The role of biobanks in genetic discovery and precision medicine

K Hveem, MD, PhD, Professor,
PI of HUNT Biobank,
Head of K.G. Jebsen Center for Genetic Epidemiology
National Node Director, Biobank Norway/BBMRI.no
Biobanks – a strategic asset in medical research for > 20 years

• in studies of common and complex diseases
• for large scale genetic studies (genetic epidemiology)
  – To identify rare genetic variants/non-coding mutations
  – Polygenic risk scores
• for optimal use of high-throughput omics platforms
• in precision medicine
• in drug discovery and biomarker validation
Biobank driven genetic discoveries in population studies
The HUNT study

- **1984-86**
  - **HUNT1** (20yrs+)
    - 75 000 (88%)

- **1995-97**
  - **HUNT2** (13yrs+)
    - 75 000 (72%)

- **2006-08**
  - **HUNT3** (13yrs+)
    - 50 800 (54%)

- **2017-19**
  - **HUNT4** (13yrs+)
    - 55 600 (54%)

Collecting disease endpoints by linkage to clinical and national registries

- 46 000 (H1-H2)
- 37 000 (H2-H3)
- 34 000 (H3-H4)
- 26 100 (H2-H4)
- 20 000 (H1-H4)

200 completed PhD-degrees, 100 peer review papers/year
• 9000 included in a cognitive test program
• 7000 - all HUNT participants

Deltakelse over 70 år

- 70-79 år: 8379
- 80-89 år: 3300
- 90-99 år: 735
- 100+ år: 18

Aldersgrupper

- Under 20 år: 64%
- 20-29 år: 35%
- 30-69 år: 53%
- Over 70 år: 55%
The HUNT Biobank
(The European Research Biobank 2013)

- H2-4, DNA: 100 000
- Serum/plasma: 100 000
- Urine: 25 000
- RNA: 15 000
- Vital frozen cells: 40 000
- CONOR DNA: 230 000
- H4 Saliva (17 000), Fecal samples (15 000), One Health (6000)
The HUNT genes all-in project

- Genome wide genotyping of 70,000 HUNT-participants (Human Core Exome),
- 604,000 genetic markers including 60,000 custom “HUNT SNPs”
- Imputed up till 28 million genetic markers
- CVD as main focus, > 60 sub-studies on various other disease categories
- Challenging ethics with variants such as FH, BRCA2 – return of results

### CVD main

<table>
<thead>
<tr>
<th>CVD</th>
<th>Endo</th>
<th>Gastro</th>
<th>Lung</th>
<th>Neuro</th>
<th>Pharma</th>
<th>Kidney</th>
<th>Reuma</th>
<th>Infection</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVD main</td>
<td>Thyroid</td>
<td>IBD</td>
<td>Asthma, KOLS</td>
<td>Common psych. disorders</td>
<td>CV Pharmaco-genomics</td>
<td>CKD</td>
<td>AS</td>
<td>Sepsis</td>
<td>PCOS</td>
</tr>
<tr>
<td>Afib</td>
<td>BMD,</td>
<td>IBS</td>
<td>Pain</td>
<td>Antikoagul.</td>
<td>RA</td>
<td>RA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>Vit D</td>
<td>Reflux</td>
<td>Sleep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTE</td>
<td>T2D</td>
<td>CRC</td>
<td>Head Ache</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FH</td>
<td>LADA</td>
<td></td>
<td>Parkinson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phys activity</td>
<td></td>
<td></td>
<td>Stroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise</td>
<td></td>
<td></td>
<td>Eating disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TnI</td>
<td></td>
<td></td>
<td>Low back pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrhythmia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- More than 150 collaborating clinicians (phenotype experts)
- EthiCert - an ethical certificate (OL Holmen, S. Gartmann, L Ursin)
Genome Wide Association studies (GWAS)
GWAS-analyses has so far been conducted on > 1600 binary or quantitative traits based on > 7000 unique variables from HUNT Data bank and ICD-codes retrieved from Electronic Health Records

Recent publications:


