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 Biobank timeline

- **Vision of PerMed**
- **Legal framework**
- **Recruitment**

**Timeline:**
- 1999: Vision of biobank framework
- 2000: Legal framework
- 2002: Recruitment
- 2011: 52,000
- 2013: Vision of PerMed
- 2017: Broad feedback
- 2018: 150,000
- 2019: 200,000 genotyped
- 2021: >500 0 NGS

- **Recruitment 2002 - 2011**
- **Re-contacting 2008 - 2014**
- **Genetic data return 2014 - ...**

Slide: Katrin Männik
§ 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.

/____________/

[Logo: Estonian Genome Center, University of Tartu]
Questionnaire of EGCUT

Personal data
- Place of birth
- Place of residence
- Nationality
- Education
- Occupation

Genealogy
- Parents
- Children
- Siblings
- Grandparents

Health behavior
- Smoking and alcohol
- Personality inventory NEO-PI-3
- Physical activity EQ-5D
- Nutrition
- Health self-assessment
- Chronotype questionnaire MCTQ

Diseases
- Diagnosis ICD – 10
- Treatment ATC
- Psychiatry module M.I.N.I. and SSP
- Questions about diabetes
- Questions about c/v diseases

Objective data
- Height & weight
- Blood pressure
- Pulse
- Handedness
- Waist
- Hip
210 000 gene donors in the EstBB

Representative sample of Estonians

- Total adults: 1 070 375
- Total adult gene donors: 104 317
- 9.75 %
- Age at the date of the agreement
- Date: 2019-06-17
**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

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

- Pooldab projekti
- Pole projektist kuulnud
- Äraootaval seisukohal
- Soovib rohkem informatsiooni
- Projekti vastu
- Ei oska öelda
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
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

![Bar chart showing counts of diagnoses before and after study.](chart.png)
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
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
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?


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

GRS quartile:
- 4 (top 25%)
- 3 (50-75%)
- 2 (25-50%)
- 1 (bottom 25%)

Whole cohort

Läll et al (2018)
Breast cancer: population vs top 5% Based on Polygenic Risk Score

Cumulative incidence on breast cancer among women

Läll K, et al. BMC Cancer 2019
Incident breast cancer cases in high-PRS group (preliminary data)

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

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

• **Flagship 7:** Alongside the ‘Genomic for Public Health’ project, the European Initiative to Understand Cancer ([UNCAN.eu](http://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

1-3. Metoprololum
4. Tamoxifenum

L27.0 = Generalized skin eruption due to drugs and medicaments taken internally

1. Sertralinum
2. Venlafaxinum

M60.8 = Other myositis

1. Escitalopramum

T88.7 = Unspecified adverse effect of drug or medicament

Y57.5 Drugs, medicaments and biological substances causing adverse effects in the exposed individual
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

**Common disorders (PRS)**
- T2D
- BrCa
- CAD, myocardial infarction

**High-risk actionable variants**
- HBOC
- Lynch syndrome, polyposes
- FH
- Arrhythmogenic right ventricular cardiomyopathy

**Risk factors with moderate effect**
- Alfa-1 antitrypsin insufficiency
- Thrombophilia
- Glaucoma (exfoliative)
- Hypolactasia

**Other topics of interest**
- Early menopause
- Ancestry (planned)

**Carrier status**
- Cystic fibrosis
- Wilson’s disease

**Pharmacogenetics**
- 11 genes
- 30 compounds

5000 people have received the feedback based on their genomic data
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

- above 115 kg
- 100-115 kg
- below 83 kg

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.
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.

Slide from Prof. Lili Milani
 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
Feelings before and after return of results

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

Very much – 4, Moderately – 3, Somewhat – 2, Not at all – 1
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
Decision regret 6 months later (n=305)

**REGRET**

- Agree: 96%
- Somewhat agree: 96%
- Somewhat disagree: 96%
- Disagree: 93%

**CAUSED HARM**

- Agree: 96%
- Somewhat agree: 96%
- Somewhat disagree: 96%
- Disagree: 93%

**RIGHT DECISION**

- Agree: 96%
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.
Thank you!

andres.metspalu@ut.ee

www.biobank.ee
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 …