From Biobanking to personalized medicine.

Andres Metspalu
Estonian Genome Center
University of Tartu

Personalised Medicine Conference 2016

May 31, 2016
Brussels
1. Electronic health care

E-Health in Estonia
X-Road, ID-card, State IS Service Register

- State Agency of Medicines
  - Coding Centre
  - Handlers of medicines
- Health Care Board
  - Health care providers
  - Dispensing chemists
- Population Register
- Business Register

- Hospitals 2009
- Family Doctors 2009
- Pharmacis 2010 January
- School Nurses 2010 September
- Emergency Medical Service 2011

- Patient Portal 2009
- X-Road Gateway Service 2009
- Pharmacies and Family Doctors 2009
- Nation-Wide Health Information Exchange Platform 2008 December
- Prescription Centre 2010 January

Estonian Genome Center
University of Tartu
The Estonian ID card

• The ID card is a **mandatory** ID document for all Estonian residents from the age of 15
• Enables secure digital authentication and signing
• A digital signature has the same legal consequences as a hand-written signature
• Does not have any additional information
  - No bank account, no health information etc.
• Active cards: **1 221 948** (08.09.2014)
  - Digital signatures: 174 385 946
  - Estonian Population 1 286 540 (01.01.2013)
  - Estonia has been issuing electronic ID cards from January 1st, 2002
  - Also Mobile-ID
1.1 mio persons medical data
### Hospital Clinical Labs

<table>
<thead>
<tr>
<th>Profile</th>
<th>#N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basophils</td>
<td>35806</td>
</tr>
<tr>
<td>Eosonophils</td>
<td>35784</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>31725</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>36424</td>
</tr>
<tr>
<td>Red Blood Cells</td>
<td>64019</td>
</tr>
<tr>
<td>White Blood Cells</td>
<td>65203</td>
</tr>
</tbody>
</table>

- **Trajectory**

![Graph showing trajectories of different blood parameters](image-url)
2. Population biobank

- Molecular description of the population (genetic variations etc.)
- Biobanks are needed for all
Background of the Estonian Biobank

• EGC is the Research institute of the University of Tartu, which keeps the Estonian Biobank

• Longitudinal, prospective, population based biobank, HGR Act from 2000

• 52,000 gene donors recruited (5% of the adult population), follow-up is on-going

• Supported by the government (HGR Act)
§ 3. Chief processor of Gene Bank

- (1) The chief processor of the Gene Bank is the University of Tartu, whose objectives as the chief processor are 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.
Population pyramid (50155 participants)

Female
EGCUT
Estonia 2007

Male
EGCUT
Estonia 2007

Count Estonia

Age

Count Estonia

Count Estonia

# Population pyramid (50155 participants)

Female
EGCUT
Estonia 2007

Male
EGCUT
Estonia 2007

Count Estonia

Age

Count Estonia

Count Estonia

estonian genome center university of tartu
Disease trajectory (for all 52,000 subjects in the Estonian biobank)

e.g. male, born 1970, age 37, joined 2007, high school, no sports, walking 2 h/week, smoking 20 cigarettes per day, 2 strong drinks per week (2x40cl), height 178, weight 87, BMI 27.5, BP 130/78 mmHg
P21 (4.68%)
PC1 (8.65%)

Austria
Bulgaria
Czech Republic
Estonia
Finland (Helsinki)
Finland (Kuusamo)
France
Northern Germany
Southern Germany
Hungary
Northern Italy
Southern Italy
Latvia
Lithuania
Poland
Russia
Spain
Sweden
Switzerland

PC2 23.8%
PC1 36.6%
Public opinion and awareness of the EGCUT 2001-2014

In favor of the idea of EGCUT:
- June, 2001: 18%
- Sept, 2001: 16%
- Feb, 2002: 13%
- March, 2002: 19%
- March, 2003: 18%
- Sept, 2004: 16%
- Dec, 2004: 22%
- May, 2007: 33%
- July, 2009: 59%
- June, 2010: 55%
- Sept, 2014: 70%

Never heard of EGCUT:
- June, 2001: 38%
- Sept, 2001: 39%
- Feb, 2002: 40%
- March, 2002: 36%
- March, 2003: 39%
- Sept, 2004: 44%
- Dec, 2004: 33%
- May, 2007: 33%
- July, 2008: 35%
- June, 2009: 36%
- July, 2010: 36%
- June, 2011: 30%
- July, 2013: 33%
- Sept, 2014: 27%

Against it:
- June, 2001: 4%
- Sept, 2001: 4%
- Feb, 2002: 3%
- March, 2002: 4%
- March, 2003: 4%
- Sept, 2004: 3%
- Dec, 2004: 2%
- May, 2007: 2%
- July, 2008: 3%
- June, 2009: 2%
- July, 2010: 2%
- June, 2011: 2%
- July, 2013: 2%
- Sept, 2014: 1%
3. Deep (30x) Whole Genome Sequencing of 2500 subjects

