Loss-of-function genomic variants highlight potential therapeutic targets for cardiovascular disease

Jonas B. Nielsen et al.#

Pharmaceutical drugs targeting dyslipidemia and cardiovascular disease (CVD) may increase the risk of fatty liver disease and other metabolic disorders. To identify potential novel CVD drug targets without these adverse effects, we perform genome-wide analyses of participants in the HUNT Study in Norway (n = 69,479) to search for protein-altering variants with beneficial impact on quantitative blood traits related to cardiovascular disease, but without detrimental impact on liver function. We identify 76 (11 previously unreported) presumed causal protein-altering variants associated with one or more CVD- or liver-related blood traits. Nine of the variants are predicted to result in loss-of-function of the protein. This includes ZNF529:p.K405X, which is associated with decreased low-density-lipoprotein (LDL) cholesterol (P = 1.3 × 10^{-8}) without being associated with liver enzymes or non-fasting blood glucose. Silencing of ZNF529 in human hepatoma cells results in upregulation of LDL receptor and increased LDL uptake in the cells. This suggests that inhibition of ZNF529 or its gene product should be prioritized as a novel candidate drug target for treating dyslipidemia and associated CVD.

#A list of authors and their affiliations appears at the end of the paper.
Cardiovascular disease (CVD) – in particular cerebrovascular and ischemic heart diseases – is the leading cause of death globally. Lowering circulating lipids is an important treatment strategy to reduce risk. However, some pharmaceutical mechanisms of lipid lowering and CVD risk reduction may unfortunately increase risk of fatty liver disease or other metabolic disorders.

The vast majority of novel candidate drugs that enter clinical testing fail to demonstrate sufficient safety and efficacy to gain regulatory approval. This is largely due to poor predictive value of preclinical models of disease along with a lack of knowledge about the long-term consequences of targeting specific biological processes in humans. It has been estimated, however, that drugs with genetic support of efficacy are twice as likely to have success in clinical testing.

We aim to identify novel candidate pharmaceutical strategies for CVD risk reduction that, importantly, are unlikely to increase the risk of liver disease, diabetes, or other metabolic disorders. To attain this, we conduct a large data-driven genomic discovery effort to identify presumed causal protein-altering variants with impact on lipids and other liver-related blood traits. In particular, we are interested in identifying presumed causal protein-altering variants associated with a more favorable lipid profile without being associated with elevated liver enzymes or vice versa.

We analyze 9 liver-related blood traits in close to 70,000 participants in the Trøndelag Health (HUNT) Study. The HUNT Study is a large population-based health survey conducted in a geographically confined region in Norway. The examined traits are related to: (i) blood lipid levels which impact cardiovascular, neurological and eye-related diseases: total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and triglyceride (TG) levels; (ii) C-reactive protein (CRP; only values <15 mg/L were included) which is predictive of cardiovascular disease; (iii) enzymes which primarily reflect liver function: alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP) and gamma-glutamyltransferase (GGT).

To maximize chances of discovery of presumed causal protein-altering variants associated with the 9 blood traits, we combine a number of genomic approaches, including: (i) low-coverage (5x) whole-genome sequencing (WGS) of a subsample of HUNT Study participants (N = 2202) to identify region-specific rare variants, (ii) targeted genotyping, also including rare region-specific variants identified by WGS, (iii) deep genotype imputation based on the TOPMed multi-ethnic reference panel consisting of 60,039 deeply sequenced genomes, (iv) genome-wide association analyses (GWAS) in up to 69,479 HUNT Study participants (ranging from 21,528 for AST to 69,479 for TG), followed by (v) stepwise conditional analyses and, for the purpose of further fine-mapping loci identified in HUNT, we perform (vi) trans-ancestry meta-analyses in up to 203,476 people of Norwegian, Japanese, and Sardinian ancestry (ranging from 128,794 for CRP to 203,476 for TC) (see Supplementary Fig. 1 for a study design overview).

We assume any protein-altering variant to be likely causally related to the trait of interest if the variant (i) was the most statistically significant variant (lowest P) in a genomic region (i.e., the locus index variant) in any of the GWAS or (ii) if the variant was independently associated with the trait of interest in GWAS stepwise conditional analyses.

Results
Genomic discovery of presumed causal protein-altering variants. We imputed 26 million genomic variants with sufficient quality and at least 10 minor allele copies into 69,479 participants in the HUNT Study (Supplementary Data 1). Using a linear mixed model to account for relatedness among study participants, we tested for genome-wide association (P < 5 × 10−8) with 9 liver-related blood traits. We identified 201 genomic regions (i.e., loci) associated with one or more of the traits. At 24 of the 201 loci, the locus index variant alters (n = 22) or results in loss-of-function (LoF) (n = 2) of the protein. We consider these 24 variants as presumed causally related to the trait of interest (Supplementary Fig. 2 and Supplementary Data 2). Stepwise conditional analyses resulted in identification of an additional 150 independently associated variants within the 201 loci. These include 28 additional protein-altering variants, hereof 2 LoF variants, which are significantly and independently associated with one or more of the liver-related blood traits (Supplementary Data 3).

For the purpose of further fine-mapping of loci and identification of additional presumed causal protein-altering variants, we performed trans-ancestry meta-analyses by combining summary statistics based on the primary discovery effort in HUNT with additional GWAS statistics from Sardinia (Sardinia cohort) and Japan (Biobank Japan). The combined meta-analyses comprised up to 203,476 participants (N range 128,794 for CRP to 203,476 for TC) and 31.5 million unique variants (n range 24.7–31.5 million) (Supplementary Fig. 3). The analyses resulted in identification of an additional 86 loci and 351 independent variants, including 13 presumed causal protein-altering variants. One of the protein-altering variants comprise a previously reported LoF variant in HBB (p.Q40X) associated with decreased TC (Table 1 and Supplementary Data 4).

