



Clinical implications of population based genetic discoveries

K Hveem, MD, PhD,
Professor in clinical epidemiology,
PI of HUNT Biobank,
National node director, BBMRI.no (Biobank Norway)
Head of K.G. Jebsen Center for Genetic Epidemiology.
Norwegian University of Science and Technology (NTNU)

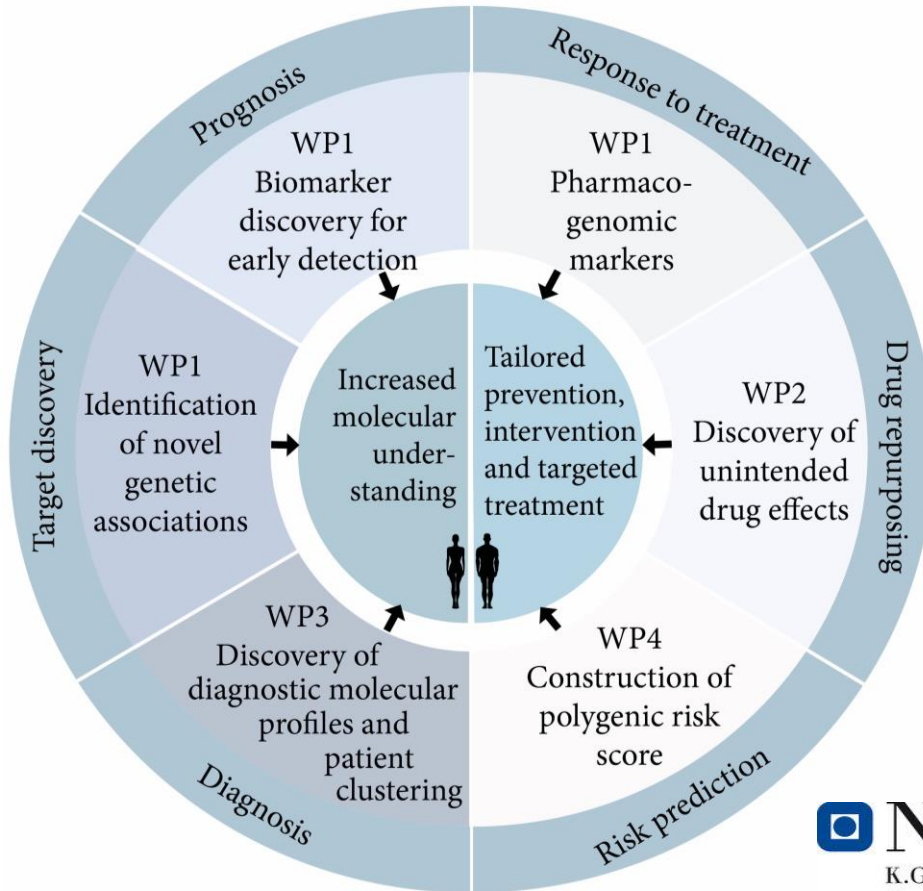
From populations to the clinic

WP 1 - Targeted follow-up of drug discovery

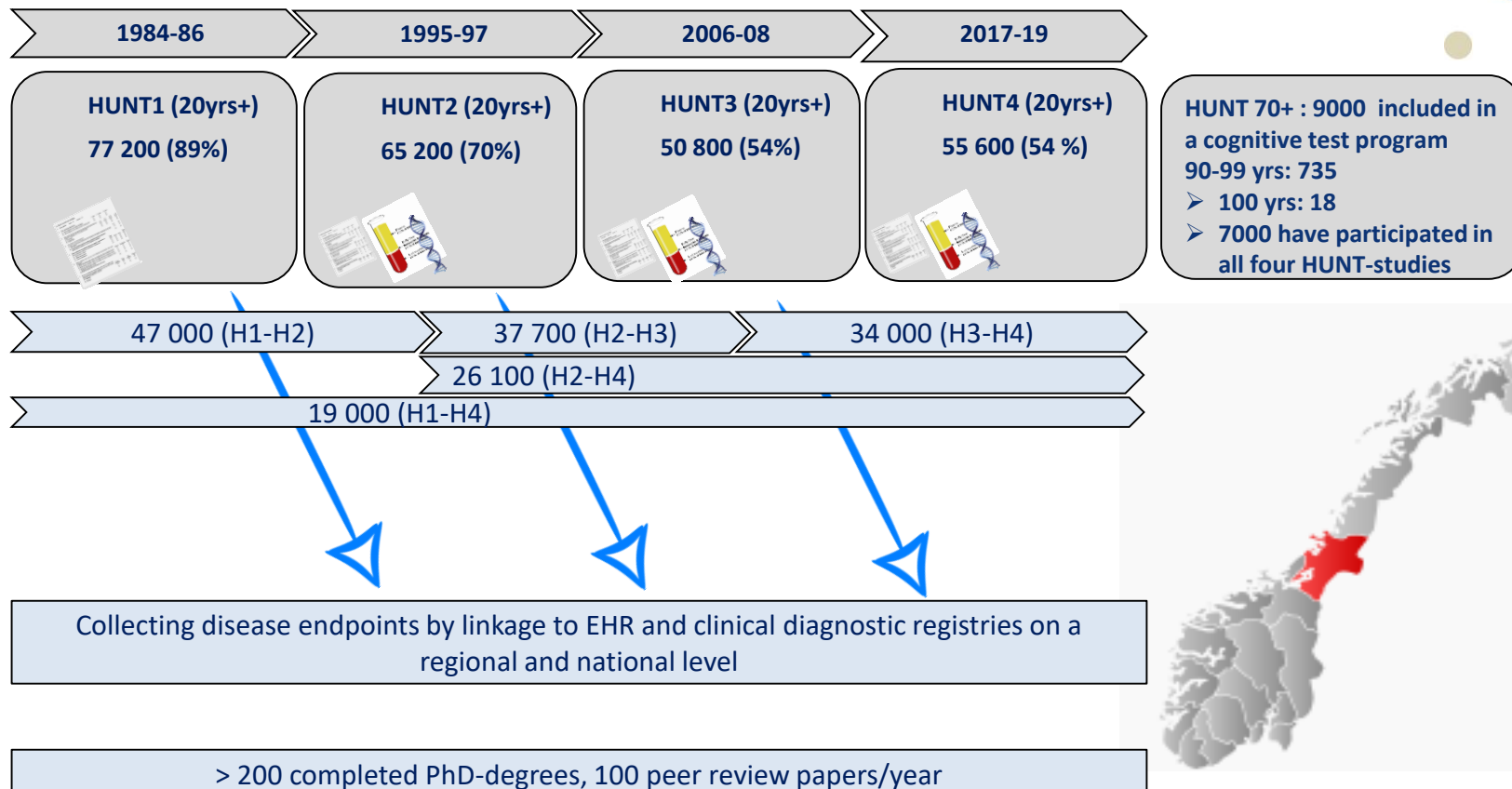
WP2 - Identify modifiable causes of disease

