**Supplementary Materials**

Table of Contents

[Exome Sequencing in the MDC, MPP and UKB 1](#_Toc158717730)

[Figures 2](#_Toc158717731)

[Regeneron Banner List 3](#_Toc158717732)

[References 7](#_Toc158717733)

# Exome Sequencing in the MDC, MPP and UKB

Exome sequencing was available for 29,295 individuals in MDC and 469,787 participants in the UKB. In MPP, imputed genotype data, was available in 5,232 participants.

High coverage whole exome sequencing was performed in both MDC and UKB by Regeneron Genetics Center as described in detail elsewhere(1).

In short, the sequencing was performed on balanced DNA pools using 75 base paired-end reads, using Illumina NovaSeq instruments. Reads were aligned to the human reference genome (build Grch38) using BWA-mem, and variants annotated according to SNPEff or Ensemble´s definition (v85). For each gene, protein-coding variants were annotated for their functional effect on the gene and predicted deleteriousness: frameshift, stop-gain, stop-loss, splice-variants, splice acceptor, splice donor, which are grouped in the predicted loss-of-function category; and missense, the one not falling in the previous classification. The deleteriousness of missense variants was defined according to different 5 in silico prediction tools: LRT (2), MutationTaster (3), SIFT (4), Polyphen2 HDIV (5), and Polyphen2 HVAR (5).

MPP genome sequencing was performed using *Infinium Global Screening Array* at the Genomic core facility of Erasmus University Medical Center (Rotterdam).

Analyses were performed in the three cohorts separately, results for outcome phenotypes which were present in more than one cohort, were meta-analysed using inverse variance-weighted method (IWV).

**1.1 Quality Control**

Prior to the association analyses, quality control steps were performed using plink tool (plink v2.0) (6): high-LD genomic regions, individuals with more that 0.1% missingness rate (--mind 0.1) or SNPs with a missingness higher than 0.1% (--geno 0.1) or with a Hardy Weinberg equilibrium below 1e-15 (--hwe) were filtered out. Analysis was restricted to European and not related individuals having consistency between estimated genetic sex and reported sex and not affected by sex chromosome aneuploidy. Principal components (PCAs) were computed from genotype data, and the first 10 used as covariates in association analysis to control population stratification.

**1.2 REGENIE**

REGENIE algorithm consists of two steps, a first step in which genetic markers from array genotype data are used to fit a whole regression model to capture the fraction of the phenotype variance attributable to genetic effects (13). The second step is represented by the linear association, as previously described, and it is conditioned upon the prediction score of first step.

# Figures

Manhattan and QQplots were created using Locus Zoom (7).

Summary statistics were available for 450,265 exonic variants included in the analysis.

26 variants resulted to significantly associated with PAM-AMA at the GWAS significance level (p<5x10E-5). Of this, 2 were annotated as missense or LoF. Conditional iterative analysis was used to determine whether these were independent signals. Eventually the two aminoacidic substitutions, Ser539Trp (rs78408340, 5:103003035 GRCh38) and Asp563Gly (rs35658696, 5:103003107 GRCh38) were identified as lead and independent signals.

In S Figure 1 and S Figure 2 the Manhattan Plot and QQPlot are represented, respectively.

With the estimated lambda values for each allele frequency range.

Figure S 1 Manhattan Plot of the Exome-wide analysis of PAM-AMA in the MDC-CC

**A graph with blue and grey squares

Description automatically generated**

On the y axis the logarithmic of the p-value of the association with PAM-AMA; on the x axis the chromosomic positions. The dashed line indicates the GWAS significance threshold.

**A screen shot of a graph

Description automatically generated**

Figure S 2 QQplot and Lambda value for the EWA

# Regeneron Banner List

**RGC Management & Leadership Team**

Aris Baras, Gonçalo Abecasis, Adolfo Ferrando, Michael Cantor, Giovanni Coppola, Andrew Deubler, Aris Economides, Luca A Lotta, John D Overton, Jeffrey G Reid, Alan Shuldiner, Katherine Siminovitch, Jason Portnoy, Marcus B Jones, Lyndon Mitnaul, Alison Fenney, Jonathan Marchini, Manuel Allen Revez Ferreira, Maya Ghoussaini, Mona Nafde, William Salerno.

**Sequencing & Lab Operations**

John D Overton, Christina Beechert, Erin D Brian, Laura M Cremona, Hang Du, Caitlin Forsythe, Zhenhua Gu, Kristy Guevara, Michael Lattari, Alexander Lopez, Kia Manoochehri, Prathyusha Challa, Manasi Pradhan, Raymond Reynoso, Ricardo Schiavo, Maria Sotiropoulos Padilla, Chenggu Wang, Sarah E Wolf.