(and many more OMICS data)
Average human

• 3 million mutations
  – 5000 unique
    – Loss of function
      • 120 heterozygous
      • 18 homozygous
4. Do we have enough information to start with the precision medicine?
Pharmaco-genomic study

Electronic health records
- ADR diagnoses (ICD10 codes)

Biobank questionnaire
- Health status & lifestyle

Health Insurance Fund data
- Drug prescriptions (ATC codes)

Genome and Exome Sequences
- LoF & non-syn variants

Recontact subjects for phenotyping
- Drug metabolism enzyme and transporter activity

Pharmacokinetic profiling
Return of results
16p11.2 del/dup

In collaboration with prof. A. Reymond,
Center for Integrative Genomics
University of Lausanne
<table>
<thead>
<tr>
<th>CNV</th>
<th>Phenotypic features¹</th>
</tr>
</thead>
</table>
| 16p11.2 micro-deletion | - Macrocephaly  
|                     | - Obesity (43-fold increased risk of obesity in adulthood)  
|                     | - Mean global cognition decreased by 2 standard deviations compared to the general population. Global functioning ranges from normal to mild – moderate developmental delay/intellectual disability.  
|                     | - Learning difficulties with specific language problems (phonology) and executive dysfunction  
|                     | - Autism spectrum disorders (20%)  
|                     | - Vertebral malformations  
|                     | - No specific dysmorphism |
| 16p11.2 micro-duplication | - Microcephaly  
|                     | - Underweight (8.3-fold increased risk of being underweight in adulthood)  
|                     | - Highly variable global cognition ranging from normal range to severe developmental delay/intellectual disability.  
|                     | - Motor delay  
|                     | - Learning difficulties  
|                     | - Autism spectrum disorders (20%)  
|                     | - Schizophrenia and psychotic disorders  
|                     | - Epilepsy |

¹ These features are based on the literature and may vary among individuals.
Reporting of a genomic finding to biobank participants

- 16p11.2 CNV carriers identified among population biobank participants

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>16p11.2 del/dup carriers detected (n = 11)</td>
<td>2 had deceased</td>
</tr>
<tr>
<td>Contact</td>
<td>Invitation letter mailed to participate in the project (n = 9)</td>
<td>1 was unable to attend at the time</td>
</tr>
<tr>
<td>Validation</td>
<td>1st visit for informed consent and blood sample for validation testing (n = 8)</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>2nd visit for return of incidental findings and cognitive tests (n = 8)</td>
<td>1 did not reply, 2 did not consent</td>
</tr>
<tr>
<td>Follow-up</td>
<td>Summary letter sent (n = 8), Consent for physician contact (n = 5)</td>
<td>3 did not reply</td>
</tr>
<tr>
<td>Survey</td>
<td>Survey sent (n = 8)</td>
<td></td>
</tr>
<tr>
<td>Question group</td>
<td>Q</td>
<td>Statement</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Interests and choices</td>
<td>Q1</td>
<td>I am glad that the Biobank contacted me about the genetic finding</td>
</tr>
<tr>
<td></td>
<td>Q2</td>
<td>I wish I would have been informed earlier about the genetic finding and the potential health risks</td>
</tr>
<tr>
<td>Information comprehension</td>
<td>Q3</td>
<td>Information provided was understandable</td>
</tr>
<tr>
<td></td>
<td>Q4</td>
<td>Information provided was interesting</td>
</tr>
<tr>
<td></td>
<td>Q5</td>
<td>Information provided was informative</td>
</tr>
<tr>
<td></td>
<td>Q6</td>
<td>Information provided was valuable</td>
</tr>
<tr>
<td></td>
<td>Q7</td>
<td>I can explain what the condition means to people in my family</td>
</tr>
<tr>
<td></td>
<td>Q8</td>
<td>I do not believe that the genetic finding is heritable</td>
</tr>
<tr>
<td></td>
<td>Q9</td>
<td>I don't know where to go to get the medical help I / my family needs</td>
</tr>
<tr>
<td></td>
<td>Q10</td>
<td>I understand the impact of the condition on my child(ren)/any child I may have</td>
</tr>
<tr>
<td>Emotional response</td>
<td>Q11</td>
<td>After counseling I felt clarity</td>
</tr>
<tr>
<td></td>
<td>Q12</td>
<td>After counseling I felt relief</td>
</tr>
<tr>
<td></td>
<td>Q13</td>
<td>After counseling I felt indifference</td>
</tr>
<tr>
<td></td>
<td>Q14</td>
<td>After counseling I felt confusion</td>
</tr>
<tr>
<td></td>
<td>Q15</td>
<td>After counseling I felt worry</td>
</tr>
<tr>
<td>Perceived impact</td>
<td>Q16</td>
<td>I am able to cope with having this condition in my family</td>
</tr>
<tr>
<td></td>
<td>Q17</td>
<td>I now have better access to health care / specialists</td>
</tr>
<tr>
<td></td>
<td>Q18</td>
<td>I feel that my treatment and/or condition has improved</td>
</tr>
<tr>
<td></td>
<td>Q19</td>
<td>The information received has somehow changed my life</td>
</tr>
</tbody>
</table>

*a Survey question/statement numbers to which also the Q’s in the text refer to.

