAS of NAD levels.

COVARIATE-AWARE GENOMIC PREDICTION OF BLOOD METABOLITE PROFILES USING MULTI-TASK NEURAL NETWORKS

Merve Nur Güler[1], Estonian Biobank Research Team[1], Maris Alver[1], Toomas Haller[1], Flora Jay[2], Luca Pagani[1,3], Lili Milani[1], Burak Yelmen[1]

[1]Institute of Genomics, University of Tartu, Tartu, Estonia
[2]CNRS, INRIA, LISN, Paris-Saclay University, Orsay, France
[3]Department of Biology, University of Padova, Padova, Italy

E-mail address of presenting author: merve.nur.guler@ut.ee

Background: Circulating metabolites reflect physiological processes and are linked to many complex diseases. Although genome-wide association studies have identified numerous metabolite-associated variants, these associations do not directly show how well metabolite profiles can be predicted from genotype and routinely available covariates.

Study question: Can multi-task neural networks improve blood metabolite prediction compared with standard linear models, and which sources of information drive any improvement?

Methods: Using Estonian Biobank genotype data at 13,049 selected variants, routinely available covariates and NMR metabolomics data, we developed a three-stage multi-task neural network that separately models covariate, genetic and residual joint covariate-genotype effects. We compared it with elastic net, single-task neural networks and an activation-free multi-task model using held-out data.

Main results: The multi-task neural network achieved the highest mean predictive performance across 109 metabolites (R^2=0.219), compared with the single-task neural network (R^2=0.211), elastic net (R^2=0.207) and activation-free multi-task model (R^2=0.191). Decomposition showed that most of the gain came from nonlinear covariate modelling, while genetic and joint covariate-genotype contributions were smaller and varied across metabolites.

Limitations: The analysis was performed in a single biobank and metabolomics platform, so external validation is needed. The selected genetic predictors may also miss part of the polygenic signal.

Wider implications: Multi-task learning provides a compact framework for metabolite prediction, but the results also show that improved performance does not necessarily come from more complex genetic effects. Separating covariate and genetic contributions may make deep-learning models easier to interpret and evaluate in genomic prediction studies.

REDUCED VIRAL DIVERSITY AND ENTEROCOCCACEAE-PHAGE ENRICHMENT DISTINGUISH THE PRETERM INFANT GUT VIROME

Radko Avi, Taavi Päll, Anni Lember, Ene-Ly Jõgeda, Eveli Kallas, Merit Pauskar, Tuuli Metsvaht, Irja Lutsar, Hiie Soeorg, Kristi Huik

Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

E-mail address of presenting author: radko.avi@ut.ee

Background: Bacterial gut colonisation differs between preterm and term infants, yet the corresponding neonatal virome remains poorly characterised. Bacteriophages can shape gut bacterial composition, but their differentiation between these groups is largely unmapped.

Study question: Do the gut viromes of preterm and term neonates differ in composition, diversity, and bacterial host-family associations?

Methods: We analysed gut viromes from faecal bulk shotgun metagenomes in 57 neonates (term n=20, preterm n=37; at week 3-4; 30 mln Illumina PE150 reads each). Viral contigs were identified with VirSorter2, VIBRANT, geNomad, quality-filtered with CheckV, and clustered to form a study vOTU catalogue. Hosts and taxonomy were assigned with iPHoP and UHGV, respectively. Analyses included alpha- and beta-diversity (PERMANOVA, PERMDISP) and per-vOTU prevalence testing (Fisher exact).

Main results: The vOTU catalogue consisted of 823 viral contigs. Term infants harboured over twice as many vOTUs per sample as preterm (median 29 vs 13; $p<0.05$). Beta-diversity separated groups (Jaccard $R^2=4.2\%$, Bray-Curtis $R^2=8.5\%$, both $p=0.0001$), with preterm samples more dispersed (PERMDISP $p<0.05$). Fisher testing identified 26 differential vOTUs (BH $p<0.05$), 21 term-exclusive. Termassociated vOTUs were depleted in Enterobacteriaceae phages ($p<0.05$); the two preterm-enriched vOTUs were Enterococcaceae phages ($p=0.017$). Most term-specific vOTUs were taxonomically unassigned.

Limitations: The cohort is cross-sectional with one sample per infant, precluding temporal inference; many vOTUs remain unassigned and host predictions are computational.

Wider implications: Preterm and term viromes differ quantitatively and qualitatively. Preterm signals were dominated by phages targeting NICU-associated Enterococcus, with potential as dysbiosis biomarkers in at-risk neonates.

INVESTIGATING THE EVOLUTIONARY DYNAMICS UNDERLYING DNA VIRUS IMMUNE RESPONSE

Rutvi Rajpara[1], Erik Abner[1], Monika Karmin[1], Irene Gallego Romero[1,2], Michael Dannemann[1]

[1]Institue of Genomics, University of Tartu, Estonia
[2]Human Genomics and Evolution, St Vincent's Institute of Medical Research, Fitzroy, Australia

E-mail address of presenting author: rutvi.rajpara@ut.ee

Background: DNA viruses such as herpes viruses, and polyomaviruses, can persist in humans and establish infections for the lifetime. Although these infections are usually asymptomatic and the amount of viral DNA present in blood can vary widely between individuals. Higher viral loads have been linked to immune dysfunction and disease risk. Recent large biobank studies using whole-genome sequencing (WGS) have shown that viral load is shaped by demographic factors and host genetics, especially variation in the HLA region (Kamitaki N et al. 2026). Our group's recent analyses have shown that archaic introgression has played a role in shaping in virus load associated loci, with multiple Neandertal and Denisovan derived haplotypes on chromosome 17 and within chromosome 6 MHC region being associated with DNA virus levels in participants of the UK Biobank (Rajpara et al. 2026).

Study question and methods: The objective of this project is to apply this viral-load and host-genetics framework to the Estonian Biobank using long-read sequencing data. By integrating viral load quantification with GWAS and archaic haplotype mapping, we aim to test whether how population-specific evolutionary histories have shaped host–virus interactions in Northern Europe.

Main results: The findings of this research will yield novel insights into the relationship between DNA viruses, viral load, and disease-associated variants in present-day Northern European individuals, highlighting how historical evolutionary events shape human health and disease susceptibility in this population.

Wider implications: Understanding the evolutionary context of immune variation can help future studies in personalized medicine and infectious disease risk prediction.

REORGANISED, NOT EXPANDED: GENOME-RESOLVED METABOLIC POTENTIAL AND VIROME ACROSS AGES 12–109

Taavi Päll, Radko Avi, Anni Lember, Ene-Ly Jõgeda, Eveli Kallas, Merit Pauskar, Epp Sepp, Imbi Smidt, Tiiu Rööp, Jelena Štšepetova, Siiri Kõljalg, Irja Lutsar, Reet Mändar, Kristi Huik

Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

E-mail address of presenting author: taavi.pall@ut.ee

Background: Gut microbiota composition is considered a determinant of lifespan. Extreme old age carries a reproducible signature (Lachnospiraceae depleted, *Akkermansia* and methanogens enriched), and a functional gain is assumed but rarely checked against genomes.

Study question: Does the ageing gut community gain metabolic genes or merely exchange them? Do the credited taxa encode them, and do phage genes replace losses?