To identify rare HUNT-specific presumed causal protein-altering variants, we performed genome-wide association testing for the 9 liver-related blood traits in up to 57,060 HUNT Study participants (N range 15,520 for AST to 57,060 for TG) based on directly genotyped variants that were included as custom content on the array and not part of the primary GWAS (see Materials and Methods for details). In brief, variants were selected for custom content genotyping if they were (i) identified by HUNT-specific low-coverage (5x) WGS (N = 2202 HUNT Study participants), (ii) observed in Norwegian clinics for familial hypercholesterolemia, or (iii) not-previously-observed variants predicted to introduce a premature stop codon in any of 56 genes in which protein-altering variants are deemed clinically actionable by The American College of Medical Genetics and Genomics (ACMG56). This resulted in identification of an additional 11 protein-altering variants, including 4 LoF variants, which are associated with one or more of the 9 traits. Five of the 11 variants originated from region-specific WGS and all variants are rare (ranging from 1 in 178 to 1 in 6313 individuals) (Table 1 and Supplementary Data 5).

Combining all discovery strategies, we identified a total of 674 unique independent variants within 287 loci associated with at least one of the 9 quantitative liver-related blood traits. Of the 287 loci, 92 have not previously been associated with the trait of interest (Supplementary Fig. 4). We identified genome-wide significant associations with at least one trait at 76 presumed causal protein-altering variants, of which 9 result in loss-of-function (LoF) of a specific protein – 3 frameshift indels and 6 premature stop codons (Table 1 and Fig. 1). Eleven of the 76 protein-altering variants, including 1 LoF variant, do not fall within a previously reported locus with respect to the trait of interest.

Loss-of-function variants with impact on liver-related blood traits. After combining results across all samples and discovery strategies, we were particularly interested in 9 variants annotated...
| Gene variant | Chr:pos | Minor/MAF | Rs ID | Trait | N | Beta | SD | SE | P | Discovery method | Novelty of variant-trait association |
|--------------|---------|-----------|-------|-------|---|------|----|----|---|----------------|-----------------------------------|
| APOB:–2:21011431 A/T | – | – | – | LDL- | – | – | – | – | – | – | – |
| K1813X | 15 Custom array | – | – | LDL- | predicted | – | – | – | – | – | Novel (in known LDL-C locus; known gene) |
| R1333X | 53 familial hypobetalipoproteinemia | – | – | LDL- | predicted | – | – | – | – | – | Novel (in known LDL-C locus; known gene) |
| W3087X | 107 HUNT locus index variant | C | 107 | ALP | 48,578 | −0.87 | 0.039 | 2.2 × 10⁻¹⁰ | – | Novel (in known ALP locus; gene experimentally linked to liver disease) |
| V815Sfs| 46 (imputed from TOPMed) | HBB | p.Q40X | LDL- | – | – | – | – | – | Known (previously associated with beta thalassemia, which again is known) |
| G247Afs| 116 Trans-ancestry meta-analysis | LIPC | p.S474X | HDL- | – | – | – | – | – | Known variant at known locus, experimentally shown to result in gain-of-function of LPL gene, known to result in gain-of-function of LIPC and ZNF529 and gain-of-function alleles in LIPC and ZNF529 do not cause liver damage, prioritizing these genes as potential drug targets to reduce blood lipids and CVD without liver-damaging side effects. |

Functional characterization of ZNF529:p.K405X. We expect that protein-altering variants that are the most strongly associated variants in a region represent functional variants that pinpoint biologically relevant genes and potential drug targets. We also sought not previously described genes that decreased cardiovascular risk factors (such as LDL-C) without increasing risk of liver disease or impact liver enzymes. Thus, we focused on the not previously reported association between ZNF529:p.K405X and LDL-C (beta = 0.6, P = 1.3 × 10⁻⁸) since this variant is neither associated with liver enzymes (P = 0.4–0.9 for all 4 liver enzyme traits, N range 21,530–48,569) nor non-fasting blood glucose (P = 0.93, N = 54,093 individuals) in HUNT.

Zinc finger 529 (ZNF529) does not have a homolog in rodents. To experimentally assess the consequence of ZNF529 LoF on cholesterol metabolism, we transiently knocked-down ZNF529 in human hepatoma HepG2 cells using siRNA (90.1% reduction, P = 2.7 × 10⁻⁸, Fig. 3a, Supplementary Data 6) and conducted an unbiased analysis of gene expression using RNA sequencing. Principal component analysis revealed a distinct gene expression pattern in HepG2 cells following ZNF529 knockdown with a total of 476 differentially expressed genes identified (Supplementary Fig. 6), including a significant upregulation of the LDL receptor (LDLR, FDR = 7.8 × 10⁻⁷) (Supplementary Data 7). Pathway analysis revealed enrichment of general metabolism pathways, drug metabolism pathways, and lipid-related pathways (statin pathway, plasma lipoprotein remodeling and plasma lipoprotein assembly, remodeling, and clearance, P = 6.5 × 10⁻⁴, P = 2.2 × 10⁻², and P = 3.9 × 10⁻², respectively) (Supplementary Data 8). We confirmed the upregulation of LDLR mRNA by qPCR (90.6% increase, P = 2.9 × 10⁻⁸, Fig. 3b, Supplementary Data 9) and protein by western blot (83.0% increase, P = 0.001, Fig. 3c, d, Supplementary Data 10).

