From Biobanking to Personal Medicine: an Estonian case

Andres Metspalu

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

Nationales Biobanken-Symposium 2023
Berlin, 25. May 2023

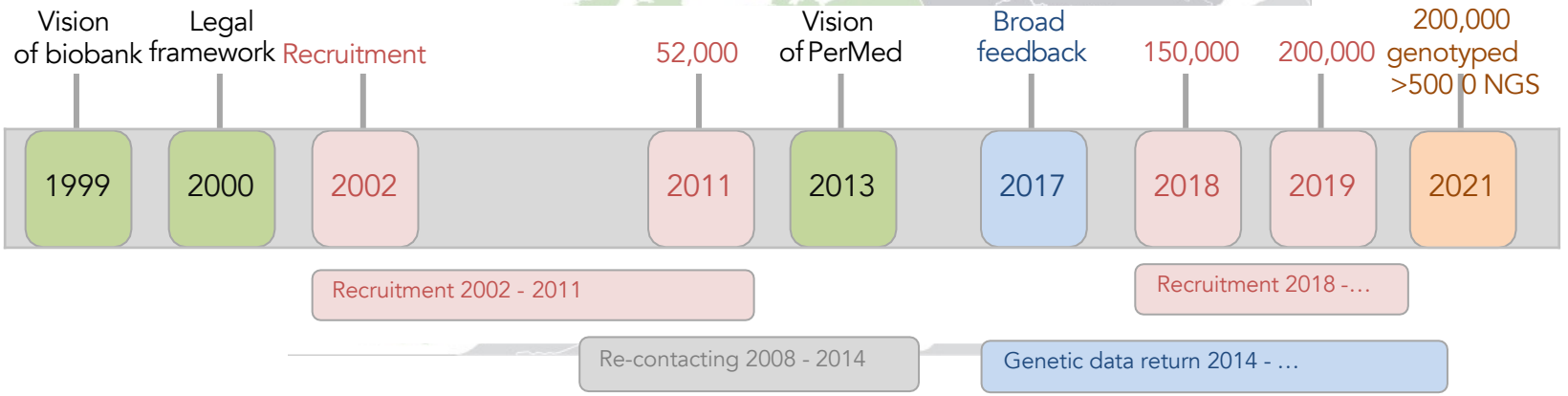
estonian genome center
university of tartu

Estonian Biobank (started in 1999!)

1. Prospective, longitudinal, volunteer-based
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,
4. **210 000 individuals = 20% of the adults (18 years and up) population of Estonia, all are genotyped by GSA and data are imputed against Estonian WGS ref panel of 2300 individuals**

estonian genome center
university of tartu

Estonian Biobank timeline



estonian genome center
university of tartu

Slide:
Katrin Männik

Human Genes Research Act Passed 13.12.2000

RT I 2000, 104, 685

§ 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 – vision for personalized medicine

Consent Form

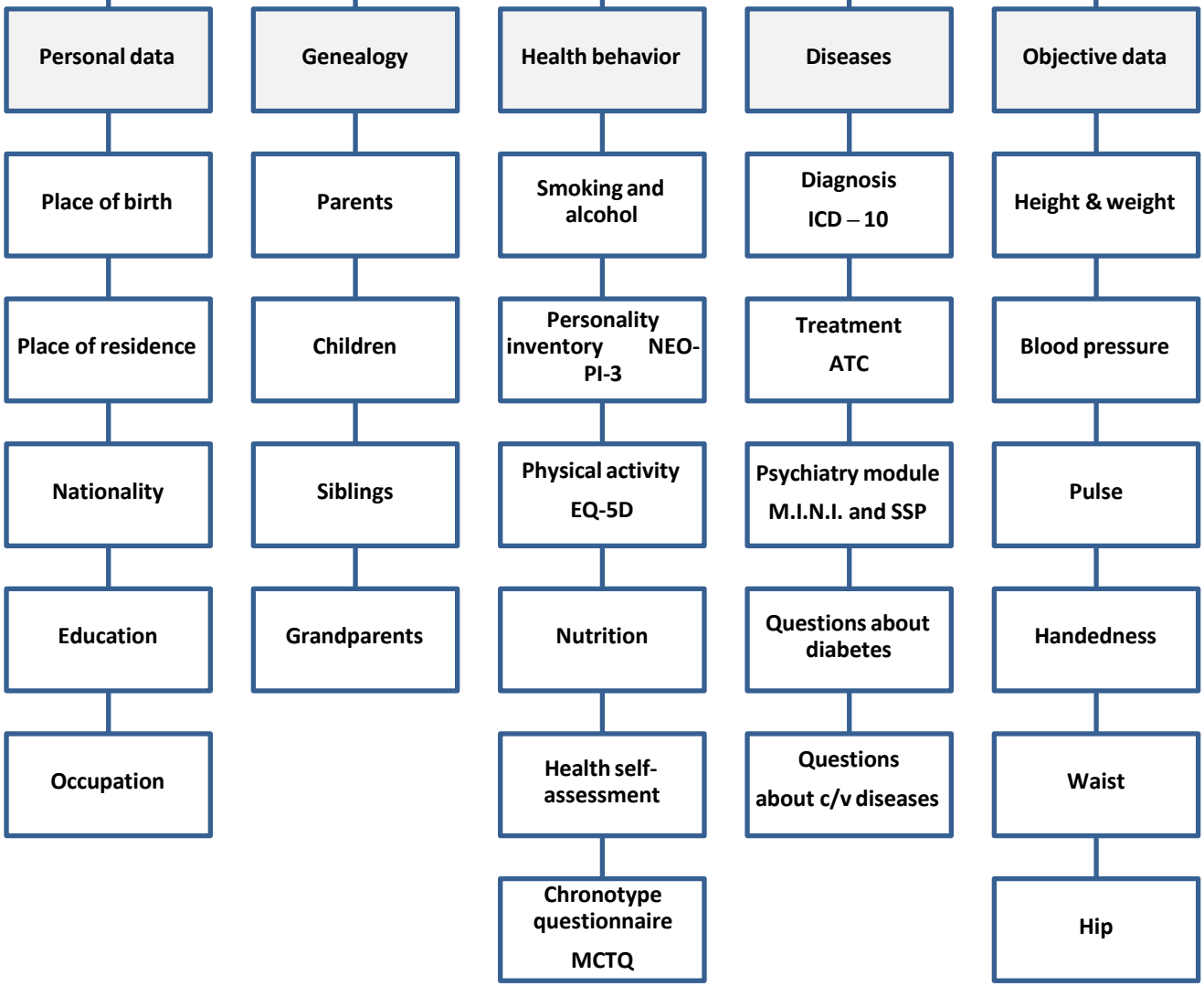
/_____/

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.

/_____/

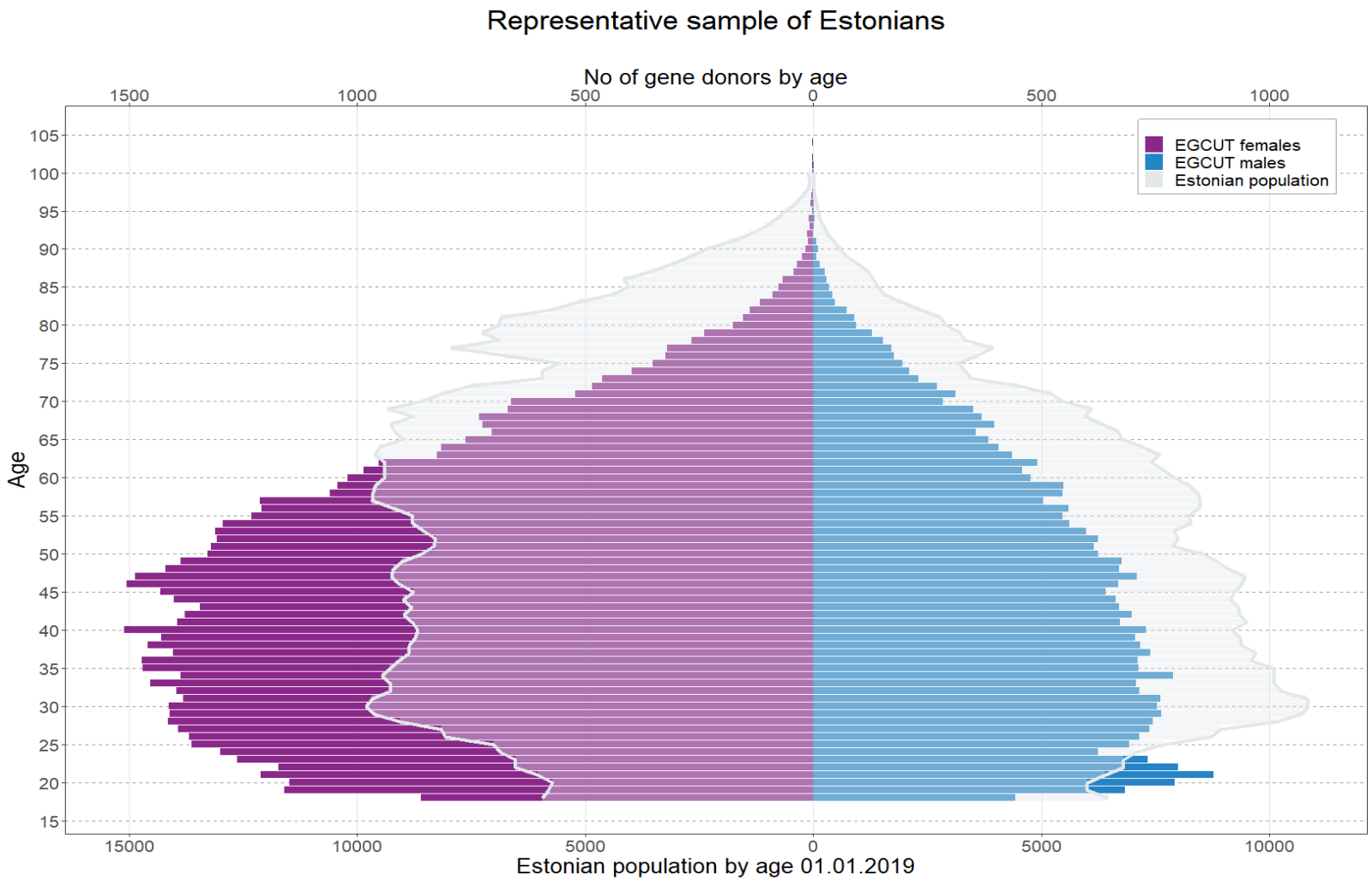
estonian genome center
university of tartu