Methods: We assembled 620 gut metagenomes from 42 studies: 315 aged 12–64, 246 aged 65–99 and 59 centenarians (≥100). We annotated KEGG orthologues, CAZymes and butyrate synthases on metagenome-assembled genomes, and recovered the virome de novo. Age was modelled continuously, adjusted for depth and study.

Main results: The canonical signature reproduced. Richness is depth-dominated at every rank; within study at matched depth only family-level richness rises in the oldest-old (1.18), not species. 57/102 taxon–function pairs validated. Metabolic potential was reshaped, not expanded. Fibre degradation fell in 8 of 13 carbohydrate classes, none rising, specifically the plant fibres (xylan, pectin, cellulose), while mucin and peptidoglycan held. B-vitamin biosynthesis fell in 9 of 26 genes, steepest at the pyridoxal-5′-phosphate (B6) route gut anaerobes use, the opposite of the predicted cobalamin gain. Acetate handling rose, butyrogenesis stayed flat. Every virome contrast was flat, with no fibre or vitamin auxiliary genes, instead mirroring bacterial structure (Mantel r=0.65).

Limitations: Cross-sectional and survivorship-biased; centenarians from three countries. Gene content indicates potential, not flux.

Wider implications: The ageing gut loses fibre degradation and B-vitamin biosynthesis; both are diet-modifiable, making them candidate targets for nutritional support in old age. The passenger virome will not buffer them.

POLYGENIC RISK LINKS TO THE GUT MICROBIOME: THE MAGI CATALOG

Oliver Aasmets[1], Jelisaveta Džigurski[1], Nele Taba[1], Merli Koitmäe[1,2], Estonian Biobank Research Team[1], Triin Laisk[1], Elin Org[1], Reedik Mägi[1], Kristi Läll[1]

[1]Institute of Genomics, Estonian Genome Centre, University of Tartu, Estonia
[2]Institute of Mathematics and Statistics, University of Tartu, Estonia

E-mail address of presenting author: oliver.aasmets@ut.ee

Background: Polygenic risk scores (PRSs) effectively identify individuals at risk for various health conditions, yet their associations with the gut microbiome remain uncharacterized.

Study question: What is the association between cumulative genetic liabilities across diverse traits and diseases and the composition of the gut microbiome?

Methods: We systematically analyzed associations between 4,794 PRSs covering 615 traits and gut microbiome profiles in the Estonian Microbiome Cohort (N>2,500). Linear regression models evaluated specieslevel abundance and diversity, followed by bidirectional mediation analysis to explore causal architecture.

Main results: Microbiome diversity was associated with 62 distinct PRSs across 10 traits. At the species level, 282 associations were identified across 100 PRSs for 34 traits, with triglyceride measurements, glucose regulation, and chronotype PRSs showing the strongest signals. Mediation analysis revealed that host genetic liability primarily influences microbiome composition, such as genetically driven hypertriglyceridemia suppressing microbial richness, though reciprocal secondary microbiome-mediated effects also occur.

Limitations: Available sample size constraints limit the detection of smaller effect sizes. Furthermore, potential sample overlap across external cohorts and cross-sectional microbiome measurements complicate replication and restrict direct causal interpretations.

Wider implications: These results help define early microbial biomarkers and explain inter-individual variability in microbiome composition. The findings are accessible through the interactive MAGI Catalog to advance personalized risk estimation and support future causal microbiome research.

PARTICIPANT RECALL TO EXPLORE THE GUT MICROBIOME-ENDOTOXEMIA AXIS AS A DRIVER OF CHRONIC INFLAMMATION IN ARTHRITIS

Kreete Lüll, Ene Reimann, Galadriel Luzia Silva, Reedik Mägi, Elin Org

Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

E-mail address of presenting author: kreete.lull@ut.ee

Background: Arthritis is characterized by chronic inflammation, but the mechanisms driving persistent immune activation remain incompletely understood. Alterations in the gut microbiome may contribute to systemic endotoxemia through translocation of bacterial products such as lipopolysaccharide (LPS), thereby promoting inflammation. Understanding interactions between the gut and oral microbiomes, endotoxemia, and inflammatory pathways may reveal novel biomarkers and disease mechanisms in arthritis.

Study question: What relationships exist between the gut and oral microbiomes, circulating LPS levels, and markers of chronic inflammation in individuals with arthritis, and which microbial features are associated with future changes in systemic endotoxemia?

Methods: Participants with arthritis (n=1,000) are being recalled from the Estonian Microbiome (EstMB) cohort within the Estonian Biobank. Stool, saliva, and blood samples, along with questionnaire data, are collected at recall, providing microbiome- and endotoxemia-related measurements at two time points. Microbiome composition and function, circulating LPS, and inflammatory markers will be analyzed using longitudinal approaches.

Main results: Participant recall and sample collection are currently underway. The study will generate a longitudinal dataset integrating clinical information with oral and gut microbiome profiles, circulating LPS concentrations, and inflammatory markers.

Limitations: As recruitment and sample collection are ongoing, no analytical results are currently available. The observational design will allow identification of associations but may limit causal inference regarding microbiome-related mechanisms contributing to endotoxemia and inflammation.

Wider implications: This study may identify microbial features associated with future endotoxemia and provide insights into microbiome-related mechanisms underlying chronic inflammation in arthritis, supporting the development of novel biomarkers and therapeutic targets.

MICROBIAL SIGNALS PRECEDING DEPRESSION ONSET: PRELIMINARY FINDINGS FROM THE ESTONIAN MICROBIOME COHORT

Kadi Vaher[1], Annabel Klemets[1], Reidar Andreson[2], Oliver Aasmets[1], Elin Org[1]

[1]Institute of Genomics, University of Tartu
[2]Institute of Molecular and Cell Biology, University of Tartu

E-mail address of presenting author: kadi.vaher@ut.ee

Background: Depression is a leading cause of disability worldwide. Depression associates with gut microbiome composition, but existing evidence comes from cross-sectional case-control studies, which cannot distinguish cause from consequence, or effects of treatment and lifestyle changes. This leaves an important gap: which microbial features, if any, predict future depression onset, and could potentially be causally involved in depression pathophysiology.

Study question: Which microbiome features are associated with future, incident depression onset?

Methods: Using microbiome sequencing and prospective electronic health record data from the Estonian Microbiome Cohort, we applied time-to-event models to test associations between microbial species, gut-brain modules, and gut metabolic modules, and incident depression (ICD-10 diagnosis codes F32/F33/F41.2 and/or antidepressant purchase [ATC N06A] after sampling). Individuals with depression, bipolar disorder, psychosis, or antidepressant use prior to sampling were excluded. Sensitivity analyses used a stricter case definition that excluded cases identified solely by antidepressant purchase.

Main results: No individual species were associated with incident depression after multiple-testing correction. Several functional modules, including those involved in propionate and glutamate metabolism, were nominally associated. Urea degradation was the only module significantly associated with incident depression after multiple-comparison correction, with a consistent negative association across case definitions and models.

Limitations: Depression case status was not validated via clinical interviews, functional modules were inferred from sequencing data rather than measured directly, and the observational cohort design cannot establish causality.

Wider implications: Microbial metabolic functional capacity, rather than abundances of individual species, may be more relevant to depression risk, and implicate urea metabolism as a candidate pathway. Replication in external Nordic cohorts and meta-analysis are underway, alongside machine learning analyses to identify predictive combinations of microbial features.