We used 1,1'-dioctadecyl-3,3',3'-tetramethylindocarbocyanine perchlorate (DiI)-labeled LDL to assess the effects of ZNF529 LoF on LDL uptake in HepG2 cells. First, we confirmed that DiI-LDL is taken up by the cells in a dose-dependent manner, resulting in saturated uptake at 25 µg/ml (Supplementary Fig. 7). Next, we assessed the specificity of the binding in competition experiments
between Dil-labeled and unlabeled LDL. Pretreatment of cells with 25-fold excess amounts of unlabeled LDL inhibited the uptake of Dil-LDL (Fig. 3e). Additionally, we noted a 2.2-fold increase in intracellular cholesterol following ZNF529 knockdown which was suppressed in the presence of excess amounts of unlabeled LDL (Fig. 3e, Supplementary Fig. 8). After con

**Clinical implications of ZNF529:p.K405X.** Individuals heterozygous for ZNF529:p.K405X ($N = 109$, minor allele frequency of 0.1%) had a mean LDL-C level of 2.58 mmol/L (99.8 mg/dL) vs. 3.44 mmol/L (133.0 mg/dL) in non-carriers. This reduction in LDL-C of 25% in heterozygous carriers is in the range of what is seen for treatment with 40 mg of statin$^{20}$, and usually correlates to a relative risk reduction of major cardiovascular events by 20–25%$^{2}$. We only observed one homozygous female carrier. Despite being obese and hypertensive, she was alive at age >90 years and had no diagnosis of cardiovascular disease, liver disease, or diabetes, and had an LDL-C level slightly below average for her age group (3.45 mmol/L [133.4 mg/dL] vs. mean 3.8 mmol/L [147.0 mg/dL] for women >90 years old). This one individual with a natural absence of both copies of ZNF529 suggests that homozygous knockout of this gene is compatible with survival. We sought to replicate ZNF529:p.K405X outside the HUNT Study, but found that the allele count was too low for meaningful association analyses (e.g., 1 copy in 26,638 alleles in the Michigan Genomics Initiative [MGI] and 1 copy in 125,568 alleles in TOPMed). The existence of ZNF529:p.K405X in HUNT was confirmed by Sanger sequencing of the homozygous sample and 9 heterozygous samples (100% match between genotyping and Sanger sequencing) (Supplementary Fig. 9).

**Highlight of protein-altering variants with high effect size.** We highlight 17 protein-altering variants with an impact >1 standard deviation on the trait (Fig. 1 and Supplementary Data 12). For lipids, protein-altering variants in APOB, LDLR, and PCSK9 that impact LDL-C, and in CETP that impact HDL-C, are well known$^{17}$. However, for liver enzyme traits, TNK1:p.G574V is a new finding to complement genes previously known to impact or encode liver enzymes including ALPL (with ALP)$^{21}$, GPLD1 (with ALP), and GPT (with ALT)$^{22}$. This rare TNK1 variant, present in 46 individuals (1 in 814 individuals), was first identified in Norwegian sequenced samples, then genotyped using the custom array, and observed to have a large impact on ALP (beta = 1.2, $P = 1.7 \times 10^{-14}$).

**Gene-based burden tests.** Gene-based burden results are, in contrast to single variant tests, independent of nearby signals and may point to the functional gene in a region. To identify genes functionally involved in the 9 liver-related blood traits, we performed gene-based burden tests using SKAT-O as implemented in SAIGE-GENE$^{23}$. We included all protein-altering variants with frequency below 0.5% in the HUNT dataset. Although we found 33 unique genes to be significantly associated ($P < 2 \times 10^{-7}$) with
at least one of 9 liver traits (Supplementary Data 13), in only two cases was the gene-based evidence for association substantially stronger than the strongest single variant. This comprised 10 variants in the gene GPT associated with ALT ($P = 2.35 \times 10^{-60}$ vs. $P = 6.43 \times 10^{-25}$ for rs147998249) and 6 variants in the gene ALPL associated with ALP ($P = 2.9 \times 10^{-239}$ vs. $P = 3.99 \times 10^{-67}$ for rs138587317). These data suggest there are multiple, functional coding rare variants in each of these two genes. The gene-based burden test also identified well-known genes such as CETP and ABCA1 associated with HDL-C, PCSK9, LDLR and APOB associated with LDL-C; the CRP gene associated with CRP; and GPLD1 associated with ALP (Supplementary Data 13). Additionally, burden tests for the 3 LoF APOB variants indicated no association with liver enzymes for heterozygous LoF carriers (Supplementary Data 14), highlighting APOB as another potentially valuable pharmaceutical target for blood lipid lowering.

Cross-trait analyses for evaluating potential consequences of gene targeting. To expand our understanding of the 76 presumed causal protein-altering variants, to investigate their impact on disease, and to evaluate potential consequences of targeting the implicated gene or its gene product, we imputed the TOPMed reference panel into the UK Biobank and performed a phenome-wide association study (PheWAS) across 1342 ICD code-defined disease groups, in 408,961 people of white British ancestry. Sixty-four of the 76 protein-altering variants could be imputed sufficiently well ($R^2 > 0.3$). Of 64 variants assessed, we found that 24 variants are associated with one or more diseases at a phenoeme-wide significance level ($P < 3.5 \times 10^{-5}$) (Fig. 4, Supplementary Fig. 10, and Supplementary Data 15).

To identify potentially useful pharmaceutical strategies that may reduce blood lipid levels and risk of coronary artery disease (CAD) and type 2 diabetes (T2D) without increasing the risk of fatty liver disease, we attempted to identify variants that decreased LDL-C or TG and decreased the risk of cardiometabolic disease, but were not associated with changes in liver enzyme levels ($P > 0.05$, Figs. 2, 4, Supplementary Fig. 10 and Supplementary Data 12–15), suggesting that liver function was not altered. The 2 variants with this pattern of association include (i) COBLL1:p.N497D which is associated with decreased TG levels and decreased risk of T2D in both HUNT and UK Biobank and of liver disease in HUNT and (ii) LPL:p.S474X which is associated with decreased risk of CAD, T2D, and hypertension. HDL-C-associated ANGPTL4:p.E40K appears to decrease risk of T2D, CAD, and hypertension but is also associated with an increased risk of ankylosing spondylitis (Supplementary Fig. 10 and Supplementary Data 15), hence a potential complication of targeting this gene.

From the PheWAS, we observed other interesting associations. For example, we found that the variant SERPINA1:p.E366K, which is known to cause alpha-1-antitrypsin deficiency, often complicated by severe liver and pulmonary disease, is also associated with a decreased risk of myocardial infarction. This finding supports clinical evidence that individuals with alpha-1-antitrypsin deficiency may be protected against coronary artery disease. The association, however, also suggests caution in ongoing efforts to treat acute myocardial infarction with...
S1PR2 variant p.Y257C (chr19:10224136T>C, MAF = 0.003) which we found to be associated with decreased LDL-C and a >50% reduction in the risk of myocardial infarction, without being associated with liver enzyme traits (Supplementary Data 12).