**Clinical Informatics**

Michael Cantor, Amelia Averitt, Nilanjana Banerjee, Dadong Li, Sameer Malhotra, Justin Mower, Mudasar Sarwar, Deepika Sharma, Jeffrey C Staples, Sean Yu, Aaron Zhang.

**Genome Informatics & Data Engineering**

Jeffrey G Reid, Mona Nafde, George Mitra, Sujit Gokhale, Andrew Bunyea, Krishna Pawan Punuru, Sanjay Sreeram, Gisu Eom, Sujit Gokhale, Benjamin Sultan, Rouel Lanche, Vrushali Mahajan, Eliot Austin, Sean O'Keeffe, Razvan Panea, Tommy Polanco, Ayesha Rasool, William Salerno, Xiaodong Bai, Lance Zhang, Boris Boutkov, Evan Edelstein, Alexander Gorovits, Ju Guan, Lukas Habegger, Alicia Hawes, Olga Krasheninina, Samantha Zarate, Adam J Mansfield, Evan K Maxwell, Suganthi Balasubramanian, Suying Bao, Kathie Sun, Chuanyi Zhang.

**Analytical Genetics and Data Science**

Gonçalo Abecasis, Manuel Allen Revez Ferreira, Joshua Backman, Kathy Burch, Adrian Campos, Lei Chen, Sam Choi, Amy Damask, Liron Ganel, Sheila Gaynor, Benjamin Geraghty, Arkopravo Ghosh, Salvador Romero Martinez, Christopher Gillies, Lauren Gurski, Joseph Herman, Eric Jorgenson, Tyler Joseph, Michael Kessler, Jack Kosmicki, Nan Lin, Adam Locke, Priyanka Nakka, Jonathan Marchini, Karl Landheer, Olivier Delaneau, Maya Ghoussaini, Anthony Marcketta, Joelle Mbatchou, Arden Moscati, Aditeya Pandey, Anita Pandit, Charles Paulding, Jonathan Ross, Carlo Sidore, Eli Stahl, Maria Suciu, Timothy Thornton, Peter VandeHaar, Sailaja Vedantam, Scott Vrieze, Jingning Zhang, Rujin Wang, Kuan-Han Wu, Bin Ye, Blair Zhang, Andrey Ziyatdinov, Yuxin Zou, Olivier Delaneau, Maya Ghoussaini, Jingning Zhang, Kyoko Watanabe, Mira Tang.

Therapeutic Area Genetics

Adolfo Ferrando, Giovanni Coppola, Luca A Lotta, Alan Shuldiner, Katherine Siminovitch, Brian Hobbs, Jon Silver, William Palmer, Rita Guerreiro, Amit Joshi, Antoine Baldassari, Cristen Willer, Sarah Graham, Ernst Mayerhofer, Mary Haas, Niek Verweij, George Hindy, Jonas Bovijn, Tanima De, Parsa Akbari, Luanluan Sun, Olukayode Sosina, Arthur Gilly, Peter Dornbos, Juan Rodriguez-Flores, Moeen Riaz, Manav Kapoor, Gannie Tzoneva, Momodou W Jallow, Anna Alkelai, Giovanni Coppola, Ariane Ayer, Veera Rajagopal, Sahar Gelfman, Vijay Kumar, Jacqueline Otto, Neelroop Parikshak, Aysegul Guvenek, Jose Bras, Silvia Alvarez, Jessie Brown, Jing He, Hossein Khiabanian, Joana Revez, Kimberly Skead, Valentina Zavala.

**Research Program Management & Strategic Initiatives**

Marcus B Jones, Esteban Chen, Michelle G LeBlanc, Jason Mighty, Nirupama Nishtala, Nadia Rana, Jennifer Rico-Varela, Jaimee Hernandez.

**Senior Partnerships & Business Operations**

Alison Fenney, Randi Schwartz, Jody Hankins, Samuel Hart.

**Business Operations & Administrative Coordinators**

Ann Perez-Beals, Gina Solari, Jaimee Hernandez, Johannie Rivera-Picart, Michelle Pagan, Sunilbe Siceron.

Affiliations:

1. Regeneron Genetics Center, Tarrytown, NY, USA.

# References

1. Akbari P, Sosina OA, Bovijn J, Landheer K, Nielsen JB, Kim M, et al. Multiancestry exome sequencing reveals INHBE mutations associated with favorable fat distribution and protection from diabetes. Nat Commun. 2022 Aug 23;13(1):4844.

2. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009 Sep;19(9):1553–61.

3. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010 Aug;7(8):575–6.

4. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm | Nature Protocols [Internet]. [cited 2023 Jun 13]. Available from: https://www.nature.com/articles/nprot.2009.86

5. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248–9.

6. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

7. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010 Sep 15;26(18):2336–7.