NICHE OCCUPATION BY UNCULTIVATED RFN20 BACTERIA IN THE HUMAN GUT

Kateryna Pantiukh, Elin Org

Institute of Genomics, Estonian Genome Centre, University of Tartu, Tartu, Estonia

E-mail address of presenting author: pantiukh@ut.ee

Background: Metagenome-resolved genome reconstruction has substantially expanded our understanding of bacterial diversity by recovering genomes from previously unknown and uncultivated microorganisms. One example is the RFN20 lineage within the Bacilli, which currently comprises exclusively uncultivated bacteria known only from reconstructed genomes. In parallel, the rapid expansion of population biobanks provides access to large microbiome datasets linked to extensive phenotypic and clinical metadata, creating new opportunities to investigate the ecology of uncultivated microorganisms.

Study question: We investigated which ecological and genomic features may explain niche occupation advantage of uncultivated bacteria from RFN20 lineage.

Methods: Genome reconstruction and downstream analyses were performed using 1,878 short-read and 16 long-read metagenomic samples from the Estonian microbiome cohort.

Main results: RNF20 relative abundance was tested against age, sex, body mass index, disease status, and dietary variables. Significant associations were detected with age and bread consumption. Long-read sequencing enabled reconstruction of eight high-quality, single-contig genomes. We characterised their core metabolic potential and screened for antiphage and defence systems, virulence factors, and antimicrobial-resistance genes. Comparative genomic analyses were used to assess within-lineage variation and identify differences from closely related taxa.

Limitations: Genome-based analyses mainly generate hypotheses and require experimental validation in pure cultures.

Wider implications: Understanding how RFN20 bacteria colonise gut communities may reveal mechanisms of niche occupation and persistence and improve engraftment of beneficial bacteria.

MULTI-SITE MICROBIOME MAPPING ALONG THE COLORECTAL NEOPLASIA CONTINUUM

Kertu Liis Krigul[1], Andri Jääger[2,3], Kalle Kisand[4], Annabel Klemets[1], Hendrik Laja[5], Heigo Reima[2], Jaan Soplepmann[2,3], Riina Salupere[4,5], Margus Lember[6,7] and Elin Org[1]

[1]Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia
[2]Department of Surgical and Gynaecological Oncology, Tartu University Hospital, Estonia
[3]Department of Hematology and Oncology, Institute of Clinical Medicine, University of Tartu, Estonia
[4]Department of Gastroenterology, Institute of Clinical Medicine, Faculty of Medicine, University of Tartu, Estonia
[5]Department of Gastroenterology, Tartu University Hospital, Estonia
[6]Department of Internal Medicine, Institute of Clinical Medicine, University of Tartu, Estonia
[7]Department of Internal Medicine, Endocrinology and Rheumatology, Tartu University Hospital, Estonia

E-mail address of presenting author: kertu.krigul@ut.ee

Background: Colorectal cancer (CRC) is a major public health challenge and successful treatment depends on the stage at diagnosis. CRC-specific stool microbiome signatures have been proposed as potential markers for CRC detection. However, how CRC-associated microbial patterns vary across anatomical sites within the same individual and along the adenoma–carcinoma sequence remains incompletely understood.

Study question: Can integrated profiling of saliva, stool, tumor, and paired adjacent tissue identify site-specific and disease-associated microbial signatures in healthy controls, patients with precancerous lesions, and patients with CRC?

Methods: The clinical cohort included 148 well-phenotyped participants (19 healthy controls, 74 patients with precancerous lesions, 55 patients with CRC) from Estonia. In total, 510 saliva, stool, non-tumor colonic tissue, and tumor tissue microbiome samples were collected. Differences in microbiome community composition and taxon abundance were evaluated using long-read 16S gene sequencing across sample types and disease stages, including within-subject multi-site comparisons. A curated set of previously reported CRC-associated taxa was used to assess site- and disease stage-specific patterns.

Main results: Distinct microbial community patterns were observed across disease groups and sample types. Previously reported CRC-associated genera showed site- and cancer-stage-specific abundance patterns. Paired tumor and adjacent non-tumor tissue sampling enabled assessment of lesion-associated microbial differences while accounting for inter-individual variation.

Limitations: Sample numbers were unequal across disease groups and sample types, and some paired CRC tissue comparisons had limited statistical power.

Wider implications: Multi-site microbiome profiling provides a more comprehensive view of microbial changes accompanying colorectal neoplasia and may help identify complementary non-invasive and tissue-associated microbial markers for future CRC risk stratification and early detection.

DRUG-NAIVE DE NOVO PARKINSON'S DISEASE MICROBIOME PROFILING BY 16S RRNA SEQUENCING

Gizi Golt[1], Kairi Koort[1], Toomas Toomsoo[1,2]

[1]School of Natural Sciences and Health, Tallinn University, Tallinn, Estonia
[2]AS Arstikeskus Confido (Confido Medical Centre), Tallinn, Estonia

E-mail address of presenting author: gizigolt@tlu.ee

Background: Gut dysbiosis is implicated in Parkinson's disease (PD), but most microbiome studies use treated cohorts, where dopaminergic therapy, disease duration, constipation, diet and lifestyle can independently reshape the microbiota. Reported PD-associated differences may partly reflect treatment rather than disease. Drug-naive, de novo cohorts capture the earliest stage, before confounding.

Study question: In treatment-naive, newly diagnosed PD patients, is the gut microbiome organized into discrete community types, and is composition associated with clinical or lifestyle variables?

Methods: We analysed fecal 16S rRNA V3-V4 amplicon sequencing from 35 drug-naive de novo PD patients, relating genus-level taxonomy to clinical metadata. Alpha diversity used Chao1, Faith's, Shannon and Simpson indices; Bray-Curtis beta diversity was assessed by NMDS ordination, PERMANOVA with beta-dispersion testing, and genus-level clustering.

Main results: The cohort showed inter-individual heterogeneity (median Chao1 443.3, Shannon 3.84). Composition was continuous rather than clustered: partitioning around medoids gave a single optimal cluster (maximum silhouette 0.21) and hierarchical clustering showed no separable subgroups. Age was the strongest univariate correlate ($R^2=0.054$, $p=0.009$), but no model survived false-discovery-rate correction ($q>=0.16$). Blautia, Faecalibacterium and Lachnospiraceae dominated, with patient-specific enrichment of rarer genera.

Limitations: The cohort is small, cross-sectional, and lacks healthy-control, diet and body-mass-index models. 16S V3-V4 gives genus-level resolution only, missing strain-level SCFA production and levodopa-metabolising capacity. Several beta-dispersion tests were significant, so PERMANOVA signals need cautious interpretation.

Wider implications: Rather than a single dysbiosis signature, early untreated PD shows a heterogeneous, continuously structured microbiome. Because these measurements precede dopaminergic therapy, they provide a treatment-free reference against which changes reported in medicated cohorts can be re-examined.