Further studies are obviously warranted to uncover the biological mechanisms underlying the associations described here, however, each of them could help inform clinical implications of targeting the underlying gene or gene product.

**Discussion**

By using complementary approaches for genomic association discovery: sequencing, imputation, array-based genotyping, stepwise conditional analyses and trans-ancestry meta-analyses, we identified >650 independent genomic variants associated with quantitative liver-related disease precursors. This included 76 protein-altering variants that we assume to be causally related to one or more of the traits. By broadly considering associations between these protein-altering variants, quantitative traits, and disease endpoints, we prioritize several genes as potential pharmaceutical targets for preventing or treating CVD.

The newly uncovered association and in vitro studies indicate that ZNF529 LoF is associated with lower plasma LDL-C, which could be explained by induction of LDLR in hepatic cells and increased LDL uptake. While these findings indicate a therapeutic potential for lowering plasma LDL-C by ZNF529 inhibition, further studies are warranted to elucidate the mechanisms by which ZNF529 regulates LDLR and LDL uptake in the liver. Considering that ZNF529 does not have a homolog in rodents, the use of animal models for such studies is limited.

Another interesting finding to highlight is the presumed causal low-frequency protein-altering variant S1PR2:p.Y257C that we found to be associated with decreased LDL-C and a >50% reduction in the risk of myocardial infarction, without being associated with altered liver function or non-fasting blood glucose. The variant was identified via low-pass sequencing and custom content genotyping in HUNT. S1PR2 encodes sphingosine-1-phosphate receptor 2, which seems to play a critical role in the endothelial inflammatory response. Several animal models have already indicated that inhibition of S1PR2 could be a valuable pharmaceutical target for vascular recovery in...
coronary artery disease and stroke.\textsuperscript{30–32} The associations that we report here represent the first direct human data supporting that S1PR2 might play an important role in ischemic heart disease.

Taken together, we demonstrate that identifying rare protein-altering variants and careful consideration of multiple phenotypes in well-powered studies may point to promising pharmaceutical drug targets. We used a variety of approaches to identify rare protein-altering variants, and we found that if exome sequencing in well-powered studies may point to promising pharmaceutical
development.

**Methods**

**The HUNT Study.** The Tromsø Health Study (HUNT) is a population-based health survey conducted in the county of Tromsø, Norway, since 1984.\textsuperscript{33} Participation in the HUNT Study is based on informed consent and the study has been approved by the Data Inspectorate and the Regional Ethics Committee for Medical Research in Norway (REK: 2014/144). We included a total of 69,479 individuals with values for at least one of the traits examined (ALT, ALP, AST, CRP, GGT, HDL-C, LDL-C, TC, and TG). Genotyping was performed using the Illumina HumanCoreExome v1.3 array with 70,000 additional custom content beads.\textsuperscript{33,34} Variants were selected for genotyping if they were: protein-altering (n = 3,618); modestly associated with lipids in HUNT but not tested in large consortia (n = 960); identified as causing familial hypercholesterolemia in Norwegian patients (n = 110); or predicted to result in a loss-of-function of one of the 56 ACMG genes (n = 2,144, Supplementary Data 16). Additionally, we selected missense variants with 2 or more copies (n = 8,720) and nonsense variants with 1 or more copy (n = 756) identified from low-pass whole-genome sequencing of 2,202 HUNT samples. Please see Supplementary Data 17 for a summary of selected custom content variants.

**Imputation.** Imputation was performed from 60,039 TOPMed reference genomes using Minimac3 and variants with imputation quality >0.3 were retained. To account for relatedness within the sample, we performed association testing using the linear mixed model with genetic relationship matrix as implemented in SAIGE [https://github.com/weizhouUMICH/SAIGE]. Birth year, sex, and PC1-4 were included as covariates. Conditional analysis was performed with the same analysis tools and command line options as the association analysis by adding the lead-SNP(s) in a step-wise manner as covariate(s) into the SAIGE step1 parameter estimation until the variant with smallest P value in the locus was >5 × 10\textsuperscript{-8}.

**Biobank Japan.** Biobank Japan (BJI) is a multi-institutional hospital-based registry of ~200,000 individuals from 66 Japanese hospitals collected from 2003 to 2007. Written informed consent was obtained from all participants, as approved by the ethics committees of RIKEN Center for Integrative Medical Sciences and the Institute of Medical Sciences, the University of Tokyo. Genotype, imputation, and QC were performed as described previously.\textsuperscript{13} Briefly, samples were genotyped with Illumina HumanOmniExpressExome or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips and imputed using 1000 Genomes Project Phase 1 version 3 East Asian reference haplotypes. Publicly available summary statistics from linear regression assuming an additive model for quantitative measures of ALP, ALT, AST, CRP, GGT, HDL-C, LDL-C, TC, and TG were used. Quantitative traits were adjusted for age, sex, top 10 PCs of genetic ancestry, and disease status for 47 target diseases. Sample sizes for traits ranged from 70,567 to 134,182.\textsuperscript{14}

**Sardinia.** 6,602 individuals from four villages in the Lanusei valley on Sardinia (>60% of the adult population) were genotyped on four different Illumina Infinium arrays: OmniExpress, Cardio-MetaboChip, Immunochip, and Exome Chip.
Low-depth (~4x coverage) whole-genome sequencing on 3839 individuals was performed, of which 2340 were also genotyped. Imputation of 1.1 million indels and 24,1 million biallelic single nucleotide variants was performed using Minimac37 and markers with imputation quality >0.3 (or >0.6 if MAF<1%) were retained. Samples, genotyping, sequencing and variant calling have been previously described38.

