

The HUNT Study: a population-based cohort for genetic research

Ben M. Brumpton^{1,2,3*#}, Sarah Graham^{4*}, Ida Surakka^{4*}, Anne Heidi Skogholt¹, Mari Løset^{1,5}, Lars G. Fritsche⁶, Brooke Wolford^{7,8}, Wei Zhou^{9,10}, Jonas Bille Nielsen¹¹, Oddgeir L. Holmen^{1,2}, Maiken E. Gabrielsen^{1,2}, Laurent Thomas^{1,12,13}, Laxmi Bhatta¹, Humaira Rasheed¹, He Zhang⁴, Hyun Min Kang⁶, Whitney Hornsby⁴, Marta R. Moksnes¹, Eivind Coward¹, Mads Melbye¹, Guro F. Giskeødegård¹, Jørn Fenstad², Steinar Krokstad^{2,14}, Marit Næss^{2,14}, Arnulf Langhammer^{2,14}, Michael Boehnke⁶, Gonçalo R. Abecasis¹⁵, Bjørn Olav Åsvold^{1,2,16}, Kristian Hveem^{1,2*#}, Cristen J. Willer^{1,4,6,17*#}

Affiliations

¹ K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, 7030, Norway

² HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, 7600 Norway

³ Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, 7030, Norway

⁴ Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI 48109, USA

⁵ Department of Dermatology, Clinic of Orthopaedy, Rheumatology and Dermatology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

⁶ Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

⁷ Department of Computational Medicine and Bioinformatics, Ann Arbor, Michigan, USA

⁸ Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA

⁹ Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

¹⁰ Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

¹¹ Department of Epidemiology Research, Statens Serum Institute, Copenhagen, Denmark

¹² Department of Clinical and Molecular Medicine, NTNU Norwegian University of Science and Technology, Trondheim, Norway

¹³ BioCore - Bioinformatics Core Facility, NTNU Norwegian University of Science and Technology, Trondheim, Norway; Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

¹⁴ Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

¹⁵ Regeneron Genetics Center, Tarrytown, NY 10591, USA.

¹⁶ Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, 7030, Norway.

¹⁷ Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA

*These authors contributed equally

#Corresponding authors:

ben.brumpton@ntnu.no

kristian.hveem@ntnu.no

cristen@umich.edu

Summary

The Trøndelag Health Study (HUNT) is a population-based cohort of ~229,000 individuals recruited in four waves beginning in 1984 in Trøndelag County, Norway. ~88,000 of these individuals have available genetic data from array genotyping. HUNT participants were recruited during 4 community-based recruitment waves and provided information on health-related behaviors, self-reported diagnoses, family history of disease, and underwent physical examinations. Linkage via the Norwegian personal identification number integrates digitized health care information from doctor visits and national health registries including death, cancer and prescription registries. Genome-wide association studies of HUNT participants have provided insights into the mechanism of cardiovascular, metabolic, osteoporotic and liver-related diseases, among others. Unique features of this cohort that facilitate research include nearly 40 years of longitudinal follow-up in a motivated and well-educated population, family data, comprehensive phenotyping, and broad availability of DNA, RNA, urine, fecal, plasma, and serum samples.

Keywords

biobank, complex disease, genetics, cardiovascular health, longitudinal, public health, genetic epidemiology, big data, electronic health records

Background

Norway, like other Nordic countries, has characteristics that are uniquely favorable for recruitment to population studies, establishing biobanks, and identifying clinical outcomes and prospective disease trajectories. This includes a unique personal identification number applied throughout the life span, a universal and digitized public health care system, and accessible harmonized electronic health records. In addition, seventeen mandatory and validated national health registries are used for health analysis, administration, and emergency preparedness, and fifty-two national medical quality registries provide disease specific data on diagnosis and treatment parameters. Finally, Norwegians are an altruistic, highly motivated population for participating in biomedical research, as reflected in survey response rates of up to 89%. These factors have supported the establishment and maintenance of the Trøndelag Health Study (HUNT), a large population-based prospective Norwegian cohort, linked to registries and biobanks dating back more than 50 years (**Figure 1**).

To understand the genetic basis of diseases, as well as follow individuals with genetic and epidemiological risk factors in a well-ascertained county in Norway, we established a comprehensive collaboration in 2005 between the HUNT study at the Norwegian University of Science and Technology, Norway and the University of Michigan, USA. This paper presents the history and status of this collaboration by describing the study population, the strategy incorporating genotyping, sequencing, and imputation-based approaches in HUNT, the vast phenotype data collected by decades of HUNT researchers, the linkage to the digitized public health care system and key findings to date.



Figure 1: The Trøndelag Health Study (HUNT), Trøndelag, Norway. The country of Trøndelag is shaded light blue and the orange point indicates the location of the HUNT Research Center at Levanger.

Study population

HUNT is an ongoing population-based health study in Trøndelag County, Norway. The study collects health-related data from questionnaires, interviews and clinical examinations from individuals within this geographical region (**Figure 2**). More than 229 000 adults (20 years or older at recruitment) have participated in the study to date, of whom 95 000 have provided at least one biological sample (<https://www.ntnu.edu/hunt/hunt-samples>)¹⁻⁴. The periodic survey design includes four recruitment waves. HUNT1 (1984-86), HUNT2 (1995-97), HUNT3 (2006-08) and HUNT4 (2017-19) concentrated primarily on the North-Trøndelag area, where all adults (age ≥ 20 years) were invited. In addition, HUNT4 expanded to collect basic questionnaire data from the adult population of South-Trøndelag (105,797 additional participants)³. ~19 000 adults have participated in all four HUNT waves, thus having longitudinal questionnaire and physical exam information spanning over 35 years. Complementing the surveys in adult participants, four separate Young-HUNT surveys gathered data from ~25 000 adolescents in junior high and high school, concurrent with HUNT2-4. No genotyping has been performed on Young-HUNT, however 4 212 have sequentially participated in the adult version of HUNT. The HUNT Study has a high level of participation (ranging from 54% to 89% between surveys among those invited) making the cohort a good representative of the general Norwegian population. The HUNT and Young-HUNT cohorts are described in more detail elsewhere¹⁻⁵.