Screen of summary statistics

Phe-was analyses running in parallel presented on HUNT pheweb

APOE:p.Arg176Cys associated with

- Alzheimer
- Fasting blood lipids
- Dyslipidemia
The GWAS-catalogue

2018 Apr

Associations: 69,885
Studies: 5,152
Papers: 3,378

- GWAS-significant associations on almost 70,000 traits/phenotypes (p=5x10^{-8})
Non coding or Loss of function mutations (LoF)

- 100-150 LoF/pr genome
- Are rare variants, presented as alternative forms of a gene with a minor allele frequency (MAF) of less than 1%
- Increased numbers in founder populations

LoF-mutation in the PCSK9-gene
Reduced risk for MI by 50-90 %, low levels of cholesterol /TG

Cohen JC. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. NeJM 2006
Sabatine MS et al. Efficacy and safety of evolocumab (PCSK9-inhibitor) in reducing lipids and cardiovascular events. NeJM 2015
GWAS on lung cancer – Susceptibility locus in region 15q25

**Protective gene against MI**

- Assessment of genome wide coding variation based on an exome array genotyping of ~ 80,000 coding variants in 5643 subjects from HUNT Biobank
- Identified a LoF causal variant in TM6SF2 affecting lipid levels and risk of MI
- Through replicated in 4666 participants from the Tromsø study, 10 variants confirmed to be associated with a lipid trait $p < 5 \times 10^{-8}$
- TM6SF2 has also shown an increased risk of fatty liver disease and T2D, so not likely to be the best drug target

**Protective gene against type 2 diabetes**

- 150,000 across 5 ancestry groups, 6000 from HUNT
- 12 variants in SLC30A8
- A common protein truncating variant (p. Trp325Arg) was assoc. with risk of T2D, glucose and proinsulin levels
- Carriers had a 65% reduced T2D risk
Biobank driven drug discoveries
Development of therapeutics in 2018

• Only 1 of 10 drug candidates reach the market
• Most failures occur in Phase II clinical trials
  - 50% due to lack of efficacy
  - 25% due to toxicity
• Pre-clinical models may be poor predictors of clinical benefit
• Compounds supported by human genetics evidence are 2.5x more likely to succeed
• The total costs of one successful drug is ~ $2.8 billion
Health and biobank data in drug R&D
– across the life cycle from target ID to RWE

Target ID | Target Val | Hit | Lead | Lead Optim | Preclinic | Phase I | Phase II | Phase III | BoD | HE | Safety | Effectiveness
---|---|---|---|---|---|---|---|---|---|---|---|---
Research | Discovery | Development | Real World Evidence

- Genetic guided drug discovery
- Biomarkers for stratification
- New models for RCT/RWD
  - Recall studies
  - rRCT
- RWE – drug utilisation
- RWE – safety
- RWE – effectiveness
- RWE – burden of disease
- RWE – Health economics

Some use areas
Drug companies turning to biobanks as a route to fast-tracking drug discovery may result in a 3-fold increase in biobank revenue in 10 years.

$19 billion in annual revenue

Figure 3.9 Biobanking for Therapeutic Use: Revenue Forecast ($bn), 2017 to 2027

Biobanking Market Forecasts 2017-2027
Visiongain.com
A use case for genetic based drug discovery

- A plan to sequence 2 million genomes recruited from biobanks world wide over the next 10 years was launched by Regeneron Genetics Center in 2014 to search for novel targets that lie buried in the human genome.
- Presently 400 000 exomes are sequenced, including the first 50 000 from UK Biobank (500 000 to be sequenced)
- Dozens of potential targets have so far been identify and validated
- Six new drugs have been taken to the market based on genetic discovery

Genomics is finally ready for pharmaceutical prime time.
The Potential for Human Genetics to Accelerate Target Identification, Validation and Drug Development

Precision Medicine
Numbers need to treat one person (NNT)

We need

- Increased diagnostic accuracy
- More targeted drugs
- Fewer side effects
- Reduced costs

