From biobanking to polygenic risk scores and personal prevention

Andres Metspalu

Estonian Genome Centre, IG and IMCB, University of Tartu, ESTONIA

Seminar on Science Diplomacy,

24. March, 2026,

Tallinn

eesti geenivaramu

15 biobanks with largest impact

Global distribution of 15 biobanks

From: Biobanking with genetics shapes precision medicine and global health
https://doi.org/10.1038/s41576-024-00794-y

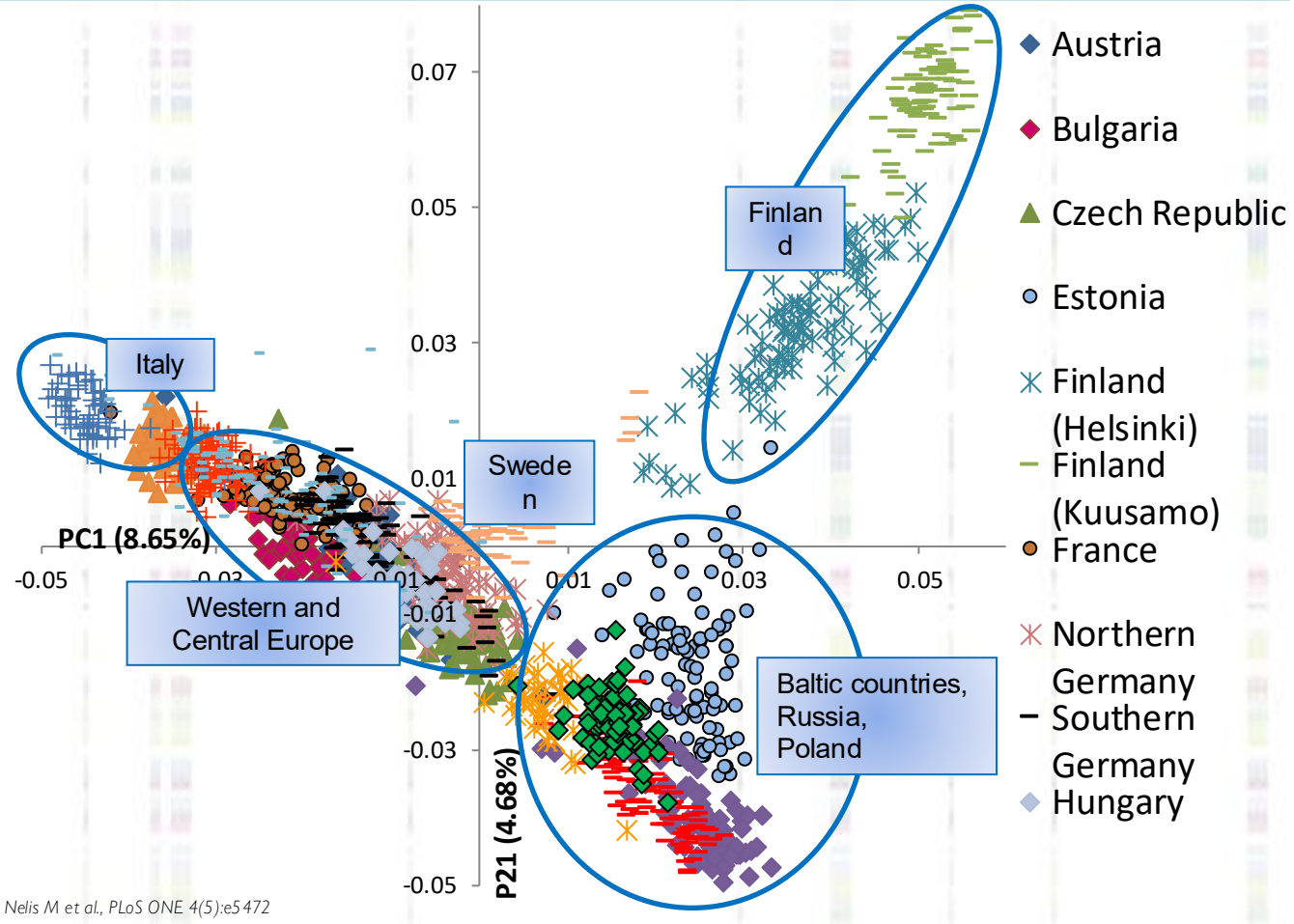
eesti geenivaramu

Estonian Biobank (started in 1999!)

1. Prospective, longitudinal, volunteer-based, **213 000 individuals = 20% of the adults (18 years and up) population of Estonia**
2. Health records, diet, physical activity, etc. DNA, plasma, 3000 WGS, 2500 WES, for all GSA array and NMR data for 250 molecules
3. Open for research and development: Clear access rules, broad informed consent, HGR Act.