Questionnaire of EGCUT



estonian genome center
university of tartu

210 000 gene donors in the EstBB



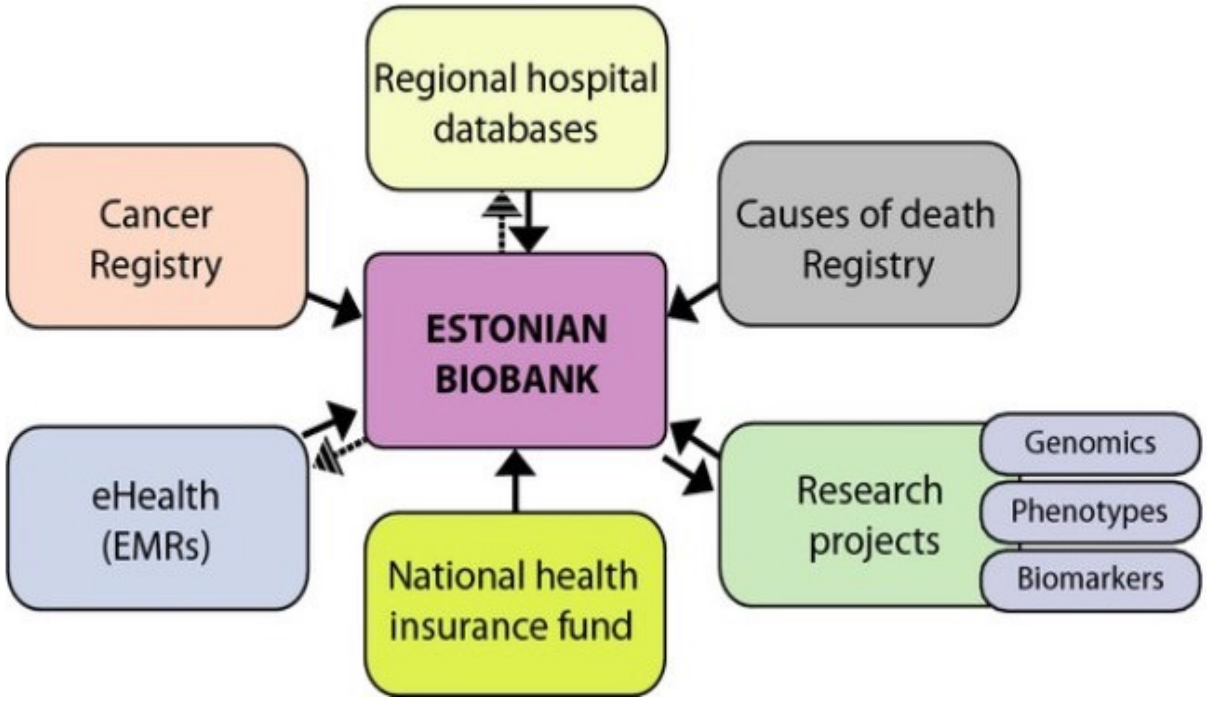
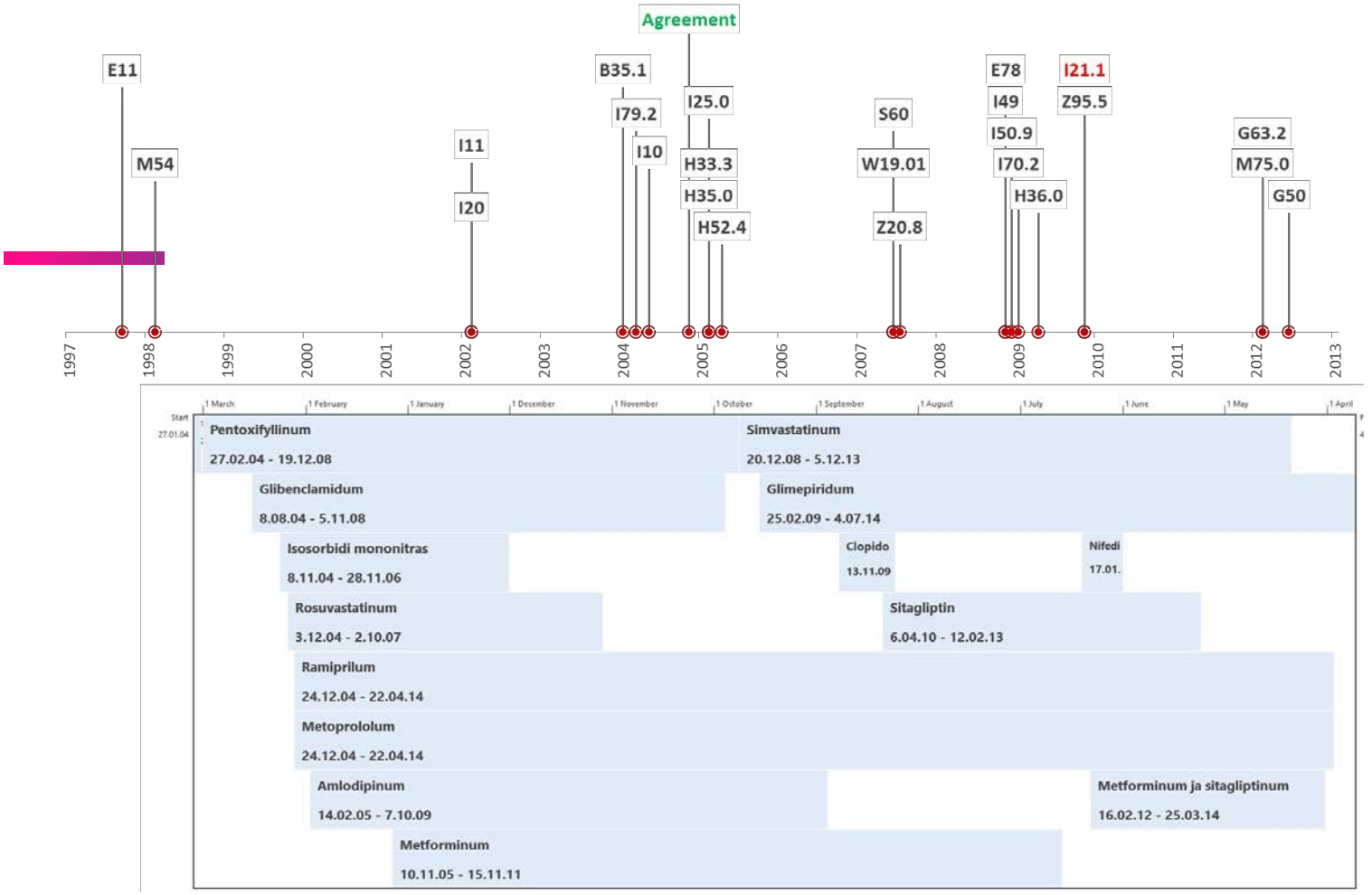


Figure 3. National registries and databases for enrichment of phenotype data in the Estonian Biobank. The schematic diagram illustrates the different layers of information available in the database of the Estonian Biobank, which is continually being updated by queries to the Estonian Causes of Death Registry, the Estonian Cancer Registry and the Digital Prescription Database of the Estonian Health Insurance Fund, as well as electronic medical records (EMRs) from the databases of the two major hospitals in Estonia. Data generated through research projects must be returned to the Biobank within 5 years of the original data release from the Biobank.

Disease trajectories + treatment info for people in the biobank

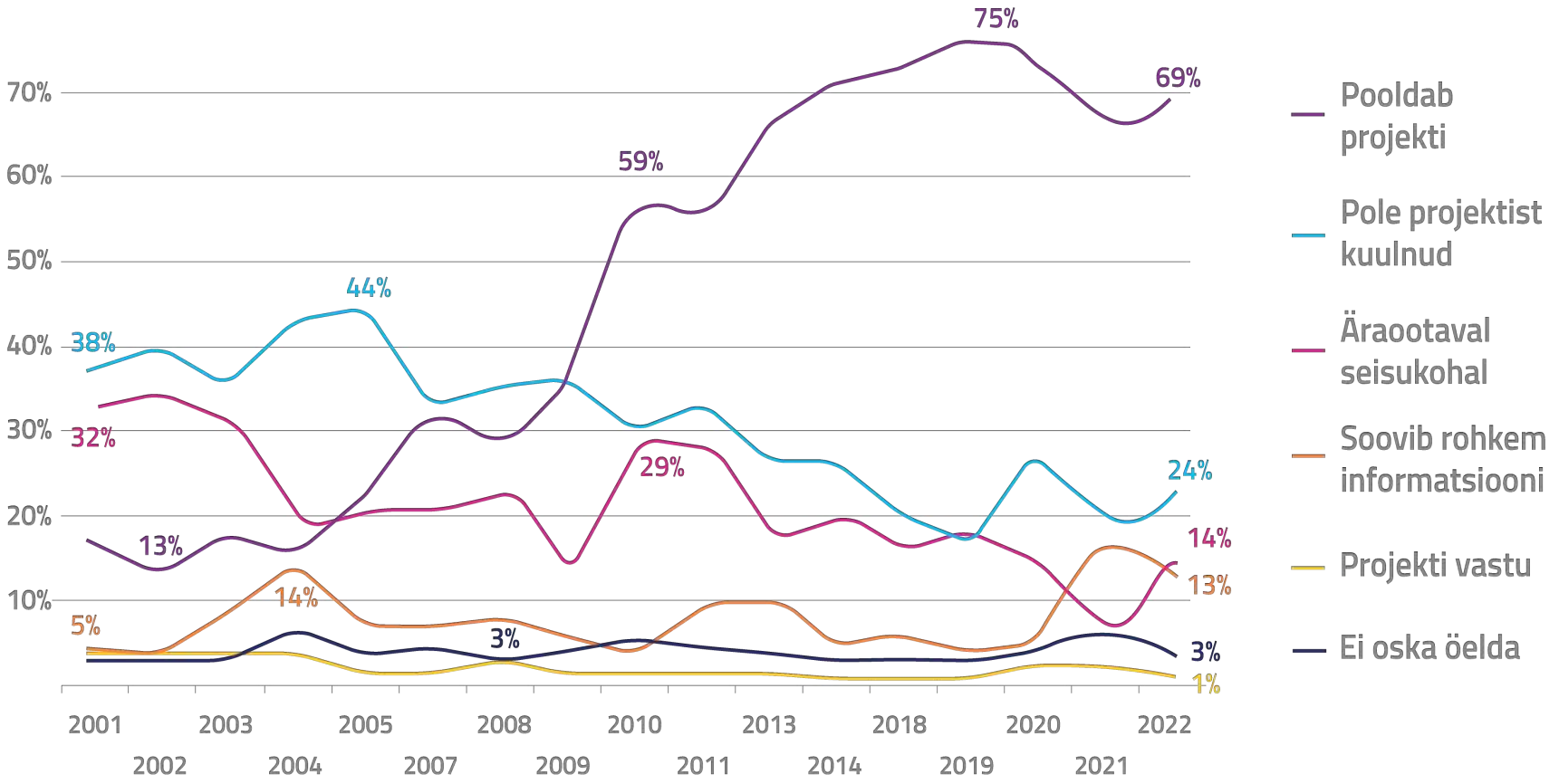


Male, born 1944

Estonian biobank 205 000 subjects: omics profiling

Method	Sample size
Whole genome sequencing (30X)	3,000
Whole exome sequencing	2,500
Genome-wide genotyping arrays	205,000
Genome-wide methylation arrays	700
Genome-wide expression arrays	1,100
mRNA sequencing	600
Total RNA sequencing	50
Metabolomics (NMR – Nightingale Health)	200, 000
Metabolomics (MS/MS)	1,100
Telomere length	5,200
Clinical biochemistry	2,700
Microbiome	2,500
IgG glycosylation	1,000