WP3 - Identifying molecular signatures of disease

WP4 – From genetic discovery to clinical translation

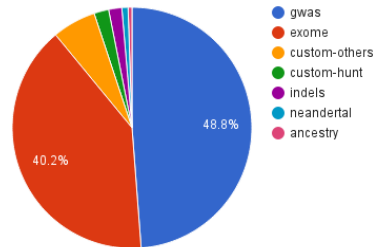


The HUNT study



- Genome wide genotyping of 70 000 HUNT-participants (Human Core Exome), ~100 000 by June 2020
- WGS of 2200 – low pass
- 604 000 genetic markers including 60 000 custom “HUNT SNPs”, imputed up till 28 mill (HRC/TopMed)
- CVD as main focus, > 60 sub-studies and 150 collaborating clinicians addressing other disease outcomes
- > 3000 GWAS-analyses completed
- Challenging ethics with actionable variants such as Familial Hypercholesterolemi (FH), BRCA2 related to return of results

Human Core Exome

[illegible]

Our analytic environment (HUNT Cloud)

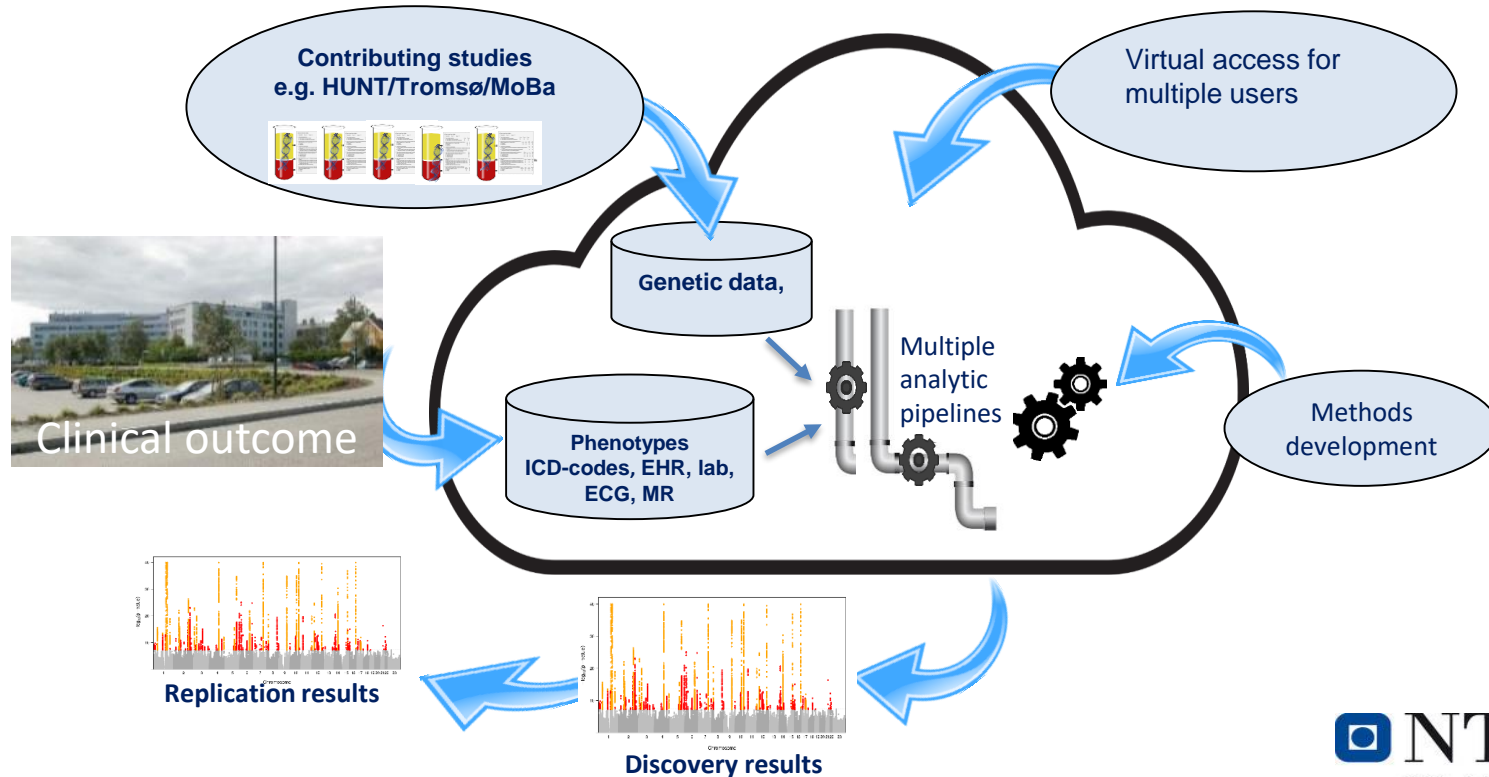


Illustration inspired by Brody JA, et al

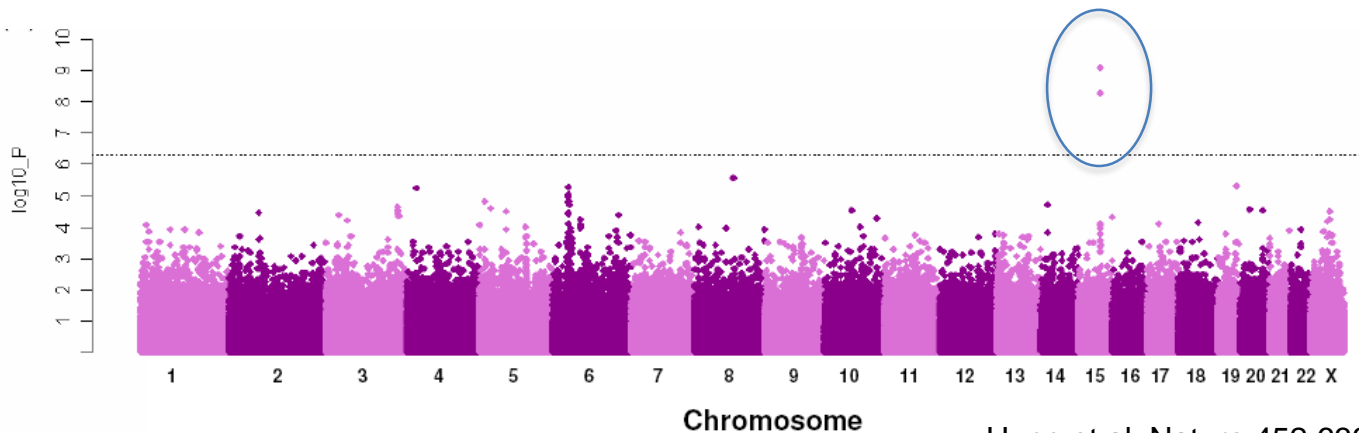
Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology,
Nat Genet. Oct, 2017



HUNT-involvement in biomarker discoveries

The first HUNT GWAS:

A susceptibility locus for lung cancer that maps to a nicotinic acetylcholine receptor subunit genes on 15q25



Hung et al. Nature 452:633-637 (2008)

GWAS-analyses has later been conducted on > 3000 binary or quantitative traits based on > 7000 unique variables from HUNT Data bank and > 2000 ICD-codes retrieved from medical records

Protective gene against MI

- 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
- Replicated in 4666 participants from the Tromsø study, 10 variants confirmed to be associated with a lipid trait $p < 5 \times 10^{-8}$
- Transient overexpression or knockdown in a mouse models altered serum lipid profiles
- *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

ARTICLES

nature
genetics

Systematic evaluation of coding variation identifies a candidate causal variant in *TM6SF2* influencing total cholesterol and myocardial infarction risk