eesti geenivaramu

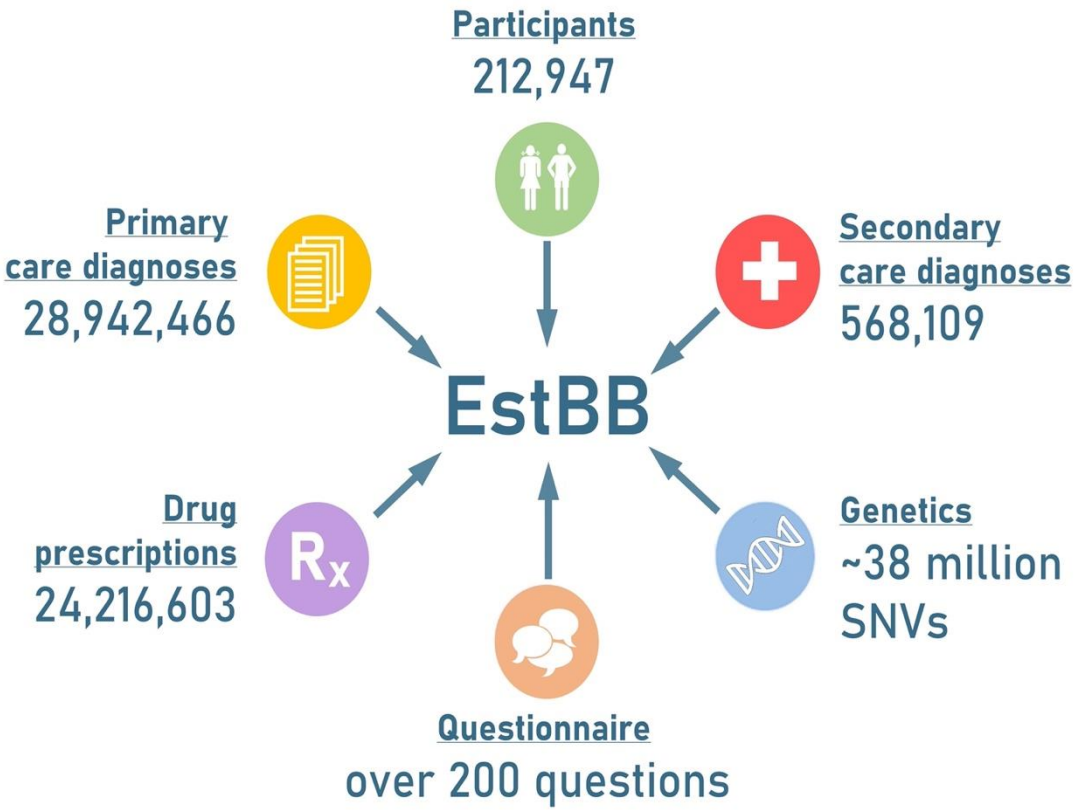
European genetic map



Nelis M et al., PLoS ONE 4(5):e5472

eesti geenivaramu

EstBB Baseline Data

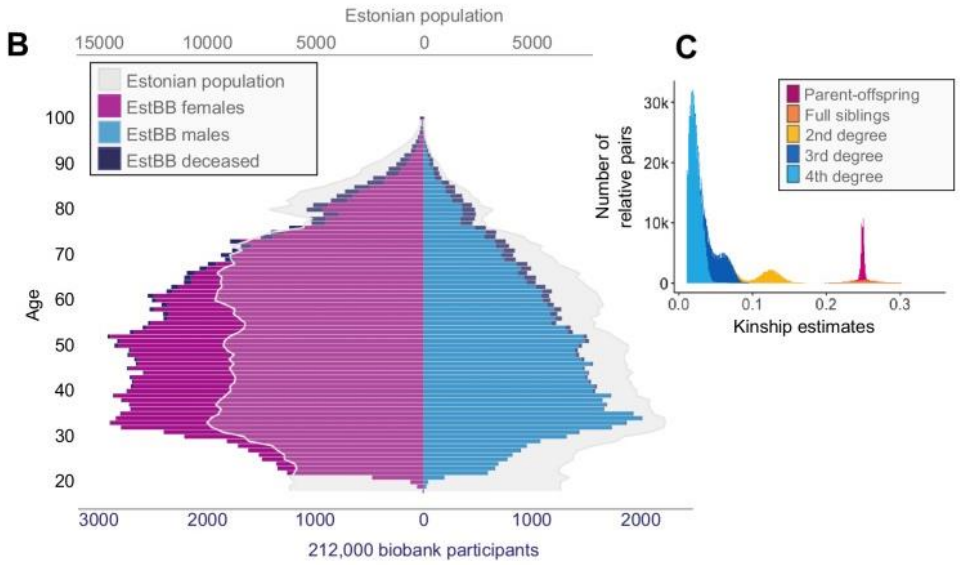


+10,000 PacBio long-read genomes

eesti geenivaramu

Overview of the Estonian Biobank

213 000 gene donors in the EstBB



B Overview of the age and sex distribution of EstBB participants and comparison to the whole Estonian population in 2023. Different colored bars correspond to male, female, and deceased participants (blue, purple, and navy, respectively) in each age category, while the grey outline corresponds to the age and sex distribution in the whole Estonian population **C** Distribution of inferred relatedness in the EstBB cohort. Relatedness was inferred using the KING software v2.2.7

eesti geenivaramu

Human Genes Research Act
Passed 13.12.2000 RT I 2000, 104, 685,
updated in March, 2026

§ 3. Chief processor of Gene Bank

(1) The chief processor of the Gene Bank is the University of Tartu whose objective as the chief processor of the Gene Bank is to:

1) promote the development of genetic research;

2) collect information on the health of the Estonian population and genetic information concerning the Estonian population;

3) use the results of genetic research to improve public health

Access rules: send application to EstBB -> scientific committee-> ethical review, ->release the data

"Tools to the data" rule, all computing in our servers in special space just for the customer. Fees apply.