*Responses on a five point Likert Scale where 5 - agree, 4 - slightly agree, 3 - unsure, 2 - slightly disagree, 1 – disagree.

c Not all 5 respondents answered this question.
Returning the variants according to the ACMG 2013 gene list on incidental findings

- WES, WGS feedback minimally for 56 genes and 24 conditions
- Mostly AD or SD
- Does not include metabolic disorders (that are also considered actionable)
- Presumably on 1% individuals
- To report regardless of age and sex
- Opt out for the tested persons and their representatives
- Normal tissue not tumor analysis
- Known and expected pathogenic mutations only
- Be aware of limitations in testing and interpretation
- Lab should also report the classification basis for variants
Results from the data analysis

- Initially 55 variants in 28 genes selected
- After manual QC, 16 variants discarded as sequencing artifacts
- 6 variants likely benign according to clinical databases (BIC), by frequency, by preserved splice site, etc.
- 33 variants in 20 genes known pathogenic by clinical evidence / expected or possibly pathogenic.
- These 20 genes causative in 14 clinical conditions
- Findings in 45 individuals (2% of the total sequenced)
- 14 known high-risk and 2 expected high-risk mutations in BRCA1 and BRCA2 genes (0.7%), corresponds to routine clinical testing of >100 patients and family members (16% mutations found in patients with respective family history)
- Recontacting of the patients is underway
Familial Hypercholesterolemia (FH) Project

• Collaboration
  – Estonian Genome Center
  – Broad Institute
  – Estonian clinical cardiologists

• Main aim
  – Determination of FH phenotype and further clinical management based on genome-wide study findings (WGS, WES and chip data)

First FH patient was examined today!
5. Common complex disease?
Economic burden of T2 diabetes (GER) 2000-2007

- Prevalence of treated diabetes rose continuously from 6.5 to 8.9% (+36.8%)
- Direct costs per patient with diabetes rose from € 5 197 to € 5 726 (+10.2%)
- Total direct cost burden of diabetes in Germany grew from € 27.8 billion to € 42.0 billion (+51.1%)
- Per-capita costs were € 2 400 in 2000 and € 2 605 in 2007 (+8.5%)
Polygenic risk scores

Calculated as  \[ S = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k, \]

\( X_2, \ldots, X_k \) - allele dosages for \( k \) independent markers (SNP-s), typically the ones with strongest effect

\( \beta_1, \beta_2, \ldots, \beta_k \) – effect estimates (logistic regression parameters, \( \ln \) OR) from a GWAS meta-analysis

Polygenic risk score for type II diabetes:
histogram of the score in 7462 individuals (Estonian Biobank)

Läll et al., 2016 Genomic Medicine, (submitted)
Incident T2D: analysis of 264 incident cases in overweight individuals free of T2D at recruitment

Cumulative risk of T2D in 3421 individuals
BMI at recruitment >24 kg/m², age 35-74

- Highest genetic risk score quintile (>80%)
- Average risk score quintiles (20-80%)
- Lowest risk score quintile (<20%)

No significant sex difference, genetic risk score is the strongest predictor after BMI.

(fasting glucose and insulin measurements are not available for this cohort)
Type II diabetes risk for men depending on age and BMI
6. Clinical Decision Support Software (CDSS)?

- CDSS provides clinicians with knowledge presented at appropriate times.
- It encompasses a variety of tools such as computerized alerts, clinical guidelines, and order sets.
- CDSS has the potential to provide the necessary level of personalized guidance to providers at the point of care that will be necessary in the era of genomic medicine.
- This is a tool to advise PCP (like radiology report).
Virtuous Cycle of Clinical Decision Support

Registry → Measure

Practice → Guideline

CDS

http://www2.eerp.usp.br/Nepien/DisponibilizarArquivos/tomada_de_decis%C3%A3o.pdf
Future of the biobanking


The Institute for Personalized Medicine at Mount Sinai  From Prof. Böttinger
Challenges and issues

• Awareness *executives*, doctors and patients
• New technologies and data *empower patient* with more possibilities to manage own health
• Ethical issues
  – Right to know and right not to know
  – Treatable and non-treatable conditions
  – Big data, *cloud*, sharing data
• Knowledge about associations between DNA variants and diseases is not equally good in all areas, *but improving rapidly*
• Reimbursement should value more prevention
• Large work-load to keep *database of known risk variance updated*, needs international collaboration
Conclusions

• Estonia has great potential to plan and implement personalized medicine solutions for the whole country, starting with the pilot project for 50,000 gene donors

  – Genetic research with 5% of population genetic and continuously updated phenotype information
  – Nation wide Health Information Exchange platform
  – 10+ years of experience of national level e-services (PKI, X-Road, ID-card, security framework)
  – High level public trust and acceptance
Thank you!
www.biobank.ee