Oddgeir I. Holmen^{1,2,15}, He Zhang^{3,15}, Yanbo Fan^{3,15}, Daniel H. Hovelson^{3,4}, Ellen M. Schmidt^{3,4}, Wei Zhou³, Yanhong Guo³, Ji Zhang³, Arnulf Langhammer¹, Maja-Lisa Lochen³, Santhi K. Ganesh^{3,6}, Lars Vatten⁷, Frank Skorpen⁸, Håvard Dalen^{9,10}, Jifeng Zhang³, Subramaniam Pennathur¹¹, Jin Chen³, Carl Platou⁹, Ellisiv B. Mathiesen^{12,13}, Tom Wilsaard³, Inger Njolstad³, Michael Boehnke¹⁴, Y. Eugene Chen³, Gonçalo R. Abecasis¹⁴, Kristian Hveem^{1,9} & Cristen J. Willer^{3,4,6}

Blood lipid levels are heritable, treatable risk factors for cardiovascular disease. We systematically assessed genome-wide coding variation to identify new genes influencing lipid traits, fine map known lipid loci and evaluate whether low-frequency variants with large effects exist for these traits. Using an exome array, we genotyped 80,137 coding variants in 5,643 Norwegians. We followed up 18 variants in 4,666 Norwegians and identified ten loci with coding variants associated with a lipid trait ($P < 5 \times 10^{-8}$). One variant in *TM6SF2* (encoding p.Glu167Tyr), residing in a known genome-wide association study locus for lipid traits, influences total cholesterol levels and is associated with myocardial infarction. Transient *TM6SF2* overexpression or knockdown in mice alters serum lipid profiles, consistent with the association observed in humans, identifying *TM6SF2* as a functional gene within a locus previously known as *NCAN-CILP2-PBX4* or *19p13*. This study demonstrates that systematic assessment of coding variation can quickly point to a candidate causal gene.

lipid levels are heritable, treatable, risk factors for cardiovascular disease. Systematic assessment of association between blood lipid levels and disease, a leading cause of death globally^{1,2}, coding variants has several potential benefits. First, it could implicate

nature
genetics

Nat. Genet. 2014 Apr;46(4):357-63. doi: 10.1038/ng.2915. Epub 2014 Mar 2.