Public opinion



estonian genome center
university of tartu

Vision: Genomics of the (common) disease

(FH, T2D, BrCa, PGx)
"PRS - *Estonian approach*"

1. Sequence (WGS) ca 0.1% - 1% of the population and capture maximum amount of the genomic variation and use it for imputations (get common variants which are not on the array).
2. Use SNP-arrays for the major part of the population and impute the arrays
3. Use the imputed SNP data for PRS and pharmacogenetics
4. We spent ca 50€ per individual to recruit, acquire health data and genotype
5. **Population scale Personal Prevention**

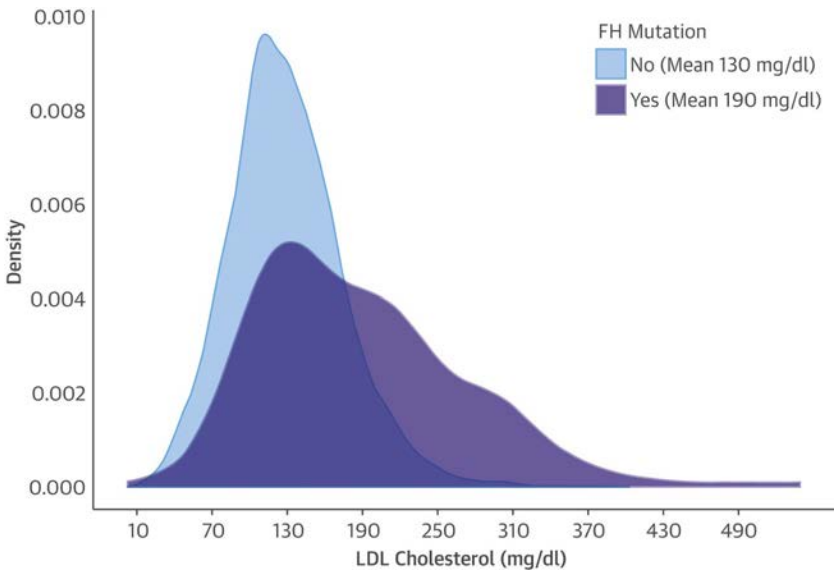
estonian genome center
university of tartu

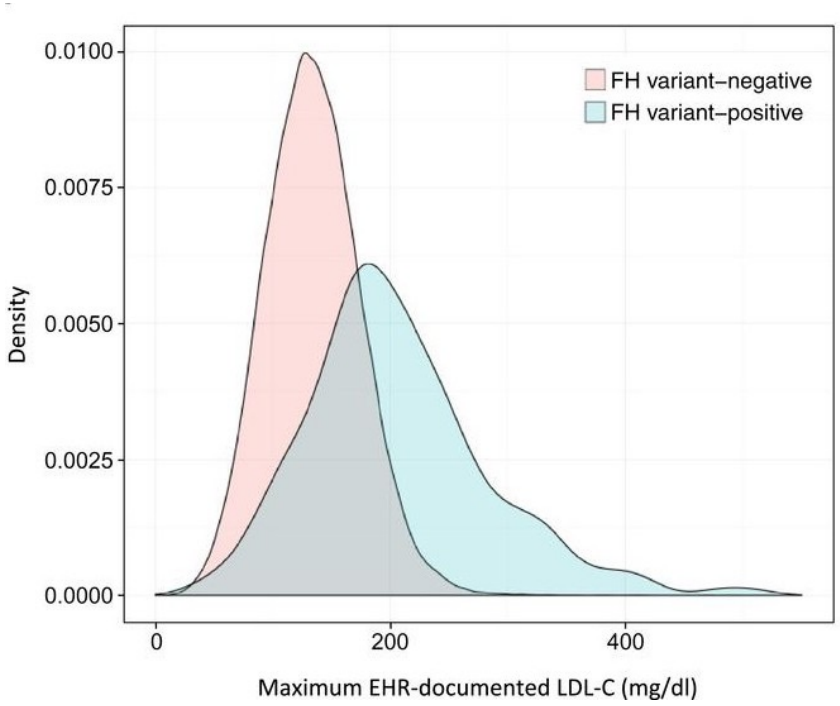
3 examples:

1. Familial hypercholesterolemia (FH) - Alver *et al.* (2018), GIM Genetics first" approach;
2. Breast cancer – Läll *et al.*, BMC Cancer (2019);
3. Pharmacogenomics

Familiar hypercholesterolemia - FH

FH-linked variant (*LDLR*, *APOB*, *PCSK9* gene) carriers display **50 mg/dl** (1.3 mmol/L) and **greater** and a **wide spectrum** of LDL-C level