Consent Form

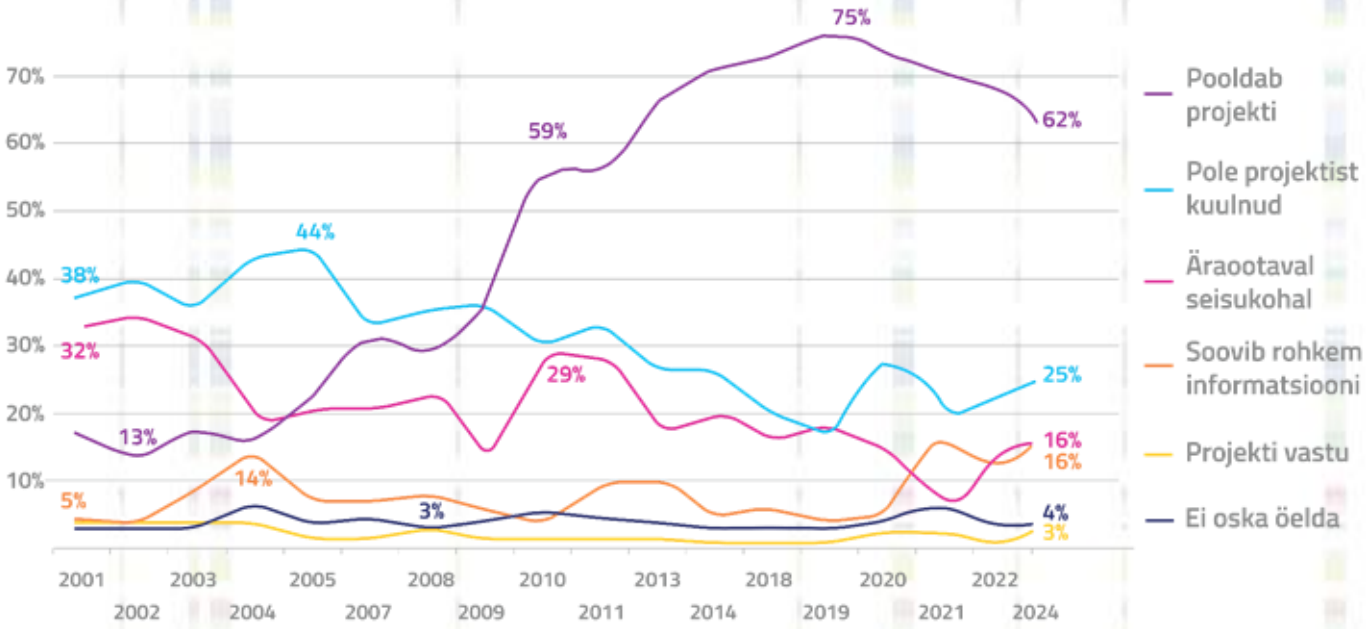
/------------------/

3) I may not demand a fee for providing a tissue sample, for the description of my state of health or genealogy, or for the use of the research results. I am aware of the fact that my tissue sample may have some commercial value and research and development institutions as well as commercial enterprises may receive anonymous data about gene donors. The right of ownership of the tissue sample, of the description of my state of health and of other personal data and genealogy shall be transferred to the University of Tartu, the chief processor of the Estonian Genome Center.

/------------------/

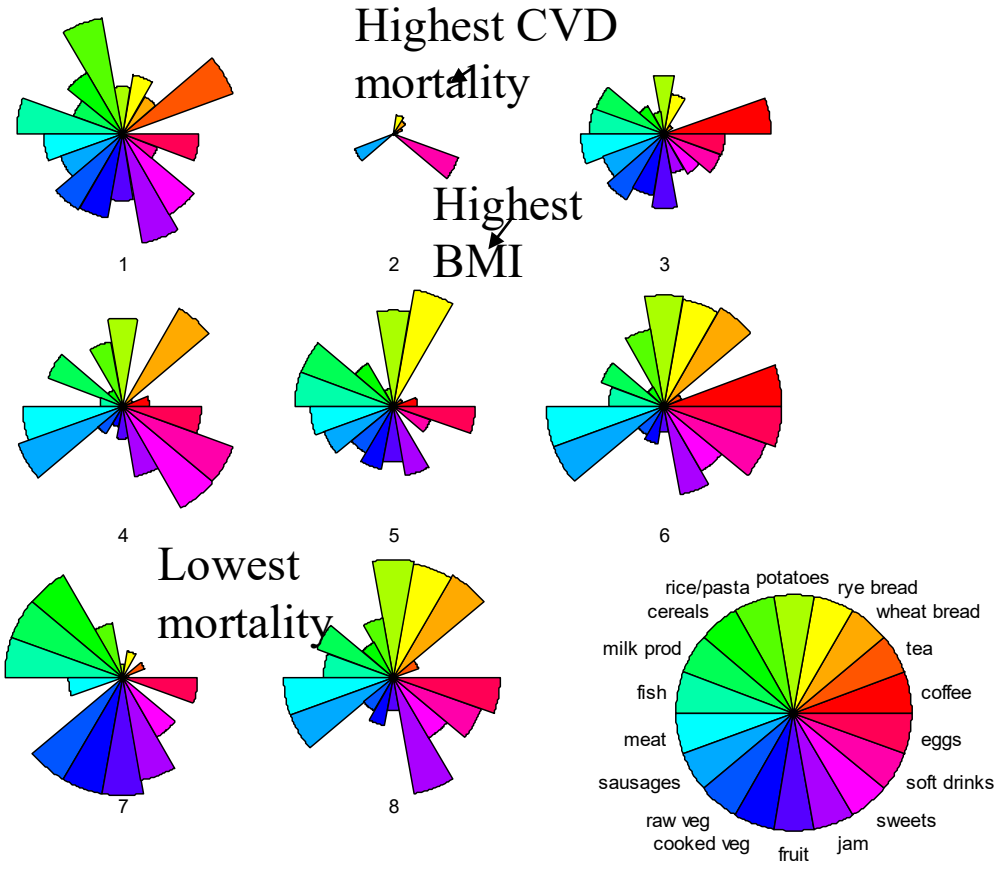
estonian genome center
university of tartu