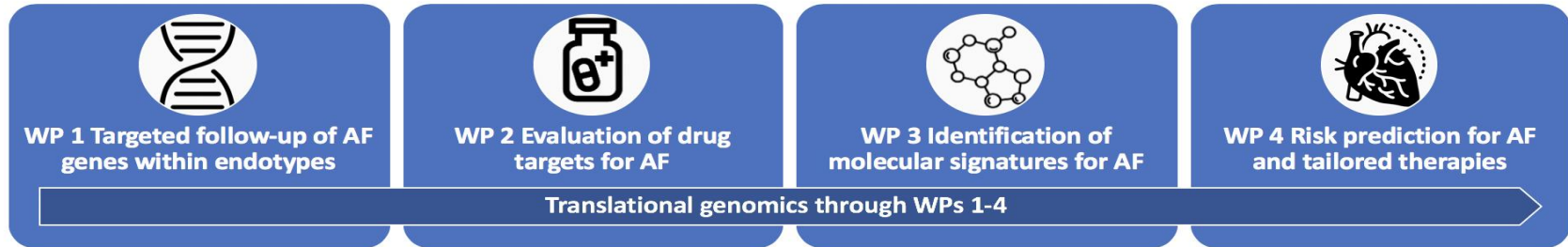
Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes.

Flannick J¹, Thorleifsson G², Beer NL³, Jacobs SB⁴, Grarup N⁵, Burtt NP⁶, Mahajan A⁶, Fuchsberger C⁷, Altmann G⁸, Benediktsson R⁹, Blangero J¹⁰, Bowden DW¹¹, Brandslund L¹², Brosnan J¹³, Burslem E¹⁴, Chambers J¹⁵, Cho YS¹⁶, Christensen C¹⁷, Douglas DA¹⁸, Duqirala R¹⁹, Dyrdek Z⁴, Farouq Y⁴, Fennell T⁴, Fontanillas P⁴, Forsén T¹⁰, Gabriel S⁴, Glaser B²⁰, Gudbjartsson DF⁴, Hanis C²¹, Hansen T²², Hreidarsson AB⁴, Hveem K²³, Ingelsson E²⁴, Isomaa B²⁵, Johansson S²⁶, Jørgensen T²⁷, Jørgensen ME²⁸, Kathiresan S²⁹, Kong A³, Kooper C³⁰, Kravic J³¹, Laakso M³², Lee JY³³, Lind L³⁴, Lindgren CM³⁵, Linneberg A³⁶, Masson G³⁷, Mettinger T³⁷, Mohike KI³⁸, Molven A³⁹, Morris AP⁴⁰, Pollut S⁴¹, Rauramaa R⁴², Ribel-Madsen R⁴³, Richard AM⁴⁴, Rolsh T⁴⁵, Salomaa Y⁴⁶, Seckl AV⁴⁷, Skärstrand H⁴⁸, Steinthorsdottir V⁴⁹, Stringham HM⁵⁰, Sulem P⁵¹, Tai ES⁵², Teo YY⁵³, Teslovich T⁵⁴, Thorsteinsdottir L⁵⁵, Trimmer JK⁵⁶, Tuomi T⁵⁷, Tuomilehto J⁵⁸, Vaziri-Sani F⁵⁹, Voight BF⁶⁰, Wilson JG⁶¹, Boehnke M⁶², McCarthy M⁶³, Njolstad PR⁶⁴, Pedersen O⁶⁵, Go-T2D Consortium, T2D-GENES Consortium, Groop L⁶⁶, Cox DR⁶⁷, Stefansson K⁶⁸, Altshuler D⁶⁹.


Atrial fibrillation



- 60 k cases and 930 k controls
- Identified 142 independent risk variants at 111 loci (80 novel), explaining 11.2% of the variation in atrial fibrillation.
- For functional follow-up,
 - integrated information on tissue and cell specific gene expression,
 - genomic regulatory elements and electrocardiogram (ECG) parameters from >62,976 Icelanders in sinus rhythm.
- Prioritized *MYH6* and *MYH7* as two of the most likely functional genes, supported by experiments in rabbits.

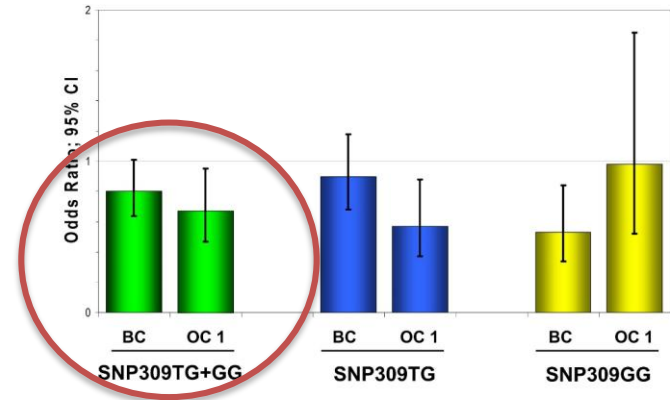


Nielsen JB et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology, Nat Genet 2018

- 
- To identify causal relations
 - *Budu-Aggrey A et al Evidence of a **causal relationship between body mass index and psoriasis**: A mendelian randomization study. PLoS Med 2019*
 - *Brumpton BM et al Variation in Serum PCSK9, Cardiovascular Disease Risk, and an Investigation of **Potential Unanticipated Effects of PCSK9 Inhibition**. Circ Genom Precis Med. 2019*
 - Methodological development
 - *Brumpton B et al. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. bioRxiv 602516; doi: <https://doi.org/10.1101/602516>*
 - *Zhou W et al. Efficiently controlling for **case-control imbalance and sample relatedness** in large-scale genetic association studies. Nat Genet 2018*