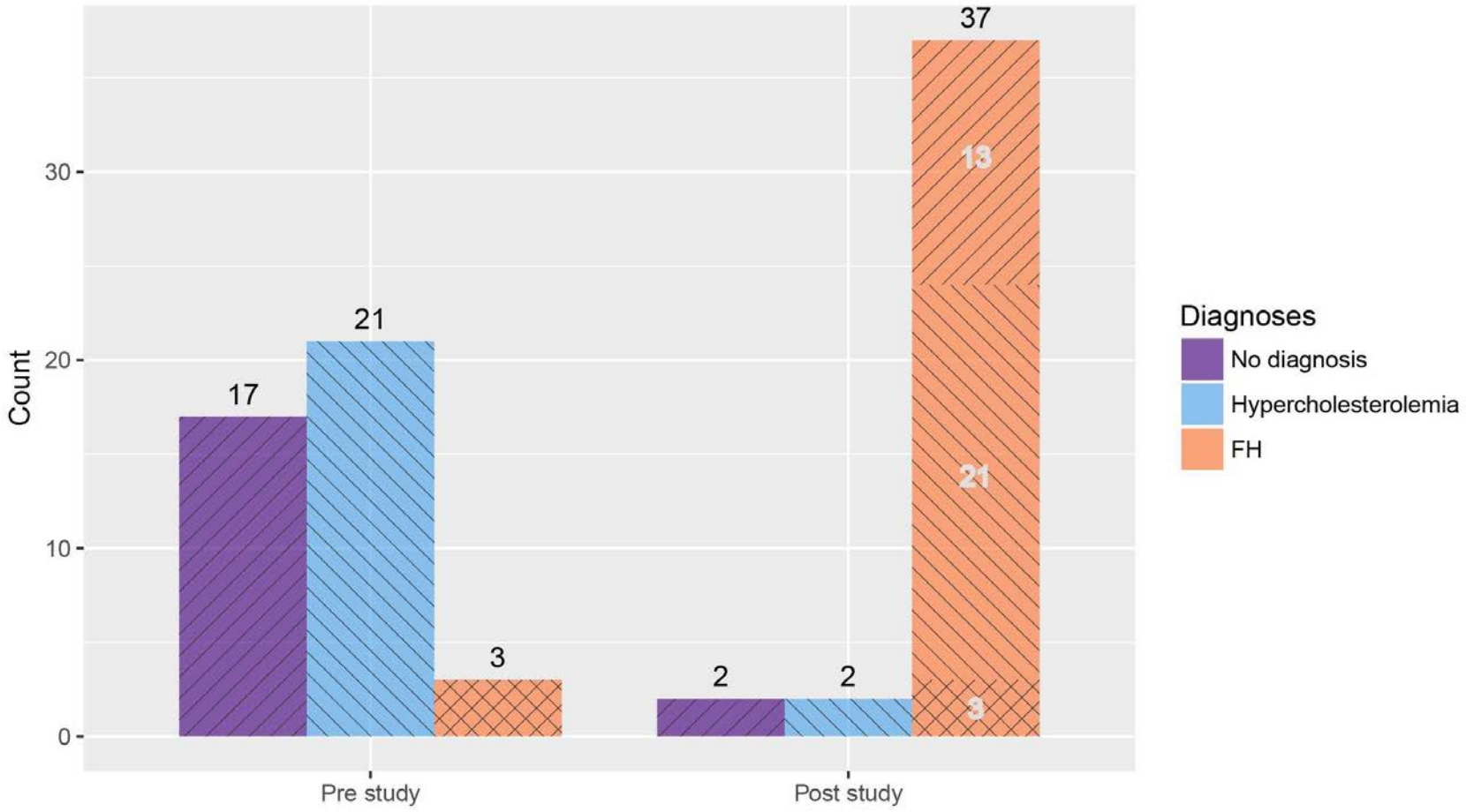


Diagnostic LDL-C level cut-off for FH cases >4.9 mmol/L

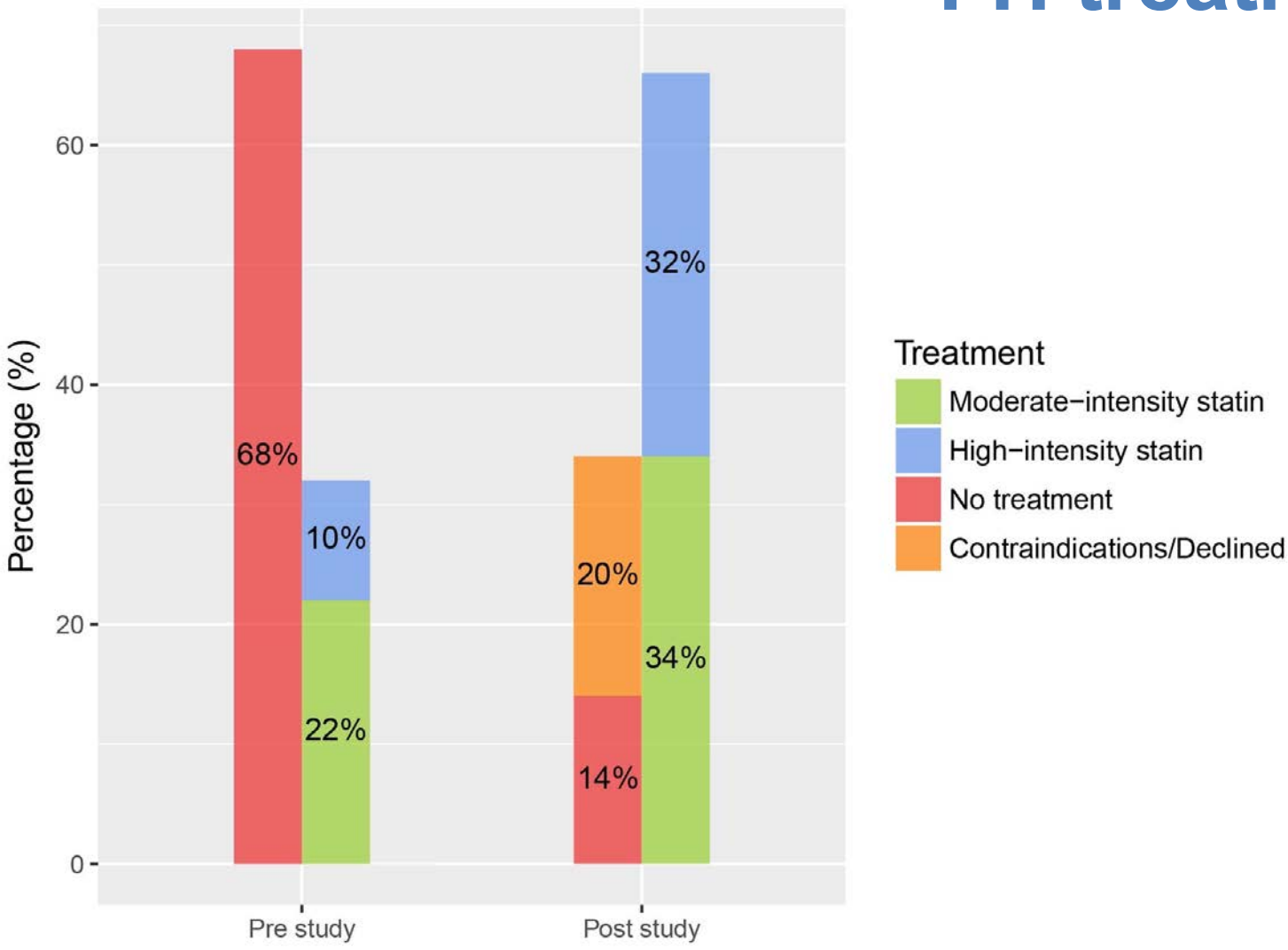
Khera et al. J Am Coll Cardiol. 2016
Abul-Husn et al. Science 2016

Alver *et al.* (2018) Genetics in Medicine

FH diagnose


Count
30
20
10
0
17
21
3
2
2
37
16
21
3
Pre study
Post study
Diagnoses
No diagnosis
Hypercholesterolemia
FH

FH treatment



FH Summary

Under-diagnosis and under-treatment

- reclassified 51% from having non-specific hypercholesterolemia to having FH, half of them were on statins, but none had LDL-C below treatment goals
- identified 32% who had gone unrecognized by the medical system
- Reliable identification of new FH cases and people with high GRS which has direct impact on family members

Insensitivity of current criteria used in FH diagnosis

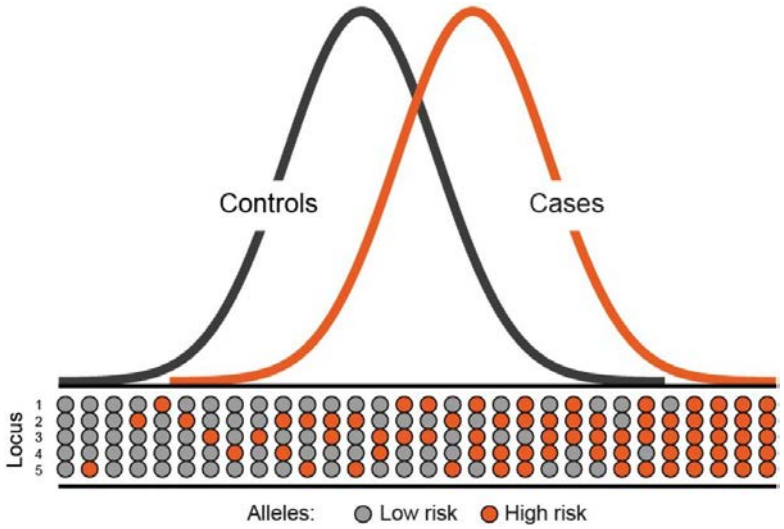
- wide spectrum of LDL-C levels
  - 34% had LDL-C levels ≤4.9 mmol/L
- visible accumulations of lipid deposits detected in 5% only
- heterogeneity in clinical expression
- Cascade

estonian genome center
university of tartu

Polygenic risk scores (PRS)

- Most of the associated loci identified in GWAS have very small effects

- Polygenic risk score can be constructed by combining the effects of all associated loci
  - –unweighted: sum of all risk alleles
  - – weighted: sum of all risk alleles weighted by their effect size



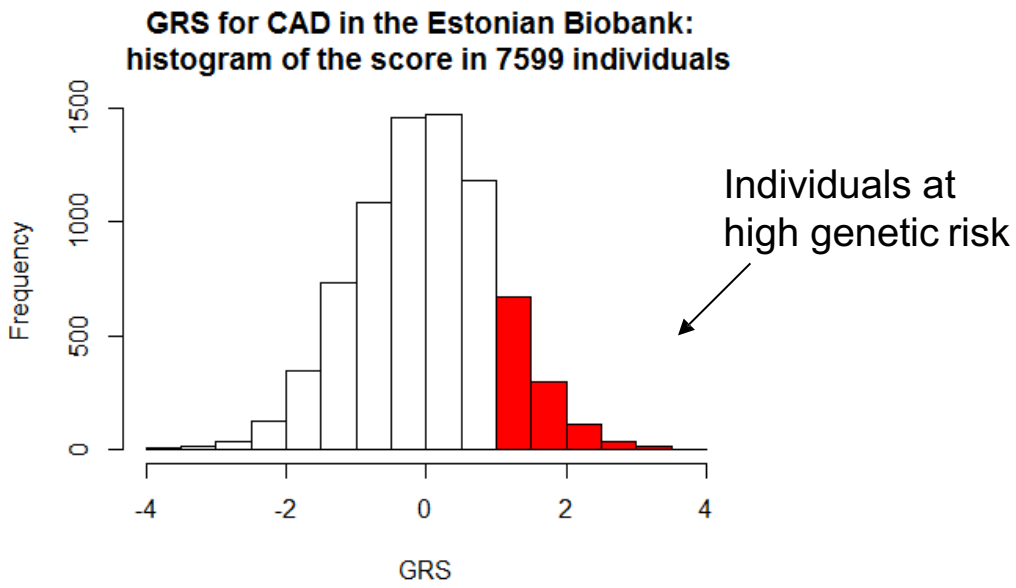
estonian genome center
university of tartu

Polygenic risk scores (PRS) weighted: sum of all risk alleles weighted by their effect size

Calculated as $S = w_1X_1 + w_2X_2 + \ldots + w_kX_k$,

$X_1, \ldots, X_k$ - allele dosages for k independent markers (SNP-s),

$w_1, w_2, \ldots, w_k$ – weights



Methodological questions:

A) How to select the SNPs – how many and what are the selection criteria?

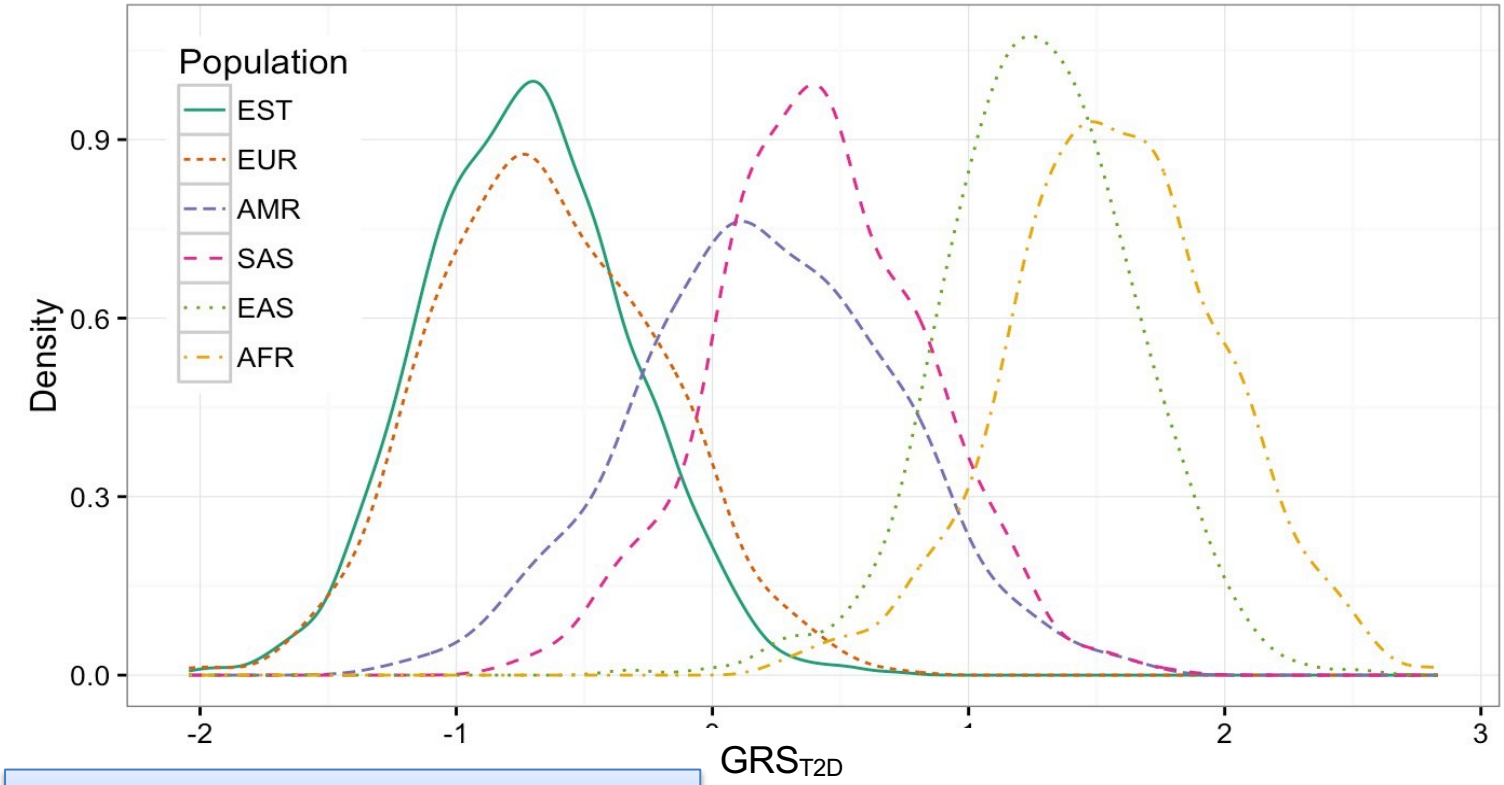
B) How to select the optimal weights?

K. Läll …. & K. Fischer, GM, 2016

GWAS – SNP data source

How much does a risk model depend on the population where it is developed?