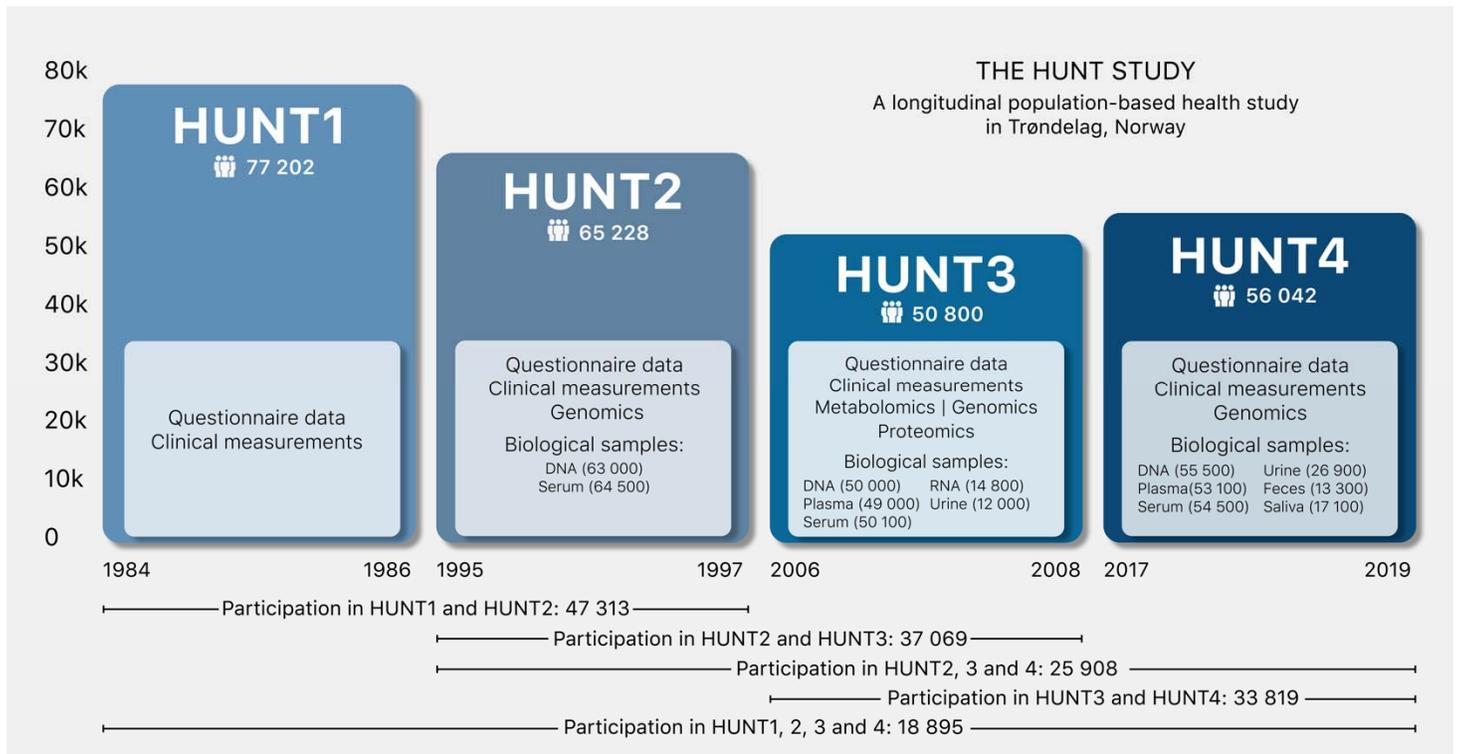


Figure 2: Sample sizes across the HUNT1-4 surveys and details of key data and biological samples.

DNA: Deoxyribonucleic acid, HUNT: Trøndelag Health Study, RNA: Ribonucleic acid

Genotyping and imputation study design in HUNT

~88 000 individuals provided DNA for medical research during at least one of the HUNT recruitment periods. Initially, our efforts were focused on identifying genetic variants associated with myocardial infarction (MI)⁶⁻⁸. Towards this goal, we genotyped exome variants and performed low-pass whole-genome sequencing (4.7x average coverage) in 2014 on 2,201 samples from HUNT2 and HUNT3 (HUNT-WGS) (**Supplementary Table 1**), including early-onset MI cases and equal numbers of sex- and age-matched controls. Although no novel significant associations were found, likely due to the limited sample size, this set of low-pass sequences provided important insights into genetic variants present in the Norwegian population and contributed Norwegian reference sequences to the Haplotype Reference Consortium (HRC) imputation panel⁹. We next undertook genome-wide genotyping on all HUNT2-3 participants (n=71 860) with available DNA (**Figure 3**). Motivated by a goal of capturing high-quality, common- and low-frequency and Norwegian-specific variants, we used a variety of approaches to observe or estimate genotypes: 1) direct genotyping using standard and customized HumanCoreExome arrays from Illumina; 2) genotyping and imputation with a merged HRC and HUNT-WGS imputation panel; and 3) imputation with the TOPMed imputation panel (**Figure 4**). After genotyping 12 864 with standard HumanCoreExome arrays (HumanCoreExome 12 v1.0 and v1.1), we performed genotyping on the remaining samples using a customized HumanCoreExome array (UM HUNT Biobank v1.0) which included protein-altering variants observed in HUNT-WGS. This resulted in genotyping

358 964 polymorphic variants. We next used the 2 201 sequenced samples (HUNT-WGS) for joint imputation with the HRC panel¹⁰. We previously showed that imputation with a HUNT-specific reference panel improved imputation of low-frequency and population-specific variants compared to either using the 1000 Genomes or HRC reference panels alone¹¹. Lastly, we imputed 25 million variants from the TOPMed imputation panel (minor allele count greater than 10), which resulted in slightly lower imputation quality compared to the population-specific reference panel but captured a larger number of variants (**Supplementary Figure 1**). These two imputed datasets can be used separately in downstream analysis; we recommend using the HRC and HUNT-WGS imputation for the investigation of the Norwegian specific variants. Together, the imputations resulted in 33 million variants in 70 517 individuals from HUNT2 or HUNT3 of which 3.3 million variants are not found in UK Biobank. Finally, 18 722 new samples from HUNT4 have recently been genotyped using the same approaches (HumanCoreExome array, UM HUNT Biobank v2.0) and following imputation will create a new, larger data freeze of ~88,000 individuals from HUNT2-4. Further details of the quality control and imputation in HUNT can be found in the **Supplementary Note**.