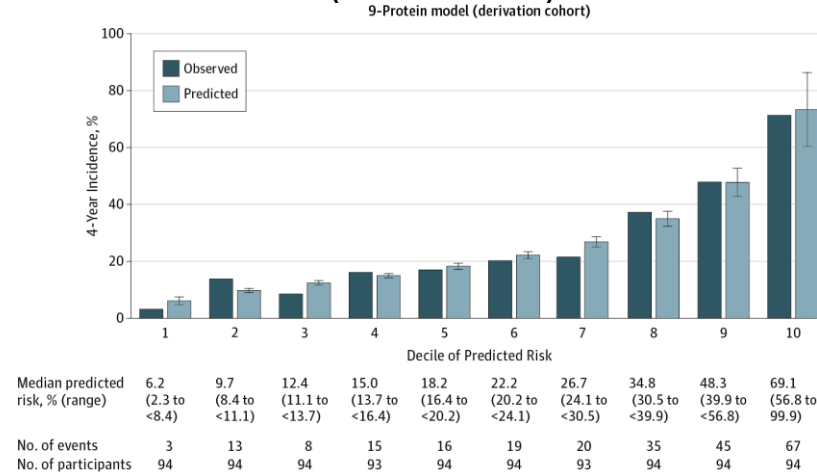
***Knappskog S et al.* The *MDM2* Promoter SNP285C/309G Haplotype Diminishes Sp1 Transcription Factor Binding and **Reduces Risk for Breast and Ovarian Cancer in Caucasians**
Cancer Cell 2011 19, 273-282**

- A MDM2 promoter polymorphism, SNP285G > C, is residing on the SNP309G allele.
- 7.7% (95% CI 7.6%–7.8%) of healthy individuals carry the SNP285C/309G haplotype.
- **Study population recruited from CONOR**
 - Ovarian cancer (n = 1993)
 - Breast cancer (n = 1973)
 - Healthy controls (n = 3646)
- SNP285C reduced the risk in both
 - ovarian cancer (**OR 0.74; CI 0.58–0.94**) and
 - breast cancer (**OR 0.79; CI 0.62–1.00**)



	HC	BC	OC 1	HC	BC	OC 1	HC	BC	OC 1
SNP285GG									
n	1232	1038	486	993	813	406	239	225	80
(%)	(86.3)	(88.9)	(90.7)	(88.3)	(89.3)	(92.9)	(78.6)	(87.2)	(80.8)
SNP285GC									
n	196	130	50	131	97	31	65	33	19
(%)	(13.7) ¹	(11.1) ²	(9.3) ³	(11.7)	(10.7)	(7.1)	(21.4) ¹	(12.8) ²	(19.2) ³
OR		0.79 ⁴	0.67		0.91 ⁴	0.57		0.55 ⁴	0.98
CI (95%)		0.62	0.47		0.69	0.37		0.35	0.52
		1.00	0.95		1.20	0.88		0.86	1.85
P-value ⁵		0.031	0.011		0.525	0.007		0.006	0.606

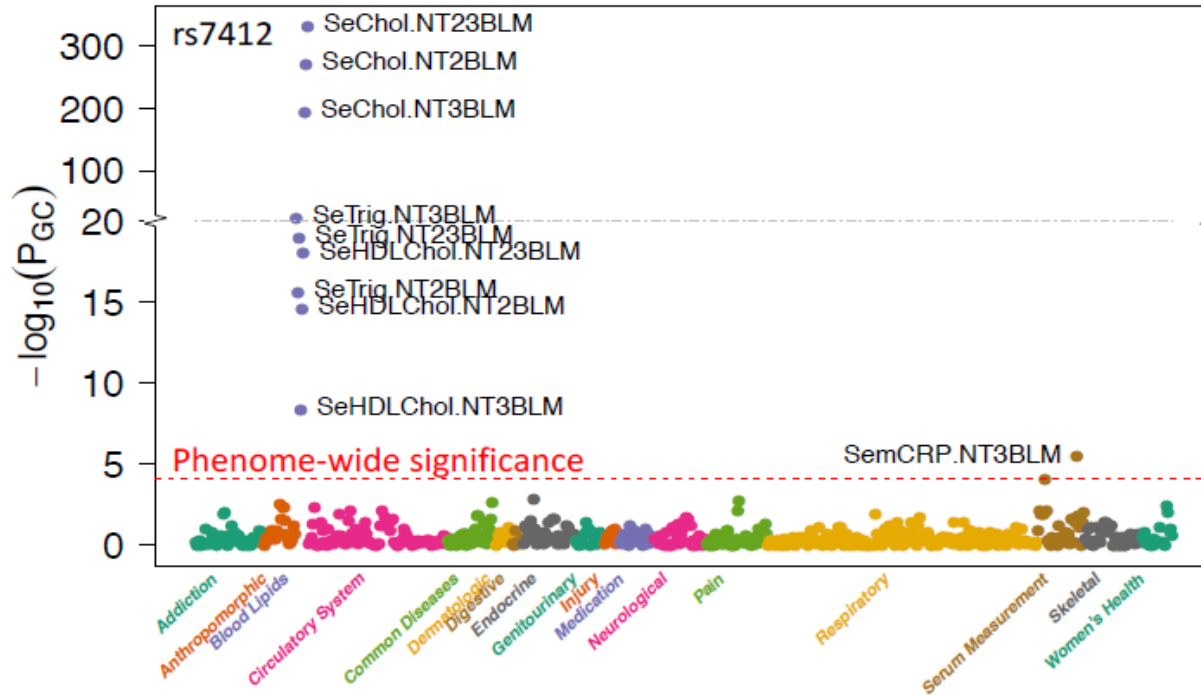
Predicting CVD-events on protein based biomarkers (Somascan)



Agreement Between Observed vs Predicted 4-Year Incidence of Myocardial Infarction, Stroke, Heart Failure, and Death With the 9-Protein Model

Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, Sterling DG, Williams SA.
Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes
Among Patients With Stable Coronary Heart Disease
JAMA. 2016;315(23):2532-2541.