Liver (ALT, AST, CRP, GGT, HDL-C, TC, and TG) from the first visit were used and LDL-C was computed using the Friedewald Equation13. Association analyses were performed for liver traits assessed in 5570–5942 individuals (median N = 5917) using age, age2 and sex-adjusted inverse-normalized residuals of the outcomes as input to the Efficient Mixed Model Association eXpedited (EMMAX) single-variant test (https://github.com/statgen/EMMAX) with kinship matrix as implemented in the EPACTS [https://github.com/statgen/EPACTS]. Genomic control correction was not applied as the lambda values were not inflated (range 0.97–1.02). Supplementary Data 18). The study, including the protocols for subject recruitment and assessment, the informed consent for participants (or consent from their legally authorized representative for those 14–17y), and the overall analysis plan was reviewed and approved by institutional review boards for the Istituto di Ricerca Genetica e Biomedica (IRGB: Cagliari, Italy), for the MedStar Research Institute (responsible for intramural research at the National Institutes of Aging, Baltimore, Maryland, United States), and for the University of Michigan (Ann Arbor, Michigan, United States).

Meta-analyses. As the summary statistics from SardiNIA and Biobank Japan were in Human Genome Build hg19, the positions were mapped to Human Genome Build hg38 using liftOver [https://genome.ucsc.edu/cgi-bin/hgLiftOver]. The genomic control corrected summary statistics from the contributing cohorts were combined with METAL40 using inverse variance weighted meta-analysis. Meta-analysis included SardiNIA, Biobank Japan, and HUNT for all traits, with the exception of ALP which was only available from HUNT and Biobank Japan.

Definition of independent loci. Independent loci were defined as genetic markers >1 Mb apart in physical distance with at least one genetic variant associated with the trait of interest at a genome-wide significance threshold of P < 5 × 10−8. Loci borders were defined as the highest and lowest genomic positions within the locus reaching genome-wide significance plus an additional 1 Mb on either side.

Novelty of identified genomic loci. A locus was classified as known if a variant previously published to be associated with the trait of interest fell with 1Mb of the locus lead variant that we identified. Otherwise, a locus was classified as novel. Previously published variants were extracted from papers and the GWAS catalog [https://www.ebi.ac.uk/gwas/] at the time of analyses.

PheWAS in UK Biobank. Association results for 1342 trait groups (PheCodes)24 in UK Biobank were generated using SAIGE24. Phenotypes were grouped by combined ICD-9 and ICD-10 codes of closely related traits following previously published methodologies. The expression is in gene fold-change. Circulating protein levels were measured in the white British population of UK Biobank after imputation with the TOPMed reference panel. Sex, birth year, and 4 principal components were included as covariates. Significance was determined based on Bonferroni correction for the number of traits tested (P < 3.5 × 10−5). Participation in the UK Biobank is based on informed consent41.

Gene-based SKAT-O tests. The exome-wide gene-based SKAT-O tests were performed using SAIGE-GENE v3641 for all 9 liver traits based on the TOPMed-imputed HUNT data. Missense and stop-gain variants annotated by ANNOVAR42 with MAF ≤ 0.005% were included. Conditional analyses were performed to condition on the most significant single variant association signal within 500 kb of the gene. We selected a significance threshold of 2.5 × 10−8 accounting for 20,000 genes.

Replication attempt of ZNF529-p.K405X in MGI. The Michigan Genomics Initiative (MGI) is a repository of electronic medical record and genetic data at Michigan Medicine (N=58,000 participants). MGI participants were enrolled during pre-surgical encounters at Michigan Medicine and provided consent to study genetic and electronic health record data for research. The MGI study was approved by the Institutional Review Board of the University of Michigan Medical School. DNA was extracted from blood samples and participants were genotyped using Illumina Infinium CoreExome-24 bead arrays, which includes the same custom content as the HUNT Study. Genotype data were imputed to the Haplotype Reference Consortium using the Michigan Imputation Server, providing 17 million imputed variants after standard quality control and filtering. Only European individuals were used for analysis. We attempted to replicate the association with ZNF529-p.K405X in 13,319 MGI participants with LDL-C measurements, however, only 1 participant was heterozygous for ZNF529-p.K405X so the power to detect association was near zero. In contrast, we identified 110 heterozygous individuals in the HUNT discovery study.

HepG2 cells. The HepG2 human hepatoma cell line was obtained from the American Type Culture Collection (ATCC) and cultured at 37 °C and 5% CO2 in Dulbecco’s modified Eagle’s Medium (DMEM). Cells were grown with 10% fetal bovine serum (FBS, Sigma-Aldrich) and 1% Penicillin-Streptomycin (Pen-Strep, Gibco).

ZNF529 gene silencing using small interfering RNA (siRNA). siRNA targeting zinc finger protein 529 (siZNF529: GCACUUGAGAUAUGAAGatt) and non-targeting control siRNA (siCTL) were obtained from Ambion (siRNA, ETL) and AM4611, respectively. HepG2 cells were transfected with 20 nM of siZNF529 or siCTL using Lipofectamine RNAiMAX (Invitrogen) in Opti-MEM reduced medium (Gibco) in accordance with the manufacturer’s protocol45. Cellular lipid or protein extraction, RNA isolation or LDL uptake assays were conducted 48 h post transfection.

RNA sequencing. Total RNA was purified from HepG2 cells using the QiAGEN’s RNeasy kit (QIAGEN). Library preparation and sequencing were performed by the University of Michigan DNA Sequencing Core. RNA was assessed for quality using the TapeStation (Agilent, Santa Clara, CA). All samples had RNA integrity numbers (RINs) >8.5. Samples were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7700L) with Poly(A) mRNA Magnetic Isolation Module (NEB, E7490L) and NEBNext Multiplex Oligos for Illumina Unique dual (NEB, E6440L), where 10 ng-1 μg of total RNA were subjected to mRNA polyA purification. The mRNA was then fragmented and copied into first strand cDNA using reverse transcriptase and dUTP mix. Samples underwent end repair, dA tailing and ligation of PCR primer. The libraries were purified and enriched by PCR to create the final cDNA library. Final libraries were checked for quality and quantity by TapeStation (Agilent) and qPCR using Kapa’s library quantification kit for Illumina Sequencing platforms (Kapa Biosystems, KK435). Libraries were paired-end sequenced on a NovaSeq 6000 Sequencing System (Illumina).