ANTICANCER AND ANTIMICROBIAL PROPERTIES OF SILYMARIN FROM MILK THISTLE (*SILYBUM MARIANUM*): PHYTOCHEMICAL ANALYSIS AND BIOACTIVITY EVALUATION

Mohmmad Asif Gawhari, Baljinder Kaur

Department of Biotechnology and Food Technology, Punjabi University, Patiala, Punjab, India

E-mail address of presenting author: mohdasif_rs24@pbi.ac.in

Background: Silymarin is a polyphenolic flavonoid mixture isolated from milk thistle (*Silybum marianum*) and is responsible for the plant's hepatoprotective effects. Silymarin is a hepatoprotective flavonoid drug available as a biomarker in *Silybum marianum*. It is used to treat various liver diseases of different origins due to these properties. Phytochemicals are playing a vital role in treating a range of diseases and are used in both traditional and modern medicine. Phytochemical analysis of milk thistle seed extract shows the plant is rich in secondary metabolites, including high levels of total phenolics, flavonoids, and antioxidants.

Study question: This study seeks to determine whether the ethanol seed extract of *Silybum marianum* exhibits significant antibacterial activity against Gram-positive and Gram-negative bacteria, and to what extent it demonstrates anticancer effects on liver cancer cell lines. It also aims to elucidate the phytochemical basis underlying these bioactivities.

Aim: The study aims to evaluate the anticancer and antibacterial activities of the ethanol seeds.

Methods: The antibacterial activity of the ethanol seed extract was tested against Gram-positive (*Staphylococcus aureus*) and Gram-negative (*Salmonella enterica*) bacteria using a modified Kirby-Bauer disc diffusion technique to determine the zone of inhibition. Anticancer activity of milk thistle was evaluated on liver cancer cell lines (HepG2).

Main results: The ethanol seed extract of milk thistle showed a strong inhibitory zone against both bacteria compared to the control. Additionally, the seed extract demonstrated significant anticancer activity.

Limitations: The primary limitation of this study is the focus on in vitro experiments, which may not fully replicate in vivo conditions. Additionally, the use of a single extraction method and limited bacterial and cancer cell models may restrict the generalizability of the findings.

Wider implications: Milk thistle may be developed into affordable, safe, standardized herbal products and could serve as a source of new, broad-spectrum anticancer and antimicrobial agents.

Keywords: Anticancer, Antioxidant, Antibacterial, Phenolic components, Flavonoid.

References

1. Kaur, A. K., Wahi, A., Brijesh, K., Bhandari, A., and Prasad, N., 2011. "Milk thistle (*silybum marianum*): A review." *International Journal of Pharma Research and Development*, vol. 3, pp. 1-10.
2. Gurib-Fakim, A., and Mahomoodally, M. F., 2013. "African flora as potential sources of medicinal plants: Towards the chemotherapy of major parasitic and other infectious diseases: A review." *Jordan Journal of Biological Sciences*, vol. 147, pp. 1-8.
3. Abed, I. J., Al-Moula, R., and Abdulhasan, G. A., 2015. Antibacterial effect of flavonoids extracted from seeds of silybum marianum against common pathogenic bacteria." *World Journal of Experimental Biosciences*, vol. 3, pp. 36-39. Available: https://doi.org/10.13140/RG.2.1.4632.6240.
4. Sayyah, M., Boostani, H., Pakseresht, S., and Malayeri, A., 2010. "Comparison of silybum marianum (l.) gaertn. With fluoxetine in the treatment of obsessive–compulsive disorder." *Progress in Neuro Psychopharmacology and Biological Psychiatry*, vol. 34, pp. 362-365.
5. Waweru, W. R., Osuwat, L. O., and Wambugu, F. K., 2017. "Phytochemical analysis of selected indigenous medicinal plants used in Rwanda." *Journal of Pharmacognosy and Phytochemistry*, vol. 6, pp. 322-324.

Acknowledgment: I sincerely acknowledge the support and guidance of my supervisor and the Department of Biotechnology and Food Technology, Punjabi University, Patiala, for facilitating my research work. I also express my sincere gratitude to the Indian Council for Cultural Relations (ICCR) for funding my search.

SPATIAL ORGANIZATION OF CANCER-ASSOCIATED FIBROBLAST SUBTYPES IN GASTROINTESTINAL TUMORS

Violeta Lisovska[1], Helvijs Niedra[1], Marta L. Springe[1], Katrine Tiltina[1], Raitis Peculis[1], Rihards Saksis[1], Jurijs Nazarovs[2,3], Arturs Ozolins[2], Sofija Vilisova[2], Natalja Senterjakova[2], Aija Gerina[2], Ilze Konrade[4,5], Aldis Pukitis[2], Vita Rovite[1]

[1]Department of Molecular and Functional Genomics, National Institute of Research and Innovation, Riga, LV-1067, Latvia
[2]Pauls Stradins Clinical University Hospital, Riga, LV-1002, Latvia
[3]Department of Pathology, Riga Stradins University, Riga, LV-1007, Latvia
[4]Department of Endocrinology, Riga East Clinical University Hospital, Riga, LV-1079, Latvia
[5]Department of internal diseases, Riga Stradins University, Riga, LV-1007, Latvia

E-mail address of presenting author: violeta.lisovska@niri.lv

Background: Cancer-associated fibroblasts (CAFs) are key components of the tumor microenvironment that influence tumor progression, immune responses, and therapeutic outcomes. However, their spatial organization and functional diversity in gastroenteropancreatic neuroendocrine tumors (GEP-NETs) remain poorly understood.

Study question: Do GEP-NETs from different anatomical sites share conserved stromal architecture and CAF transcriptional states?

Methods: Eight treatment-naïve primary GEP-NETs (five pancreatic, one appendiceal, one colorectal, and one biliary) were profiled using 10x Genomics Visium HD spatial transcriptomics. Following image-based cell segmentation, quality control, and integration of stromal compartments, fibroblasts were classified into myofibroblastic (myCAF), complement-secreting (csCAF), inflammatory (iCAF), and antigen-presenting (apCAF) states using literature-derived gene signatures.

Main results: Stromal abundance varied from 8–38% across tumors. Despite this heterogeneity, integrated analysis identified ten conserved stromal clusters shared between tumors. myCAFs predominated across most samples, with csCAFs representing the second-largest population, while iCAFs and apCAFs were less abundant. Spatial mapping localized myCAFs to collagen-rich stroma and csCAFs to tumor-adjacent stromal interfaces, indicating distinct functional niches.

Limitations: The study is limited by its small cohort.

Wider implications: GEP-NETs exhibit conserved stromal transcriptional programs despite marked inter-tumoral heterogeneity. These findings provide a framework for understanding NET stromal biology and support future investigation of CAF-targeted therapeutic strategies.

HIGH-RESOLUTION SPATIAL TRANSCRIPTOMICS IN HUMAN ATHEROSCLEROTIC PLAQUES: DISSECTION OF ATHEROSCLEROTIC PLAQUE MICROANATOMY

Aydin Bölük[1], Mari Taipale[1], Uma Thanigai Arasu[1], **Tiit Örd**[1,2], Nadja Sachs[3], Jessica Pauli[3], Lars Maegdefessel[3], Minna Kaikkonen-Määttä[1]

[1]A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland
[2]Institute of Genomics, University of Tartu, Tartu, Estonia
[3]Institute of Molecular Vascular Medicine, TUM School of Medicine and Health, Munich, Germany

E-mail address of presenting author: tiit.ord@uef.fi

Background: Atherosclerotic disease remains a leading global cause of mortality. While single-cell studies have highlighted the heterogeneity of plaque cells, the lack of spatial context limits our ability to understand the cellular neighborhoods that underpin tissue pathology.