PheWas analyses are presented on HUNT pheweb based on summary statistics from multiple GWAS



- APOE:p.Arg176 Cys (rs7412) associated with
- Alzheimers disease,
 - fasting blood lipids and
 - dyslipidemia

Pre-competitive collaborations to promote large-scale analyses (population based)

- **Sequencing**
 - WES of 15 000 -100 000 in the planning face
- **Metabolomics**
 - 20 000 within next 12 months (*Nightingale*)
 - 78 000 (26 000x3 – H2,H3,H4) – in the planning face (*Metabolon*)
- **Proteomics**
 - Somascan – aptamer-based assay optimized for protein biomarker discovery, 3000 CVD-related subjects based on a both protein based arrays (1500/5000 proteins pr array) (*SomaLogic*)
 - Microbiota -14 000 fully characterized within 6 months (*Bio-Me*)
- **Other biomarkers**
 - Troponin I – 10 000 HUNT2, 6000 HUNT3, 35 000 HUNT4 (*Abbot*)

Polygenic Risk Scores (FinnGen)

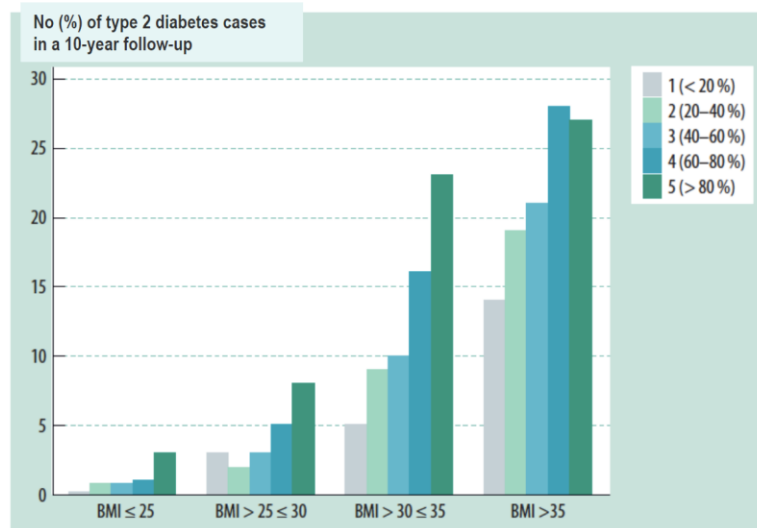
COMPARISON OF PRS AND BMI IN TYPE 2 DIABETES



Genetic risk brings added personalized value to the disease prevention regardless of lifestyle



10/06/19

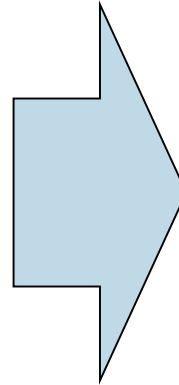


markus.perola@thl.fi

Perola et al Duodecim 2019

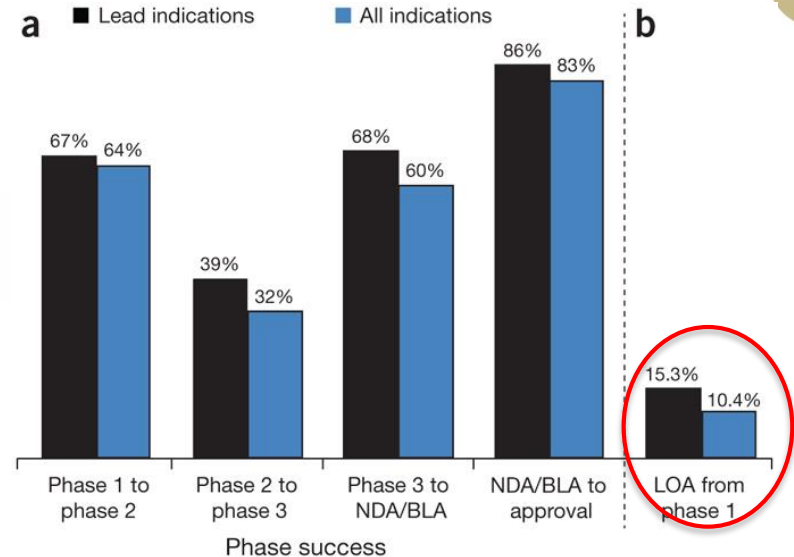
11

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





Potential drug targets based on HUNT-related genetic discoveries

MEPE-gene,

associated with reduced bone mineral density and risk of fracture

- *Ida Surraka et al. Loss-of-function mutation in the MEPE gene is associated decreased bone mineral density and increased risk for fractures and osteoporosis—submitted*
- *Cristopher V et al . Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. bioRxiv 572347; doi: <https://doi.org/10.1101/572347>*

ZNF529-gene,

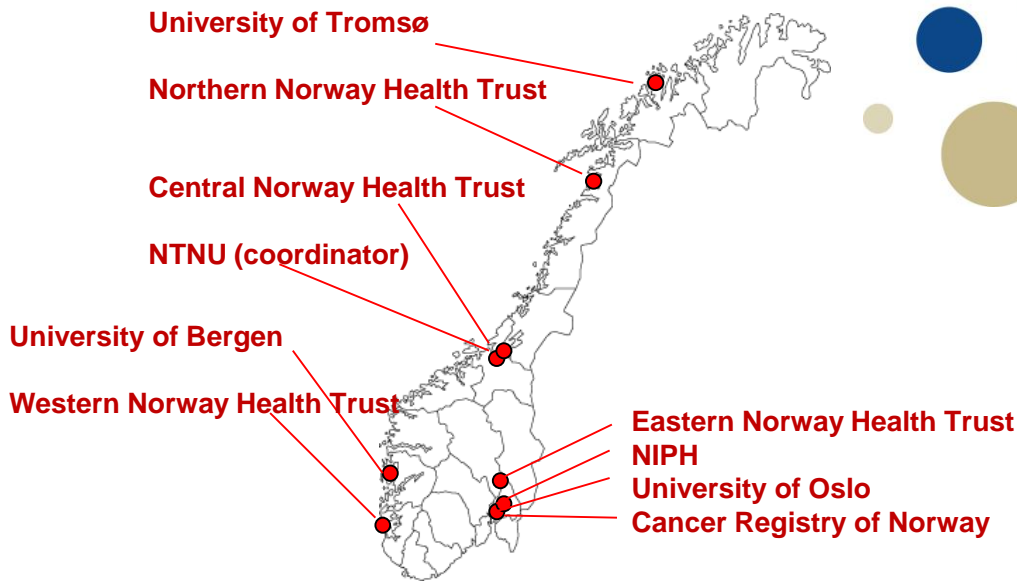
A novel LoF variant, *ZNF529*:p.K405X, associated with decreased levels of LDL-C ($P=1.3 \times 10^{-8}$) but demonstrated no association with liver enzymes or non-fasting blood glucose levels.