Public opinion in 2024



eesti geenivaramu

Cluster analysis of diet data from the FFQ, no genetic data



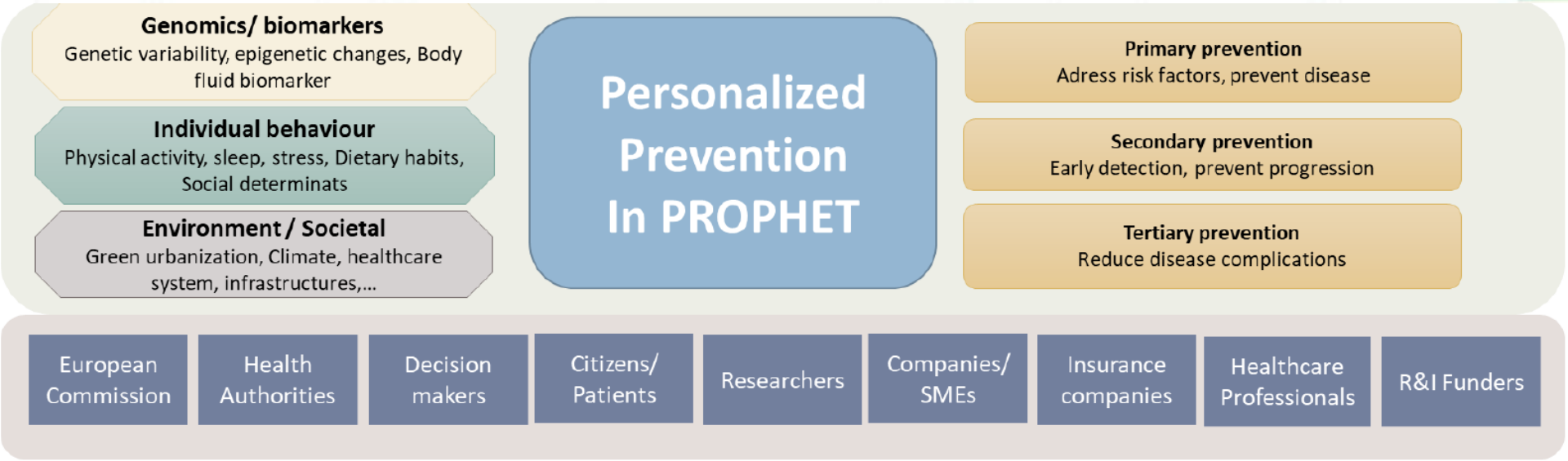
Why personalized prevention?

Chronic diseases remain the main cause of morbidity and mortality in Europe and the problem is getting worse from year to year:
Not enough medical doctors and of course, money.

The traditional reactive healthcare model, which focuses on treating established diseases, must evolve into personalized prevention, which prioritizes early diagnosis and risk reduction.
1.5% of the health care budget is not simply enough to prevent disease, starting from early detection of the disease risks.

We have too many patients because we do not prevent disease as much as we could using current research evidence on genomics , technologies (arrays and WGS) and IT/AI solutions.

Definition of Personalized Prevention



SRIA – Strategic Research and Innovation Agenda

- Challenge 1 : Continuous evidence synthesis system supporting personalised prevention
- Challenge 2: The PROPHET Framework implementation
- Challenge 3: Data collection and integration, and Data Infrastructure
- Challenge 4: Responsible Research and Innovation
- Challenge 5: Community Engagement and trust
- Challenge 6 : Health Professionals and Policy Makers involvement
- Challenge 7: Regulatory aspects and synergy with private sector
- Challenge 8: Access, Equity and Coverage
- Challenge 9: Ethical, Legal, Social Issues (ELSI)
- Challenge 10: Changing behaviour

The SRIA outlines the major areas that must be addressed in order to fully realize the potential of personalised prevention.

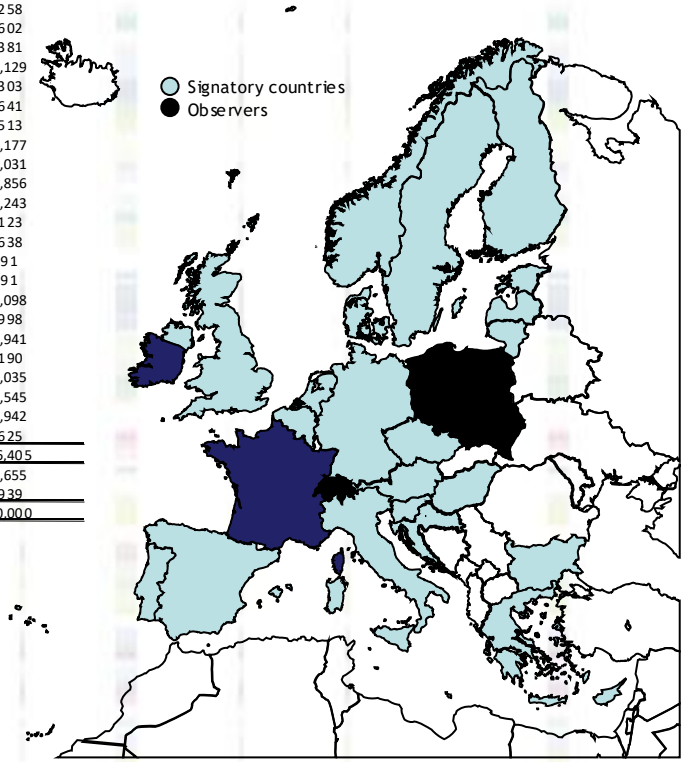
eesti geenivaramu

The European 1+ million genomes project: Genome of Europe reference genomes

1+MG Declaration of Cooperation 2018
cross-border access to genomic data (i.e., whole genome sequences; WGS) of 1+ million European citizens by 2027