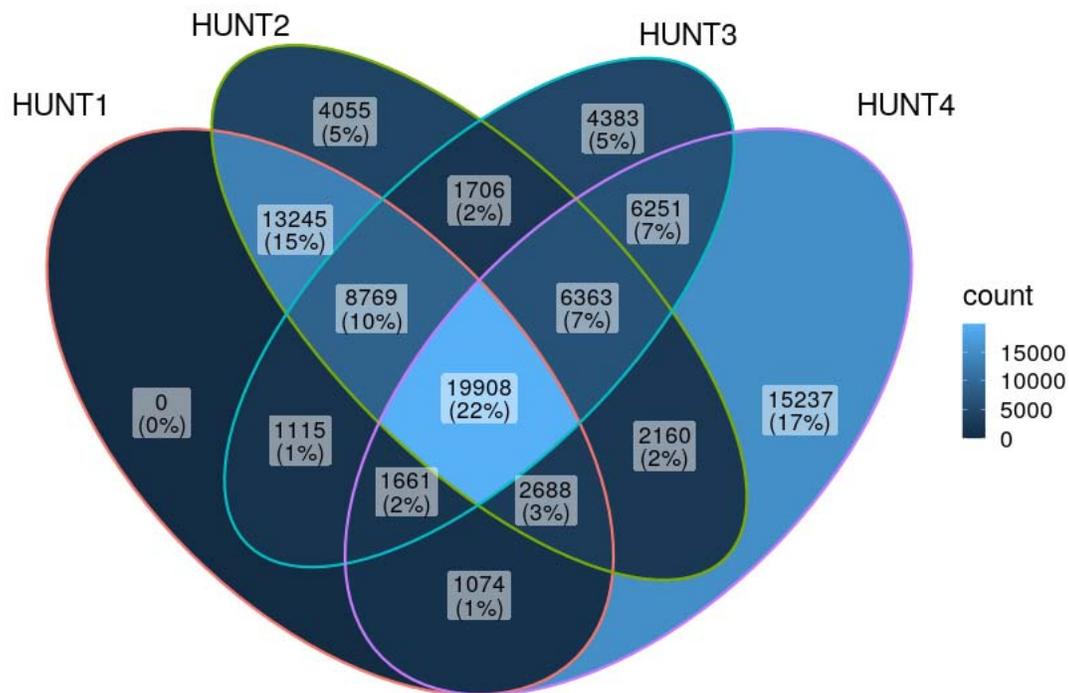


Figure 3: Genotyped samples from HUNT available from the different HUNT surveys (n=88 615).
HUNT: Trøndelag Health Study

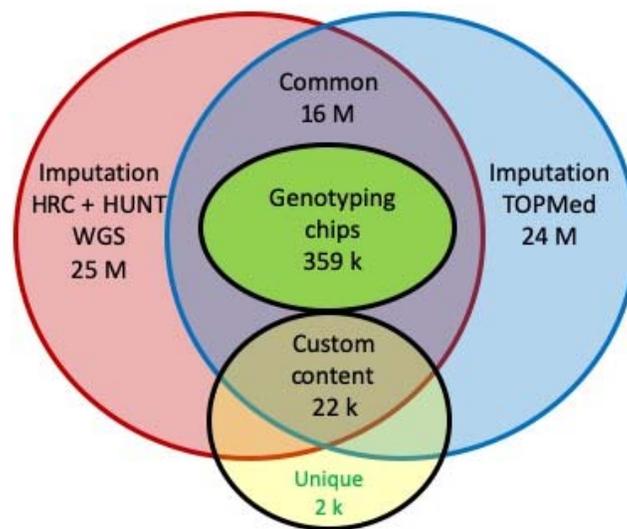


Figure 4: Genotyping and imputation-based approach in HUNT

HUNT: Trøndelag Health Study, HRC: Haplotype Reference Consortium, k: Thousand, M: Million, TOPMed: Trans-Omics for Precision Medicine, WGS: Whole-genome sequencing

Phenotypes

A broad range of phenotypes are available for HUNT participants based on laboratory tests, clinical examinations, and self-reported questionnaires. These include non-fasting blood lipids and glycaemic traits; history (including age of diagnosis) of a range of diseases including cardiovascular events; basic demographics including sex and participation age; anthropometrics including weight, height, BMI, and waist-to-hip ratio; blood pressure measurements; and lifestyle information including smoking status (**Table 1**). HUNT data categories have been previously described^{2, 3}, and are described in detail on the HUNT databank website (<https://www.ntnu.edu/hunt/databank>). Importantly, many measurements and questionnaire items have been intentionally kept identical or similar across HUNT surveys to enable longitudinal analyses.

Table 1: HUNT cohort demographics for all attendees at HUNT1-4 clinical examinations (n=123 219), and among those genotyped (n=88 615).

	All (HUNT1-4)				Genotyped (HUNT2-4)			
	N	Total	Male	Female	N	Total	Male	Female
Number of Individuals (%)	123 219		59 121 (48%)	64 098 (52%)	88 615		41 482 (47%)	47 133 (53%)
Age at ascertainment, yrs (range 18-90+; mean ± SD)*	123 219	43.8 ± 17.7	43.8 ± 17.3	43.9 ± 18.0	88 566	39.1 ± 14.0	39.2 ± 13.8	39.0 ± 14.1
Follow-up time, years	123177	22.4 ± 12.8	22.2 ± 12.9	22.5 ± 12.7	88548	24.1 ± 12.4	24.2 ± 12.4	24.0 ± 12.5
Quantitative Measurements (mean ± SD) §								
BMI, kg/m²	119 888	26.8 ± 4.6	26.9 ± 4.1	26.8 ± 5.1	88 345	27.2 ± 4.7	27.3 ± 4.1	27.0 ± 5.1
SBP **, mm Hg	120 448	136.8 ± 23.7	138 ± 21.2	135 ± 25.6	88 420	133 ± 21.3	135 ± 19.2	131 ± 22.8
LDL-C#, mg/dL	93 835	3.4 ± 1.1	3.3 ± 1.1	3.4 ± 1.1	87 163	3.3 ± 1.1	3.3 ± 1	3.4 ± 1.1
Creatinine, µmol/l	95 361	80.3 ± 22	88.9 ± 22.7	72.7 ± 18.3	88 527	79.6 ± 22.0	88.3 ± 22.9	72.0 ± 18.0
Glucose###, mmol/L	78 429	5.6 ± 1.7	5.7 ± 1.8	5.5 ± 1.6	71 790	5.6 ± 1.7	5.7 ± 1.8	5.5 ± 1.6
Thyroid stimulating hormone, mIU/L	71 213	1.4 ± 1.5	1.5 ± 1.5	1.4 ± 1.6	70 541	1.4 ± 1.5	1.5 ± 1.5	1.4 ± 1.6
Blood haemoglobin, g/dL	54347	14.6 ± 1.3	15.4 ± 1.2	14.0 ± 1.0	51 892	31 ± 1.7	31.1 ± 1.6	30.8 ± 1.8
FEV1	18 854	3.1 ± 1.0	3.6 ± 1.1	2,7 ± 0.8	17687	3.1 ± 1.0	3.6 ± 1.1	2.7 ± 0.8
BMD total hip T-score HUNT3###	11435	0.1 ± 0.9	- 0.1 ± 0.9	0.2 ± 0.9	11281	0.1 ± 0.9	-0.1 ± 0.9	0.2 ± 0.9

*Age at first attendance of HUNT is reported. §Measurements at last attendance of HUNT is reported. **Mean of first and second measurements in HUNT1 and mean of second and third measurement in HUNT2,3, and 4. #Friedewald equation was used to estimate LDL-C. ###Non-fasting glucose. ##~70% of participants from HUNT4 also have BMD measured in total hip which are undergoing quality control.