Paired-end reads (101 bp) from RNA sequencing of 4 siCTL and 4 siZNF529 samples were aligned to hg38 reference genome using TopHat2 (v2.0.13 11/5/19 7:2900 PM)48 with default parameters. In each sample, 94.3–95.4% reads could be aligned. All valid alignments were used for downstream analysis.

GENCODE release 29 was used to obtain gene boundaries of 19,840 protein coding genes. We used Samtools (v1.9)49 and bedtools (v2.22.0)46 coverage function to count numbers of reads aligned in each genic bin. Genes with >3 counts per million (CPM) in 8 samples were used for further analysis. EdgeR50 library in R was used to identify differentially expressed genes (Supplementary Data 7). We used glmFit function from Limma to identify significant changes in expression with FDR < 0.05. We used ConsensusPathDB48 to identify enriched pathways from list of differentially expressed genes (Supplementary Data 8).

RNA isolation, RT-PCR and qPCR. Total RNA was purified from HepG2 cells using the QiAGEN’s RNeasy kit (QIAGEN). cDNA was synthesized using SuperScript III (Invitrogen), and qPCR was performed using SYBR green reagents (Bio-Rad). Gene analysis was done as fold-change. RNA iso-

ated from control cells by the comparative CT (2−ΔΔCT) method using GAPDH as the reference gene49,50. Primer pairs used for qPCR were obtained from Integrated DNA Technologies and are available in Supplementary Data 19.

Protein extraction and western blot. Cells were lysed in radioimmunoprecipitation assay lysis buffer (RIPA buffer, Thermo Scientific) supplemented with a protease inhibitor cocktail (Roche Applied Science). Proteins were resolved in 4% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose membranes (Bio-Rad). The membranes were blocked for 1 h at room temperature in tris-buffered saline–tween 20 (TBS) containing 5% fat-free milk and incubated with primary antibody at 4 °C overnight. The following primary antibodies were used: rabbit monoclonal anti- LDLR antibody (Abcam, ab52818, working dilution 1:1000) and mouse monoclonal anti-beta actin antibody (Cell signaling, 8H11D010, working dilution 1:2000). After TBST washing, membranes were incubated with secondary antibodies (LI-COR Bio-
technology, donkey anti-rabbit IRDye 926-68072, working dilution 1:1000 for 1 h at room temperature. After TBST washing, bands were visualized and quantified using an Odyssey Infrared Imaging System (LI-COR Biosciences, version 2.1)49.

Dil-LDL uptake assay. 1,1-diotadecyl- 3,3,3’,3’-tetramethylindocarbocyanine perchlorate-low-density lipoprotein (Dil-LDL Alfa Aesar, 65330 or Kalen Bio-
medical, 770230-9) was used to evaluate the cellular uptake of LDL33 in accordance with previously published methodology. 48 h following RNA transfection, HepG2 cells were washed with PBS (x2) and changed to serum-free DMEM supplemented with 0.1% bovine serum albumin (BSA, Sigma-Aldrich). The cells were then incubated with Dil-LDL (~1.25 μg/ml) for 5 h at 37 °C in the dark. In some experiments, cells were pre-treated for 30 min with 25-fold excess amounts of unlabeled LDL (Alfa Aesar, J65330) to assess the specificity of the uptake. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI, Cayman Chemical Company, 14285). After incubation, the cells were washed with PBS (x2) and
changed to serum- and probe-free DMEM. Finally, the cells were visualized using a fluorescent microscope (Olympus, IX71). For each experiment two random fields were chosen and photographed in a blinded fashion. Dil-LDL and DAPI images were merged using ImageJ software (v.1.52k, NIH).

Lipid extraction and cholesterol quantification. The lipids of HepG2 cells were extracted using hexane (≥99%, 32293, Sigma-Aldrich) and isopropanol (≥99.5%, A426-4, Fisher Chemicals) at a 3:2 ratio (v/v), and the hexane phase was left to evaporate for 48 h. The remaining cells in the plates were disrupted in 0.1 M NaOH for 24 h and an aliquot was taken for measurement of cellular protein using the Bradford protein assay (Bio-Rad). The content of cellular cholesterol was determined spectrophotometrically using a commercially available kit (Wako Chemicals, 999-02601). Cholesterol data were normalized to cellular protein levels[44,42].

Statistical analyses for in vitro studies. Statistical analyses were performed using SPSS 24.0 software (SPSS Inc. IBM). Unless indicated otherwise, values are presented as mean ± SD showing all points. All data were tested for normality and equal variance. If the data passed those tests, Student t test was used for comparisons between the two groups. If the data did not pass those tests, nonparametric Mann–Whitney U test was used. P < 0.05 was considered statistically significant.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All relevant data supporting the key findings of this study are available within the article and its Supplementary Information or Source data files or from the corresponding author upon reasonable request. The genome-wide association scan results are available for download at http://csg.sph.umich.edu/willer/public/hunt-lipids-liver-2020/ or at http://bimdata.biobankinternational.org/. The raw RNA sequence reads (accession number PRJNA549711) are available for download at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA549711. Source data are provided with this paper.

Received: 21 January 2020; Accepted: 17 September 2020;
Published online: 18 December 2020

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Lond. Engl. 385, 117–171 (2015).

2. Navarrete, E. P. et al. Association between baseline LDL-C level and total and cardiovascular mortality after LDL-C lowering: a systematic review and meta-analysis. JAMA 319, 1566–1579 (2018).

3. Lund, E. G., Menke, J. G. & Sparrow, C. P. Liver X receptor agonists as therapeutic agents for the treatment of hypercholesterolemia. Arterioscler. Thromb. Vasc. Biol. 23, 1169–1177 (2003).

4. Patzer, N. et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. Lancet Lond. Engl. 375, 752–742 (2010).

5. Kirchgeessner, T. G. et al. Beneficial and adverse effects of an LXR agonist on human lipid and lipoprotein metabolism and circulating neutrophils. Cell Metab. 24, 223–233 (2016).