- 26 signatory countries; 2 observers; 12 working groups

	Country	Population Size	Samples in GoE by population size	samples in GoE by population size - including observers
Signatory countries	AUSTRIA	8,901,064	10,566	9,417
	BELGIUM	11,521,238	14,514	12,936
	BULGARIA	6,520,000	9,265	8,258
	CROATIA	4,284,889	6,285	5,602
	CYPRUS	896,000	1,550	1,381
	CZECHIA	11,023,631	14,731	13,129
	DENMARK	5,850,189	7,072	6,303
	ESTONIA	1,308,739	1,841	1,641
	FINLAND	5,533,793	7,308	6,513
	GERMANY	83,020,000	101,177	90,177
	GREECE	10,635,997	14,620	13,031
	HUNGARY	9,937,628	13,302	11,856
	ITALY	59,236,213	78,811	70,243
	LATVIA	1,893,223	2,382	2,123
	LITHUANIA	2,795,680	4,082	3,638
	LUXEMBOURG	634,730	1,000	891
	MALTA	525,285	1,000	891
	NETHERLANDS	17,475,415	22,550	20,098
	NORWAY	5,415,166	6,730	5,998
	PORTUGAL	10,340,000	14,519	12,941
	SLOVENIA	2,107,007	2,457	2,190
	SPAIN	47,394,223	60,626	54,035
	SWEDEN	10,379,295	12,953	11,545
	FRANCE	66,732,539	85,206	75,942
	IRELAND	5,042,151	6,311	5,625
		389,404,095	500,858	446,405
Observers	POLAND	38,081,000	48,980	43,655
	SWITZERLAND	8,670,300	11,152	9,939
		436,155,395	560,990	500,000



(since 2021) Working Group 12 (coordinator A.Uitterlinden):
"Genome of Europe" : Creation of "reference genomes" (WGS 30x) of >500,000 European citizens (>40 subgroups by country of origin). Starting in 2024

eesti geenivaramu

Genome of Europe

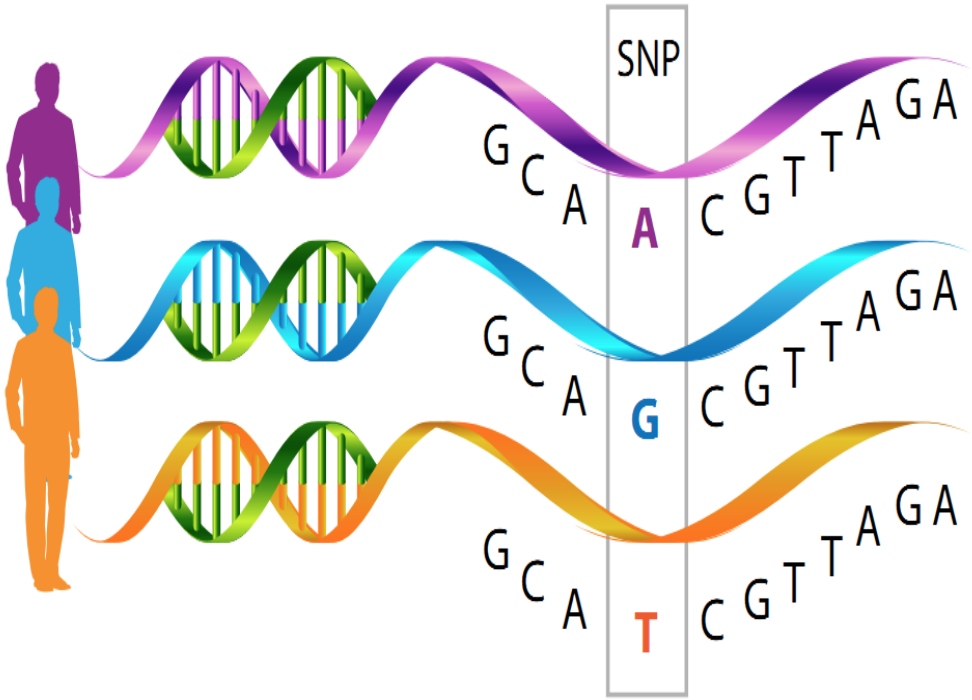
- Focus on general citizens as “normal control population”
- Focussed on creating overview of genetic variation across Europe: across country borders and as inclusive as possible
- A reference genome is important for ALL use cases: what is “normal” commonly seen genetic variation, in a particular ethnic/population background? (As compared, e.g., to a mutation for a rare disease, a tumour mutation, susceptibility alleles, clinical variants)
- Minimal phenotype is required (e.g., age, sex, country of origin). Focus is on genetic variation, not on linking it to disease/phenotypes, which happens in other 1+MG WG’s (e.g., WG 8, 10, 11, 12)
- Various sequencing technologies (short read, long read), various sequencing centres
- GoE is a crucial starting point and requirement for implementing genetics broadly into health care in hospitals, screening programs, and in society
- Project is ongoing, started in November 2024.