FEV1: Forced expiratory volume in the first second, HUNT: Trøndelag Health Study, BMI: Body mass index, SBP: Systolic blood pressure, LDL-C: Low-density lipoprotein cholesterol, SD: standard deviation

Linkage to regional and national health registries

Using the unique personal identification number given to all Norwegian citizens allows for longitudinal follow-up by linkage between HUNT data, regional and national registries and electronic health records. Norway currently has 17 national health registries (<https://helsedata.no/no/>) that are mandatory and cover the entire population (**Supplementary Table 2**). Commonly used national registries linked with HUNT include the Norwegian Cause of Death Registry (established 1951), the Cancer Registry of Norway (established 1952), the Medical Birth Registry of Norway (established 1967), and the Norwegian Prescription Database (established 2004). Another 52 national disease-specific medical quality registries hold detailed information on treatment and responses at an individual level (<https://www.kvalitetsregistre.no/registeroversikt>) (**Supplementary Table 2**). Electronic health records from the local hospitals hold International Statistical Classification of Diseases and Related Health Problems (ICD) codes back to 1987. Potential linkage to non-health related registries expands the data resource, that among others includes Statistics Norway, recording income and wealth statistics for individuals and households, and the Norwegian Armed Forces Health Registry (<https://helsedata.no/no/>). Together, the listed registries provide opportunities to integrate a breadth of data from multiple time points to obtain high quality phenotypes and related information on, for example, environmental and socioeconomic factors. Time-stamped data allows studies of disease development and progression. Some selected disease endpoints are presented in **Table 2**.

Table 2: ICD codes captured in the local hospital register from 1987 to 2021 for selected diseases and the observed case numbers in genotyped HUNT participants.*

ICD Chapter	ICD-9	ICD-10	Cases genotyped
Infectious and parasitic diseases			
COVID-19, virus identified		U07.1	39
Personal history of COVID-19		U08	112
Post COVID-19 condition		U09	8
Neoplasms			
Malignant neoplasm of colon, rectosigmoid junction, rectum, anus and anal canal	153, 154	C18, C19, C20, C21	3291
Malignant neoplasm of bronchus and lung	162	C34	2521
Malignant melanoma of skin	172	C43	2345
Malignant neoplasm of breast	174, 175	C50	2047
Malignant neoplasm of prostate	185	C61	3153
Endocrine, nutritional and metabolic diseases			
Hypothyroidism	240, 241, 242, 243, 244, 245	E00, E01, E02, E03	7250
Type 2 diabetes mellitus	250	E11	11444

Mental and behavioural disorders			
Dementia	290, 294, 331	F00, F01, F02, F03, G30, G31.1	4509
Mood (affective) disorders	296, 298, 300, 301, 311	F30, F31, F32, F33, F34, F38, F39	9267
Diseases of the nervous system			
Parkinson's disease	332, 333	G20, G21, G22, F02.3	5124
Epilepsy	345	G40	11480
Migraine	346	G43	1728
Diseases of the eye and adnexa			
Glaucoma	365	H40	4223
Diseases of the circulatory system			
Essential (primary) hypertension	401	I10	18263
Angina pectoris	413.9	I20	8388
Acute myocardial infarction	410	I21	10144
Atrial fibrillation and flutter	427	I48	14538
Heart failure	428	I50, I09, I11	7273
Intracerebral hemorrhage	431, 432	I61	2901
Aortic aneurysm and dissection	441	I71	2491
Diseases of the respiratory system			
Chronic obstructive pulmonary disease	496	J44.8, J44.9	6529
Asthma	493	J45	6511
Post-inflammatory pulmonary fibrosis	515	J84.1, J84.8	954
Diseases of the digestive system			
Crohn's disease	555	K50	586
Ulcerative colitis	556	K51	2288
Coeliac disease	579	K90.0	1152
Diseases of the skin and subcutaneous tissue			
Atopic dermatitis	691.8	L20	1067
Psoriasis	696	L40	3135
Diseases of the musculoskeletal system and connective tissue			
Gout	274	M10	6178
Ankylosing spondylitis	720	M45	2064
Diseases of the genitourinary system			

Chronic kidney disease	585	N18	4668
Pregnancy, childbirth and the puerperium			
Gestational hypertension	342	O13	1666
Prescription data [#]			
Low dose aspirin	-	-	22500
Statin	-	-	2200

*Numbers are from a data query (August 8, 2021) from the Nord-Trøndelag Hospital Trust including Namsos and Levanger Hospitals of participants selected for genotyping from HUNT2-4. The register is ongoing and therefore the number of cases continue to increase over time.

[#]Approximate numbers from the prescription register and restricted to those genotyped in HUNT2-3 only.

HUNT: Trøndelag Health Study, ICD: International Statistical Classification of Diseases and Related Health Problems

Analytical approaches with related samples

97% of HUNT participants are of Norwegian ancestry⁴. Using principal components of ancestry projected onto the Human Genome Diversity Project, we typically exclude samples of non-European ancestry (<2%) (**Supplementary Figure 2**) due to limited power. We have observed fine-scale differences between North- and South-Trøndelag and between individuals born closer to the coast versus the border with Sweden¹². Additionally, because of high ascertainment from a single county in Norway (Trøndelag), there are many related individuals within the cohort. 79 551 (89%) out of 88 615 HUNT2-4 participants have at least one 2nd degree or closer relative who also participates in HUNT (**Supplementary Figure 3, Supplementary Table 3**). High degree of participant relatedness in the dataset on one hand allows for unique data-analysis methods using nuclear or extended families, but can result in bias when using methods that assume unrelated individuals or power loss if related individuals are excluded. An early effort to use extended families and genetic data in HUNT was for the analysis of rare coding variants¹³, where family samples can provide more power to detect associations when sample sizes were limited and only a modest fraction of all trait-associated variants were identified¹³.

Previously, methods had been developed to account for relatedness for analysis of quantitative traits¹⁴, but methods to properly account for relatedness and control for unbalanced case-control ratios for binary traits were lacking. We therefore developed statistical methods to allow for the analysis of all individuals, and to control for case-control imbalance of binary phenotypes, which is commonly observed in biobanks such as HUNT. These methods, which are computationally efficient in biobank-scale data, allowed us to perform association testing in HUNT for both single variants (using SAIGE) and gene-based burden tests (using SAIGE-GENE), while accounting for sample relatedness with a sparse identical by state (IBS) sharing matrix^{13, 15-17}. These methods account for case-control imbalance of binary phenotypes, typical in a population-based sample, by using the saddlepoint approximation to calibrate unbalanced case-control ratios in score tests

based on logistic mixed models [Add SAIGE reference]. We demonstrated a vast improvement in reducing type I error rates when analyzing unbalanced case control ratios with SAIGE in HUNT. For example, venous thromboembolism with 2 325 cases and 65 294 controls and a case:control ratio of 0.036 had substantial inflation of type I error with methods available prior to the development of SAIGE (**Supplementary Figure 4**). To demonstrate the application of SAIGE-GENE, we investigated 13 416 genes, with at least two rare ($MAF \leq 1\%$) missense and/or stop-gain variants that were directly genotyped or imputed from the joint HRC and HUNT-WGS reference panel among 69 716 Norwegian samples from HUNT2-3 with measured high-density lipoprotein. We identified eight genes with p-values below the exome-wide significance threshold ($P \leq 2.5 \times 10^{-6}$), seven of which remained significant after conditioning on nearby single-variant associations, suggesting independent rare coding variants within these genes¹⁶. Importantly, using SAIGE and SAIGE-GENE, we were able to use all samples, account for sample relatedness case-control imbalance, and maintain well controlled type I error rates.