Study question: This research seeks to define the spatial organization of cells and their subtypes in human atherosclerotic plaques and potentially uncover links between tissue architecture and patient outcomes.

Methods: Plaque samples were collected from endarterectomy surgery patients and subjected to imaging-based spatial transcriptomics (10x Genomics Xenium Prime) using a 5,000 gene panel supplemented with 100 custom genes. Cell segmentation was based on transcripts (Baysor) and multimodal tissue staining.

Main results: We identified all major cell types expected in atherosclerotic vascular tissue, with smooth muscle cells (SMC-s) and macrophages dominating the dataset. Transcriptomically defined cell subtypes typically also displayed specific spatial localization patterns. From the outer periphery of the plaque towards the necrotic core, we resolved a gradual phenotypic transition of SMC-s from contractile (healthy) to osteochondromyocyte (highly-modulated) and a progression of macrophage phenotype from resident-like to SPP1-high/MMP9-high macrophages, expressing genes involved in lipid handling, matrix degradation and calcification.

Limitations: While the plaque samples cover multiple morphological stages, they are derived from late-stage plaques collected during surgery and thus the early stages of disease are insufficiently studied.

Wider implications: These findings provide a high-resolution map of the plaque microenvironment. Integrating spatial transcriptomics with GWAS summary statistics will enable localizing trait-associated cell populations and candidate genes within the diseased tissue.

PUBLIC UNDERSTANDING AND READINESS FOR PERSONALISED GENETIC PREVENTION: GENOME OF EUROPE (GOE) BASELINE SURVEY IN 5 EUROPEAN COUNTRIES

Andres Metspalu[1], Annely Allik[1,2], Helen Ray-Jones[3], Jeroen van Rooij[3], Marlies Saelaert[4], Chloé Mayeur[4], Emmanuelle Genin[5], Frederique Nowak[5], Ines Amado[5], George Patrinos[6,7], Astrid Moura Vicente[8,9], Maria Luis Cardoso[8,9], Jernej Kovac[10], Mirjam Repše[10], The Genome of Europe Consortium, André Uitterlinden[3], **Merit Kreitsberg**[1]

[1]Institute of Genomics, University of Tartu, Tartu, Estonia
[2]Estonian Research Council, Tartu, Estonia
[3]Laboratory of Population Genomics, Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, Netherlands
[4]Sciensano, Brussels, Belgium
[5]INSERM-National Institute of Health and Medical Research, Brest, France
[6]Department of Pharmacy, University of Patras, Patras, Greece
[7]Laboratory of Innovative Therapeutics and Personalized Medicine, Hellenic Pasteur Institute, Athens, Greece
[8]Department of Health Promotion and NCDs Prevention, National Health of Institute Doutor Ricardo Jorge, Lisbon, Portugal
[9]BioISI–Biosystems and Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, Lisbon, Portugal
[10]University Medical Centre Ljubljana, Ljubljana, Slovenia

E-mail address of presenting author: merit.kreitsberg@ut.ee

Background: The Genome of Europe project is creating a pan-European genomic reference database of modern Europe. This will characterise the genetic diversity across populations supporting personalised medicine applications. The successful adoption of these applications depends not only on scientific progress but also on public awareness, trust, and understanding of the role of genetics in health.

Study question: How do European citizens understand genetics and personalised medicine, and what factors influence their willingness to use genetic information for healthcare and prevention?

Methods: A harmonised online survey was conducted in 5 European countries: Estonia, France, Greece, Portugal, Slovenia; >5,000 adults aged 18–75 participated. The questionnaire assessed knowledge of genetics and personalised medicine, attitudes

towards genomics-based healthcare, willingness to get tested, information needs, and preferred sources of information among with sociodemographic characteristics.

Main results: Basic awareness of genetics was high across all countries, whereas awareness of personalised medicine was substantially lower. Despite this, willingness to use genetic information consistently exceeded awareness, creating an "awareness-willingness paradox". Education emerged as the strongest and most consistent predictor of knowledge and engagement. However, significant country-specific differences were observed, particularly regarding concerns. Across all countries, people seek clear scientific explanations and guidance from healthcare professionals.

Limitations: The study collected self-reported responses through an online survey, which may not fully reflect actual behaviour or populations with limited internet access.

Wider implications: Increasing genomic literacy may support informed uptake of personalised medicine. However, substantial national differences indicate that communication strategies should be adapted to local contexts rather than relying on a single Europe-wide approach.

TRUST IN THE FUTURE GENOMIC DATA - BIOBANK PARTICIPATION AND GOVERNANCE PREFERENCES IN ESTONIA

Piret Hirv[1], Tanel Mällo[2], Liis Leitsalu[1]

[1]Institute of Genomics, University of Tartu
[2]Cybernetica AS

E-mail address of presenting author: piret.hirv@ut.ee

Background: As secondary use of health and genomic data expands, governance frameworks must address expectations regarding data users, conditions of use, and individual control.

Study question: We examined associations between self-reported biobank participation and preferences for data uses, consent types, and withdrawal options across sociodemographic groups. Among participants, we examined perceived risk and preferences for post-mortem data use.

Methods: A quota-controlled online survey of 1,463 Estonian adults was conducted in 2023. Associations were analysed using chi-square tests and analysis of variance.

Main results: Participants were more willing than non-participants to participate in biobanking in the future (82.4% vs 36.8%; Cramér's V=.427). Differences in willingness to share data were primarily explained by data-user type (η^2=.370), whereas differences by participation status were comparatively small (η^2=.014). Willingness was highest for hospitals and the Estonian Biobank and lowest for local authorities and technology companies. Trust, understanding research purpose, and withdrawal rights were the highest-rated conditions for data reuse. Project-specific consent was the most preferred overall; participants more often preferred longer-term consent (Cramér's V=.234). Governance preferences varied by sociodemographic characteristics, although effect sizes were small. Among participants, perceived risk was low. Most respondents (58.1%) supported continued use of post-mortem data, whereas 31.2% preferred transfer to heirs; post-mortem preferences varied by ethnicity, education, and income.

Limitations: The cross-sectional design, self-reported measures, and online-panel recruitment limit causal inference and generalisability.

Wider implications: Participation does not resolve governance questions. Governance frameworks should define acceptable data users, consent arrangements, oversight, withdrawal mechanisms, and postmortem policies to maintain public legitimacy and engagement.

ANDMERING: BUILDING A NATIONAL LEARNING HEALTH SYSTEM FROM ROUTINE HEALTH AND GENOMIC DATA USING THE OMOP COMMON DATA MODEL

Riina Raudne[1,2,3], Sulev Reisberg[4], Raivo Kolde[4], Kerli Mooses[4], Artur Novek[5], Monika Prits[6], Taavi Lai[7], Patrick Pihelgas[8], Jaak Vilo[4], Alar Irs[3,9]

[1]TeamPerMed, Tartu
[2]University of Tartu, Institute of Genomics
[3]Tartu University Hospital, Heart Clinic
[4]University of Tartu, Institute of Computer Science (OHDSI Estonia)
[5]Health and Welfare Information Systems Centre (TEHIK)
[6]Estonian Health Insurance Fund (Tervisekassa)
[7]INTEGRIA
[8]Ministry of Social Affairs
[9]University of Tartu, Institute of Clinical Medicine

E-mail address of presenting author: riina.raudne@ut.ee

Background: Estonia has extensive routine healthcare data and mature digital infrastructure but no mechanism that continuously converts these data into better care. National implementation of the OMOP Common Data Model creates the opportunity for a learning health system based on recurring measurement, clinical validation and feedback. Familial hypercholesterolaemia (FH), a common inherited disorder with established genomic confirmation pathways, is an illustrative case.