---

**IMPRECISION MEDICINE**

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

<table>
<thead>
<tr>
<th>1. ABILIFY (aripiprazole)</th>
<th>2. NEXUM (esomeprazole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia</td>
<td>Heartburn</td>
</tr>
<tr>
<td>![Blue: 3 people]</td>
<td>![Red: 24 people]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. HUMIRA (adalimumab)</th>
<th>4. CRESTOR (rosuvastatin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis</td>
<td>High cholesterol</td>
</tr>
<tr>
<td>![Blue: 3 people]</td>
<td>![Red: 24 people]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. CYMBALTA (duloxetine)</th>
<th>6. ADVAIR DISKUS (fluticasone propionate)</th>
<th>7. ENBREL (etanercept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>Asthma</td>
<td>Psoriasis</td>
</tr>
<tr>
<td>![Blue: 3 people]</td>
<td>![Red: 24 people]</td>
<td>![Red: 24 people]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. REMICADE (infliximab)</th>
<th>9. COPAXONE (glatiramer acetate)</th>
<th>10. NEULASTA (pegfilgrastim)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crohn's disease</td>
<td>Multiple sclerosis</td>
<td>Neutropenia</td>
</tr>
<tr>
<td>![Blue: 3 people]</td>
<td>![Red: 24 people]</td>
<td>![Red: 24 people]</td>
</tr>
</tbody>
</table>

Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr7Sl.
Nordic biobanks can play an important role to drive life science R&D and improve R&D productivity
It’s a numbers game
Nordic Precision Medicine Initiatives
(supported by NordForsk)

**Finland**
- >150 000 GWAS
- 5 400 WGS, 30 000 WES
- FinnGen – 500 000 GWAS by 2021

**Norway**
- ~ 200 000 GWAS
- WGS 5000, WES 10 000

**Iceland**
- deCODE >160 000 GWAS, 40 000 WGS
- AGES - Reykjavik Heart study
- The SAGA Cohort – 100 000 to be included

**Denmark**
- 240 000 GWAS,
- 20 000 WES/WGS
- GENLIFE- SSI initiative for genotyping 1 million

**Sweden**
- ~ 200 000 GWAS, ~ WGS 3000

**Estonia**
- ~ 100 000 GWAS
- WGS 2500, WES 2500
A precompetitive model for public-private partnership

FINNGEN RESEARCH PROJECT IS AN EXPEDITION TO THE FRONTIER OF GENOMICS AND MEDICINE

Important discoveries could be found on a single sample from any one of Finland’s 500,000 biomedical patients.

Read more

FINNGEN BRINGS TOGETHER THE NATION-WIDE NETWORK OF FINNISH BIOBANKS.

Every Finn can be a part of the FinnGen study by giving a biobank consent.

Samples from biobanks
The FinnGen study will utilise samples collected by a nationwide network of Finnish biobanks. The...

Take part
Samples are needed from all over Finland in the thousands, because each one is the key of...

Everybody benefits
The genomic data produced during the project will be owned by the Finnish biobanks and remain...

Collaboration is the key
The collaborative nature of the FinnGen research project is exceptional; financed on many...

www.ntnu.no
Precision medicine initiative, Biobank Norge 3

BIOBANK NORWAY PARTNERS
UIT, NTNU, UiB, UiO, NIPH, HN, HV, HMN, HSØ, KREG

Population-based studies ~400k
- MoBa ~265k, HUNT ~90k, TU ~40k
  HUSK ~30k = 425k in total

Existing hospital disease-specific biobanks ~30k
- Neuro, Psychiatry, Lung, Gastro, Cancer, Cardiovascular, Metabolic etc. TBD

Broad consent hospital based biobanking ~200k
- Sampling of ~200k subjects from the clinic nation-wide

Step-wise approach:
1. Existing WGGT
2. Complete WGGT
3. Imputed WGGT
4. Large-scale WES/WGS
5. Other omics

Step-wise funding:
1. BN3
2. Governmental
3. Industry
4. Precomp. PPP

Research & Innovation

WGGT=whole genome genotyping, WES=Whole exome sequencing, WGS=Whole genome sequencing
Digitizing biobanks – the future perspectives