A traditional way of using related samples is linkage analysis, which however has computational challenges in the era of whole-genome genetics. To allow for linkage testing in datasets with millions of genetic markers, faster and computationally scalable linkage analysis method have been developed, Population Linkage¹⁸. Population Linkage uses a Haseman-Elston regression (originally used for sibling pair linkage analysis) to estimate variance components from pairwise relationships and IBD estimates. Using HUNT data, Zajac et. al. observed 25 significant linkage peaks with $LOD > 3$ across 19 distinct loci for the four traits (high-density lipoprotein, low-density lipoprotein, total cholesterol and triglycerides), where 5 peaks with $LOD > 3$ were not replicated at genome-wide significance in a genome-wide association study (GWAS) of 359,432 genotyped variants in HUNT. However, after imputing the dataset with the HRC and HUNT-WGS reference panel to cover more variants or meta-analysis in GLGC, significant associations in all 5 linkage peaks were observed. This study demonstrates one of the benefits of linkage analysis over GWAS, that is the ability to test for linkage in regions that are difficult to genotype such as rare variants, structural variants, copy number variants or variants in highly repetitive regions, as long as identical by descent segments in the region can be identified¹⁸. Finally, linkage analysis may improve statistical power when investigating rare risk variants which segregate within families and reduce confounding effects of population stratification.

The high degree of relatedness in the HUNT Study participants has enabled analysis methods are tailored to this study design. These include GWAS by proxy^{19,20}, in which the phenotypes of non-genotyped family members of genotyped HUNT participants can be used to identify proxy-cases, individuals with a proportion (0.5 for first degree relatives) of the genetic risk of cases. These proxy-cases can be appropriately modelled to increase statistical power in GWAS. For example, the power to detect an allele with an odds ratio of 1.1 and MAF of 0.21 at an alpha of 5×10^{-8} increases from 0.419 to 0.644 when proxy-cases are appropriately modelled instead of used in controls as in standard GWAS (**Supplementary Figure 5**).

Genetic discoveries from HUNT

The wealth of phenotypic and genetic data available in the HUNT cohort has led to the discovery of many new genetic associations across a broad range of traits (**Table 3**). Early genetic studies of HUNT participants used exome arrays and focused on cardiovascular disease. We identified a novel coding variant in *TM6SF2* associated with total cholesterol, MI, and liver enzymes⁶ and replicated known MI associations at the 9p21 locus and a low-frequency missense variant in the *LPA* gene (p.Ile1891Met)⁷. Following the genotyping of nearly 70 000 participants in HUNT2 and HUNT3 and the development of a combined HRC and HUNT-WGS imputation reference panel, we extended our analyses to a genome-wide search (**Figure 5**). Through imputation of indels called from low-pass HUNT-WGS, we discovered a rare mutation in the *MEPE* gene, enriched in the Norwegian population (0.8% in HUNT, 0.1% in non-Finnish Europeans), that was associated with low forearm bone mineral density and increased risk of osteoporosis and fractures²¹. Although this region had been previously identified as associated with bone mineral density²², the association in HUNT with replication in the UK biobank²³ pin-pointed *MEPE* as the likely causal gene in the region by identifying an insertion/deletion polymorphism that likely resulted in a loss-of-function protein. In another study we paid special attention to loss-of-function mutations associated with favorable blood lipid profiles (reduced LDL cholesterol and reduced CAD risk) which were not associated with altered liver enzymes or liver damage. We additionally found one elderly individual with homozygous *ZNF529* loss-of-function variant showing no signs of cardiovascular disease or diabetes, suggesting that the full knock-out of this gene is viable. This highlighted *ZNF529* as a potential therapeutic target for lipids²⁴ identified from sequencing and custom content genotyping.

On top of the association studies performed using HUNT data only, we have contributed to many international consortium efforts aimed at aggregating GWAS data across cohorts. By performing GWAS meta-analyses that included HUNT and other cohorts, efforts driven by our research team have identified genetic variants associated with atrial fibrillation which may act through a mechanism of impaired muscle cell differentiation and tissue formation during fetal heart development²⁵ and cardiac structural remodeling²⁶; variants associated with estimated glomerular filtration rate exhibiting a sex-specific effect^{27, 28}; and variants associated with thyroid stimulating hormone that revealed an inverse relationship between TSH levels and thyroid cancer²⁹. Later studies using the TOPMed reference panel³⁰ identified variants associated with circulating cardiac troponin I level and investigated its role as a non-causal biomarker for MI using Mendelian randomization³¹, and identified variants associated with iron-related biomarker levels and explored their relationship with all-cause mortality³².

Table 3: Genetic discoveries across HUNT genotyping and analysis strategies.

Strategy	Frequency range/ Number of variants	Benefits	Exemplary papers
Genotyping with custom exome content designed from HUNT sequenced samples (UM HUNT Biobank Array)	Rare-Common/ 80,137-358,964	Identify low-frequency variants not amenable to imputation-based approaches	Identified likely causal gene, <i>TM6SF2</i> , associated with TC and MI ³³
			Found LOF variant in <i>ZNF529</i> that leads to lower LDL-C ²⁴
HRC and HUNT-WGS Imputation from Human CoreExome Array	Low-Common/ 22 million	Include population-specific variants through improved imputation	Population-enriched variant in <i>MEPE</i> pinpoints causal gene for fracture risk ²¹
			Identified variants associated with thyroid function ²⁹ , kidney function ²⁷ serum PCSK9 ³⁴ and atrial fibrillation ^{25, 26}
TOPMed Imputation from Human CoreExome Array	Low-Common/ 25 million	Expand number of available variants for association testing	Identified variants associated with troponin and serum iron in the general population ^{31, 32}
Family-based design, > 15,000 sibling pairs, >35,000 parent-offspring	Any/ 10-1000	Improve effect size estimates and test traits among unstudied relatives	Introduced new analysis methods, including SAIGE ³⁵ , within-family Mendelian randomization and GWAS ^{36, 37}
Mendelian Randomization	Any/ 10-1000	Identify causal links between environmental factors (genetically determined traits) and outcomes	Explored the role of lipids and apolipoproteins on kidney-function ^{38, 39}
			Demonstrated an inverse association between thyroid stimulating hormone and thyroid cancer ²⁹

Note: Rare variants <1% minor allele frequency (MAF), Low frequency variants 1-5% MAF, Common variants >5% MAF