Study question: Can routine health data support clinician-validated cardiovascular quality indicators, and can recurring measurement reveal priorities for improving both care and data?

Methods: Andmering ("data loop") runs recurring learning cycles: clinician-defined indicators are computed with standard OHDSI tools on OMOP-transformed data. Specifications (cohort, numerator, denominator, observation window) are version-controlled. Results are reviewed with clinicians, then returned as structured feedback to providers and specialty societies; each cycle identifies priority data gaps. The FH pathway develops and clinically evaluates a computable phenotype from routine care, Estonian Biobank pathogenic-variant recall and clinical genetics.

Main results: Retrospective assessment across five domains—acute coronary syndrome, atrial fibrillation, heart failure, valvular disease and FH—indicates that most denominators and principal outcomes are computable from existing OMOP resources. Remaining gaps concern echocardiographic variables, discharge medication and procedural timing. A nationally supported pilot is planned for early 2027.

Limitations: Indicator definitions remain under clinical and technical validation, and genomic confirmation is limited to research participants.

Wider implications: The infrastructure required for European Health Data Space secondary use also supports continuous learning in care delivery. Each cycle strengthens quality assessment and produces a prioritised roadmap for improving the data infrastructure on which learning depends.

BRIDGING BIOBANKS: INTRODUCING JAPAN'S NATIONAL INFRASTRUCTURE AND USER-CENTRIC INITIATIVES FOR ADVANCING GENOMIC RESEARCH

Mizuki Morita

Okayama University Hospital Biobank (Okadai Biobank), Okayama University, Japan

E-mail address of presenting author: mizuki@okayama-u.ac.jp

Background: While Japan has established robust biobanks, including national flagship cohorts (BBJ, ToMMo, NCBN) and clinical biobanks with deep phenotyping, fragmented infrastructure and varied governance have historically hindered seamless access for both domestic and international researchers.

Study question: How can we construct a unified, user-centric ecosystem to enhance the discoverability, accessibility, and interoperability of Japanese bioresources and genomic data for the global scientific community?

Methods: Japanese biobank stakeholders and funding agencies have collaboratively developed a multilayered support infrastructure. This includes implementing a federated cross-search platform across major biobanks and publishing comprehensive directories. Furthermore, we created standardized "Biobank Utilization Handbooks" and established the Society of Clinical Biobanking (SCB) to educate users, align standard operating procedures, and promote global quality standards like ISO 20387.

Main results: These initiatives successfully shifted the landscape from isolated collections to an integrated network. This foundational work directly catalyzed the launch of JAMBO (2026-), a project building a virtual platform integrating over one million genomic profiles, mirroring Europe's 1+MG initiative.

Limitations: Despite these initiatives, they have not yet fully translated into the desired outcome of significantly increased biobank utilization. Furthermore, obtaining ISO 20387 certification is financially and logistically challenging for small-scale biobanks; targeted efforts are required to ensure they can still reap the benefits of standardization. Cross-border access remains constrained by metadata language barriers.

Wider implications: By ensuring interoperability, Japan is positioned to seamlessly collaborate with European networks (e.g., BBMRI-ERIC). Sharing underrepresented Asian genomic data enhances global diversity and accelerates international precision medicine.

MATERNAL CONTACTS ACROSS THE BALTIC SEA – THOUSANDS OF MITOGENOMES REVEAL MATERNAL GENETIC STRUCTURE AND RECENT EXPANSIONS IN THE CIRCUM-BALTIC REGION

Hovhannes Sahakyan[1], Maere Reidla[1], Ene Metspalu[1], Ivan A. Kuznetsov[2], Rodrigo Flores[2], Kadri Irdt Rezzakioğlu[1], Olga Utevska[1,3], Ajai K. Pathak[1,4], Raimonds Rescenko-Krums[5], Vita Rovite[5], Janis Klovins[5], Adam Ameur[6,7], Ulf Gyllensten[6,7], Estonian Biobank research team[8], **Anne-Mai Ilumäe**[1]

[1]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia
[2]Institute of Genomics, University of Tartu, Tartu 51010, Estonia
[3]Genetics and Cytology Department, V.N. Karazin Kharkiv National University, Kharkiv 61022, Ukraine
[4]Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
[5]Latvian Biomedical Research and Study Centre, Riga LV-1067, Latvia
[6]Department of Immunology Genetics and Pathology, Uppsala University, Uppsala 75185, Sweden
[7]Science for Life Laboratory, Uppsala University, Uppsala 75185, Sweden
[8]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia

E-mail address of presenting author: ami@ut.ee

Background: Present-day European mitochondrial diversity has been interpreted as relatively homogenous, reflecting Mesolithic postglacial expansions from glacial refugia. In contrast, Y-chromosomal diversity shows evidence of pronounced Neolithic bottlenecks and Bronze Age expansions involving a limited number of successful patrilineages.

Study question: To resolve fine-scale maternal demographic history in the Circum-Baltic region.

Methods: We assembled a large dataset comprising more than 9,000 complete mitogenomes from six Circum-Baltic populations and integrated ~5,000 additional phylogenetically related mitogenomes from published sources. Using these sequences, we reconstructed maximum-parsimony phylogenetic trees, identified high-frequency terminal branches and estimated their coalescence ages within the Bayesian phylogenetic analysis framework.

Main results: Maternal diversity in the Circum-Baltic region was substantially shaped by Bronze and Iron Age demographic processes. Temporal changes in maternal effective population size revealed distinct demographic trajectories across populations, with Sweden showing a pronounced decline during the Bronze Age, followed by a steep rise in the late Iron Age. We uncovered elevated sharing of matrilineages between Estonia and Finland, pointing to a maternal genetic component underlying their linguistic affinity that had previously been attributed primarily to shared paternal history. Phylogenetic links in several subclades suggest long-distance east–west interactions across the East European Forest Belt, involving Bronze Age populations in the forest-steppe contact zones near the Ural Mountains.

Limitations: Limited representation of populations from the Eastern European Plain and the southern Baltic region.

Wider implications: Large population-based mitogenome datasets are essential to resolve fine-scale maternal demographic structure and distinguish population-specific maternal histories, complementing insights from paternal lineages.

42

A GENETIC BRIDGE OVER THE GULF OF FINLAND: TRACING THE ORIGIN OF GENETIC CONNECTEDNESS BETWEEN FINNS AND ESTONIANS

R. Didonna[1], L. Saag[1], T. Kivisild[1, 2], A. Kushniarevich[1], M. Tõrv[3], L. Varul[4], M. Malve[1,3], M. Niinesalu-Moon[3], A. Lillak[3], H. Valk[3], M. Mägi[4], V. Lang[3], A. Kriiska[3], K. Tambets[1], M. Metspalu[1]

[1]Institute of Genomics, University of Tartu, Tartu, Estonia
[2]Department of Human Genetics, KU Leuve, Leuven, Belgium
[3]Institute of History and Archaeology, University of Tartu, Tartu, Estonia
[4]School of Humanities, Tallinn University, Tallinn, Estonia

E-mail address of presenting author: roberto.didonna@ut.ee

Background: The Finnish population is a unique example of a genetic isolate affected by a recent founder event. Previous studies have suggested that the ancestors of Finnic-speaking Finns and Estonians reached the circum-Baltic region by the 1st millennium BCE. However, high linguistic similarity points to a more recent split of their languages.