Silencing of *ZNF529* in human hepatocytes resulted in upregulation of LDL receptor (LDLR) and increased LDL-C uptake in the cells, suggesting that inhibition of *ZNF529* or its gene product could be used for treating hypercholesterolemia and hence reduce the risk of CVD.

- *Jonas B. Nielsen et al. Loss-of-function genomic variants with impact on liver-related blood traits highlight potential therapeutic targets for cardiovascular disease. bioRxiv 597377, doi: <https://doi.org/10.1101/597377>*



A National Biobank Infrastructure since 2010



- Observer state of BBMRI-ERIC since 2013
- Full member state of BBMRI-ERIC since 2016
- Funded with 215 mill NOK by the Research Council of Norway

BIOBANK NORWAY PARTNERS

UiT, NTNU, UiB, UiO, NIPH, HN, HV, HMN, HSØ, OUS, Cancer Registry

Step-wise approach:

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

Population-based studies ~400k

- MoBa ~265k, HUNT ~90k, TU ~40k, HUSK ~30k
- (200 000 presently genotyped)

Existing hospital disease-specific biobanks ~30k

- Neuro, Psychiatry, Lung, Gastro, Cancer, Cardiovascular, Metabolic etc. TBD

Broad based consent hospital biobanking ~200k

- Sampling of ~200k subjects from the clinic nation-wide
- Sequencing of clinical samples

extraction of pathology and molecular data to

- >30k new incident cancer cases per year

Step-wise funding:

1. BN3
2. Governmental
3. Industry
4. Precomp. PPP

PRIVACY-BY-DESIGN DATA PLATFORM AND OPEN ACCESS GOVERNANCE MODEL

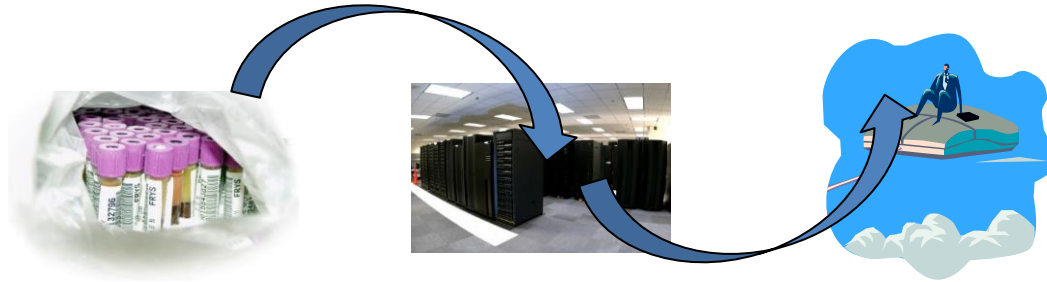


Research & Innovation



Data storage, access and analyses

Digitizing biobanks – the future perspectives



A shift from samples to data (digitalization)

- Reduced costs for access to larger sample sizes (omics-driven analyses) as a trade off for significant return of results to the biobank
- Increased costs for data storage
- Increased focus on data security (GDPR)
- Reduction of data export, researchers will be granted virtual access to data clouds e.g. HUNT Cloud, TSD, SAFE, HAP..... , also offering computational capacity
- Biobanks will play a stronger role in precision medicine
- Access to annotated biobank samples and national registry data will be centralized to publicly governed Health data platforms and Health analyses platforms

Ethical considerations and return of results

Ethics, openness and dissemination



- Attendance rate and public support is highly dependant on trust
- All projects are approved by both the REC and HUNT-Data Access Committee
- Transparency is encouraged by active use of webpages (e.g. www.hunt.no) with updates on on-going research projects, project resumé, recent and previous publications and new findings.
- Communcation with the donor community is encouraged and a “my page” for research participants on www.helsenorge.no is established for all HUNT4-participants.

- 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
 - **84 % yes**

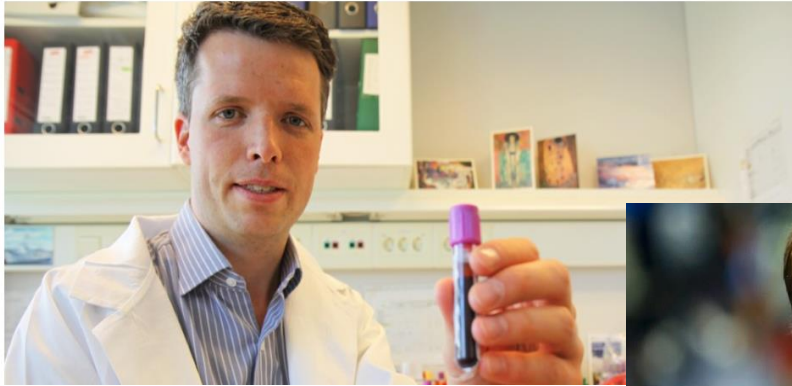
Return of results



- International recommendations:
Genetic information/risk must be "actionable" to trigger a feedback.
 - Provided good opportunities for prevention, or even treatment, the situation most commonly is referred to as actionable
 - The researchers must then plan for feedback (Biotechnology act
- Based on WES 3,5 % have actionable results (Geisinger)
- Other markers? Genetic risk scores?