HUNT: Trøndelag Health Study, HRC: Haplotype Reference Consortium, LDL-C: Low density lipoprotein cholesterol, LOF: Loss-of-function, MI: Myocardial infarction, TC: Total cholesterol, TOPMed: Trans-Omics for Precision Medicine, UM: University of Michigan, WGS: Whole-genome sequencing

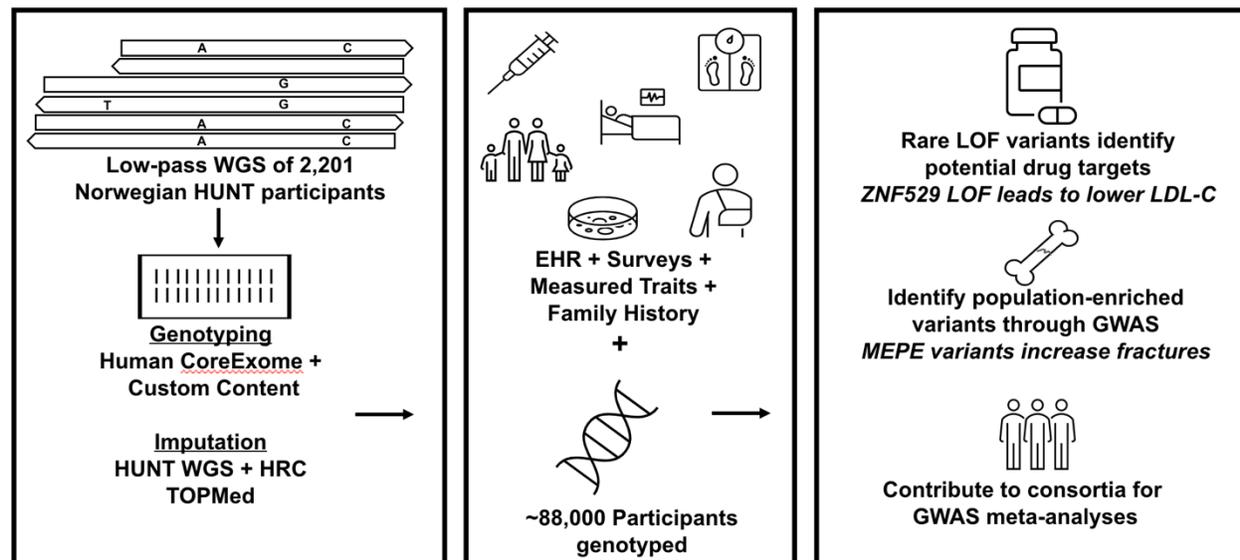


Figure 5: Overview of HUNT genetics, phenotypic resources and discoveries.

EHR: Electronic health records, GWAS: Genome-wide association study, HUNT: Trøndelag Health Study, WGS: Whole genome sequencing, HRC: Haplotype Reference Consortium, LDL-C: Low density lipoprotein cholesterol, LOF: Loss-of-function, TOPMed: Trans-Omics for Precision Medicine

Causal inference and family effects

The high degree of relatedness in the HUNT Study offers a unique opportunity to use family-based designs to investigate causal associations. Mendelian randomization (MR), which uses genetic variants as instrumental variables to investigate modifiable (non-genetic) factors, was first proposed using parent-offspring⁴⁰. Alleles that are inherited from each parent are randomly determined during the meiotic process. This random allocation is essential to providing reliable comparisons in MR studies. However, due to the lack of genotyped family data, previous studies applied MR on the population-level, where the random allocation of alleles is only approximate. We were able to use the ~15 000 families in HUNT to perform MR as originally proposed - in family-based designs³⁶. Using this approach in HUNT, we showed empirically that MR estimates from samples of unrelated individuals for the association of taller height and lower BMI increase educational attainment, were likely induced by population structure, assortative mating or dynastic effects. We observed no clear associations in within-family MR analyses in HUNT or in a replication cohort of 222 368 siblings from 23andMe³⁶. This approach has since grown in popularity and, together with HUNT, many cohorts now contribute to the investigation of causal associations with family-based designs³⁷.

Further leveraging the family structure information in HUNT, we have performed and have future opportunities to investigate causal effects between family members, for example parent-offspring effects^{41, 42}, assortative mating and sibling effects⁴³. These study designs have not been previously possible due to the lack

of genotyped family data and this has both limited causal inference (as mentioned above) and the ability of typical GWAS to distinguish between direct and indirect genetic effects³⁷. HUNT data allows for study designs to disentangle these sources of genotype-phenotype associations in humans. In one such example, we used 26 057 mother–offspring and 9 792 father–offspring pairs to investigate whether adverse environmental factors *in utero* increased future risk of cardiometabolic disease in the offspring. We observed that adverse maternal intrauterine environment, as proxied by maternal SNPs that influence offspring birthweight, were unlikely to be a major determinant of late-life cardiometabolic outcomes of the offspring⁴¹.

Contribution to collaborative studies

We contribute to genetic studies worldwide through participation in consortia focused on a variety of diseases including cardiovascular disease^{44, 45}, lipids^{46, 47}, type 2 diabetes⁴⁸, osteoporosis⁴⁹, decline in kidney function⁵⁰, Alzheimer’s disease⁵¹, bipolar disease⁵², intracranial aneurysms⁵³, insomnia⁵⁴, respiratory health⁵⁵ and sleepiness⁵⁶. We also contributed HUNT data to studies of anthropometric traits⁵⁷, alcohol and nicotine use^{58, 59}, COVID-19⁶⁰, phenome-wide discovery⁶¹, and genetic risk prediction⁶², among others. We believe that team science by consortia⁶¹ fulfills the goals of the HUNT study and moves the science fastest towards new discoveries and improved human health.

Summary

Together, the multifaceted genetic discovery strategy incorporating genotyping, sequencing, and imputation-based approaches in HUNT has aided the identification of likely causal genes and variants for disease and human traits. It has also proved to be a valuable resource for genetically informed methods of causal inference, supporting the identification of modifiable risk factors. We owe this success to the willingness and high participation rates of the people of Trøndelag, the vast phenotyping collected by decades of HUNT researchers, and access to digitized public health care systems. We anticipate that the rich data collection will continue to be a unique dataset for future opportunities in longitudinal and family-based designs, genetic discoveries, Mendelian randomization, meta-analysis and polygenic score validation, well into the future.

HUNT Data Access

Researchers associated with Norwegian research institutes can apply for the use of HUNT data and samples with approval by the Regional Committee for Medical and Health Research Ethics. Researchers from other countries may apply if collaborating with a Norwegian Principal Investigator. Information for data access can be found at <https://www.ntnu.edu/hunt/data>. The HUNT variables are available for browsing on the HUNT databank at <https://hunt-db.medisin.ntnu.no/hunt-db/>. Use of the full genetic data set requires the use of an approved secure computing solution such as the HUNT Cloud (<https://docs.hdc.ntnu.no/>). Data linkages between HUNT and health registries require that the principal investigator has obtained project-specific

approval for such linkage from the Regional Committee for Medical and Health Research Ethics, Norway and each registry owner. GWAS summary statistics from publications including HUNT are available from NTNU Open Research Data (<https://dataverse.no/dataverse/root>) and the Willer lab (<http://csg.sph.umich.edu/willer/public/>).