Study question: We set out to study genetic connectedness between Finns and Estonians directly. In a study published in 2021, we searched for long shared allele intervals (LSAIs; similar to identity-by-descent segments) in unphased data for >143,000 present-day Estonians, 99 Finns, and 14 imputed ancient genomes from Estonia. We found unexpectedly high levels of individual connectedness between Estonians and Finns for the last eight centuries, contrasting their clear differentiation by allele frequencies.

Methods: Here, to shed more light on the origins of the connectedness between Estonians and Finns, we extracted and sequenced ancient DNA from 16 additional prehistoric individuals from Estonia dating to 1,400 calBCE to 700 calCE, including the first genomes from 300–700 calCE.

Main results: We show that already the 300–700 calCE individuals had higher levels of individual connectedness with modern Estonians and Finns than individuals from earlier time periods, including preceding individuals from 600–100 cal BCE.

Limitations: There are very few samples available for this specific period, likely due to the increasing prevalence of cremation practices.

Wider implications: Understand the development and diversification of Finno-Ugric languages and shed further light on how Iron Age populations interacted.

SCALING AND DEMOCRATISING STRUCTURE-BASED PROTEIN FUNCTION PREDICTION WITH METAGENOMIC-DEEPFRI

Valentyn Bezshapkin[1,*], **Filip Schymik**[2,3,*], Piotr Kucharski[4], Paweł Szczerbiak[2,3], Jakub W. Wojciechowski[2], Lukasz M. Szydlowski[2,5], Tomasz Kosciolek[2,3]

[1]Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zurich, Zurich, Switzerland
[2]Sano Centre for Computational Medicine, Krakow, Poland
[3]Faculty of Computer Science, AGH University of Krakow, Krakow, Poland
[4]Aiformatics, Krakow, Poland
[5]IRA 3P Medicine Lab; Center for Space Biomedicine and Radiological Protection, Gdansk Medical University, Gdansk, Poland

[*]Equal contribution

E-mail address of presenting author: f.schymik@sanoscience.org

Background: The human gut microbiome encodes far more genes than the host genome, and this gene pool shapes metabolism, immunity, and cognition. Yet annotation lags behind sequencing: Swiss-Prot, the curated part of UniProtKB, covers only ~0.3% of known sequences. Most microbiome proteins remain uncharacterized "dark proteins", and assigning them function at scale, without orthology, is key to understanding the microbiome in human health.

Study question: Can structure-informed function prediction be scaled to large metagenomic datasets without computing a structure for every query sequence?

Methods: DeepFRI predicts Gene Ontology (GO) terms from both sequence and structure, where structure improves confidence and specificity. We developed Metagenomic-DeepFRI, which extends it by retrieving homologous structures from reference databases (AlphaFold DB, ESM Atlas) instead of folding each query. It uses Foldcomp for compact reference storage, MMseqs2 for similarity search, and PyOpal for global alignment and contact-map construction.

Main results: Homology-based retrieval achieved structural coverage of ~90% for metagenomic sequences. Adding structure to sequence features improved GO-term confidence and Information Content by up to 50%. Against eggNOG-mapper at similar throughput, metagenomic-DeepFRI reached nearly 90% functional coverage versus ~3% for eggNOG-mapper when restricted to GO terms.

Limitations: Metagenomic-DeepFRI predicts only GO terms, which may be insufficient to fully characterize a sample, potentially requiring additional tools. Future work could add structure quality checks or investigate sequences that pass MMseqs2 search but fail global alignment.

Wider implications: The structure-retrieval heuristic could be incorporated into other tools using 3D protein information, to cheaply obtain a good approximation. Explicitly combining sequence and structure can improve function prediction.

F64: RAPID CUSTOMIZABLE SEQUENCING DATA FILTER

Alar Aints

Department of Immunology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

E-mail address of presenting author: Alar.Aints@ut.ee

Background: Metagenomic and meta-transcriptomic sequencing data, when obtained in a clinical setting, are frequently dominated by human sequence reads. Downstream analysis is greatly facilitated by efficient removal of the host organism reads early in the analysis pipeline. Currently, dedicated, customizable tools for sequence filtration are lacking. Existing host-depletion and sequence-filtering workflows typically rely on general-purpose aligners or taxonomic classification tools and often require costly database rebuilding.

Study question: To create a tool for customisable removal of reads belonging to a specified dataset.

Methods: The program was evaluated on simulated genomic reads of human, mouse and E.coli, and human clinical metagenome sequencing data.

Main results: In the evaluated datasets, F64 achieved filtration accuracy exceeding that of established alignment-based and k-mer-based approaches (classification errors 717 reads per million) while reducing processing time substantially (0.85 - 2.2 M reads per second).

Limitations: F64 stores raw hash values rather than employing compressed minimizer-based representations, resulting in larger reference libraries than those used by many taxonomic classification systems. The software is designed specifically for sequence filtration and therefore does not provide taxonomic classification, genomic alignment, or variant analysis.

Wider implications: F64 provides a dedicated framework for high-throughput sequence filtration that combines rapid processing with flexible reference library management. The ability to incrementally update, merge and refine user-defined libraries distinguishes F64 from existing host-depletion workflows and may simplify maintenance of filtration databases in routine sequencing applications.

A NOVEL METHOD OF SINGLE CELL WHOLE TRANSCRIPTOME SEQUENCING FROM FFPE SAMPLES

Ritika Kulshreshtha

E-mail address of presenting author: ritika.kulshreshtha@singleron.bio

Background: Formalin-fixed, paraffin-embedded (FFPE) tissue archives represent one of the largest and most clinically annotated biological resources available today. However, RNA degradation and chemical modification have greatly limited the application of single-cell RNA sequencing to FFPE samples, restricting transcriptomic analysis largely to bulk or spatially averaged approaches.

Study question: Here, we present a novel single-cell RNA analysis technology specifically designed to enable robust whole transcriptome profiling from FFPE-derived single nuclei.

Methods: By combining optimized nuclei extraction from FFPE samples, random RT priming and subsequent tagging of the cDNA, this approach overcomes key technical barriers associated with single cell analysis in archived samples. The random-priming strategy in RT step enables coverage across gene body and makes mutation detection possible. Furthermore, our method can sequence not only mRNA but also noncoding RNAs with single cell resolution.

Main results: We demonstrate the performance of this technology across multiple FFPE tissue types, highlighting reproducibility, sensitivity, and biological interpretability. Importantly, we focus not only on technical metrics, but on the ability to extract actionable biological insights—such as accurate cell-type annotation, disease-relevant expression profiles, and treatment-related cell composition changes—from samples previously considered incompatible with single-cell analysis. The method is automated and suitable for routine testing.

Wider implications: This work expands the scope of single-cell transcriptomics to clinically relevant archived specimens, enabling retrospective studies, translational research, and new connections between molecular phenotypes and long-term clinical outcomes. By bridging the gap between archived pathology samples and single-cell resolution, this technology opens new opportunities for discovery in both research and clinical translational settings.