Familial Hypercholesterolemia (FH)

- 25.000 Norwegians have the same disease as Dale Oen
 - Many of us are carriers of a serious condition such as FH without knowing



© Publisert 10.06.2015, kl. 22:22



21 women had their breasts and ovaries removed – should never been operated

-

21 kvinner som fikk operert bort brystene, eggstokkene eller begge deler, ved OUS mellom 2002 og 2014, har fått beskjed om at de egentlig ikke skulle blitt operert.

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



- Population studies offer a unique prospective longitudinal design for biomarker discovery and validation
- Large scale omics analyses in combination with health data, physical examination and access to a large number of phenotypes also creates a comprehensive research platform for biomarker discovery
- The future of drug discovery and precision medicine is presently fueled by biobank based human genomic discovery
- Genetic “experiments of nature” can inform therapeutic target discovery and provide insight into new mechanism (LoF)
- “Biological recall” is accepted
- 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. Jebsen Center for Genetic Epidemiology, NTNU

- Kristian Hveem, Professor, Center leader, PI HUNT biobank
- Maiken Elvestad Gabrielsen, Center Coordinator
- Pål Sætrum, Professor Bioinformatics,
- Bjørn Olav Åsvold, Professor, Epidemiology MR/
- Eivind Almås, Professor Systems Biology
- Oddgeir Lingaas Holmen, MD, PhD, PI/Head of 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 doc, MR
- Eirin Haug, PhD, Post doc
- Ailin Falkmo Hansen, PhD, Post doc
- Gunnhild Aaberg Vie, PhD, Post doc
- Christian Jonasson, Researcher pharmacogenomics
- Laurent Thomas – Bioinformatics, Applied biostatistics
- Eivind Coward, Bioinformatics
- Endre Bakken Stovner, System developer, PhD-student
- Tom Erik Røberg – Data administrator, HUNT Cloud
- Sandor Zeestraten – Data administrator, HUNT Cloud
- Qussay Ghazeia, Quality management, HUNT Cloud
- Matus Kosut, Programmer, HUNT Cloud
- Siv Hege Stemshaug, MD, PhD student
- Ole Jørgen Bekkevold MD, PhD student
- Marta Riise Moksnes, PhD-student
- Kjartan Øvretveit, PhD-student
- Lars Ursin, PhD, Assoc prof., ethicist
- Jonas Bille Nielsen, MD, PhD, Researcher,
- Maria Brandkvist, PhD-student
- Janne Tellefsen, Communication
- Åshild Solvin, medical student/Student Research Program
- Eivind Ness-Jensen, MD, PhD, Assoc professor
- Eivor Laugsand, MD, PhD, Assoc professor

K. G. Jebsen Center for Genetic Epidemiology, Univ of Michigan

- Cristen Willer, PhD, Prof Human Genetics and Computational Medicine and Bioinformatics
- Goncalo Abecasis, Professor, Stat. Genetics, Biostatistics, affiliated professor in Biostatistics, NTNU
- Mike Boehnke, Professor, Biostatistics/Statistical genetics,
- Brooke Wolford, PhD-student, Bioinformatics
- Sarah Graham, PhD, Researcher/analyst
- Ida Surakka, PhD, Post doc,
- Lars Fritsche, PhD, Researcher
- Wei Zhou, Post doc, Broad Institute

Dept of Public Health

- Siri Forsmo Professor, Dept head
- Geir Kristiansen, HR
- Surur Taso, Chief administrator
- Johan Håkon Bjørnegaard, Prof/Head of Research

International collaborators

- Goncalo Abecasis, Professor, Statistical genetics/Biostatistics, University of Michigan (UM), affiliated professor in Biostatistics, NTNU
- Cristen Willer, Professor, Internal Medicine, Human Genetics and Computational Medicine and Bioinformatics, Univ. of Michigan
- Mike Boehnke, Professor, Biostatistics/Statistical genetics, UM
- Mads Melby, MD, Professor Epidem., SSI, Copenhagen, Stanford Univ.
- George Davey Smith, Professor, MRC Unit, Univ. of Bristol, UK
- Henning Bundgaard, MD,PHD, Prof Cardiology, Copenhagen Central Hospital
- Hilma Holm, MD, PhD, Professor, Cardiology, deCODE Genetics

Scientific Advisory Board

- Sekar Kathiresan, Dir., Prev. Cardiology, Mass. General Hospital, Prof in Medicine at Harvard Medical School.
- Eleftheria Zeggini, Professor, Head of genetics, Helmholtz Institute, Munich
- Björn Pasternak, MD. PhD, Pharmacoepidem., Karolinska Institutet, Stockholm
- Mark Daly, PhD, Prof, Director FIMM, Finland – Broad Institute, US

NTNU Genotyping Core Facility

- Sten Even Erlandsen, Senior engineer
- Tone Christensen, Lab engineer
- Tom Even Wheeler. Lab engineer
- Arnar Flatberg, Bioinformatics
- Vidar Beisvåg, Lab leader,
- Arne Sandvik, Director, professor

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 databank

- Arnulf Langhammer, Professor, Head of HUNT databank
- 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/Head of Biobank
- Trine Altø, Kristin Sætermo, Rita Skjærvø, Elin Kyllø Lab engineers
- Ann Helen Røstad, Lab engineer, Quality manager

Thank you