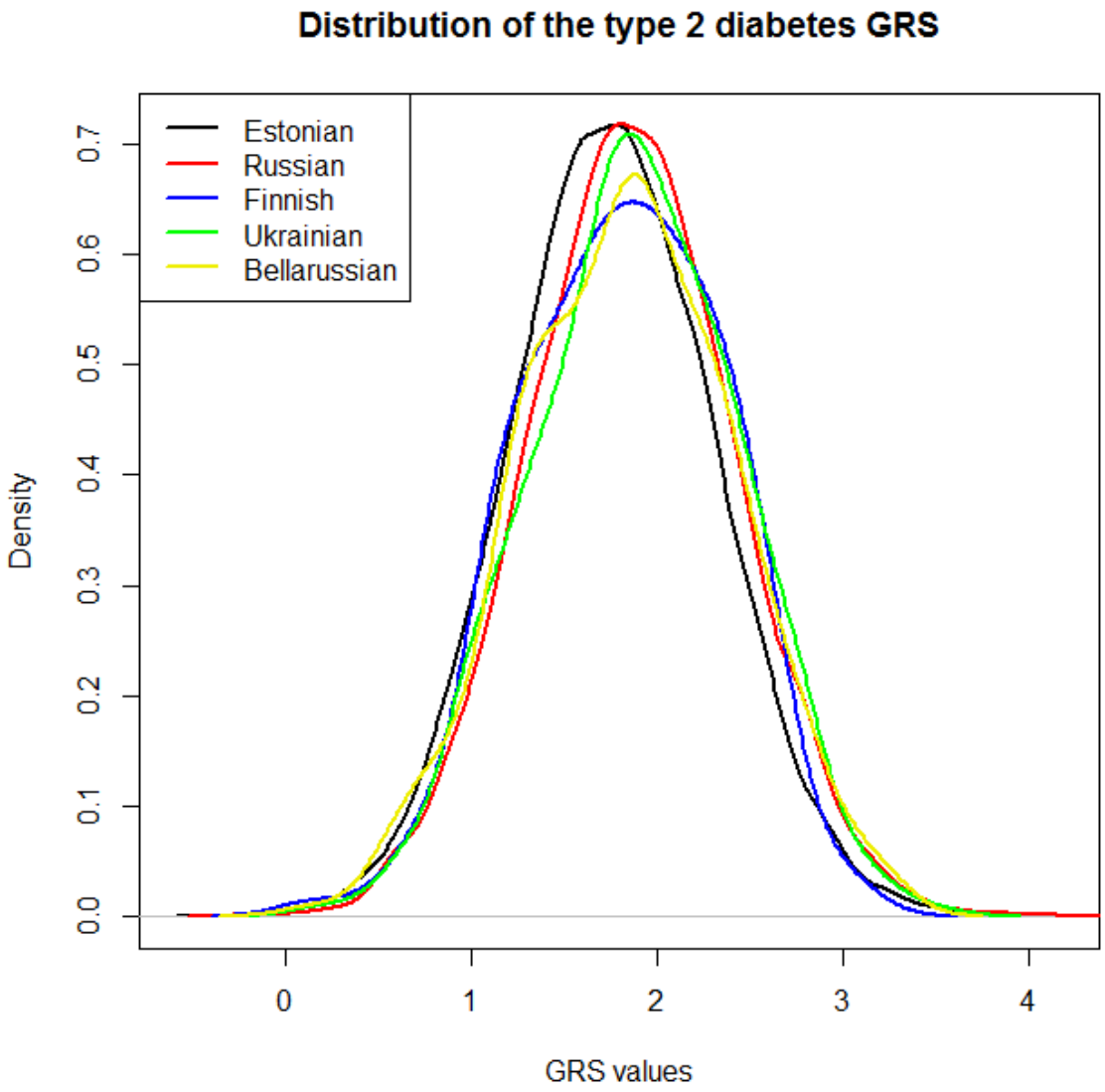
Genetic risk score distributions in different populations



Reisberg *et al*. 2017. PlosONE

estonian genome center
university of tartu

Regional PRS are rather similar



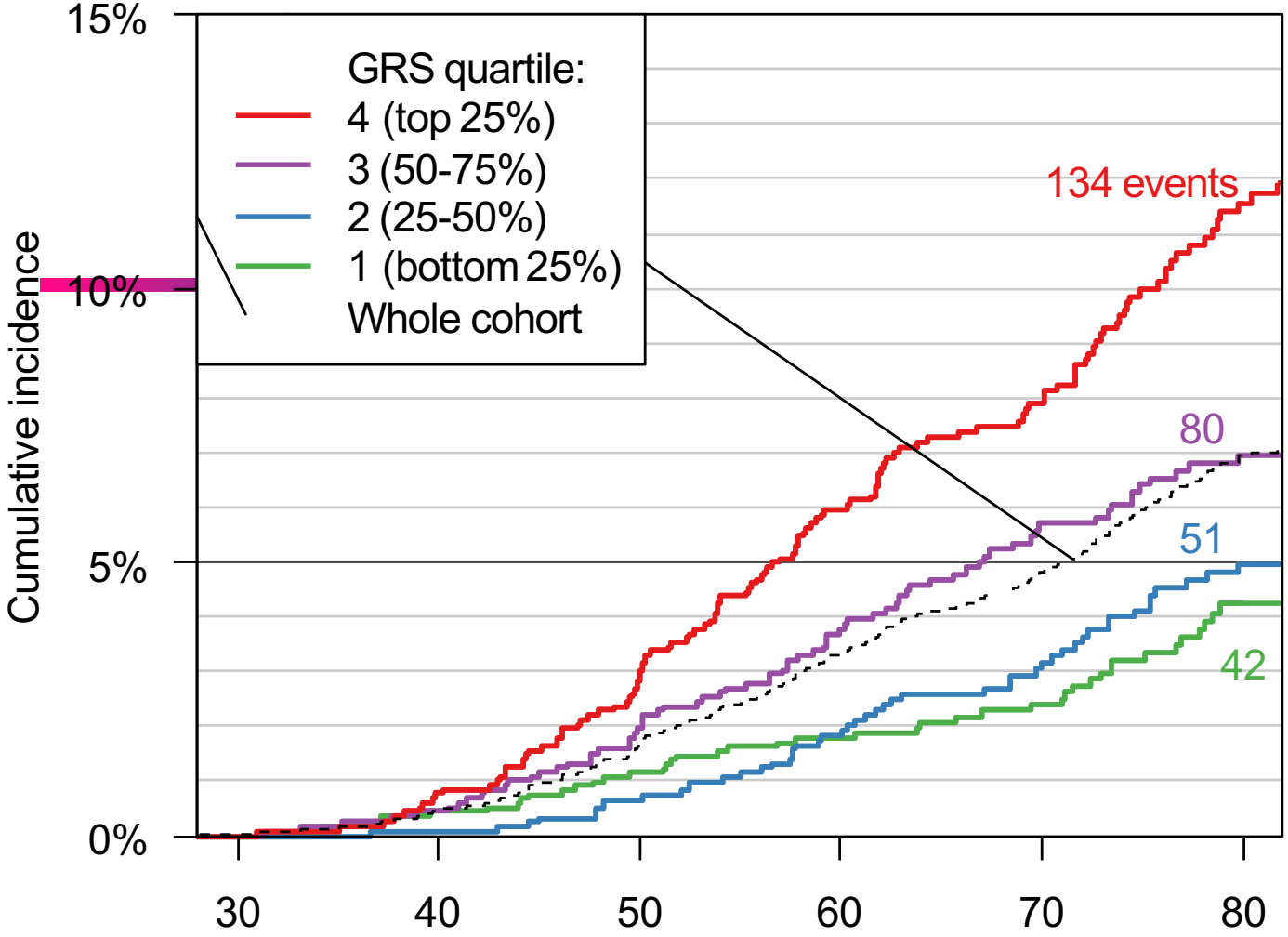
PRS of Breast Cancer

- No BRCA1 & BRCA2, but ca 900 SNP variants
- In Estonia, ~700 new BC cases every year (~10% BRCA, 630 non BRCA cases) -> extrapolation to population
  - In GRS 0%-10% - 5% of cases*630 = 32
  - In GRS 10%-90% - 74% of cases*630 = 466
  - In GRS 90%-100% - 21% of cases*630 = 132

Läll *et al.* (2019) *BMC Cancer 19, 557*

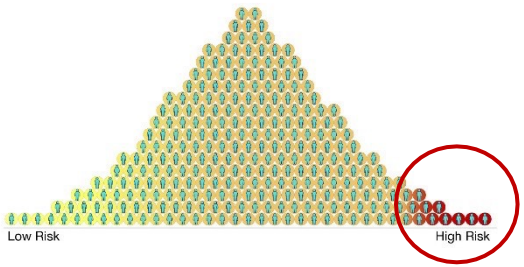
estonian genome center
university of tartu

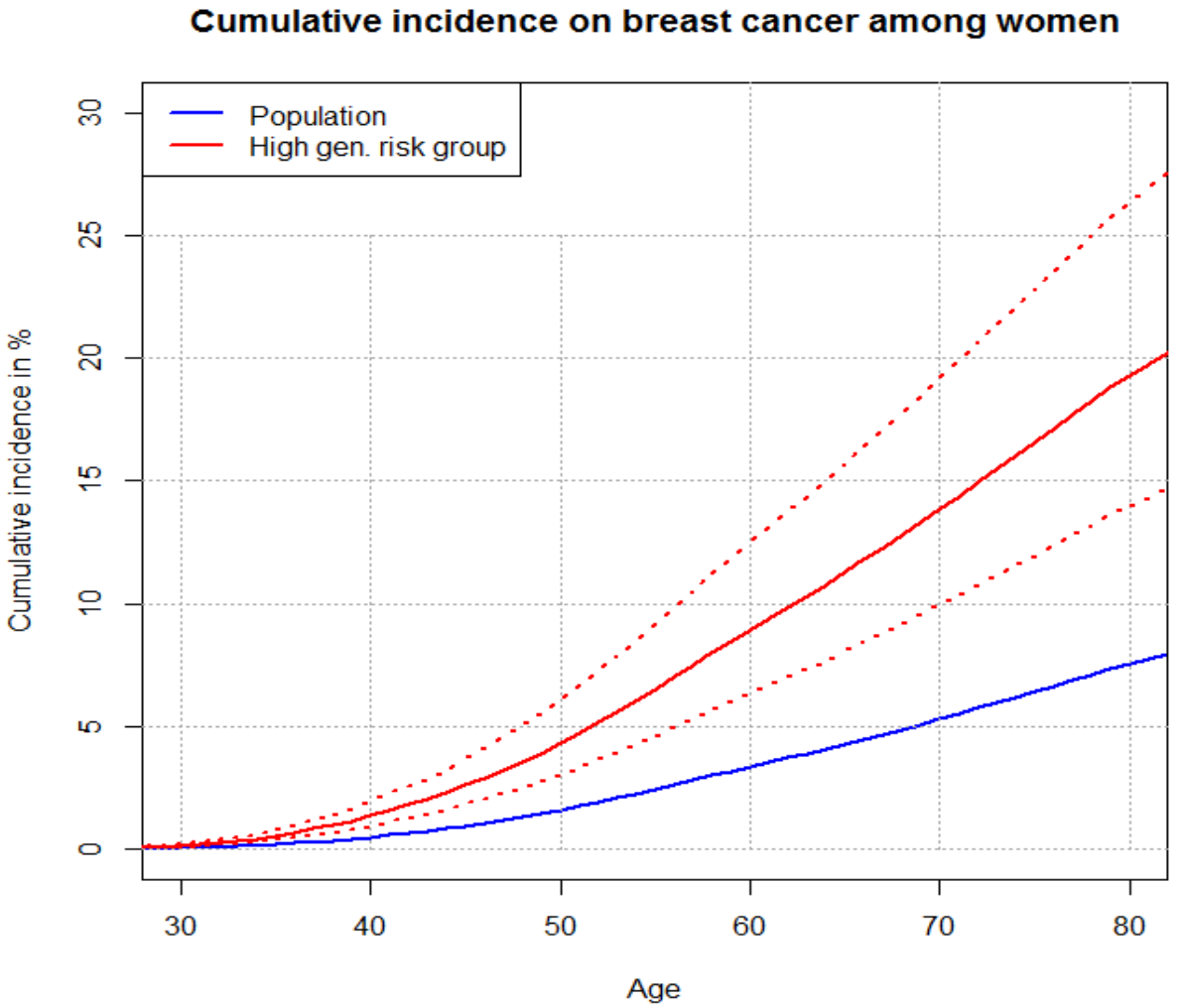
Breast Cancer risk by GRS quartile (317 incident cases in 33554 women)



