Methods: We developed a protocol for long-read targeted sequencing using capture probes from Twist Bioscience and applied this workflow to sequence 21 pharmacogenes from 41 samples with PacBio HiFi technology.

Results: In total, 41 samples had an average on target phasing of 62% (47%-73%) and the average haploblock size was 7,509bp demonstrating the large number of nucleotides in the target region that were phased. In the CYP3A locus, 1,088 unique variants were detected, of which 570 variants were located in the core regions of CYP3A4, CYP3A5 and CYP3A7. Only 27 of these variants (2%) are included in the clinically used *-allele nomenclature. Notably, 1 frameshift-, 5 missense- and 8 splice site variants which are not included in clinical nomenclature were detected. Per individual, an average of 155 unique variants were detected and 34% (5% - 86%) of nucleotides were phased in the CYP3A locus.

Conclusions: Our results indicate that a panel-based long-read sequencing approach can phase the majority of variants in complex genomic regions, revealing a high abundance of unknown but potentially impactful variants in the CYP3A locus.

Grant Reference: Not applicable

Conflict of Interest: Qinglian Zhai Full, maaike van der lee Full, Roberta Menafra: None declared, Loes Busscher: None declared, Redmar R. van den Berg: None declared, Susan Kloet: None declared, Jesse Swen Full

P19.003.C Implementation of polygenic risk scores from sequencing data towards practice by utilizing large publicly available datasets

Alejandro Correa Rojo[1;2], *Dirk Valkenborg*[2], *GOKHAN ERTAYLAN*[1]

[1]*Flemish Institute for Technological Research (VITO), Health, Mol, Belgium;* [2]*Hasselt University, Data Science Institute, Diepenbeek, Belgium*

Introduction: Polygenic risk scores (PRSs) are a potential tool for assessing an individual's genetic predisposition for diseases or traits. Current methodologies use genome-wide summary statistics from SNP-array data. Yet, the applications might be limited due to ascertainment biases of the SNP array. Next-generation sequencing technologies such as whole-exome sequencing (WES) have higher resolution of genetic variants and reduced batch effects, allowing higher interoperability with other population cohorts. We aimed to test the applicability of WES data from large-population datasets for calculating PRSs to analyze smaller sample sizes.

Methods: We developed a workflow that processes WES data from large-reference datasets to estimate PRSs in small datasets with similar population characteristics. It generates a base exome-wide summary statistics and target dataset to increase the sample size of the small cohort. We applied the workflow using the 200,643 exomes from the UK Biobank and 30 exomes from our in-house data, IAM Frontier. We calculated PRSs for blood pressure, body mass index, LDL cholesterol, and vitamin D.

Results: The workflow generated scores with significant performance (P-value $< 2 \times 10^{-6}$) for the selected traits. For the IAM Frontier cohort, the scores had a predictive performance (R^2) of up to 11.54%. The generated scores had similar performance; compared to the reported ones in the Polygenic Score (PGS) Catalog.

Conclusions: We successfully estimated PRSs using the UK Biobank WES data for the IAM Frontier cohort. To our knowledge, this is the first study that reports PRSs purely from sequencing data.

Grant references: Flemish Special Research Fund (BOF) [BOF21DOC23].

Conflict of Interest: Alejandro Correa Rojo Full time, Flemish Special Research Fund (Bijzonder Onderzoeksfonds - BOF). Grant Project Number: BOF21DOC23, Dirk Valkenborg Full time, Flemish Special Research Fund (Bijzonder Onderzoeksfonds - BOF). Grant Project Number: BOF21DOC23, GOKHAN ERTAYLAN Full time

P19.004.D A polygenic model for Japanese height implicates the importance of localization

Yoichi Sutoh[1], *Tsuyoshi Hachiya*[1], *Yayoi Otsuka-Yamasaki*[1], *Shohei Komaki*[1], *Shiori Minabe*[1], *Hideki Ohmomo*[1;2], *Atsushi Shimizu*[1;2]

[1]*Iwate Medical University, Iwate Tohoku Medical Megabank Organization, Yahaba, Japan;* [2]*Iwate Medical University, Institute for Biomedical Sciences, Yahaba, Japan*

Background/Objectives: Height is implicated in skeletal growth disorders and used as a model trait for studying polygenic inheritance. In a previous GWAS involving over 5 million individuals of diverse ancestry, 12,111 independent variants were identified as being significantly associated with height (Yengo et al. Nature, 2022). The cross-ancestry polygenic score (PGS) calculated based on this GWAS showed prediction accuracy (R^2) of up to 0.447 for the actual height of Europeans adjusted for age, sex, and 10 genetic principal components (PCs). However, the prediction accuracy of the PGS was significantly attenuated for the height of other populations, such as Africans (~0.154) and East Asians (~0.205). In this study, we aimed to improve the prediction accuracy of PGS for the height of East Asians.

Methods: We constructed models using PRS-CS (Ge et al. Nat. Commun., 2019) based on a previously reported GWAS for a height of 165,056 Japanese subjects (Sakaue et al. Nat. Genet., 2021). The best model was selected based on its correlation with the height of 9,811 Japanese individuals and included 1,031,950 variants from the 1000 Genomes Phase 3 dataset.

Results: The prediction accuracy of the PGS for the adjusted height of 53,305 Japanese individuals was 0.66, which is similar to heritability estimates of a height based on whole-genome sequence data (0.68) and pedigree estimates (0.7-0.8) (Wainschtein et al. Nat. Genet., 2022).

Conclusion: This finding may suggest the potential for optimizing a polygenic model for a specific population to achieve practically sufficient accuracy.

Grant References: 21tm0124006j0001 (AMED, Japan)

Conflict of Interest: None declared

P19.005.A Common risk factors enhance the predictive ability of polygenic scores for type 2 diabetes

Yanina Timasheva[1;2], *Zhanna Balkhiyarova*[3], *Diana Avzaletdinova*[4], *Tatyana Morugova*[4], *Gulnaz Korytina*[1], *Inga Prokopenko*[3], *Olga Kochetova*[1]

[1]*Institute of Biochemistry and Genetics UFRC RAS, Physiological Genetics, Ufa, Russian Federation;* [2]*Bashkir State Medical University, Medical Genetics and Fundamental Medicine, Ufa, Russian Federation;* [3]*University of Surrey, Section of Statistical Multi-Omics, Department of Clinical & Experimental Medicine, School of Biosciences & Medicine, Guildford, United Kingdom;* [4]*Bashkir State Medical University, Endocrinology, Ufa, Russian Federation*

We aimed to examine associations between common variants in genes implicated in the development of type 2 diabetes (T2D) and evaluate the predictive value of polygenic scores calculated from these variants alone and in combination with other risk factors, including age and sex.