6. Newman Connie, B. et al. Statin safety and associated adverse events: a scientific statement from the American Heart Association. Arterioscler. Thromb. Vasc. Biol. 39, e38–e81 (2019).

7. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov. 12, 581–594 (2013).

8. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. Nat. Genet. 47, 856–860 (2015).

9. Krookstad, S. et al. Cohort profile: the HUNT study. Nor. Int. J. Epidemiol. 42, 968–977 (2013).

10. Emerging Risk Factors Collaboration. et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction model. Nat. J. Med. 367, 1310–1320 (2012).

11. Talukdor, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Preprint at https://www.biorxiv.org/content/10.1101/563866v1.

12. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. 50, 1335–1341 (2018).

13. Pölli, G. et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet. 2, e132 (2006).
43. Fan, Y. et al. Endothelial TFEB (Transcription Factor EB) positively regulates postischemic angiogenesis. Circ. Res. 122, 945–957 (2018).
44. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
45. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformat. Oxf. 25, 2078–2079 (2009).
46. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing DNA sequences. Bioinformat. Oxf. 26, 841–842 (2010).
47. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformat. Oxf. 26, 139–140 (2010).
48. Kamburov, A. et al. ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res. 39, D712–D717 (2011).
49. Rom, O. et al. Genome-wide RNAi screen reveals ALK1 mediates development of nonalcoholic fatty liver disease in mice. ElBioMedicine 41, 62–72 (2019).
50. Villacorta, L. et al. In situ generation, metabolism and immunomodulatory signaling actions of nitro-conjugated linoleic acid in a murine model of inflammation. Redox Biol. 15, 522–531 (2018).
51. Kraehling, J. R. et al. Genome-wide RNAi screen reveals ALK1 mediates LDL uptake and transcytosis in endothelial cells. Nat. Commun. 7, 13516 (2016).
52. Xiong, W. et al. Brown adipocyte-specific pparγ (peroxisome proliferator-activated receptor γ) deletion impairs perivascular adipose tissue development and enhances atherosclerosis in mice. Arterioscler. Thromb. Vasc. Biol. 38, 1738–1747 (2018).
53. Collins, D. R. et al. Truncated variants of apolipoprotein B cause hypobetalipoproteinemia. Nucleic Acids Res. 16, 8361–8375 (1988).
54. Chalasani, N., Vuppalanchi, R., Raikwar, N. S. & Deg, M. A. Glycolipidolipidinosin-specific phospholipase d in nonalcoholic fatty liver disease: a preliminary study. J. Clin. Endocrinol. Metab. 91, 2279–2285 (2006).
55. Haghpahsah, S., Davani, M., Samadi, B., Ashrafi, A. & Karimi, M. Serum lipid profiles in patients with beta-thalassemia major and intermedia in southern Iran. J. Res. Med. Sci. J. Isfahan Univ. Med. Sci. 15, 150–154 (2010).
56. Romano, P. S. et al. Incidental finding of severe hypertriglyceridermia in children. Role of multiple rare variants in genes affecting plasma triglyceride. J. Clin. Lipid. 11, 1329–1337.e3 (2017).

Acknowledgements

The Trondelag Health (HUNT) Study is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trondelag County Council, Central Norway Regional Health Authority, and the counties. Bioinform. Oxf. 26, 841–842 (2010). The HUNT TOPMed studies (Freeze 5: AACAC, AFGen, Amish, ARIC+VTE, Asthma_Afr, Asthma_CR, CHS, COPDGene, Framingham, GeneStar, GENOA, GenSalt, GOLDN, HyperGen, Jackson, MESA, MLOF/Hemophilia, OMG-SCD, PharmHU, REDS-III-SCD, SAFHS, Sarcoïdosis, SARP, THRy, walk-PhaSST, WHI) was performed at Baylor Human Genome Sequencing Center, Broad Institute of MIT and Harvard, Illumina Genomic Services, Macrogen Corp, New York Genome Center, University of Washington Northwest Genomics Center). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-0251). Phenotype harmonization, data management, sample identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-0251). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. This research has been conducted with the UK Biobank Resource under application number 24460. The K.G. Jebsen Center for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology (NTNU) and Central Norway Regional Health Authority. Whole-genome sequencing for the HUNT Study was funded by HL109946. Whole-genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). J.B.N. is supported by grants from the Danish Heart Foundation (16-R107-A6779) and the Lundbeck Foundation (R220-2016-1434). O.R. is supported by the National Heart, Lung and Blood Institute (NHLBI) of the National Institutes of Health (NIH) grant K99HL150233. American Heart Association Post-doctoral Fellowship 19POST34300224, and the Michigan-Israel Partnership Research Grant. S.A.L. is supported by NIH grant 1R01HL139731 and American Heart Association 18FRNR34250007. I.S. is supported by a Precision Health Scholars Award from the University of Michigan Medical School. L.V. is supported by NIH grant R01-HL133333. P.N. is supported by NIH grant K08-HL140203. J.Z. is supported by NIH grant R01-HL138139. Y.E.C. is supported by NIH grants R01-HL068878 and R01-HL137214. C.I.W. is supported by NIH grants R35-HL135824, R01-HL125764, R01-HL117662-02-S1, and R01-HL130705. The SardiNIA research is supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health (NIH) (contracts N01-AG-1-2109 and HHSN27201100005C). This work is also supported by the National Institutes of Health (NHLBI grant HL117626).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-20086-3.

Correspondence and requests for materials should be addressed to J.B.N., Y.E.C., C.J.W. or K.H.