Läll *et al* (2018)

Breast cancer: population vs top 5% Based on Polygenic Risk Score

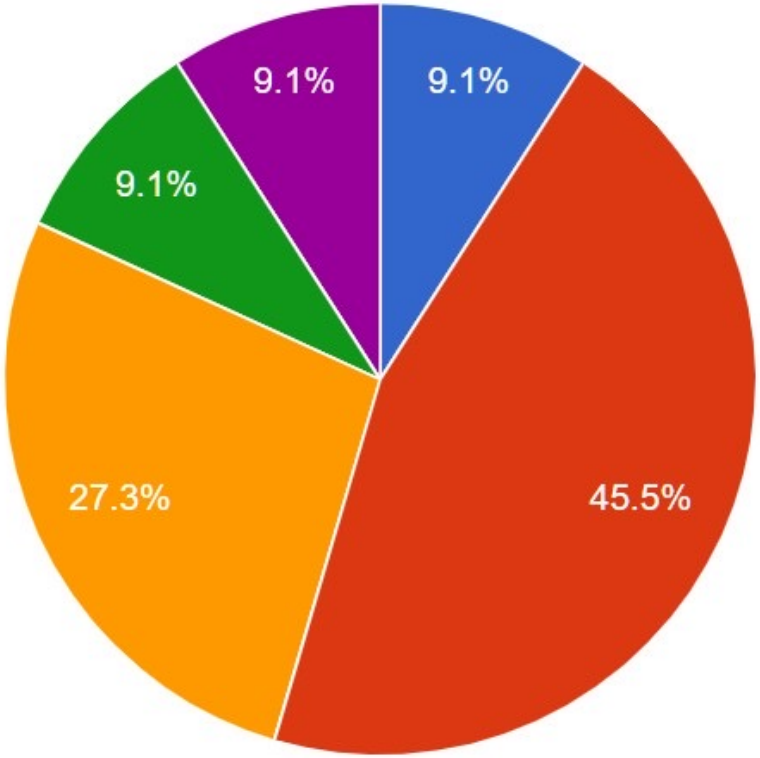




Läll K, et al. BMC Cancer 2019

Incident breast cancer cases in high-PRS group (preliminary data)

0	(1, 9.1%)
IA	(5, 45.5%)
IB	(3, 27.3%)
IIA	(1, 9.1%)
IIB	(1, 9.1%)
IIIA	(0,0,0%)
IIIB	(0,0,0%)
IIIC	(0, 0,0%)
IV	(0, 0,0%)



PRS - 1/38 (present study)
mammography 1/ 244 participants
(Kiivet RA, et al. 2015)
In PRS cases 50% were younger than 52 years

manuscript in preparation, N. Tõnisson

estonian genome center
university of tartu

Intervention?

Perhaps for the high risk group start mammography/MRI 10-15 years earlier, perform liquid biopsy

Clinical study to test PRS clinical utility is underway in TU hospital

estonian genome center
university of tartu

Europe's Beating Cancer Plan

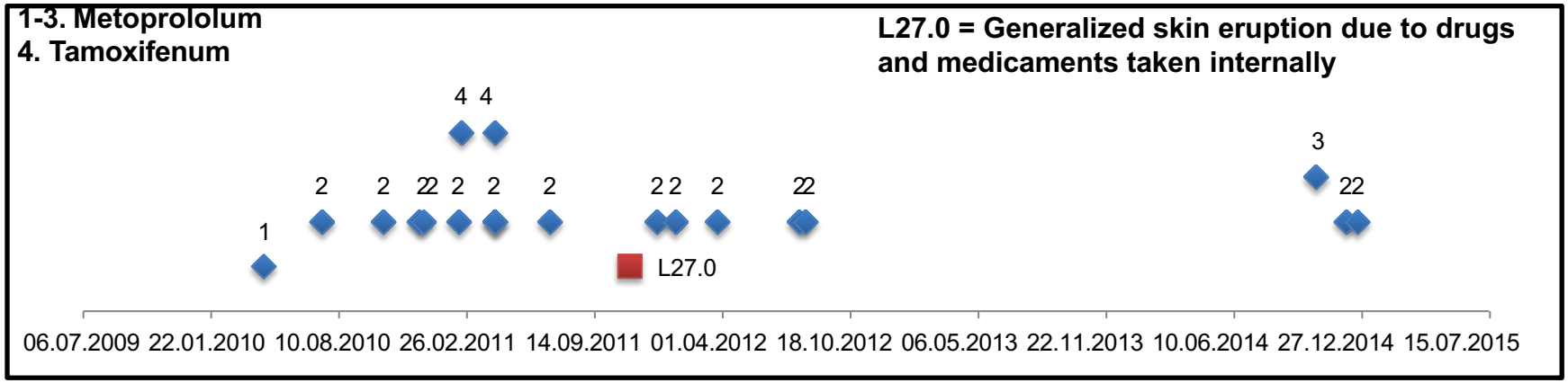
- **Flagship 7:** Alongside the 'Genomic for Public Health' project, the European Initiative to Understand Cancer (UNCAN.eu), planned to be launched under the foreseen Mission on Cancer to increase the understanding of how cancers develop, will also help identify individuals at high risk from common cancers using **the polygenic risk scores technique**. This should facilitate personalised approaches to cancer prevention and care, allowing for actions to be taken to decrease risk or to detect cancer as early as possible.

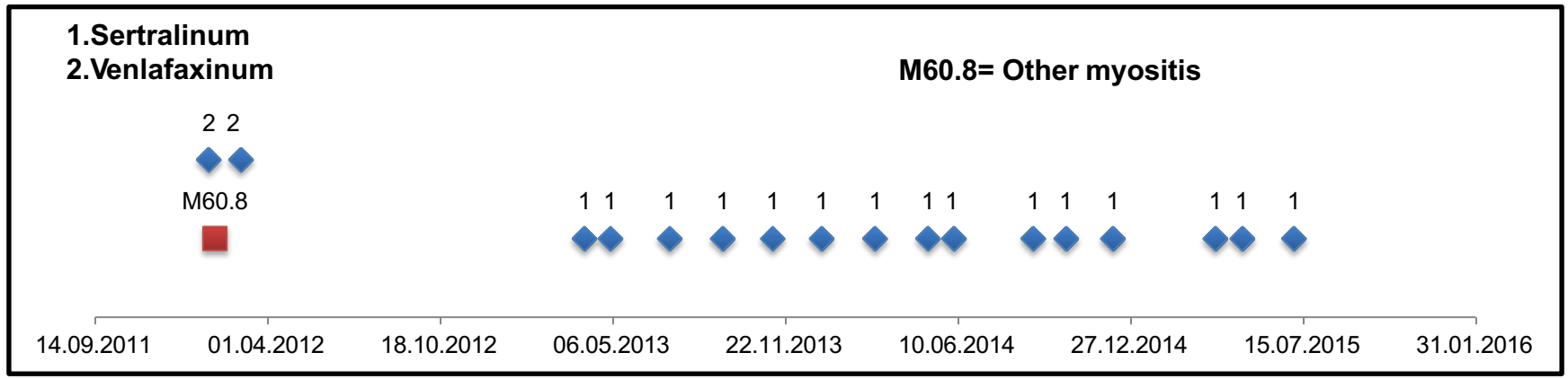
Pharmacogenetics

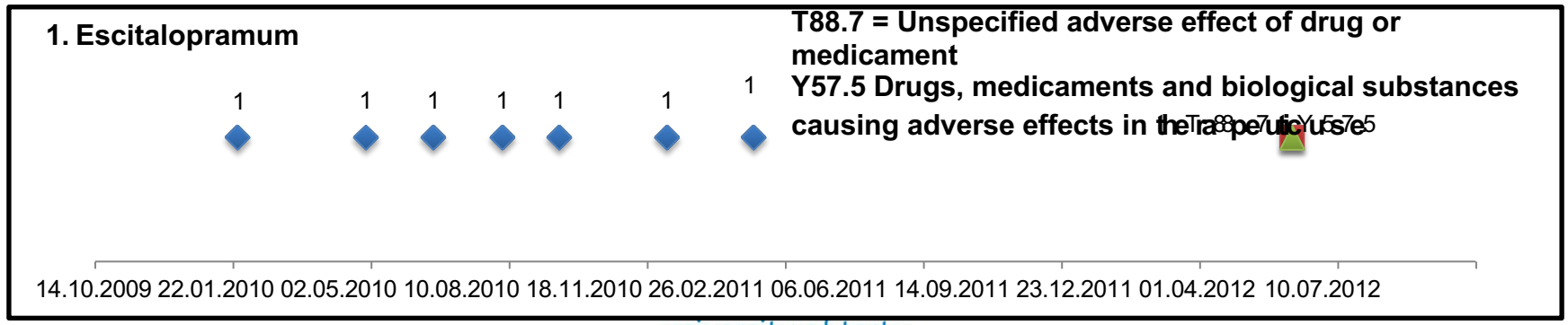
On average 5.5% of individuals in the population use at least one of the 32 drugs associated with the studied genes on a daily basis.