A shift from samples to data (digitalization)

- Enhanced by reduced costs if analyses of larger sample sizes (omics-driven analyses), a trade off for significant return of analyses to the biobank
- Reduction of data export, researchers will be granted virtual access to biobank clouds (Amazone, Google, Microsoft, Computerome.....)
- Biobanks will play a stronger role in precision medicine
- Access to annotated biobank samples and national registry data will be centralized to a publicly governed Health data platforms and Health analyses platforms

HUNT Data center and HUNT computer cloud

- ISO-9001 certified May 2017
- ISO 27001 certified by Dec 2017 (Information security standard)
Return of results
NEM about return of results

- International recommendations: Genetic information/risk must be "actionable" to trigger a feedback of results.

- Provided good opportunities for prevention, or even treatment, the situation most commonly is referred to as actionable.

- The researchers must then plan for feedback.

- BRCA-mutations, FH
~ 3.5% have actionable results (Geisinger)

<table>
<thead>
<tr>
<th>GENOMIC CONDITION</th>
<th>Number or patients diagnosed</th>
<th>CLINICAL RISK</th>
<th>DISEASE-ALTERING INTERVENTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial Hypercholesterolemia (FH)</td>
<td>1 in 250</td>
<td>Early-onset Coronary Artery Disease and Stroke</td>
<td>Targeted screening and aggressive medical management</td>
</tr>
<tr>
<td>Hereditary Breast and Ovarian Cancer Syndrome</td>
<td>1 in 400</td>
<td>Early-onset Breast, Ovarian, and Prostate Cancers</td>
<td>Targeted screening with prophylactic medical and surgical intervention</td>
</tr>
<tr>
<td>Lynch Syndrome</td>
<td>1 in 440</td>
<td>Early-onset Colon and Uterine Cancers</td>
<td>Targeted screening and management of pre-cancerous changes</td>
</tr>
<tr>
<td>TOTAL</td>
<td>&gt; 1 in 100</td>
<td>Multiple Cancers and Cardiovascular Diseases</td>
<td>Life-saving screening and intervention before development of disease</td>
</tr>
</tbody>
</table>

Other conditions: cardiomyopathy, long QT syndrome, malignant hyperthermia, arrhythmogenic right ventricular cardiomyopathy, MEN2, tuberous sclerosis, hereditary pheochromocytomas and paragangliomas
• Do you want feedback of results if the genetic information obtained may result in potential treatment or preventive measures
  – 93 % yes

• Are you willing to participate in follow-up studies based on genetic findings with no clear clinical impact
  – 88 % yes

• Other markers? Genetic risk scores?
Familiar Hypercholesterolemi (FH)

– 25.000 nordmenn har samme sykdom som Dale Oen

Mange av oss bærer på en alvorlig sykdom kalt FH, uten å være klar over det.
Familiar Hypercholesterolemia (FH) in HUNT and Geisinger (US)

**HUNT**
- An expected prevalence of 280 in the population (pr 100,000)
- 90 have been registered as heterozygous for FH
- > 200 are probably identified through HUNT
- What explains the discrepancy?
  - Clinically less penetrant in a population setting?
  - Clinical treatment is ongoing, but without genetic testing?
- Special strategy for follow-up of FH in HUNT

**Geisinger**

<table>
<thead>
<tr>
<th>Population characteristics</th>
<th>FH variant positive/total</th>
<th>Estimated prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>All DiscovEHR participants</td>
<td>229/50,726</td>
<td>1:222</td>
</tr>
<tr>
<td>Participants recruited from cardiac catheterization lab</td>
<td>57/6,747</td>
<td>1:118</td>
</tr>
<tr>
<td>Participants recruited from other sites</td>
<td>172/43,979</td>
<td>1:256</td>
</tr>
</tbody>
</table>

- FH variant negative
- FH variant positive

Currently on statin
- 38% (FH variant negative)
- 58% (FH variant positive)

Statin-treated with LDL-C < 100 mg/dl
- 46% (FH variant negative)
- 77% (FH variant positive)
21 women had their breasts and ovaries removed – should never been operated.