Peer review information Nature Communications thanks Timothy Frayling and other anonymous reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020
Jonas B. Nielsen1,2,6,53, Oren Rom1,63, Ida Surakka1,63, Sarah E. Graham1,63, Wei Zhou3,4,5,6,63, Tanmoy Roychowdhury1, Lars G. Fritsche1,2,7, Sarah A. Gagliano Taliun6,8, Carlo Sidore9, Yuhaq Liu1, Maiken E. Gabrielsen2, Anne Heidi Skogholt2, Brooke Wolford1,7, William Overton7, Ying Zhao1, Jin Chen1, He Zhang1, Whitney E. Hornsby1, Akua Acheampong1, Austen Grooms1, Amanda Schaef10, Gregory J. M. Zajac7,8, Luis Villacorta1, Jifeng Zhang1, Ben Brumpton2, Mari Leset2,11, Vivek Rai6, Pia R. Lundegaard1,2,13, Morten S. Oleen1,2,13, Kent D. Taylor1,14, Nicholette D. Palmer15, Yili-Der Chen14, Seung H. Choi3, Steven A. Lubitz3,16, Patrick T. Ellinor1,3,16, Kathleen C. Barnes1,17, Michelle Daya1,17, Nicholas Rafaels17, Scott T. Weiss18,19, Jessica Lasky-Su18,19, Russell P. Tracy20,21, Ramachandran S. Vasan22,23, L. Adrienne Cupples23,24, Rasiak Mathias25, Lisa R. Yanek25, Lewis C. Becker25, Patricia A. Peyser26, Lawrence F. Bielak26, Jennifer A. Smith26,27, Stella Aslibekyan28,29, Bertha A. Hidalgo28, Donna K. Arnett30, Marguerite R. Irvin28, James G. Wilson31,32, Solomon K. Musani32,33, Adolfo Correa32,33, Stephen S. Rich34, Xuqing Guo14, Jerome I. Rotter14, Barbara A. Konkle35, Jill M. Johnsen35, Allison E. Ashley-Koch36,37, Marilyn J. Telen37, Vivien A. Sheehan38, John Blangero39,40, Joanne E. Curran39,40, Juan M. Peralta39,40, Courtney Montgomery41, Wayne H-H Sheu42, Ren-Hua Chung43, Karen Schwander44, Seyed M. Nouraei45, Victor R. Gordeuk46, Yingze Zhang45, Charles Kooperberg47, Alexander P. Reiner47,48, Rebecca D. Jackson49, Eugene R. Bleecker50, Deborah A. Meyers50, Xingnan Li51, Sayantan Das6, Ketian Yu7, Jonathon LeFaive7, Albert Smith7, Tom Blackwell7, Daniel Taliun7,8, Sebastian Zollner7, Lukas Forer51, Sebastian Schoenherr52, Christian Fuchsberge7,8,53, Anna Pandit7, Matthew Zawistowski7, Sachin Kheterpal54, Chad M. Bremmert55, Pradeep Natarajan3,55, David Schlessinger56, Seunggeun Lee7, Hyun Min Kang7, Francesco Cucca9,57, Oddgeir L. Holmen2,58, Bjorn O. Asvold2,58,59, Michael Boehnke7,8, Sekar Kathiresan19,55,60, Goncalo R. Abecasis7,8,61, Y. Eugene Chen1,103, Cristen J. Willer1,2,6,103 & Kristian Hveem2,58,62

1Department of Internal Medicine, Cardiology, University of Michigan, Ann Arbor, MI, USA. 2K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, NTNU, Trondheim, Norway. 3Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. 4Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. 5Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. 6Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. 7Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA. 8Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA. 9Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy. 10Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. 11Department of Dermatology, St. Olav’s Hospital, Trondheim University Hospital, Trondheim, Norway. 12Laboratory for Molecular Cardiology, Department of Cardiology, Centre for Cardiac, Vascular, Pulmonary and Infectious Diseases, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark. 13Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, University of Copenhagen, Copenhagen, Denmark. 14The Institute for Translational Genomics and Population Sciences, Department of Pediatrics and Los Angeles Biomedical Research Institute, Harbor-UCLA, Torrance, CA, USA. 15Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. 16Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. 17Colorado Center for Personalized Medicine, School of Medicine, University of Colorado, Aurora, CO, USA. 18Channing Division of Network Medicine, Department of Medicine Brigham and Women’s Hospital, Boston, MA, USA. 19Harvard Medical School, Boston, MA, USA. 20Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT, USA. 21Department of Biochemistry, Larner College of Medicine, University of Vermont, Burlington, VT, USA. 22Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA. 23Framingham Heart Study, Framingham, MA, USA. 24Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. 25GeneSTAR Research Program, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. 26Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. 27Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. 28The University of Alabama at Birmingham, Birmingham, AL, USA. 2923andMe, Inc., Sunnyvale, CA, USA. 30Deans Office, College of Public Health, University of Kentucky, Lexington, KY, USA. 31Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. 32Jackson Heart Study, Jackson, MS, USA. 33Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. 34Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. 35BloodWorks Northwest, University of Washington, Seattle, WA, USA. 36Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA. 37Department of Medicine, Duke University Medical Center, Durham, NC, USA. 38Department of Pediatrics, Division of Hematology/Oncology, Baylor College of Medicine, Houston, TX, USA. 39Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. 40South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. 41Division of Genes and Human Disease, Oklahoma Medical Research Foundation, Oklahoma, OK, USA. 42Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. 43Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan. 44Section of Biostatistics,
Washington University School of Medicine, St. Louis, MO, USA. University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. University of Illinois at Chicago, Chicago, IL, USA. Department of Epidemiology, University of Washington, Seattle, WA, USA. Division of Endocrinology, Diabetes and Metabolism, Ohio State University, Columbus, OH, USA. Division of Pharmacogenomics University of Arizona, Tucson, AR, USA. Division of Genetics, Genomics and Precision Medicine, Department of Medicine, University of Arizona, Tucson, AR, USA. Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. Institute for Biomedicine, Eurac Research, Bolzano, Italy. Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA. Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. Laboratory of Genetics, National Institute on Aging, US National Institutes of Health, Baltimore, MD, USA. Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy. HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway. Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. Broad Institute, Cambridge, MD, USA. Regeneron Pharmaceuticals, Tarrytown, NY, USA. Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. These authors contributed equally: Jonas B. Nielsen, Oren Rom, Ida Surakka, Sarah E. Graham, Wei Zhou. email: jonas.bille.nielsen@gmail.com; echenum@umich.edu; cristen@umich.edu; kristian.hveem@ntnu.no