CYP2D6 Loss of function mutation and adverse drug reactions

(slide from dr. K. Krebs)







Decision support tools (DST)

Population scale genomics based on implementing the PRS is the “instrument” for disease prediction and prevention and this should start on the primary care level

GP need support in order to implement the new genomics based information

The DST should be easy to use, but PRS must base on the inform updated information in the relevant database

EGCUT broad feedback initiative

- T2D
- BrCa
- CAD, myocardial infarction

- HBOC
- Lynch sydrome, polyposes
- FH
- Arrhythmogenic right venreickle cardiomyopathy

- Cystic fibrosis
- Wilson's disease

- *11 genes*
- *30 compounds*

- Alfa-1 antitrypsiin insufficiency
- Thrombophilia
- Glaucoma (exfoliative)
- Hypolactasia

- *Early menopause*
- *Ancestry (planned)*

5000 people have received the feedback based on their genomic data

estonian genome center
university of tartu

Jaan Tamm

Your Data

Male Female Age 54

Weight 89 Height 171 Waist 86

Hypertension Myocardial Infarction

Genetic risk of type 2 diabetes

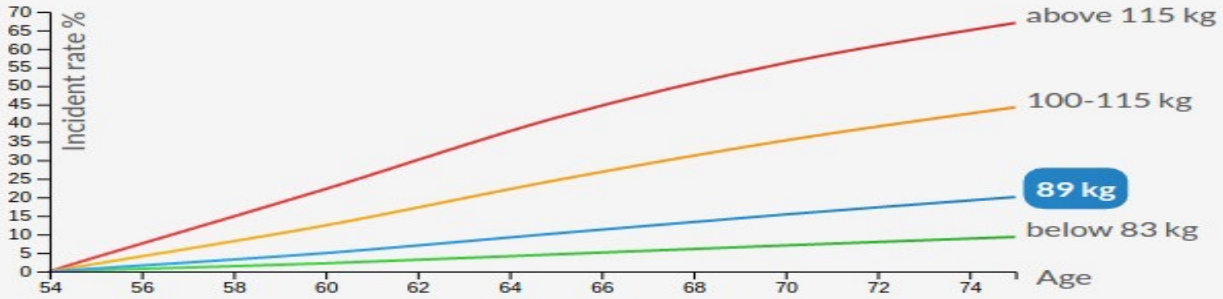
Seven persons out of 10 have lower genetic risk than you.

Two persons out of 10 have higher genetic risk than you.

Your genetic risk of type 2 diabetes is **average**. Your added lifestyle risk is **low**.

Your total risk of type 2 diabetes is **low**.

Risk of type 2 diabetes depends on body weight



Your 10-year risk of developing type 2 diabetes is **2%**. Your probability of developing type 2 diabetes before age 70 is **15%**.

An average person similar to you, but with lower body weight, has up to **50% lower** diabetes risk.

estonian genome center
university of tartu

Pharmacogenetic feedback

- 33-y female with depression
- CYP2C19 slow metabolizer, dose reduction to 50% recommended
- Sertralin and escitalopram formerly prescribed
- Both withdrawn, due to ADR – agitation, aggressiveness, pharyngitis, etc.

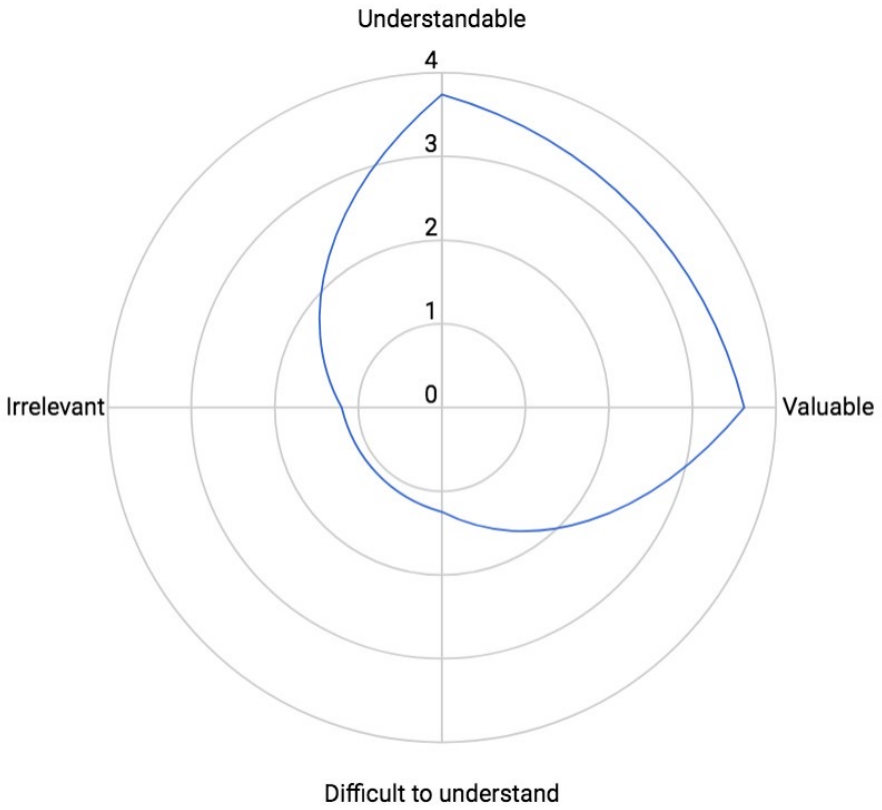
Geen	Genotüüp	Hinnang	Soovitus	Mõjutatud ravimi toimeained
CYP2C19	*2/*2	Aeglane ravimi lagundamine	!	Estsitalopraam, Tsitalopraam, Klopidogrel, Sertraliin, Vorikonasool, Esomeprasool, Lansoprasool, Pantoprasool, Omeprasool, Klomipramiin, Amitriptüliin
CYP2C9	*1/*1	Tavapärane ravimi lagundamine	+	Fenütoiin
CYP2C9; VKORC1	*1/*1; rs9923231 (AA)	Tavapärasest madalam doosisoovitus	!	Varfariin
CYP3A5	*3/*3	Aeglane ravimi lagundamine, Tavapärane muster	+	Takroliimus
DPYD	*1/*5	Tavapärane ravimi lagundamine	+	Kapetsitabiin, Fluorouratsiil
IFNL3	rs12979860 (CC)	Tavapärane ravimi toime	+	Alfa-2b-peginterferoon, ribaviriin
SLCO1B1	rs4149056 (TT)	Tavapärane müopaatia risk	+	Simvastatiin
TPMT	*1S/*1	Tavapärane ravimi lagundamine	+	Tioguaniin, Merkaptopuriin, Asatiopriin

+ - Tarvitada tavapäraselt. ! - Tarvitada ettevaatusega, võib vajada doosi muutmist. ! - Tarvitada väga ettevaatlikult, oht kõrvaltoimeteks.

Slide from Prof. Lili Milani

estonian genome center
university of tartu

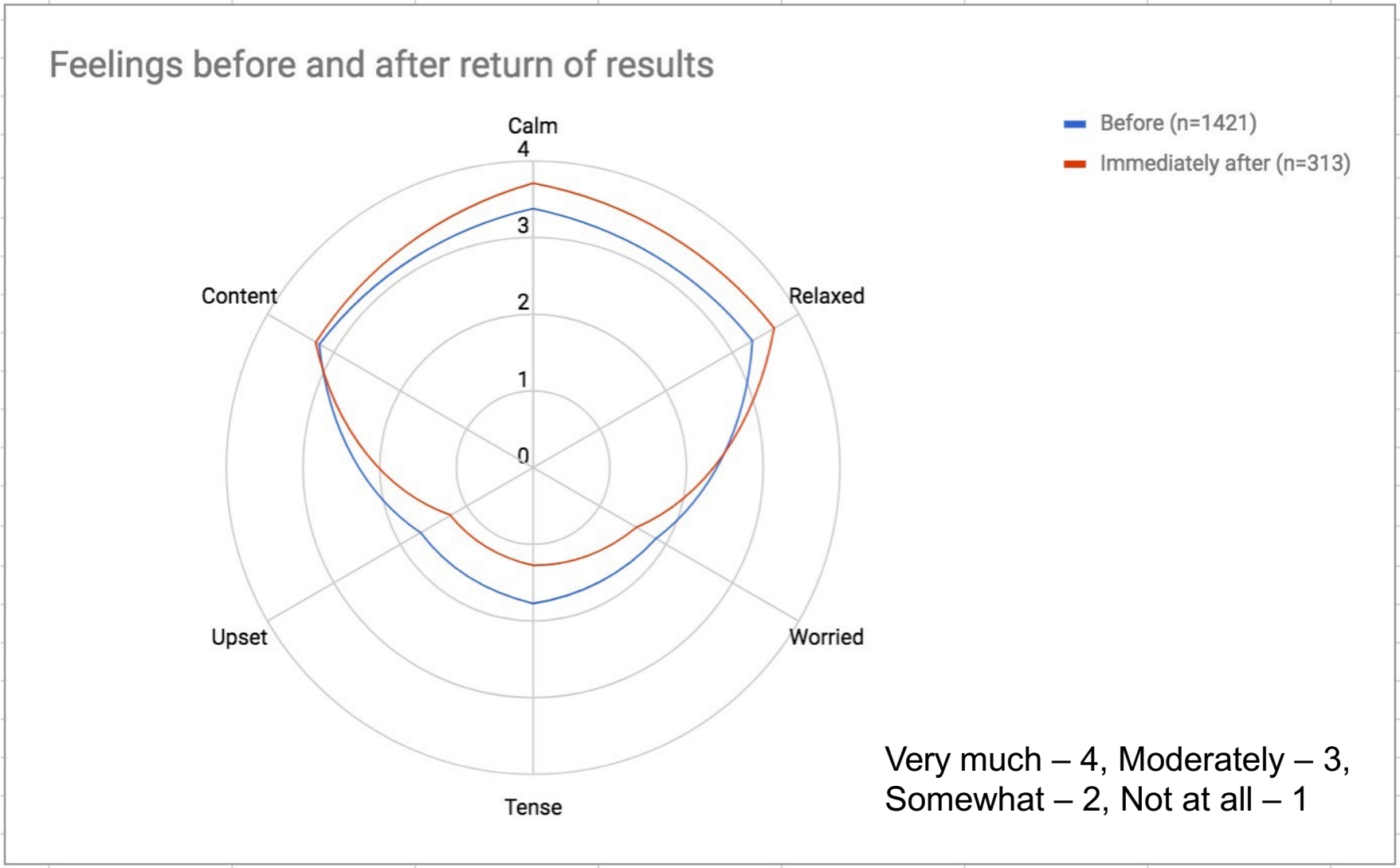
Impressions on explanations and counseling received



- Approximately 24-40 semi-structured sessions per week by 4 individuals
- Ave. length of GC session 35 min