Patients were all wrongly advised about their BRCA-mutation related breast cancer risk. A 12 - 34 year follow-up of ~40 000 women in HUNT will have the potential to describe the population risk more precisely.
Summary and conclusions

• The future of drug discovery and precision medicine is presently fueled by human genomic discovery
• Genetic “experiments of nature” can inform therapeutic target discovery and provide insight into new mechanism (LoF)
• Return of medically genetic actionable results will affect health care resources to realize downstream health and economic benefits
• Partnership between industry, academia and health care systems can accelerate genomic discovery and implementation of precision medicine
Key Personnel

K. G. Jebesen Center for Genetic Epidemiology
- Kristian Hveem, Professor, Center leader, Head of HUNT biobank
- Maiken Elvestad Gabrielsen, Center Coordinator
- Pål Sæstrom, Professor Bioinformatics
- Bjørn Olav Åsvold, Professor, Epidemiology MR/
- Eivind Almås, Professor Systems Biology
- Oddgeir Lingaas Holmen, Leader HUNT Data Center/HUNT Cloud,
- Anne Heidi Skogholt, Coordinator, Analyses group
- Ben Brumpton, Senior Researcher, MR/GWAS
- Mari Løset, Post doc/Expression analysis
- Humaira Rasheed, Post, MR
- Christian Jonasson, Researcher pharmacogenomics
- Laurent Thomas – Bioinformatics, Applied biostatistics
- Eivind Coward, Bioinformatics
- Endre Bakken Stovner, System developer,
- Almut Luetge, Researcher, Biostatistics
- Tom Erik Røberg – Data administrator, HUNT Cloud
- Sandor Zeestraten – Data administrator, HUNT Cloud
- Siv Hege Stenshaug, MD, PhD Candidate
- Morten Solberg, MD, PhD Candidate
- Lars Ursin, ethicist

NTNU Genotyping Core Facility
- Sten Even Erlandsen, Senior engineer
- Tone Christensen, Lab engineer
- Tom Even Wheeler, Lab engineer
- Arnar Flatberg, Bioinformatics
- Vidar Beisvang, Lab leader,
- Arne Sandvik, Director, professor

Department of Public Health
- Siri Forsmo Professor, Dept head
- Geir Kristiansen, HR
- Surur Taso, Chief administrator

HUNT Research Center
- Inger Holbø, Secretary
- Maria Stuifbergen, HUNT Data Access Committee
- Turid Rygg Stene, HUNT Data Access Committee
- Steinar Krokstad, Professor, Head of HUNT Research Center

HUNT Database
- Arnulf Langhammer, Professor, Head of HUNT database
- Jon Heggland, Data base/LIMS programmer
- Jørn Fenstad, Data handler
- Elin Pettersen, Data handler
- Per Bjarne Løvsletten, programming, web application

HUNT Biobank
- Marit Næss, Lab leader
- Trine Altp, Kristin Sætermo, Rita Skjærvø, Elin Kyllo Lab engineers
- Ann Helen Røstad, Lab engineer, Quality manager

International Collaborators
- Goncalo Abecasis, Professor, Statistical genetics/Biostatistics, University of Michigan (UM), affiliated professor in Biostatistics, NTNU
- Cristen Willer, Assoc Professor, Internal Medicine, Human Genetics and Computational Medicine and Bioinformatics, Univ. of Michigan
- Mike Boehnke, Professor, Biostatistics/Statistical genetics, UM
- Mads Melby, MD, Professor Epidemiol., SSI, Copenhagen, Stanford Univ.
- George Davey Smith, Professor, MRC Unit, Univ. of Bristol, UK

Scientific Advisory Board
- Sekar Kathiresan, Dir., Prev. Cardiology, Massachusetts General Hospital, Professor of Medicine at Harvard Medical School.
- Eleftheria Zeggini, Professor, Wellcome Trust Sanger Institute, UK
- Björn Pasternak, Pharmacoepidem., Karolinska Institutet, Stockholm
Thank you