eesti geenivaramu

Genome-wide DNA analysis

Whole genome sequencing: $500

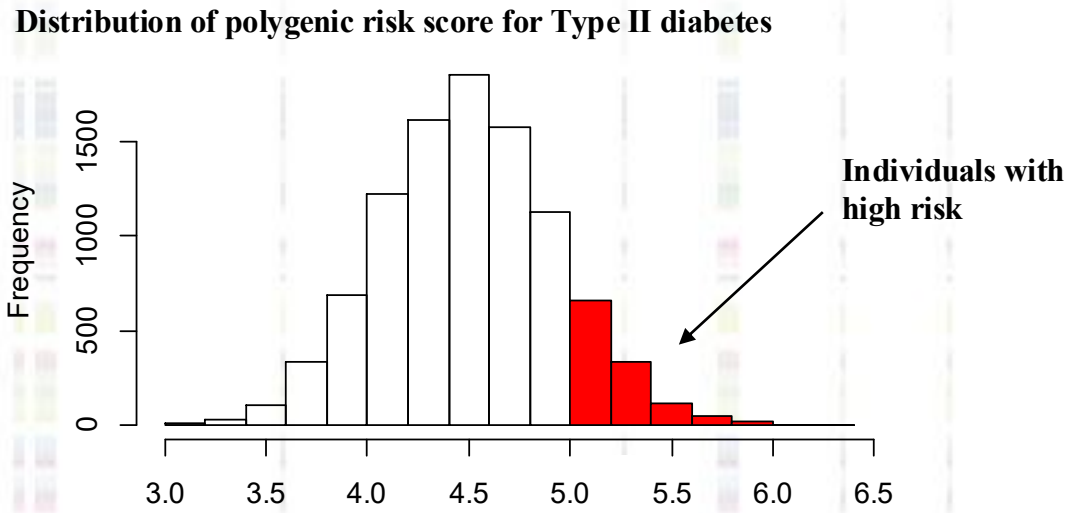
Genome-wide genotyping: $50



eesti geenivaramu

Polygenic risk scores for common diseases

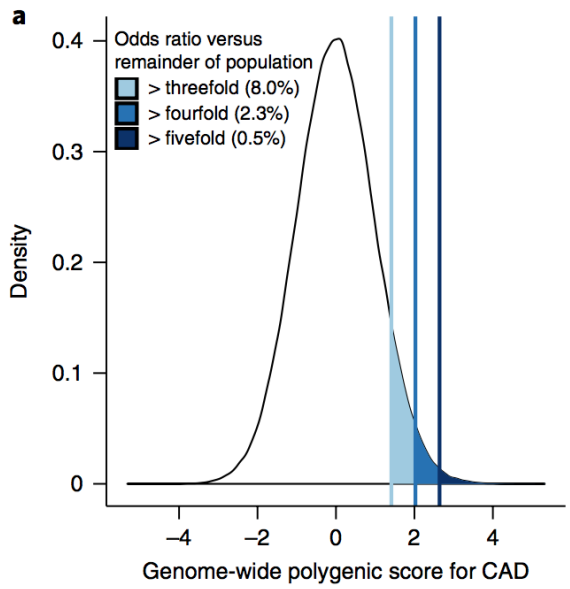
- Effect of each single variant is small
- Variants and effect sizes can be summed into **risk scores**

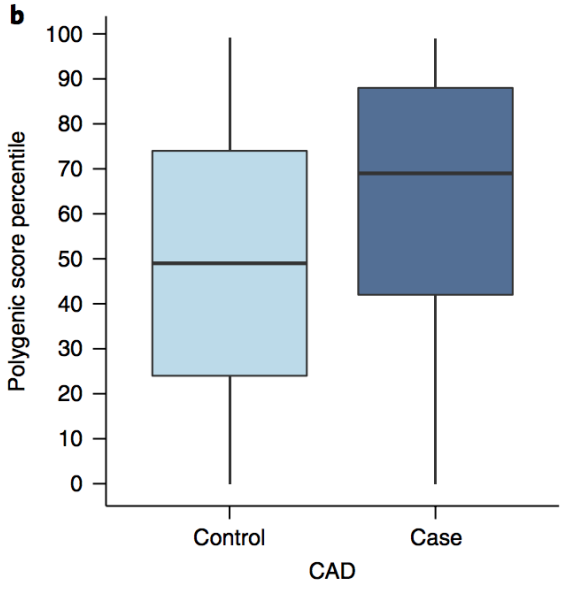


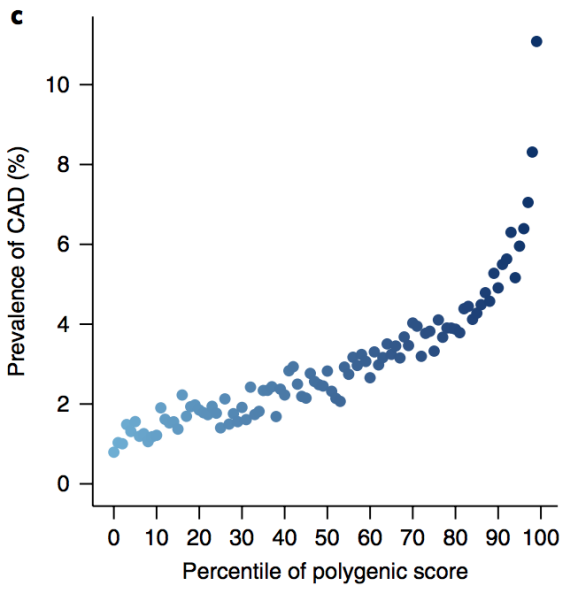
Prof Krista Fischer, Dr Kristi Läll

eesti geenivaramu

CAD polygenic risk score (PRS)