Agree – 4, Disagree – 1

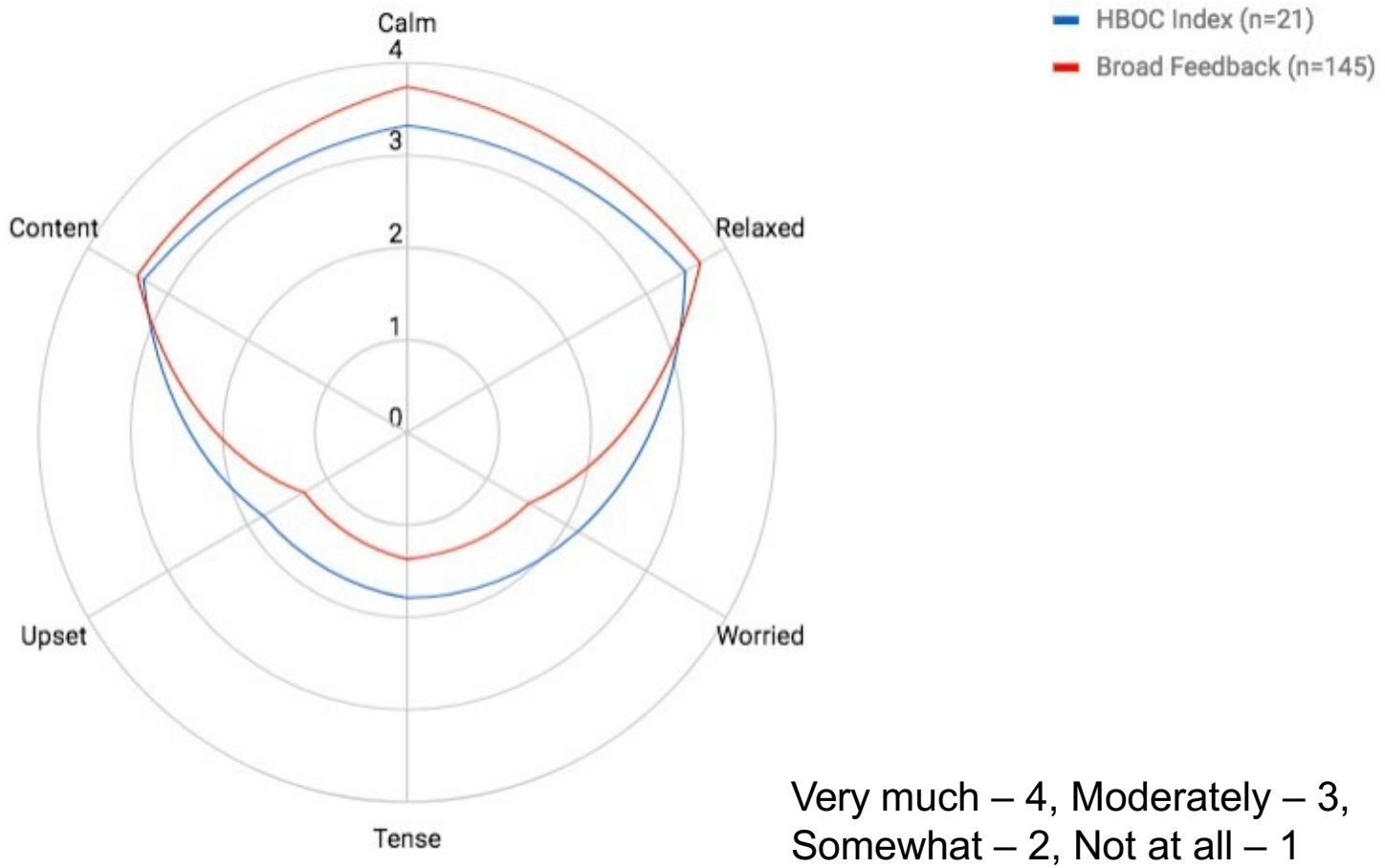
estonian genome center
university of tartu



- Will also be collected at >6-months
- STAI Y-6 item as also used in HBOC project

estonian genome center
university of tartu

Effect of high impact variants reported vs other risks

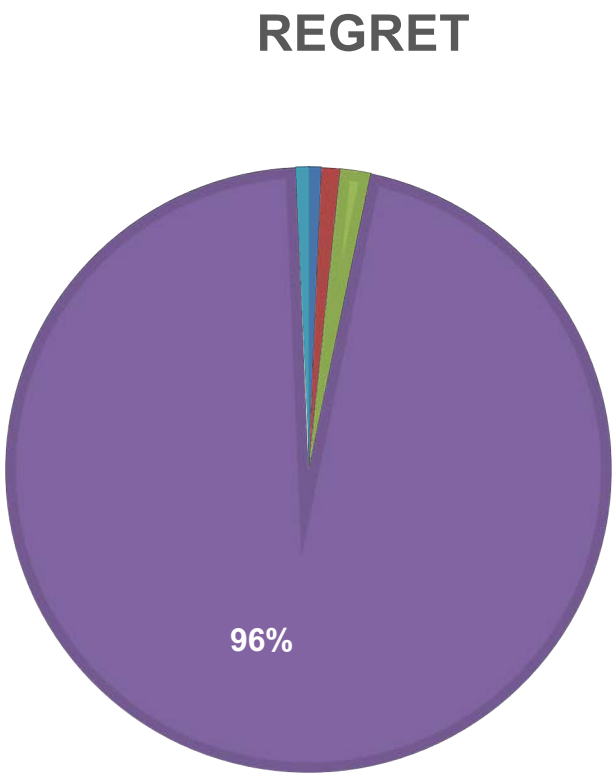


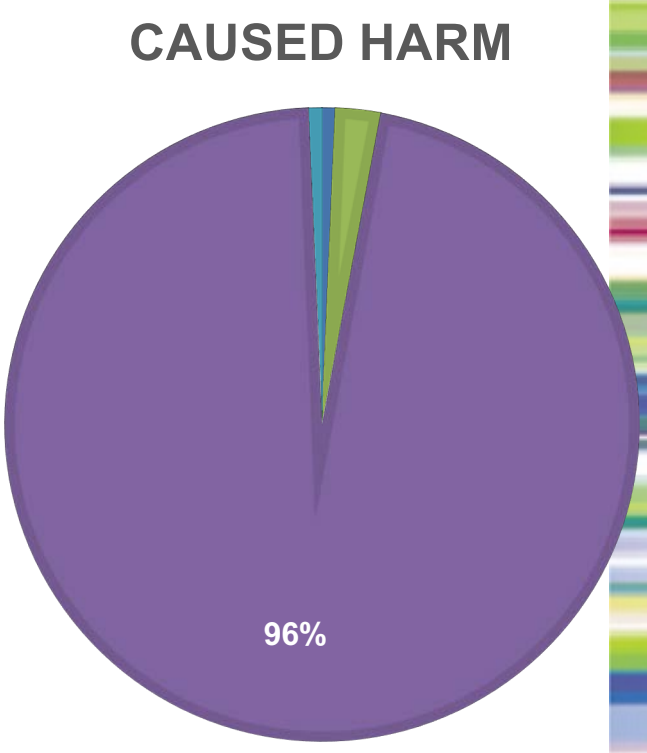
Very much – 4, Moderately – 3,
Somewhat – 2, Not at all – 1

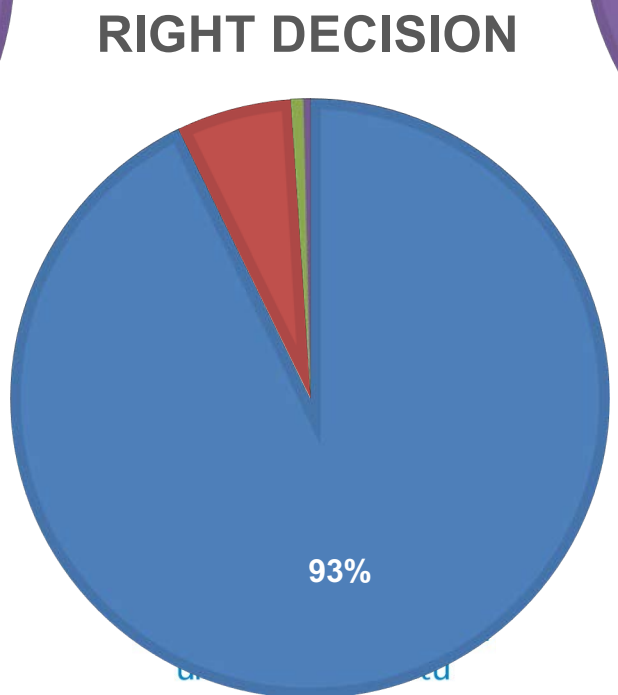
- Using STAI Y-6 item

estonian genome center
university of tartu

Decision regret 6 months later (n=305)







Problems to be solved

GWAS and PRS done mainly on GWAS chips – 700-800 000 SNPs

- Not yet clinically validated technology, but companies are working on this issue

Algorithms

- There are no standards for PRS, work is ongoing
- Imputation (methods vary, does imputation work similarly for all, can we use imputed data for variables used for individual health decisions)
- Mixed population

Regulation

- EU Medical Device Directive (2017/745/EU)(MDR)
- The European Union In Vitro Diagnostics Regulation (2017/746/EU) (IVDR)
- ISO-standards etc.
- National legislation(s)

Conclusion

Large prospective biobank cohorts make it possible to move towards personalized genetic risk prediction and to use it in general medical practice in preventing disease or ADR.

However, in the future, I hope, the whole health care infrastructure together with data (incl. genomic data) could be the basis of providing personal prevention, treatment and care as a part of the general health care.

andres.metspalu@ut.ee

estonian genome center
university of tartu

www.biobank.ee

UNIVERSITY OF TARTU

estonian genome center

Tõnu Esko, Krista Fischer, Reedik Mägi, Maris Alver, Kristi Läll, Kristi Krebs, Tõnis Tasa, Mart Kals, Tom Haller, Neeme Tõnisson, Tiit Nikopensius, Anu Reigo, Liis Leitsalu, Kristjan Metsalu, Kairit Mikkel, Mari-Liis Tammesoo …

Eesti Haigekassa

E-HEALTH
Estonian e-Health Foundation

Regionaalhaigla

Tartu Ülikool · Universitas Tartuensis · 1632

BIT IT

NEXT 10 BROAD INSTITUTE

Tartu Ülikooli Kliinikum A.D. 1804

Prof. Jaak Vilo, Hedi Peterson

QureTEC

BioMed@STACC: the health data mining team

▼ 18 people

Alexander Tkachenko · Anne-Liis Tanav · Dage Särg · Gea Pajula · Harry-Anton · Karl-Oskar Masing

Marek Oja · Margus Jäger · marje johanson · Markus Lippus · Mihkel Manguse · monika soosaar · Raul Sirel

Robert Roosalu · Sulev Reisberg · Sven Laur · Tormi Reinson · Uku R

STACC Software Technology and Applications Competence Center