RNA-SEQ LIBRARY PREPARATION BIAS AFFECTS LONG TRANSCRIPT DETECTION

Swethaa Natraj Gayathri[1,2], Victoria Lillback[1,2], Bjarne Udd[2], Peter Hackman[2], Marco Savarese[2], Ali Oghabian[2]

[1]University of Helsinki, Helsinki, Finland
[2]Folkhälsan Research Center, Helsinki, Finland

E-mail address of presenting author: swethaa.natrajgayathri@helsinki.fi

Background: RNA sequencing (RNA-seq) is used in both basic biology and precision medicine by supporting biomarker discovery, aberrant splicing detection, and variant interpretation. However, the choice of library preparation technique can influence the RNA pool captured, and, consequently, the results that are obtained. Poly(A)+ RNA-seq enriches for polyadenylated transcripts, whereas rRNA depletion captures a broader range of RNAs, including non-polyadenylated and partially processed transcripts.

Study question: How do poly(A)+ selection and rRNA depletion differ in transcript coverage, splice junction detection, and representation of long transcripts in human RNA-seq?

Methods: We compared poly(A)+ selection and rRNA depletion using human and blood RNA cohorts. We assessed transcript 5'-3' coverage, splice junctions and the representation of long genes and alternatively spliced transcripts.

Main results: rRNA depletion detected more splice junctions and more novel splicing junctions than poly(A)+ libraries. In contrast, poly(A)+ libraries showed a stronger 3' bias, with reduced coverage across long transcripts. Most importantly, rRNA depletion performed better for long genes and transcripts with complex splicing patterns, which are often clinically relevant but difficult to capture with poly(A)+ selection.

Limitations: Most comparisons in this study were performed on independent skeletal muscle cohorts, rather than a large cohort of paired samples from the same individuals. Given the variability in terms of biological expression, sex and disease related effects, this may give rise to inter-individual differences than enrichment technique alone.

Wider implications: RNA-seq library preparation is not merely a technical step. Selecting an appropriate library preparation strategy can improve the detection of clinically relevant long transcripts, hence, enhancing RNA-based diagnostic rates.

CLICK-CHEMISTRY-BASED HIGH-DEGREE SAMPLE MULTIPLEXING FOR LARGE-SCALE SINGLE CELL SEQUENCING DATASET GENERATION

Nan Fang

Singleron Biotechnologies GmbH, Gottfried Hagen Strasse 60, 51105 Cologne, Germany

E-mail address of presenting author: nan.fang@singleronbio.de

Background: Single-cell transcriptomics has become a powerful readout for perturbation screens, from compound libraries to CRISPR panels. However, scaling to the hundreds to thousands of conditions typical of high-throughput screens is often constrained by per-sample library costs, plate-to-plate batch effects, and instrument time.

Study question: Is it possible to have a method that enables high-through single cell sequencing at scale without all the above constrains? Here we describe a click-chemistry-based sample multiplexing method that can be used to significantly scale single cell sequencing experiments and reduce per sample cost.

Methods: Click chemistry offers key advantages in bioconjugation and labeling applications due to ist highly selective and effective reactions requiring only mild aqueous conditions. We developed a sample multiplexing method using modified oligos that can be bound to the amino groups of the cell surface proteins with a straightforward click chemistry reaction. Cells are first tagged with such modified oligos in a click reaction, and then pooled in one cell suspension mix and subjected to single cell sequencing. All 96 different samples are pooled and processed as one sample.

Main results: We demonstrated:

- Application case in a high-throughput drug screening study
- Scalability to 96plex in one single cell sequencing
- Good single cell sequencing performance metrics
- Reproducibility of replicates in the sample pool
- Efficient sample labeling agnostic to species and cell type
- Significantly reduced processing time and costs.

Limitations: Current method is applicable to cells, not nuclei. Further optimization is needed to make it applicable to single nucleus sequencing.

Wider implications: With all 96 conditions (or more if necessary, by simply increase number of oligos used) within a chip processed under matched reagent lots, timing, and instrument handling, condition-to-condition technical variance is constrained at the batch level rather than accumulating across manually staggered runs, substantially reducing the batch-correction burden typically required before analyzing large scale single cell dataset and AI model training.

EASIGEN-DS: DESIGNING A DISTRIBUTED RESEARCH INFRASTRUCTURE FOR ADVANCED GENOMICS TECHNOLOGIES

Mireia Vaca-Dempere[1,*], Maria Xandri[1,*], Steven McGinn[2], Jean-François Deleuze[2], Michael Forster[3], Andre Franke[3], Philine Warnke[4], Liliya Pullmann[4], Louis Colnot[5], Jessica Catalano[5], Orlando Contreras[6], Tuuli Lappalainen[7], Tuuli Reisberg[8], **Kristiina Tambets**[8], Hegel Rafael Hernández-López[9], Joris Vermeesch[9], Filip Rokić[10], Oliver Vugrek[10], Christian Conrad[11], Mònica Bayés[1], Ivo Gut[1] & EASIGEN Consortium

[1]Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain
[2]Centre National de Recherche en Génomique Humaine (CNRGH), Institut de biologie François Jacob, CEA, Evry, France
[3]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany
[4]Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, Germany
[5]CSIL Scrl, Milan, Italy; Science for Life Laboratory, Department of Gene Technology
[6]KTH Royal Institute of Technology, Solna, Sweden
[7]New York Genome Center, New York, NY, USA
[8]Institute of Genomics, University of Tartu, Tartu, Estonia
[9]Department of Human Genetics, KU Leuven, Leuven, Belgium
[10]Rudjer Boskovic Institute, Zagreb, Croatia
[11]Berlin Institute of Health at Charité, University of Medicine Berlin, Corporate Member of the Free University of Berlin, Humboldt-University of Berlin, Berlin, Germany

[*]Equal contribution

E-mail address of presenting author: kristiina.tambets@ut.ee

Background: Rapid advances in genomics technologies are transforming biomedical research. However, access to these technologies and the specialised expertise required for their implementation remains uneven across Europe, limiting scientific excellence and widening regional disparities. The EASIGEN-DS consortium is conducting a design study to establish EASIGEN, a European research infrastructure for advanced genomics technologies, with the aim of preparing a proposal for inclusion in the ESFRI Roadmap. The consortium includes thirteen state-of-the-art genomics facilities across Europe providing cutting-edge capacities in genome sequencing, single-cell and spatial genomics, multi-omics integration, advanced bioinformatics, data analysis, training, and secure data management.

Methods: To assess the needs of the European community in genomics technologies and associated services, we are conducting: 1) a landscape analysis mapping mid- to large-scale sequencing infrastructures across Europe, based on publicly available data from the public and private sectors; 2) a survey targeting academic researchers, clinicians, and bioindustry representatives to identify current gaps, future needs, and the expected impact of EASIGEN; and 3) a forward-looking analysis of emerging technology trends.

Main results: Preliminary analyses indicate: 1) fragmented genomics and multi-omics capacity; 2) unequal access to advanced technologies; 3) strong demand for advanced genomics services; 4) rapid technological evolution; and 5) growing interest from the biotech industry.

Wider implications: These findings demonstrate significant disparities and unmet needs in advanced genomics across Europe. The evidence generated is directly informing the design of a coordinated infrastructure aimed at strengthening capacity, fostering innovation, and ensuring equitable access to cutting-edge genomics technologies within the European Research Area.

“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.” Funded by the European Union under GA 101060011.