eesti geenivaramu

The PROBLEM with breast cancer screening

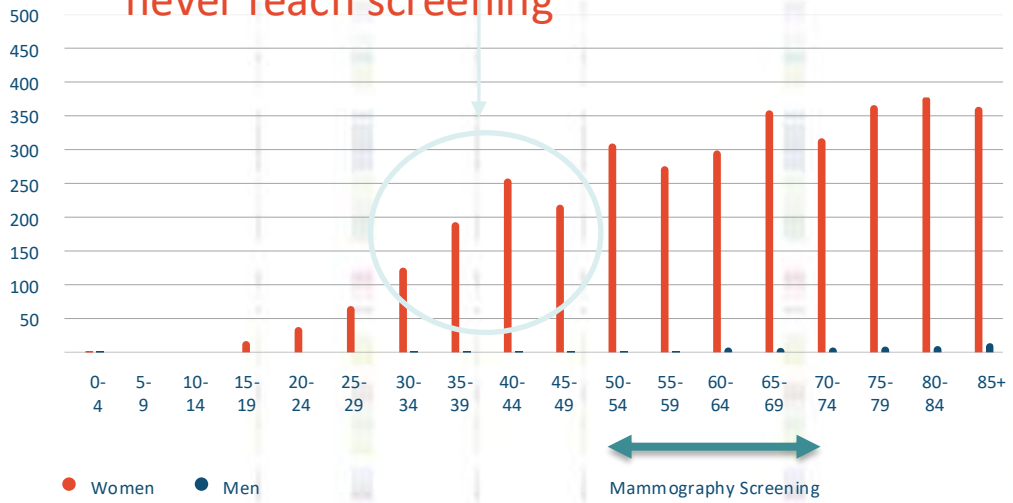
Antegenes®

- Mammography screening reduces breast cancer mortality by 20-30% in average risk level in age 50-69
- But also has negative aspects & costs
- The benefits & harms ratio, and also cost-efficiency are not satisfactory for mammography screening <u>for all</u> women under 50

<u>SOLUTION:</u>

Personalised PRS risk-based screening for women ages 30-49, earlier mammography for high PRS risk women

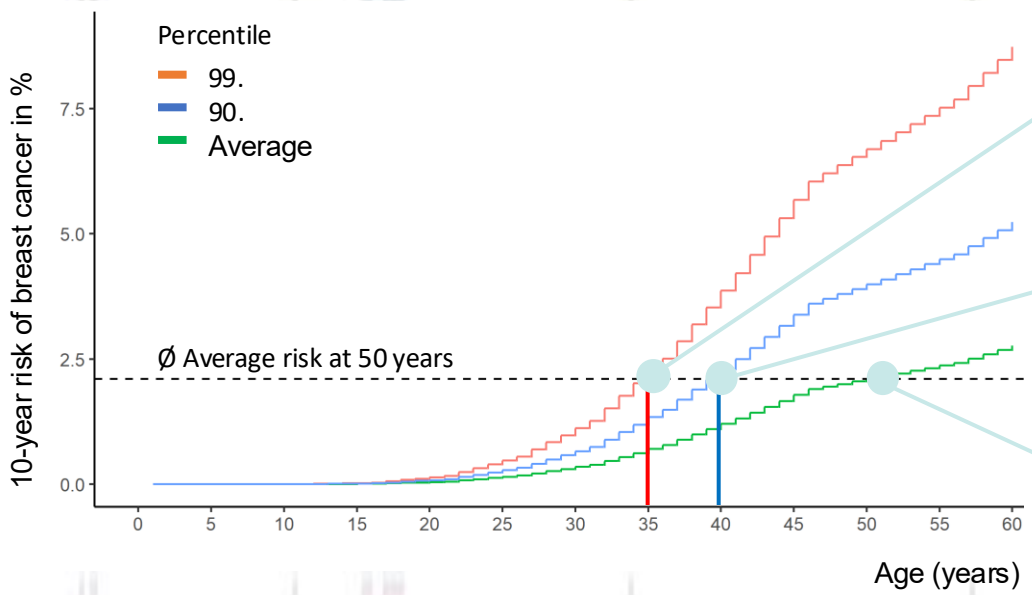
20% of breast cancer cases are diagnosed among women younger than 50, they never reach screening



eesti geenivaramu

The breast cancer PRS test AnteBC defines the age to start personalised breast cancer screening

Antegenes®



Very high risk (red, percentile 99)

1% of women with a high risk already exceed the risk threshold of an average 50-year-old woman at the age of 35.

High risk (blue, percentile 90)

10% of women at high risk exceed the risk threshold of an average 50-year-old woman at age 40.