Ethics

The genotyping in HUNT and work presented in this cohort profile was approved by the Regional Committee for Ethics in Medical Research, Central Norway (2014/144, 2018/1622, 152023). All participants signed informed consent for participation and the use of data in research.

Acknowledgement

A special thanks to all HUNT participants for donating their time, samples and information to help others.

The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was financed by the National Institutes of Health; University of Michigan; the Research Council of Norway; the Liaison Committee for Education, Research and Innovation in Central Norway; and the Joint Research Committee between St Olavs hospital and the Faculty of Medicine and Health Sciences, NTNU. The genetic investigations of the HUNT Study is a collaboration between researchers from the K.G. Jebsen Center for Genetic Epidemiology, NTNU and the University of Michigan Medical School and the University of Michigan School of Public Health. The K.G. Jebsen Center for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU, Norway.

We want to thank clinicians and other employees at Nord-Trøndelag Hospital Trust for their support and for contributing to data collection in this research project.

Thanks to Trøndelag County Council, Jon Olav Sliper for creating Figure 1, K.G. Jebsen Center Communications Officer Janne Tellefsen for creating Figure 2 and Willer Lab Research Coordinator Bethany Klunder for organizing our collaborative meetings.

The authors thank Robin Walters and Mark Daly for an internal review of this manuscript prior to submission.

We also acknowledge; **HUNT-MI Leadership:** Kristian Hveem, Cristen J. Willer, Oddgeir L. Holmen, Michael Boehnke, Gonçalo R. Abecasis, Bjorn Olav Åsvold, Ben M. Brumpton; **Scientific Advisory Committee:** Ele Zeggini, Mark Daly, Bjørn Pasternak; **HUNT Research Centre:** Jørn Sørberg Fenstad, Anne Jorunn Vikdal,

Marit Næss; **HUNT Cloud:** Oddgeir L. Holmen, Sandor Zeestraten, Tom Erik Røberg; **Data applications and registry linkages:** Maiken E. Gabrielsen, Anne Heidi Skogholt; **Low-pass whole sequencing genome bioinformatics and statistical analysis:** He Zhang, Hyun Min Kang, Jin Chen; **Array genotyping:** Sten Even Erlandsen, Vidar Beisvåg; **GWAS bioinformatics, QC, imputation and statistical analysis:** Wei Zhou, Jonas Nielsen, Lars G. Fritsche, Hyun Min Kang, Oddgeir L. Holmen, Laurent Thomas and Ben M. Brumpton; **CNV calling:** Ellen Schmidt, Ryan Mills; **Statistical methods development for analyzing HUNT data:** Wei Zhou, Shawn Lee, Hyun Min Kang; **Communications management:** Bethany Klunder.

Author Contributions

K.H., C.J.W and B.M.B wrote the first draft of the manuscript. W.Z., J.N., L.G.F., H.M.K., O.L.H., L.T and B.M.B contributed to analysis. K.H, C.JW., G.R.A., O.L.H and B.M.B contributed to funding acquisition. K.H., C.J.W., O.L.H., M.B., G.R.A, B.O.Å and B.M.B contributed to project administration. L.T., E.C., S.G., B.W. and B.M.B. contributed to visualization. All authors reviewed and edited the paper.

Declaration of interests

G.R.A. works for Regeneron Pharmaceuticals. C.J.W.'s spouse works for Regeneron Pharmaceuticals. The remaining authors have no competing interests to declare.

References

1. Holmen J, Midthjell K, Forsén L, Skjerve K, Gorseth M and Oseland A. [A health survey in Nord-Trøndelag 1984-86. Participation and comparison of attendants and non-attendants]. *Tidsskr Nor Laegeforen*. 1990;110:1973-7.
2. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, Bratberg G, Heggland J and Holmen J. Cohort Profile: the HUNT Study, Norway. *Int J Epidemiol*. 2013;42:968-77.
3. Åsvold BO, Langhammer A, Rehn TA, Kjelvik G, Grøntvedt TV, Sørgerd EP, et al. Cohort Profile Update: The HUNT Study, Norway. *medRxiv*. 2021:2021.10.12.21264858.
4. Holmen J, Midthjell K, Krüger Ø, Langhammer A, Holmen TL, Bratberg GH, Vatten L and Lund-Larsen PG. The Nord-Trøndelag Health Study 1995–97 (HUNT 2): objectives, contents, methods and participation. *Norsk epidemiologi*. 2003;13:19-32.
5. Holmen TL, Bratberg G, Krokstad S, Langhammer A, Hveem K, Midthjell K, Heggland J and Holmen J. Cohort profile of the Young-HUNT Study, Norway: a population-based study of adolescents. *Int J Epidemiol*. 2014;43:536-44.
6. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nature genetics*. 2014;46:345.
7. Holmen OL, Zhang H, Zhou W, Schmidt E, Hovelson DH, Langhammer A, et al. No large-effect low-frequency coding variation found for myocardial infarction. *Hum Mol Genet*. 2014;23:4721-8.
8. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. Meta-analysis of gene-level tests for rare variant association. *Nat Genet*. 2014;46:200-4.
9. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. 2016;48:1279.
10. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279-83.
11. Zhou W, Fritsche LG, Das S, Zhang H, Nielsen JB, Holmen OL, et al. Improving power of association tests using multiple sets of imputed genotypes from distributed reference panels. *Genet Epidemiol*. 2017;41:744-755.
12. Gabrielsen ME. Genetic Risk Factors for Lung Cancer: Relationship to Smoking Habits and Nicotine Addiction: The Nord-Trøndelag (HUNT) and Tromsø Health Studies. 2013.
13. Feng S, Pistis G, Zhang H, Zawistowski M, Mulas A, Zoledziwska M, et al. Methods for association analysis and meta-analysis of rare variants in families. *Genet Epidemiol*. 2015;39:227-38.
14. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284-90.

15. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*. 2018;50:1335-1341.
16. Zhou W, Zhao Z, Nielsen JB, Fritsche LG, LeFaive J, Gagliano Taliun SA, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat Genet*. 2020;52:634-639.
17. Zhou W, Bi W, Zhao Z, Dey KK, Jagadeesh KA, Karczewski KJ, Daly MJ, Neale BM and Lee S. Set-based rare variant association tests for biobank scale sequencing data sets. *medRxiv*. 2021:2021.07.12.21260400.
18. Zajac G. Methods and Applications for Collection, Contamination Estimation, and Linkage Analysis of Large-scale Human Genotype Data. 2020.
19. Liu JZ, Erlich Y and Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nat Genet*. 2017;49:325-331.
20. Zhuang Y, Wolford BN, Nam K, Bi W, Zhou W, Willer CJ, Mukherjee B and Lee S. Incorporating family disease history and controlling case-control imbalance for population based genetic association studies. *medRxiv*. 2021:2021.07.04.21259997.
21. Surakka I, Fritsche LG, Zhou W, Backman J, Kosmicki JA, Lu H, et al. MEPE loss-of-function variant associates with decreased bone mineral density and increased fracture risk. *Nat Commun*. 2020;11:4093.
22. Rivadeneira F, Styrkársdóttir U, Estrada K, Halldórsson BV, Hsu YH, Richards JB, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet*. 2009;41:1199-206.
23. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*. 2020;586:749-756.
24. Nielsen JB, Rom O, Surakka I, Graham SE, Zhou W, Roychowdhury T, et al. Loss-of-function genomic variants highlight potential therapeutic targets for cardiovascular disease. *Nat Commun*. 2020;11:6417.
25. Nielsen JB, Fritsche LG, Zhou W, Teslovich TM, Holmen OL, Gustafsson S, et al. Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development. *Am J Hum Genet*. 2018;102:103-115.
26. Nielsen JB, Thorolfsson RB, Fritsche LG, Zhou W, Skov MW, Graham SE, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet*. 2018;50:1234-1239.
27. Graham SE, Nielsen JB, Zawistowski M, Zhou W, Fritsche LG, Gabrielsen ME, et al. Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. *Nat Commun*. 2019;10:1847.
28. Stanzick KJ, Li Y, Schlosser P, Gorski M, Wuttke M, Thomas LF, et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat Commun*. 2021;12:4350.

29. Zhou W, Brumpton B, Kabil O, Gudmundsson J, Thorleifsson G, Weinstock J, et al. GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. *Nat Commun.* 2020;11:3981.
30. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590:290-299.
31. Moksnes MR, Røsjø H, Richmond A, Lyngbakken MN, Graham SE, Hansen AF, et al. Genome-wide association study of cardiac troponin I in the general population. *Hum Mol Genet.* 2021;30:2027-2039.
32. Moksnes MR, Hansen AF, Graham SE, Gagliano Taliun SA, Wu K-H, Zhou W, et al. Genome-wide meta-analysis of iron status biomarkers and the effect of iron on all-cause mortality in HUNT. *medRxiv.* 2021:2021.09.20.21262960.
33. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet.* 2014;46:345-51.
34. Brumpton BM, Fritsche LG, Zheng J, Nielsen JB, Mannila M, Surakka I, et al. Variation in Serum PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9), Cardiovascular Disease Risk, and an Investigation of Potential Unanticipated Effects of PCSK9 Inhibition. *Circ Genom Precis Med.* 2019;12:e002335.
35. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018;50:1335-1341.
36. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie G, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun.* 2020;11:3519.
37. Howe LJ, Nivard MG, Morris TT, Hansen AF, Rasheed H, Cho Y, et al. Within-sibship GWAS improve estimates of direct genetic effects. *bioRxiv.* 2021:2021.03.05.433935.
38. Rasheed H, Zheng J, Rees J, Sanderson E, Thomas L, Richardson TG, et al. The causal effects of serum lipids and apolipoproteins on kidney function: multivariable and bidirectional Mendelian-randomization analyses. *Int J Epidemiol.* 2021.
39. Zheng J, Zhang Y, Rasheed H, Walker V, Sugawara Y, Li J, et al. Trans-ethnic Mendelian-randomization study reveals causal relationships between cardiometabolic factors and chronic kidney disease. *Int J Epidemiol.* 2021.
40. Davey Smith G and Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology.* 2003;32:1-22.
41. Moen GH, Brumpton B, Willer C, Åsvold BO, Birkeland KI, Wang G, et al. Mendelian randomization study of maternal influences on birthweight and future cardiometabolic risk in the HUNT cohort. *Nat Commun.* 2020;11:5404.

42. Wang G, Bhatta L, Moen GH, Hwang LD, Kemp JP, Bond TA, et al. Investigating a Potential Causal Relationship Between Maternal Blood Pressure During Pregnancy and Future Offspring Cardiometabolic Health. *Hypertension*. 2021:HYPERTENSIONAHA12117701.
43. Young AI, Benonisdottir S, Przeworski M and Kong A. Deconstructing the sources of genotype-phenotype associations in humans. *Science*. 2019;365:1396-1400.
44. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Weeks EM, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *medRxiv*. 2021:2021.05.24.21257377.
45. Klarin D, Verma SS, Judy R, Dikilitas O, Wolford BN, Paranjpe I, et al. Genetic Architecture of Abdominal Aortic Aneurysm in the Million Veteran Program. *Circulation*. 2020;142:1633-1646.
46. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-1283.
47. Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021.
48. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44:981-90.
49. Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell*. 2021;184:4784-4818.e17.
50. Köttgen A and Pattaro C. The CKDGen Consortium: ten years of insights into the genetic basis of kidney function. *Kidney Int*. 2020;97:236-242.
51. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet*. 2021;53:1276-1282.
52. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*. 2021;53:817-829.
53. Bakker MK, van der Spek RAA, van Rheenen W, Morel S, Bourcier R, Hostettler IC, et al. Genome-wide association study of intracranial aneurysms identifies 17 risk loci and genetic overlap with clinical risk factors. *Nat Genet*. 2020;52:1303-1313.
54. Lane JM, Jones SE, Dashti HS, Wood AR, Aragam KG, van Hees VT, et al. Biological and clinical insights from genetics of insomnia symptoms. *Nat Genet*. 2019;51:387-393.
55. John C, Guyatt AL, Shrine N, Packer R, Olafsdottir TA, Liu J, et al. Genetic associations and architecture of asthma-chronic obstructive pulmonary disease overlap. *medRxiv*. 2020:2020.11.26.20236760.

56. Wang H, Lane JM, Jones SE, Dashti HS, Ollila HM, Wood AR, et al. Genome-wide association analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes. *Nat Commun.* 2019;10:3503.
57. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197-206.
58. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet.* 2019;51:237-244.
59. McGuire D, Jiang Y, Liu M, Weissenkampen JD, Eckert S, Yang L, et al. Model-based assessment of replicability for genome-wide association meta-analysis. *Nat Commun.* 2021;12:1964.
60. Initiative C-HG. Mapping the human genetic architecture of COVID-19. *Nature.* 2021.
61. Zhou W, Kanai M, Wu K-HH, Humaira R, Tsuo K, Hirbo JB, et al. Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. *medRxiv.* 2021:2021.11.19.21266436.
62. Surakka I, Wolford BN, Ritchie SC, Hornsby WE, Sutton NR, Gabrielsen ME, et al. Sex-specific survival bias and interaction modeling in coronary artery disease risk prediction. *medRxiv.* 2021:2021.06.23.21259247.