Average risk (green, percentile 50)

Current screening threshold risk level

eesti geenivaramu

Movement Disorders
CLINICAL PRACTICE

RESEARCH ARTICLE

Polygenic Risk Score Combined with Transcranial Sonography Refines Parkinson's Disease Risk Prediction

Mart Kals, PhD,[1] Anu Reigo, MD,[1] Maris Teder-Laving, MSc,[1] Mariliis Vaht, PhD,[1] Estonian Biobank research team, Tiit Nikopensius, PhD,[1] Andres Metspalu, MD, PhD,[1,2] and Toomas Toomsoo, MD, PhD[3,4,]*

Aim: to determine the predictive value of marker combinations for the development of PD

eesti geenivaramu

Vision: Genomics of the common disease – early detection and prevention

PRS based disease risk analysis on population scale is currently the most innovative personal prevention strategy

1. Sequence ca 0.1% - 1% of the population and capture maximum amount of the genomic variation and use it for imputations.
2. Use SNP-arrays for the major (max) part of the population and impute the arrays. Start from 40-70 years old. Genotypes should go to expert center (Te.g. EHIK?), where array imputation, GWAS analysis and PRS calculations will be carried out.
3. **Take top (e.g.1-2%) of the high PRS individuals, forward thie info to primary care doctors and they apply intervention, where appropriate.**
4. **They do it already, but without genomic information (e.g.SCORE)**
5. **Population scale Personal Prevention**

eesti geenivaramu

Conclusions

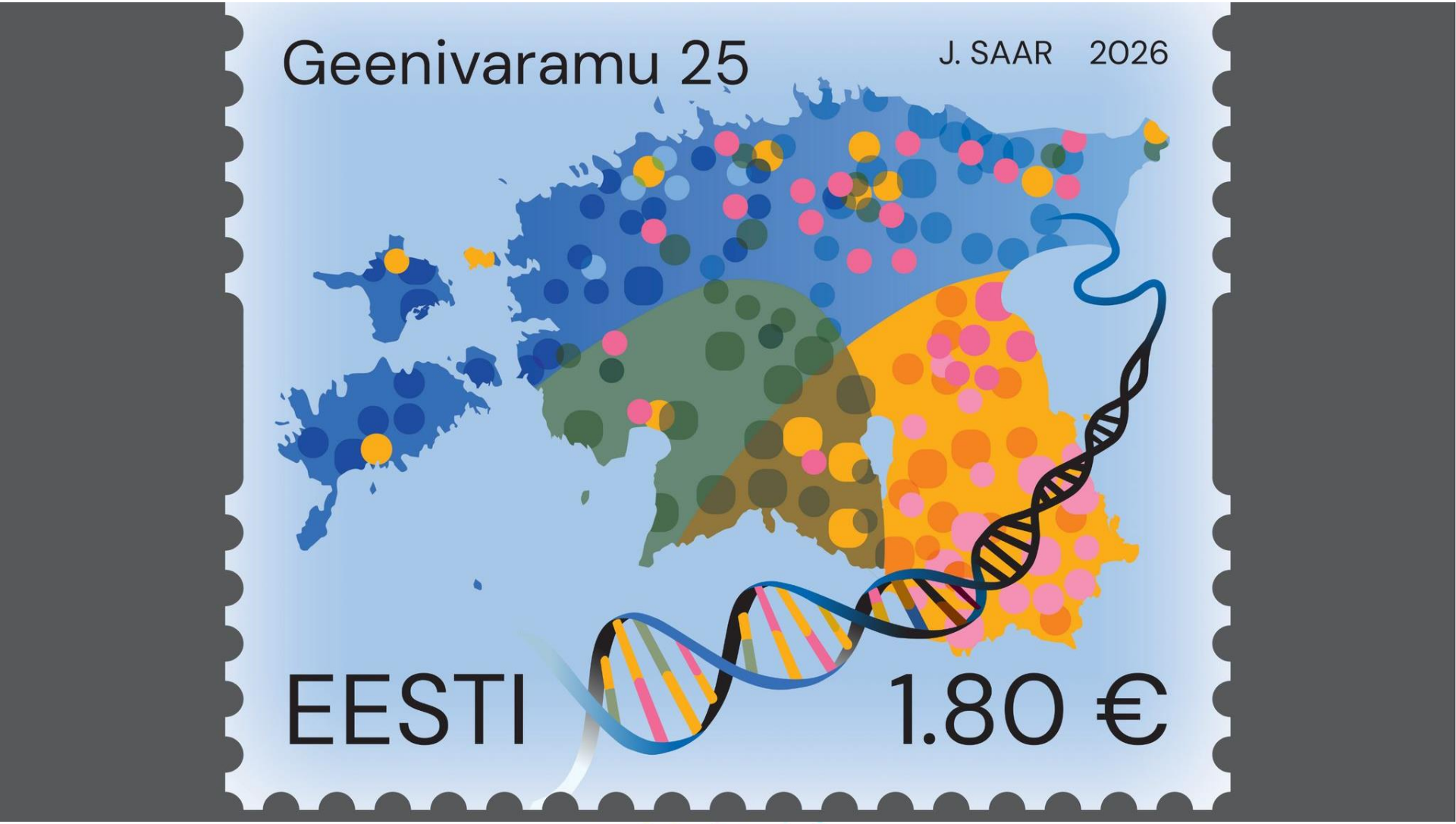
Large prospective biobanks make it possible to move towards personalized genetic risk prediction and to use it in general medical practice for preventing or postponing the disease or adverse drug reactions.

However, we should move from the biobanks (very good for research and discovery) **to the population based personal prevention based on genomic data**, but using other data (environment, behavior) as well to amplify the potential risks.

Whole population = biobank

Moreover, we could move from the PRS to real diagnostics by implementing the secondary (diagnostic!) test to high PRS individuals, which can not be used on whole population e.g. ultrasound in PD, liquid biopsy in cancer risk etc.

eesti geenivaramu



eesti geenivaramu